Frequency-Selective Channel State Feedback in Multiuser MIMO Downlink

Filippo Tosato, *Member, IEEE*

***Abstract*—In this paper we derive a novel method for encoding the MIMO channel state information (CSI) for frequency selective channels and limited feedback. The technique is based on the decomposition of the directional information of a MIMO channel in a two-coordinate system formed by the tangent space of a Grassmann manifold and the tangent space of a unitary group. The first component provides a non-redundant description of the subspace spanned by a set of channel eigenvectors, while the second component provides a non-redundant addendum of information to identify the strongest channel vector directions within that subspace. One advantage of this feedback representation is that it allows to exploit the correlation of MIMO channels by applying standard source coding techniques to the two coordinate components. We also provide simulation results of an LTE downlink system, which show that the proposed CSI feedback description offers dramatic throughput improvement over the existing feedback scheme for MU-MIMO transmission.**

***Index Terms*—Antenna arrays, communication system signaling, feedback communication, algebraic-geometric codes.**

I. Introduction

THE MIMO broadcast channel has been the subject of intensive study in the recent past. Some of the most relevant information-theoretic findings, including the duality principle with the MIMO multiple-access channel, the characterisation of the sum-rate capacity and the dirty-paper coding strategy to achieve capacity have appeared in [1]–[5]. The full description of the capacity region was given in [6], under perfect channel state information at both the receiver (CSIR) and the transmitter (CSIT).

The design of the transmitter precoder and efficient resource scheduling algorithms have also been thoroughly investigated, with the aim to operate the system at the desired point of the capacity region. A plethora of schemes have been proposed, which can be categorised in two broad families: 1) linear preprocessing schemes, including zero-forcing techniques [7], [8], near MMSE solutions [9] and many variations thereof, such as the regularised channel inversion [10] and block-diagonalisation [7], [11]. 2) Non-linear preprocessing schemes, which comprise the Tomlinson-Harashima precoding (THP) and the vector perturbation (VP) technique [12].

The most critical aspect for real-world downlink multiuser MIMO to deliver significant throughput gain is the availability of reliable CSIT. CSI at the transmitter is commonly provided through feedback from each mobile, especially for frequency-division duplexing (FDD), where channel reciprocity is hard to exploit. These feedback channels for CSI signalling are provided in all modern wireless communication systems, such as the 3GPP long-term evolution (LTE), LTE-Advanced and the IEEE 802.16 standards, to support closed-loop spatial multiplexing transmission modes. However, the maximum rate of feedback on these control channels is very limited, hence optimisation of the CSI feedback is fundamental.

The finite-rate feedback case for the MIMO broadcast channel has also been a prolific area of research in recent times with most contributions focussing on the flat-fading channel. It was acknowledged that for a MIMO transmitter the most important information about a user's channel vectors consists of their direction and that the representation of such directional information can be achieved through vector quantisation with codebooks given by unit vectors distributed on a multidimensional unit sphere. Structured spherical vector quantisers can be designed by constructing packings in Grassmann manifolds [13]–[16]. In [17] the analysis of ergodic achievable rates, under finite-rate feedback, is made possible by considering random ensembles of spherical vector quantisation codebooks and a zero-forcing precoding strategy. These results are expanded in [18] to more realistic feedback channels affected by noise, fading, delay, imperfect channel estimation at the mobile terminals and user selection. In particular, it is shown that for an $M \times 1 \times K$ system, with M transmit antennas and $K \geq M$ single-antenna receivers, the number of quantisation bits per user needs to increase with SNR as $B = \alpha(M-1)\log_2 \text{SNR}$, for some $\alpha \geq 1$, in order to achieve the optimal multiplexing gain (i.e. the pre-log factor of the sum-rate at high SNR) of M.

The frequency-selective finite-rate feedback case has received much less attention, despite its relevance for the new generation of wireless systems, which mostly rely on MIMO-OFDM schemes to boost spectral efficiency in the downlink. One of the challenges in designing a CSI feedback scheme for the frequency-selective channel is how to exploit the correlation of the channel vectors across the available bandwidth to reduce the amount of bits needed for their representation. Clearly, sending a CSI feedback message for each narrow subcarrier is wasteful of resources and impractical. One simple solution, adopted for example in the LTE standard, is to deliver separate feedback indications for blocks of a few dozens subcarriers. Albeit effective in reducing the signalling overhead, this is a rather crude way of exploiting the frequency correlation of the channel.

Paper approved by D. J. Love, the Editor for MIMO and Adaptive Techniques of the IEEE Communications Society. Manuscript received March 28, 2011; revised February 9, 2012.

The author is with the Telecommunications Research Laboratory of Toshiba Research Europe, 32 Queen Square, Bristol BS1 4ND, UK (e-mail: filippo.tosato@toshiba-trel.com).

Digital Object Identifier 10.1109/TCOMM.2012.070912.110184

In [19] the authors propose an interpolation-based method to expand the precoder matrices provided through feedback on a few pilot subcarriers to all the subcarriers in a single user MIMO-OFDM system. The interpolation is done by appropriately sampling the geodesic curve connecting the precoder subspaces at consecutive pilot positions, where the precoders are seen as points in the Grassmann manifold. Note that the precoders for non-pilot subcarriers are determined by interpolation at the transmitter, whereas the precoders calculated in the pilot positions by the receiver are fed back in a conventional way, by independently quintising each precoder matrix with a codebook shared between transmitter and receiver. Extension of this scheme to a MIMO broadcast channel does not appear straightforward: in a multiuser downlink transmission the precoder is designed by combining one or more precoding vectors from different users in such a way that spatial interference is minimised. Therefore, subspace interpolation should be applied separately to each user's precoder subspace, while maintaining the orthogonality between these subspaces as much as possible. A second algorithm proposed in [19], which can be combined with the interpolation method, consists in finding a median precoder in a cluster of subcarriers, which still lies in the Grassmann manifold and is obtained by averaging the precoders of all the subcarriers in the cluster.

In [20] a different approach is presented, inspired by the rate-distortion theory, which takes full advantage of the channel correlation. The scheme generates the CSI feedback by quantising the coefficients of the channel baseband impulse response in the time domain. By assuming that these coefficients are uncorrelated, one can simply apply a scalar uniform quantiser to the real and imaginary part of each coefficient independently, with the number of bits determined by the reverse water-filling rule, which minimises the distortion-rate function. However, there are several drawbacks in a practical implementation of such a scheme: 1) if the number of channels coefficients to quantise is too large because, for example, a receiver has multiple antennas, we may want to report only the most significant channel vector directions instead of the full channel matrix. In this case, it is not obvious how to scale down the number of time-domain coefficients to provide such selective information. 2) Quantising the CSI in the time-domain introduces unnecessary overhead if the user has to report the channel state on a few discontinuous subbands rather than the whole bandwidth. 3) In typical MIMO-OFDM systems channel estimation at the terminals is acquired in the frequency domain, hence obtaining the impulse response for a broadband channel requires an additional IDFT (inverse discrete Fourier transformation) across the entire bandwidth at the terminal side.

In this paper we propose a novel solution to the finite-rate CSI feedback problem for the frequency-selective (OFDM) case, which operates directly in the frequency domain. Once recognised that orthonormal matrices provide the structure for the MIMO channel directional information, we introduce a non-redundant representation of such matrices, i.e. a representation that requires as few real coefficients as the number of real degrees of freedom. After representing the channel directional information by a non-redundant set of real coefficients, we are able to remove the correlation across frequency by applying standard source coding techniques, for example the Karhunen-Loève transformation or the more economical DCT (discrete cosine transformation) followed by vector quantisation of the transformed coefficients.

Our proposed representation is based on a transformation of a reference system in the Stiefel manifold into a two-coordinate system given by the Cartesian product of tangent spaces in the Grassmann manifold and the unitary group, respectively. The non-redundant representation of the Grassmann coordinate component is achieved by means of the geodesic curve originating from a fix reference point (e.g. the identity point) at time[1] $t = 0$ and passing through the wanted point at time $t = 1$, where time in our context is just an auxiliary parameter. A similar procedure applies to obtain a non-redundant expression for the unitary group coordinate component.

Unlike the methods in [19], our technique does not apply interpolation nor averaging between precoder matrices. Instead, we achieve a loss-less non-redundant representation of the orthonormal matrices and apply standard source coding techniques to the set of coefficients representing these matrices at different subcarrier positions. The source compression mechanism is similar in concept to the approach taken in [20], however we apply it in the frequency domain rather than the time domain representation of the CSI.

The proposed solution to the encoding of wideband CSI feedback in MIMO systems appears to be novel in the literature. Moreover, the main original contributions of the paper are: 1) a computationally efficient method to derive a representation of orthonormal matrices, i.e. points in the Stiefel manifold, with the smallest possible number of coefficients, 2) an incremental technique to compress the directional information contained in a set of channel eigenvectors across multiple subcarriers or clusters of subcarriers. By compressing the first of the two coordinates we encode the vector subspaces spanned by a set of eigenvectors and by compressing the second coordinate we provide an addendum of information about the relative strength of the corresponding eigenvalues.

The remainder of this paper is organised as follows. In Sec. II we introduce the system model and show what aspects of the MIMO channel spatial structure are critical to ensure beam separation at the transmitter. In Sec. III we derive our two-stage CSI feedback decomposition and in Sec. IV we provide algorithmic procedures for the encoding and decoding operations, whose complexity is assessed in Sec. V. In Sec. VI we apply source coding techniques to the two CSI coordinate components to obtain the final feedback message. Finally, in Sec. VII we present extensive simulation results on a real LTE MIMO downlink, while Sec. VIII draws the conclusions.

[1]The time variable here should not be confused, in general, with the physical time of the transmission system. In fact, when representing the samples of the channel frequency response across frequency, "t" should be replaced by frequency, whereas if we consider the time variation of that frequency response, "t" correctly denotes time. However, throughout the paper we keep the "time" notation to avoid confusion when we describe a geodesic path in terms of the "velocity" and "acceleration" vectors.

II. System Model

We consider a MIMO downlink system with a single transmitter (base station) equipped with M antennas and an arbitrary number of receivers (users) each having $N \leq M$ antennas. At any useful time and frequency, K users are scheduled for transmission, with $1 \leq K \leq M$. This broadcast channel is characterised by the set of input-output equations, for $k = 1, \ldots, K$,

$$\mathbf{y}_k = \mathbf{H}_k^{\dagger}\mathbf{x} + \mathbf{n}_k \ , \tag{1}$$

where $\mathbf{H}_k$ is the $M \times N$ channel matrix containing the base-band complex coefficients, † denotes Hermitian transposition, $\mathbf{x}$ is the transmitted signal vector and $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ is an i.i.d. proper Gaussian noise vector. The input power is constrained such that $\mathrm{tr}\left(\mathrm{E}[\mathbf{x}\mathbf{x}^{\dagger}]\right) \leq P$, where P denotes the total transmitted power (energy per channel use).

We assume that a precoding matrix is applied at the transmitter to map the k-th scheduled user data vector $\mathbf{u}_k$ onto the M transmit antennas. We call *layers*, i.e. spatial streams, the columns of the precoding matrix and *rank* of user k, $L_k \leq N$, the number of layers assigned to the k-th scheduled user. We also assume that the total number of layers, $L = \sum_{k=1}^{K} L_k$, cannot exceed M.

The channel state information available at the transmitter (CSIT) is provided by each user through a short feedback message, which is used for precoder generation, user scheduling, and modulation and coding adaptation. In this work we are primarily interested in the component[2] of the CSI feedback describing the spatial "structure" of the channel matrix $\mathbf{H}_k$.

The precoding operation can be designed in many different ways, but in all cases the primary objective is that of spatially separating the transmit layers such that the system does not become interference limited. In order to do so the transmitter has to be provided with some information about the vector space spanned by the columns of $\mathbf{H}_k$, for any user k. Let us introduce the singular value decomposition (SVD) of a generic user's channel: $\mathbf{H} = \mathbf{U_H}\boldsymbol{\Sigma_H}\mathbf{V_H}^{\dagger}$, where we omit the user index for simplicity. In the ideal case, the directional information is provided by the channel left singular vectors, $\mathbf{U_H}$. On the other hand, in the simplest possible case, a directional indication can be conveyed by a precoding matrix of choice, selected by a user from a finite codebook such that it is the closest match to the channel singular vectors, in terms of providing the largest received SINR.

In general, we can distinguish two levels of accuracy in the directional information. The first level is the minimum information required by the transmitter to ensure an interference-free system, and that is knowledge of the column space of $\mathbf{U_H}$. The second level is full directional information provided by the particular orthonormal matrix $\mathbf{U_H}$, which allows the identification of the strongest directions for transmission. Note that the directional information may also be restricted to the few strongest channel directions, i.e. only a subset of the columns of $\mathbf{U_H}$, or a linear combination thereof, may be signalled to the transmitter.

Once the transmitter has obtained spatial information from each receiver in the form of a column space of $\mathbf{U_H}$ or the actual orthonormal matrix, it can generate a suitable precoder by using, for example, the zero-forcing design criterion [7], [17], combined with a user selection mechanism. An example of a simple greedy user selection strategy that can be combined with any precoding technique is described in [21]. In real OFDM systems the transmitter typically applies the same precoder to a number of adjacent subcarriers forming a cluster of one or more resource blocks. Consequently, the receivers have to provide directional CSI feedback for each cluster of the configured bandwidth by using very limited feedback resources (a few bits). Therefore, exploiting the frequency correlation by compressing the feedback information results in more accurate CSIT for the same number of feedback bits, which can improve the precoder design significantly as will become apparent in the simulation results.

III. Representation of CSIT by Tangent Spaces

The problem of CSIT representation can be described mathematically as finding an efficient representation for 1) the vector subspace spanned by the columns of an orthonormal matrix and/or 2) the orthonormal matrix itself. In the sequel, we show how these quantities can be represented by as few real coefficients as their respective number of degrees of freedom. We call this non-redundant representation. Moreover, the representation of 2) can be made incremental with respect to 1) in the sense that it consists of the same representation as 1) plus an addendum of information.

The algebraic structures needed to achieve such representations are the *complex* unitary group, the Stiefel and the Grassmann manifolds. In the following derivation we will see that a non-redundant representation of these manifolds can be achieved by association with their tangent space. The tangent space has an intuitive definition in Euclidean space, as the space formed by the tangent vectors at a point of a "surface" or manifold. The intuition suggests that the dimension of the tangent space is the same as that of the manifold and that it is possible to represent a point in the manifold by means of a velocity vector by taking a reference point in the surface and drawing the geodesic path, i.e. the shortest path connecting the reference point to the point we want to represent. More precisely, we can take as a representation of a point in a manifold the velocity vector at a reference point of origin that allows to get to the final point along the geodesic connecting the two points in a reference time interval (e.g. one unit of time) at "constant speed". The concept of "constant speed" can be rigorously defined by imposing that the acceleration vector is normal to the tangent at any point along the geodesic. This condition turns out to be a necessary and sufficient to define a geodesic path [22].

Before introducing our CSIT representation we review some fundamental differential-geometric properties, however a more complete survey of the geometry of the *real* version of these manifolds can be found, for example, in [22] or [23] and references therein.

[2]In a typical LTE CSI feedback message, for example, the main indication of the channel spatial structure (a.k.a. precoder matrix indicator) is accompanied by some indication of the preferred transmission rank (a.k.a. rank indicator) and the estimated signal-to-noise ratio measured at the decoder input for each codeword (a.k.a. channel quality indicator).

A. The Unitary Group

If the number of transmit and receive antennas is the same, full directional information is provided by an $M \times M$ unitary matrix, which can be represented as a point in the complex unitary group

$$U_M = \{\mathbf{P} \in \mathbb{C}^{M\times M} \mid \mathbf{P}^\dagger\mathbf{P} = \mathbf{I}\} . \tag{2}$$

Lemma 1: The *complex* unitary group is a manifold of real dimensions[3] $\dim_{\mathbb{R}}(U_M) = 2M^2 - M^2 = M^2$, which equals the number of real dimensions of its tangent space. At a point $\mathbf{P}$, the tangent space consists of all vectors of the form: $\mathbf{PA}$, with $\mathbf{A}$ skew-Hermitian.

Therefore, representing $\mathbf{P}$ by listing the real and imaginary parts of its matrix elements is redundant by a factor of two. By taking the point $\mathbf{I}$ as reference we can represent the tangent space as

$$\mathfrak{u}(M) = \{\mathbf{A} \in \mathbb{C}^{M\times M} \mid \mathbf{A}^\dagger = -\mathbf{A}\} . \tag{3}$$

The geodesic equation in the unitary group can be derived by imposing that the acceleration vector is in the normal space of $\mathbf{P}(t)$ at any point, which yields

$$\mathbf{P}(t) = \mathbf{P}(0)e^{\mathbf{A}t} , \tag{4}$$

with $\mathbf{A}$ skew Hermitian. Given the initial point, $\mathbf{P}(0) = \mathbf{O}$ and the point $\mathbf{P}(1) = \mathbf{P}$ of the geodesic, reached at time $t = 1$, the skew-Hermitian matrix $\mathbf{A}$ can be found by taking the eigenvalue-decomposition (EVD) of the inner product

$$\mathbf{O}^\dagger\mathbf{P} = \mathbf{Q}e^{j\mathbf{\Phi}}\mathbf{Q}^\dagger = e^{\mathbf{Q}j\mathbf{\Phi}\mathbf{Q}^\dagger} \triangleq e^{\mathbf{A}} , \tag{5}$$

such that

$$\mathbf{A} = \mathbf{Q}j\mathbf{\Phi}\mathbf{Q}^\dagger, \tag{6}$$

where $\{e^{j\phi_n}\}$, $n = 1, \ldots, M$, are the eigenvalues of $\mathbf{O}^\dagger\mathbf{P}$. In particular, by choosing $\mathbf{I}$ as the initial point, we can find $\mathbf{A}$ by taking the eigenvalue decomposition of $\mathbf{P}$ and the velocity vector at $\mathbf{I}$ along the geodesic is $\mathbf{A}$, which can be used as an equivalent representation of $\mathbf{P}$.

In conclusion, we only need M^2 real coefficients to list the lower (or upper) triangular part of $\mathbf{A}$ as opposed to the $2M^2$ coefficients needed to represent the real and imaginary part of all the entries of $\mathbf{P}$.

B. The Stiefel Manifold

If $N < M$ or, in general, the number of reported channel eigenvectors is lower than M, full directional information is conveyed by an $M \times N$ orthonormal matrix, which can be represented by a point in the complex Stiefel manifold

$$V_{M,N} = \{\mathbf{P} \in \mathbb{C}^{M\times N} \mid \mathbf{P}^\dagger\mathbf{P} = \mathbf{I}\} . \tag{7}$$

The manifold can be seen as a quotient group of the unitary group, thanks to the following coset representation: $V_{M,N} = U_M/H$, where $H \subset U_M$ is such that

$$H = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & U_{M-N} \end{pmatrix} . \tag{8}$$

[3]The real unitary group, however, has dimensions: $\dim_{\mathbb{R}}(U_M) = M^2 - \sum_{i=1}^{M}(i-1) = \frac{M(M+1)}{2}$, as found in [22].

Thus, a point in the Stiefel manifold, also referred to as an N-frame in the source coding language, can be written as $\mathbf{P} = \bar{\mathbf{P}}\begin{pmatrix}\mathbf{I}\\ \mathbf{0}\end{pmatrix}$, where $\bar{\mathbf{P}}$ is an equivalent class of the unitary group[4], i.e. $\bar{\mathbf{P}} = \left\{\begin{pmatrix}\mathbf{P} & \mathbf{P}_\perp\end{pmatrix} H\right\}$, with $\mathbf{P}_\perp\mathbf{P}_\perp^\dagger = \mathbf{I} - \mathbf{P}\mathbf{P}^\dagger$.

As for the orthogonal group, we can seek a non-redundant representation of a point $\mathbf{P}$ in the Stiefel manifold by means of the tangent at a reference point $\mathbf{O}$ along the geodesic path connecting $\mathbf{O}$ and $\mathbf{P}$. The tangent space of $V_{M,N}$ can be found by decomposing the tangent space (3) with respect to H as $\mathfrak{u}(M) = \mathfrak{h} \oplus \mathfrak{h}_\perp$, where $\mathfrak{h}$ is the tangent space of H at $\mathbf{I}$, called "vertical" component, and

$$\mathfrak{h}_\perp = \left\{ \begin{pmatrix} \mathbf{A} & -\mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{0} \end{pmatrix} \mid \mathbf{A} \in \mathbb{C}^{N\times N}, \mathbf{A}^\dagger = -\mathbf{A} , \mathbf{B} \in \mathbb{C}^{(M-N)\times N} \right\} , \tag{9}$$

is the "horizontal" component.

Lemma 2: The number of real dimensions of the *complex*[5] Stiefel manifold $V_{M,N}$ is: $\dim_{\mathbb{R}}(V_{M,N}) = \dim_{\mathbb{R}}(U_M) - \dim_{\mathbb{R}}(U_{M-N}) = N(2M - N)$, which equals the number of real dimensions of its tangent space. The quotient group representation of the tangent in the Steifel manifold is given by the horizontal tangent (9).

The geodesic is obtained by restricting the geodesic of the orthogonal group (4) to the horizontal space, which yields

$$\mathbf{P}(t) = \mathbf{P}(0)e^{\left(\begin{smallmatrix}\mathbf{A} & -\mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{0}\end{smallmatrix}\right)t} , \tag{10}$$

where the velocity vector $\left(\begin{smallmatrix}\mathbf{A} & -\mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{0}\end{smallmatrix}\right)$ can be taken as a representation of the point $\mathbf{P}(1)$, given the point of origin $\mathbf{P}(0)$. To the author's knowledge, a method for explicitly calculating $\mathbf{A}$ and $\mathbf{B}$ in (10) from $\mathbf{P}(0)$ and $\mathbf{P}(1)$, is currently unknown[6]. However, we will see in the next section how we can obtain a non-redundant representation of $\mathbf{P} \in V_{M,N}$, by using a combination of the geodesics in the Grassmann manifold and unitary group.

C. A Two-coordinate Decomposition

The CSI information associated with the vector space spanned by the columns of an $M \times N$ matrix $\mathbf{P}$ is represented by the complex Grassmann manifold

$$G_{M,N} = \{\langle\mathbf{P}\rangle \mid \mathbf{P} \in V_{M,N}\} , \tag{11}$$

where $\langle\mathbf{P}\rangle$ indicates the N-dimensional subspace of $\mathbb{C}^M$ spanned by the columns of $\mathbf{P}$. Similarly to the Stiefel manifold, $G_{M,N} = U_M/H$, $H \subset U_M$, has the structure of a quotient group of the unitary group with

$$H = \begin{pmatrix} U_N & \mathbf{0} \\ \mathbf{0} & U_{M-N} \end{pmatrix} . \tag{12}$$

Thus, a point in the Grassmann manifold can be written as a set of orthonormal matrices[4] $\langle\mathbf{P}\rangle = \{\bar{\mathbf{P}}\begin{pmatrix}\mathbf{I}\\ \mathbf{0}\end{pmatrix}\}$, where $\bar{\mathbf{P}} =$

[4]In the following, when using the quotient representation for a manifold or its tangent space we will omit the post multiplication by the subgroup H and by $\begin{pmatrix}\mathbf{I}\\ \mathbf{0}\end{pmatrix}$ whenever this does not generate confusion.

[5]The real Stiefel manifold, however, has dimensions: $\dim_{\mathbb{R}}(V_{M,N}) = \frac{M(M+1)}{2} - \frac{(M-N)(M-N+1)}{2} = MN - \frac{N(N-1)}{2}$.

[6]In [22] a method is described to calculate the Stiefel geodesic from $\mathbf{P}(0)$ and $\dot{\mathbf{P}}(0)$. However, the technique is not applicable when the points $\mathbf{P}(0)$ and $\mathbf{P}(1)$ are provided instead.

$\left\{\begin{pmatrix}\mathbf{P} & \mathbf{P}_\perp\end{pmatrix} H\right\}$ is an equivalence class in U_M, and $\mathbf{P}_\perp\mathbf{P}_\perp^\dagger = \mathbf{I} - \mathbf{P}\mathbf{P}^\dagger$.

Similarly to the previous two manifolds, a non-redundant representation can be achieved through the tangent at a reference point $\langle\mathbf{O}\rangle$ along the Grassmann geodesic connecting $\langle\mathbf{O}\rangle$ and $\langle\mathbf{P}\rangle$. By decomposing the tangent space (3) with respect to the equivalence class H, we obtain: $\mathfrak{u}(M) = \mathfrak{h}\oplus\mathfrak{h}_\perp$, where the vertical component $\mathfrak{h}$ is the tangent space of H at $\mathbf{I}$, and the horizontal component is given by

$$\mathfrak{h}_\perp = \left\{ \begin{pmatrix}\mathbf{0} & -\mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{0}\end{pmatrix} \mid \mathbf{B}\in\mathbb{C}^{(M-N)\times N} \right\}. \tag{13}$$

Lemma 3: The number of real dimensions of the *complex*[7] Grassmann manifold equals that of its tangent space and is given by: $\dim_{\mathbb{R}}(G_{M,N}) = M^2 - N^2 - (M-N)^2 = 2N(M-N)$. The horizontal tangent (13) is the quotient group representation of the Grassmann manifold tangent.

By restricting the geodesic of the unitary group (4) to the horizontal space, we obtain the geodesic in the Grassmann manifold:

$$\mathbf{P}(t) = \mathbf{P}(0)e^{\left(\begin{smallmatrix}\mathbf{0} & -\mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{0}\end{smallmatrix}\right)t}. \tag{14}$$

In order to use the matrix $\mathbf{B}$ as an equivalent representation of $\langle\mathbf{P}\rangle$, we need to fix the starting point $\mathbf{P}(0) = \mathbf{O}$ and calculate the geodesic curve impinging on a representing N-frame of $\langle\mathbf{P}\rangle$ at time 1.

To simplify the derivation, let us also assume for now that $N \le M/2$. This restriction will be removed later on. Let $\mathbf{\Theta}$ be the $N\times N$ diagonal matrix of the principal angles between two subspaces $\langle\mathbf{O}\rangle$ and $\langle\mathbf{P}\rangle$.

The geodesic curve can be calculated by taking the CS-decomposition of the inner product

$$\begin{pmatrix}\mathbf{O} & \mathbf{O}_\perp\end{pmatrix}^\dagger \begin{pmatrix}\mathbf{P} & \mathbf{P}_\perp\end{pmatrix} = \begin{pmatrix}\mathbf{Q}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{21} & \mathbf{Q}_{22}\end{pmatrix} \begin{pmatrix}\cos\mathbf{\Theta} & -\sin\mathbf{\Theta} & \mathbf{0} \\ \sin\mathbf{\Theta} & \cos\mathbf{\Theta} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}\end{pmatrix} \begin{pmatrix}\mathbf{V}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}^\dagger \tag{15}$$

where we define $\mathbf{Q}_2 = \begin{pmatrix}\mathbf{Q}_{21} & \mathbf{Q}_{22}\end{pmatrix}$ and $\mathbf{V}_2 = \begin{pmatrix}\mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}$. More on how to calculate the CS-decomposition can be found, for example, in [24, §2.6.4]. Let us introduce the rotation matrices $\tilde{\mathbf{V}}_1$ and $\tilde{\mathbf{V}}_2$, of size $N\times N$ and $(M-N)\times(M-N)$, respectively, defined as

$$\tilde{\mathbf{V}}_1 \triangleq \mathbf{V}_1\mathbf{Q}_1^\dagger \tag{16}$$

$$\tilde{\mathbf{V}}_2 \triangleq \mathbf{V}_2\mathbf{Q}_2^\dagger. \tag{17}$$

By multiplying both sides of (15) by the block diagonal unitary matrix $\left(\begin{smallmatrix}\tilde{\mathbf{V}}_1 & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{V}}_2\end{smallmatrix}\right)$ and by using the definition of the matrix exponential and its basic properties, we obtain

$$\begin{pmatrix}\mathbf{O} & \mathbf{O}_\perp\end{pmatrix}^\dagger \begin{pmatrix}\mathbf{P}\tilde{\mathbf{V}}_1 & \mathbf{P}_\perp\tilde{\mathbf{V}}_2\end{pmatrix} = e^{\begin{pmatrix}\mathbf{Q}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{21}\end{pmatrix}\begin{pmatrix}\mathbf{0} & -\mathbf{\Theta} \\ \mathbf{\Theta} & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{Q}_1^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_{21}^\dagger\end{pmatrix}} = e^{\begin{pmatrix}\mathbf{0} & -\mathbf{B}^\dagger \\ \mathbf{B} & \mathbf{0}\end{pmatrix}}, \tag{18}$$

where we have defined

$$\mathbf{B} \triangleq \mathbf{Q}_{21}\mathbf{\Theta}\mathbf{Q}_1^\dagger. \tag{19}$$

By introducing the independent time variable, we acknowledge that (18) yields the implicit geodesic (14) with $\mathbf{P}(0) = \mathbf{O}$ and $\mathbf{P}(1) = \mathbf{P}\tilde{\mathbf{V}}_1$. Hence, the $(M-N)\times N$ matrix (19) can be taken as a non-redundant representation of $\langle\mathbf{P}\rangle$ and from (18) we obtain the following explicit geodesic curve, which allows the reconstruction of the N-frame $\mathbf{P}(1) = \mathbf{P}\tilde{\mathbf{V}}_1$ from $\mathbf{B}$ (and the initial point $\mathbf{O}$)

$$\mathbf{P}(t) = \mathbf{O}\mathbf{Q}_1\cos(\mathbf{\Theta}t)\mathbf{Q}_1^\dagger + \mathbf{O}_\perp\mathbf{Q}_{21}\sin(\mathbf{\Theta}t)\mathbf{Q}_1^\dagger. \tag{20}$$

Note that (18) also provides the geodesic curve connecting the orthogonal complements $\langle\mathbf{O}_\perp\rangle$ and $\langle\mathbf{P}_\perp\rangle$ in $G_{M,M-N}$, with equation

$$\mathbf{P}_\perp(t) = -\mathbf{O}\mathbf{Q}_1\sin(\mathbf{\Theta}t)\mathbf{Q}_{21}^\dagger + \mathbf{O}_\perp\left(\mathbf{Q}_{21}\cos(\mathbf{\Theta}t)\mathbf{Q}_{21}^\dagger + \mathbf{Q}_{22}\mathbf{Q}_{22}^\dagger\right). \tag{21}$$

Therefore, if $N > M/2$, we can always switch to the orthogonal complements, apply the same derivation and use equation (21) for the geodesic. Thus, (19) can also be taken as non-redundant representation of $\langle\mathbf{P}_\perp(1)\rangle$ and (21) allows the reconstruction of the representing $(M-N)$-frame $\mathbf{P}_\perp(1) = \mathbf{P}_\perp\tilde{\mathbf{V}}_2$.

In the reconstruction of $\mathbf{P}(1)$ or $\mathbf{P}_\perp(1)$ from $\mathbf{B}$ we need to take the SVD of $\mathbf{B}$, which gives, in general, different component matrices from those in (19):

$$\mathbf{B} = \begin{pmatrix}\tilde{\mathbf{Q}}_{21} & \tilde{\mathbf{Q}}_{22}\end{pmatrix}\begin{pmatrix}\mathbf{\Theta} \\ \mathbf{0}\end{pmatrix}\tilde{\mathbf{Q}}_1^\dagger. \tag{22}$$

However, the two decompositions (19) and (22) are related as follows: $\tilde{\mathbf{Q}}_{21} = \mathbf{Q}_{21}\mathbf{L}$, $\tilde{\mathbf{Q}}_1 = \mathbf{Q}_1\mathbf{L}$, where $\mathbf{L}$ is the product of a permutation matrix and a diagonal matrix with complex exponentials on the diagonal. Moreover, the following identity holds true: $\mathbf{Q}_{22}\mathbf{Q}_{22}^\dagger = \tilde{\mathbf{Q}}_{22}\tilde{\mathbf{Q}}_{22}^\dagger$, because the two matrices span the same subspace. Therefore, all the terms in (20) and (21) are unaffected by the particular choice of SVD.

We can combine this representation in the Grassmann manifold with the representation in the unitary group described in Sec. III-A to provide a non-redundant description of a point in the Stiefel manifold. In fact, if we want to represent the particular frame $\mathbf{P}\in V_{M,N}$, rather than the subspace $\langle\mathbf{P}\rangle$, we can find the matrix exponential representation of the matrix $\tilde{\mathbf{V}}_1^\dagger \in U_N$, with respect to some reference point $\mathbf{O}_1$, such that $\tilde{\mathbf{V}}_1^\dagger = \mathbf{O}_1 e^{\mathbf{A}}$. Then, from the pair $(\mathbf{B},\mathbf{A})$ (and by knowing the fixed reference points $\mathbf{O}$ and $\mathbf{O}_1$), we can reconstruct $\mathbf{P}$ by taking the product of the two reconstructed matrices. More precisely, by taking the SVD of $\mathbf{B}$, from (20) we get

$$\mathbf{P}\tilde{\mathbf{V}}_1 = \mathbf{O}\mathbf{Q}_1\cos(\mathbf{\Theta})\mathbf{Q}_1^\dagger + \mathbf{O}_\perp\mathbf{Q}_{21}\sin(\mathbf{\Theta})\mathbf{Q}_1^\dagger, \tag{23}$$

and, by taking the EVD of the skew-Hermitian matrix $\mathbf{A}$, from (5) we obtain

$$\tilde{\mathbf{V}}_1^\dagger = \mathbf{O}_1\mathbf{Q}(\cos\mathbf{\Phi} + j\sin\mathbf{\Phi})\mathbf{Q}^\dagger. \tag{24}$$

The N-frame $\mathbf{P}\in V_{M,N}$ is simply given by the right multiplication of (23) by (24). Note that, in the complex case, we need $2N(M-N)$ real coefficients to represent $\mathbf{B}$ and

[7]The real Grassmann manifold, instead, has dimensions: $\dim_{\mathbb{R}}(G_{M,N}) = \frac{M(M+1)}{2} - \frac{N(N+1)}{2} - \frac{(M-N)(M-N+1)}{2} = N(M-N)$.

N^2 real coefficients for $\mathbf{A}$. In total, the pair $(\mathbf{B}, \mathbf{A})$ requires $N(2M-N)$ real coefficients, which is the number of real degrees of freedom of $V_{M,N}$. Hence, we have achieved a non-redundant representation for $V_{M,N}$. The same two-step representation applies in the case $N > M/2$, where we use the reconstruction formula (21) instead of (20) and we apply the exponential map to $\tilde{\mathbf{V}}_2 \in U_{M-N}$ instead of $\tilde{\mathbf{V}}_1$.

IV. Two-stage Representation of CSIT

From the above discussion we conclude that a non-redundant representation of the full channel directional information, given by the $M \times N$ matrix of channel eigenvectors, $\mathbf{U_H}$, can be achieved in two stages by using the transformation

$$\mathbf{U_H} \rightarrow (\mathbf{B}, \mathbf{A}), \tag{25}$$

where $\mathbf{B}$ is $(M-N) \times N$ and $\mathbf{A}$ is $N \times N$ skew-Hermitian. Algebraically, this transformation corresponds to the change of coordinates:

$$V_{M,N} \rightarrow \mathfrak{h}_\perp(G_{M,N}) \times \mathfrak{u}(N), \tag{26}$$

where $\mathfrak{h}_\perp(G_{M,N})$ is the tangent space in the Grassmann manifold $G_{M,N}$, defined in (13), and $\mathfrak{u}(N)$ is the tangent space in the unitary group $U(N)$, defined in (3).

This representation has the advantage of providing two levels of refinement of CSIT: if only $\mathbf{B}$ is known, we can reconstruct the subspace $\langle \mathbf{U_H} \rangle$, which is enough to generate an interference-free multiuser transmission. If the second component, $\mathbf{A}$, is also made available we can reconstruct the full MIMO spatial structure by combining the two indications.

Note that instead of providing the matrix $\mathbf{B}$, as per (19), to represent $\langle \mathbf{U_H} \rangle$, we may equivalently use

$$\mathbf{B}' \triangleq \mathbf{Q}_{21} \sin \mathbf{\Theta} \mathbf{Q}_1^\dagger, \tag{27}$$

where we have replaced the principal angles $\mathbf{\Theta}$ with their sine. This new matrix can be calculated, for $N \leq M/2$, as

$$\mathbf{B}' = \mathbf{O}_\perp^\dagger \mathbf{P} \tilde{\mathbf{V}}_1, \tag{28}$$

as can be easily verified from the geodesic expression (20) for $t = 1$. Similarly, for $N > M/2$, $\mathbf{B}'$ can be found from

$$\mathbf{B}' = -\tilde{\mathbf{V}}_2^\dagger \mathbf{P}_\perp^\dagger \mathbf{O}, \tag{29}$$

as appears from (21). The advantage of replacing $\mathbf{B}$ with $\mathbf{B}'$ is to save the calculation of N 'arccos' functions needed to find the principal angles $\mathbf{\Theta}$. However, if the coefficients of $\mathbf{B}'$ are to be later quantised to produce a feedback message, some eigenvalues of the quantised matrix may exceed one. These values need to be rounded down to one in the reconstruction phase, which may introduce some additional reconstruction error. Therefore, in the following we will always use $\mathbf{B}$ for subspace representation.

We also note that a similar replacement of the skew-Hermitian matrix $\mathbf{A}$, given by (6), with the matrix: $\mathbf{A}' = \mathbf{Q} j \sin(\mathbf{\Phi}) \mathbf{Q}^\dagger$, is not possible because the latter does not provide an equivalent representation in U_N. This is because the angles $\{\phi_n\}$ are defined in $(-\pi, \pi]$, unlike the singular values of $\mathbf{B}$, $\{\theta_n\}$ which are in $[0, \pi/2)$.

We are now ready to give an algorithmic procedure to perform the transformation (25). Let us fix as reference point the identity, $(\mathbf{O} \quad \mathbf{O}_\perp) = \mathbf{I}_N$.

1) *Calculation of* $\mathbf{B}$.
   - If $N \leq M/2$: set $\mathbf{O} = \left(\begin{smallmatrix} \mathbf{I}_N \\ \mathbf{0} \end{smallmatrix}\right)$, $\mathbf{O}_\perp = \left(\begin{smallmatrix} \mathbf{0} \\ \mathbf{I}_{M-N} \end{smallmatrix}\right)$ and $\mathbf{P} = \mathbf{U_H}$. Compute the CS-decomposition (thin version) of $\mathbf{P} = \left(\begin{smallmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{smallmatrix}\right)$, with $\mathbf{P}_1 \in \mathbb{C}^{N \times N}$ and $\mathbf{P}_2 \in \mathbb{C}^{(M-N) \times N}$. Find the matrix $\tilde{\mathbf{V}}_1$ from (16) and $\mathbf{B}$ from (19).
   - If $N > M/2$: set $\mathbf{O} = \left(\begin{smallmatrix} \mathbf{I}_{M-N} \\ \mathbf{0} \end{smallmatrix}\right)$, $\mathbf{O}_\perp = \left(\begin{smallmatrix} \mathbf{0} \\ \mathbf{I}_N \end{smallmatrix}\right)$ and $\mathbf{P}_\perp = \mathbf{U_H}$. Compute the CS-decomposition (thin version) of $\mathbf{P}_\perp = \left(\begin{smallmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{smallmatrix}\right)$, with $\mathbf{P}_1 \in \mathbb{C}^{(M-N) \times N}$ and $\mathbf{P}_2 \in \mathbb{C}^{N \times N}$. Find the matrix $\tilde{\mathbf{V}}_2$ from (17) and $\mathbf{B}$ from (19).
2) *Calculation of* $\mathbf{A}$. Set $\mathbf{O} = \mathbf{I}_N$,
   - If $N \leq M/2$: set $\mathbf{P} = \tilde{\mathbf{V}}_1^\dagger$. Compute $\mathbf{A}$ from (6).
   - If $N > M/2$: set $\mathbf{P} = \tilde{\mathbf{V}}_2^\dagger$. Compute $\mathbf{A}$ from (6).

The reverse transformation of (25) is performed by the following procedure.

1) *Reconstruction of* $\langle \mathbf{U_H} \rangle$. Calculate the SVD of $\mathbf{B}$ as per (22)[8] and compute an N-frame representative, $\tilde{\mathbf{U_H}}$, of $\langle \mathbf{U_H} \rangle$ as follows:
   - If $N \leq M/2$,
   $$\tilde{\mathbf{U_H}} = \begin{pmatrix} \tilde{\mathbf{Q}}_1 \cos \mathbf{\Theta} \tilde{\mathbf{Q}}_1^\dagger \\ \tilde{\mathbf{Q}}_{21} \sin \mathbf{\Theta} \tilde{\mathbf{Q}}_1^\dagger \end{pmatrix} \tag{30}$$
   - If $N > M/2$,
   $$\tilde{\mathbf{U_H}} = \begin{pmatrix} -\tilde{\mathbf{Q}}_1 \sin \mathbf{\Theta} \tilde{\mathbf{Q}}_{21}^\dagger \\ \tilde{\mathbf{Q}}_{21} \cos \mathbf{\Theta} \tilde{\mathbf{Q}}_{21}^\dagger + \tilde{\mathbf{Q}}_{22} \tilde{\mathbf{Q}}_{22}^\dagger \end{pmatrix} \tag{31}$$
2) *Reconstruction of* $\mathbf{U_H}$. Calculate the EVD: $\mathbf{A} = \tilde{\mathbf{Q}} j \mathbf{\Phi} \tilde{\mathbf{Q}}^\dagger$, and the unitary matrix
   $$\tilde{\mathbf{V}} = \tilde{\mathbf{Q}} (\cos \mathbf{\Phi} + j \sin \mathbf{\Phi}) \tilde{\mathbf{Q}}^\dagger. \tag{32}$$
   The matrix of channel eigenvectors, $\mathbf{U_H}$, is reconstructed by the matrix product
   $$\mathbf{U_H} = \tilde{\mathbf{U_H}} \tilde{\mathbf{V}}. \tag{33}$$

V. Complexity Analysis

A final look now at the complexity involved in the above calculations. The computational cost of a CS-decomposition (thin version) of an $M \times N$ matrix, with $M \geq N$, is essentially that of an SVD of size $\min\{N, M-N\} \times N$. This is the main cost of calculating the $\mathbf{B}$ component matrix above. The calculation of the $\mathbf{A}$ component matrix, instead requires one EVD of size $\min\{N, M-N\} \times \min\{N, M-N\}$. On the other hand, reconstructing $\langle \mathbf{U_H} \rangle$ from $\mathbf{B}$ requires one SVD of size $(M-N) \times N$, whilst reconstructing $\mathbf{U_H}$ from $\mathbf{A}$ entails one EVD of size $\min\{N, M-N\} \times \min\{N, M-N\}$. Because the computational costs of SVD's and an EVD's are comparable, we can conclude that any of the operations described above requires the equivalent complexity of one small SVD of size dictated by the number of receiving antennas. In Table I we compare the complexity of the proposed feedback scheme with a simple clustering method like the one used currently in the LTE family of standards, whereby a single codebook index is reported per cluster. We assume that F is the number of

[8]Note that $\tilde{\mathbf{Q}}_{22} = \mathbf{0}$ if $N \leq M/2$.

TABLE I
COMPLEXITY COMPARISON BETWEEN THE PROPOSED FEEDBACK AND THE LTE FEEDBACK MECHANISM FOR A FEEDBACK RANK N.

	Proposed scheme		LTE feedback
	$\mathbf{B}$ feedback	$\mathbf{A}$ feedback	
Receiver			
Calculation of $\mathbf{B}$	$O(FN^3)$	—	—
Calculation of $\mathbf{A}$	—	$O(FN^3)$	—
Transform coding (DCT)	$O(N(M-N)F\log F)$	$O(N^2F\log F)$	—
Vector quantisation	$O(2^B)$	$O(2^B)$	$O(F2^B)$
Transmitter			
Transform decoding (IDCT)	$O(N(M-N)F\log F)$	$O(N^2F\log F)$	—
Subspace reconstruction: $\langle\mathbf{U_H}\rangle$	$O(MN^2+N^3)$, for $N<(M-N)$ $O(N^3)$, for $N\geq M-N$	—	—
Matrix reconstruction: $\mathbf{U_H}$	—	$O(N^3)$	—

frequency clusters for which CSI feedback is provided and B the number of bits per feedback index. The vector quantisation operation to calculate the codebook index is similar in both schemes and its complexity grows exponentially with B. However, whilst the LTE scheme performs F codebook searches, in our scheme only one codebook index is generated for the $\mathbf{B}$ component and one for the $\mathbf{A}$ component, irrespectively of the number of clusters. The proposed compression scheme bears the additional cost of calculating the $\mathbf{B}$ and $\mathbf{A}$ matrices for the F clusters and a discrete cosine transform (DCT) for the two sets of coefficients (see Sec. VI). For the SVD decomposition we use the R-SVD algorithm developed by Chan [25], which requires $6MN^2+20N^3$ flops to compute the thin version SVD of a matrix of size $M\times N$, with $M\geq N$ (see [24, §5.4.5]). The flop count for a thin SVD of size $\min\{N, M-N\}\times N$ is $26N^3$ for $N\leq(M-N)$ and $6N(M-N)^2+20(M-N)^3<26N^3$ for $N>(M-N)$. Therefore, the complexity of this operation amounts to $O(N^3)$. Similarly we can derive the complexity for the other SVDs and EVDs. With regard to the transform encoding operation, we need to operate $2N(M-N)$ DCTs of size F for the $\mathbf{B}$ component and N^2 DCTs of size F for the $\mathbf{A}$ component (see Sec. VI), which require a computational cost of $O(N(M-N)F\log F)$ and $O(N^2F\log F)$, respectively.

VI. QUANTISATION OF THE CSIT

Once the CSIT has been encoded by the transformation (25) in the matrices $\mathbf{B}$ and $\mathbf{A}$, these have to be quantised in some way before being fed back to the transmitter on a digital control channel.

One of the advantages of the CSIT description (33) is that it allows to exploit the correlation of the channel vectors in frequency selective channels by quantising the elements of $\mathbf{B}$ and $\mathbf{A}$ across the frequency blocks in a MIMO-OFDM system. In fact, we can apply the rate-distortion theory to the quantisation of $\mathbf{B}$ and $\mathbf{A}$ in a similar fashion to [20] but without the drawbacks of dealing with the channel impulse response, as pointed out in the introduction.

In particular, if there are F frequency resource blocks for which to report CSI, one can calculate a matrix pair representation for each block, $(\mathbf{B}_1,\mathbf{A}_1),\ldots,(\mathbf{B}_F,\mathbf{A}_F)$ and remove the correlation of the samples by applying the Karhunen-Loève (K-L) transformation to the $2N(M-N)$ length-F sequences of real coefficients of $\mathbf{B}$ and N^2 length-F sequences of real coefficients of $\mathbf{A}$. Then, scalar or vector quantisation can be applied to the K-L transformed coefficients.

One practical implementation of this transform coding scheme, which does not require knowledge of the channel correlation matrix, is obtained by replacing the K-L transformation with the DCT [26]. It is well known (see, for example, [26]) that for highly correlated signals the DCT exhibits essentially the same performance as the optimum K-L transform. We then use one vector quantisation codebook to quantise all the DCT-transformed coefficients of the $\mathbf{B}$ matrix and another codebook for the vector quantisation of the transformed coefficients of $\mathbf{A}$. These codebooks can be designed, for example, by using the LBG algorithm [27]. We have used this transform coding and vector quantisation scheme in the simulation study reported in the next section.

VII. NUMERICAL RESULTS

For our simulations we consider a MIMO-OFDM downlink transmission in an LTE system [28] with a system bandwidth of 1.4 MHz, 4 transmit antennas ($M=4$), 4 receive antennas ($N=4$) and 4 users ($K=4$). The channel model is the SCM (spatial channel model) for a typical urban configuration [29]. The transmission bandwidth is divided up in 6 resource blocks (RB). A resource block is a frequency-time unit for resource allocation comprising 12 OFDM subcarriers and spanning one subframe of $1ms$ in time, which corresponds to the duration of 14 OFDM symbols.

In our simulation setup we use typical LTE parameters for the feedback granularity and overhead. In particular, we assume each user feeds back one CSI report for every subframe and every 2 RBs. We recall that each CSI feedback message consists, according to the LTE terminology, of a PMI (precoding matrix indicator), which is associated to an orthonormal matrix of channel directional vectors or precoding vectors, one RI (rank indicator), which indicates the rank of such orthonormal matrix, and one CQI (channel quality indicator) for each codeword, or independently encoded data stream, configured by the base station. In the simulations, we assume that each scheduled user receives a single codeword whose coded symbols are multiplexed across the transmit layers assigned to a user. We also fix the feedback overhead as follows. The CQI is a 4-bit indicator of the modulation and coding scheme (MCS) for a codeword that ensures a

TABLE II
SIMULATION PARAMETERS FOR THE LTE SYSTEM.

Channel model	SCM, typical urban
Carrier frequency	2GHz
eNodeB[a] antenna configuration	4 elements, 10λ spacing
UE[b] antenna configuration	4 elements, 0.5λ spacing
Number of UEs	4
No. layers per scheduled UE	1,2,3,4
No. codewords per UE per subframe	1
DL transmission bandwidth	1.4MHz
UE velocity	3 km/h
PDCCH[c]/PDSCH[d] configuration	3/11 OFDM symbols
Scheduling unit in time	1 subframe
Scheduling unit in freq.	2 RBs
Channel coding	LTE Rel-8 turbo coding
Link adaptation	LTE Rel-8 MCS
Detector	MMSE
Feedback granularity	2 RBs, 1 subframe
Feedback overhead per UE	30 bits/1ms/1.4MHz
Precoding scheme	Zero-forcing
Scheduler	maximum cell throughput
Common reference signals	LTE Rel-8 CRS
Demodulation reference signals	LTE Rel-10 DM-RS[e]
CSI feedback delay	5ms
Channel estimation for CSI feedback	2D-MMSE
Channel estimation for demodulation	2D-MMSE

[a]i.e., LTE base station
[b]user equipment
[c]physical downlink control channel
[d]physical downlink shared channel
[e]LTE Rel-10 introduced DM-RS support for more than 2 layers [31]

predetermined average frame error rate (FER), typically 1%. Hence, 16 combinations of turbo codes and QAM modulations are available for MCS selection at the transmitter. The RI is a 2-bit index corresponding to the four possible ranks of the precoding matrix between 1 and M. Finally, we assume that the PMI is a 4-bit index as in LTE Rel-8/9 for four transmit antennas. Therefore, in all our comparisons the feedback overhead for CSIT is limited, for each user, to 30 bits per millisecond for the whole bandwidth or 0.0214 bit/s/Hz. The total uplink resources needed to support all the four users' feedback is then 0.0857 bit/s/Hz, which should be a small proportion of the throughput achieved in the downlink to make a MIMO limited feedback scheme worthwhile. We assume that the feedback channel is error free, but it introduces a delay of 5ms. The precoder is zero-forcing combined with the simple greedy user selection described in [21].

Two sets of pilots are provided in the downlink frame for channel estimation. The CRS (common reference signals) are inserted after the precoding operation and used to estimate the downlink channel for CSI feedback generation. The DM-RS (demodulation reference signals) are inserted before precoding and used to estimate the combination of precoder and channel for correct signal demodulation. In all simulations we assume two-dimensional MMSE (2D-MMSE) channel estimation [30] on both the DM-RS and CRS. The main simulation parameters are summarised in Table II.

In the following results we compare the feedback scheme of Sec. IV, with and without quantisation, to the codebook-based feedback mechanism adopted in LTE. While in the LTE

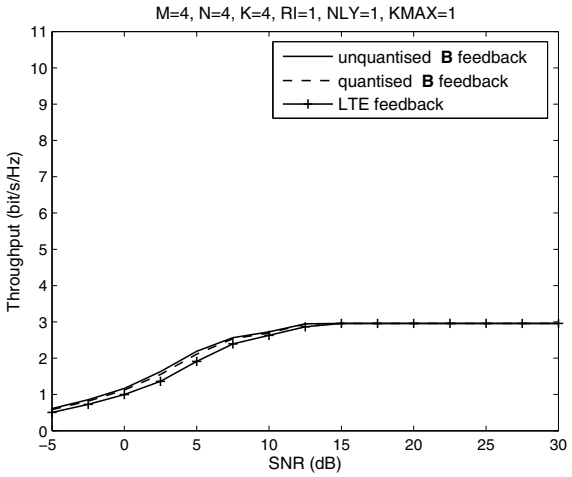

Fig. 1. Throughput comparison when the users report a rank-1 indicator (RI=1). 1 layer is allocated per user (NLY=1) and 1 user scheduled per RB (KMAX=1).

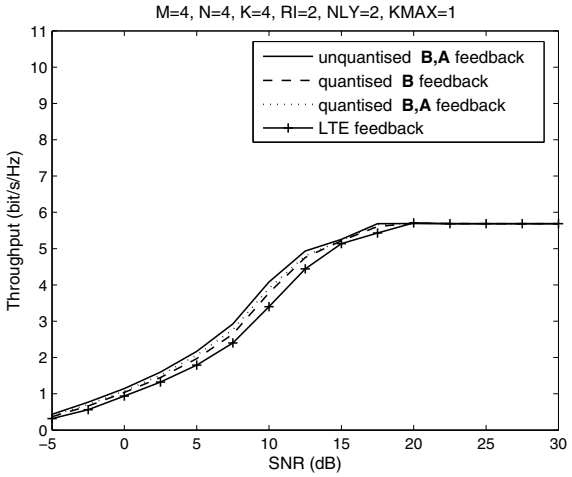

Fig. 2. Throughput comparison when the users report a rank-2 indicator (RI=2). 2 layers are allocated per user (NLY=2) and 1 user scheduled per RB (KMAX=1).

feedback each user reports separate PMIs per frequency sub-block, in our scheme one indicator is generated for the $\mathbf{B}$ matrix and one for the $\mathbf{A}$ matrix for the whole bandwidth in order to exploit the frequency correlation of the MIMO channel. Each indicator is obtained as described in Sec. VI by applying vector quantisation to the DCT-transformed coefficients. When reporting both feedback components, $\mathbf{B}$ and $\mathbf{A}$, the two indicators can be multiplexed in time such that the total signalling overhead is the same as for a single indicator feedback. In particular, we study the performance of the two feedback schemes for different values of the user's reported rank (RI), the actual number of layers assigned to each scheduled user[9], indicated by the parameter NLY, and the maximum number of users scheduled per RB, denoted by KMAX.

In Fig. 1 and 2 we look at the special case of a single user allocated per RB (SU-MIMO), for RI=1 and 2, respectively. Note that for RI=1, a single unit-norm vector is reported and full directional information is conveyed by the $\mathbf{B}$ matrix only. In fact, (30) allows to reconstruct the reported vector up to a

[9]This can be lower or equal to the reported RI, as the transmitter can override the rank indicated by a user.

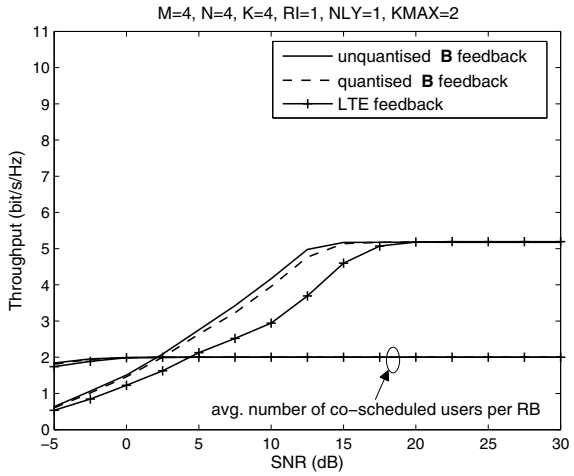

Fig. 3. Throughput comparison when the users report a rank-1 indicator (RI=1). 1 layer is allocated per user (NLY=1) and up to 2 users scheduled per RB (KMAX=2). The average number of users scheduled per RB is also plotted.

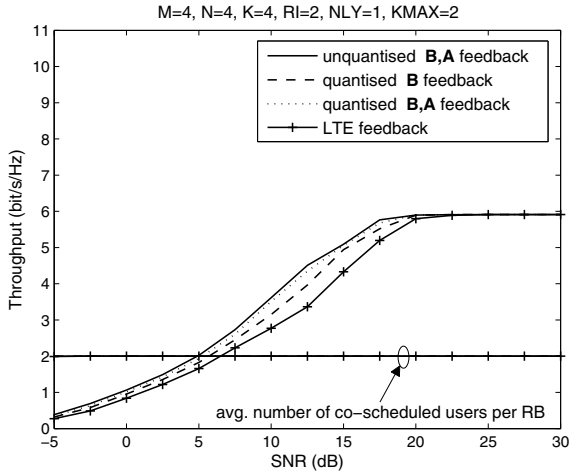

Fig. 4. Throughput comparison when the users report a rank-2 indicator (RI=2). 1 layer is allocated per user (NLY=1) and up to 2 users scheduled per RB (KMAX=2). The average number of users scheduled per RB is also plotted.

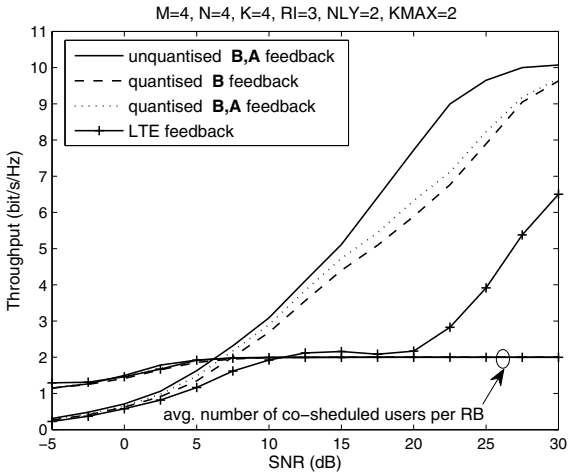

Fig. 5. Throughput comparison when the users report a rank-3 indicator (RI=3). 2 layers are allocated per user (NLY=2) and up to 2 users scheduled per RB (KMAX=2). The average number of users scheduled per RB is also plotted.

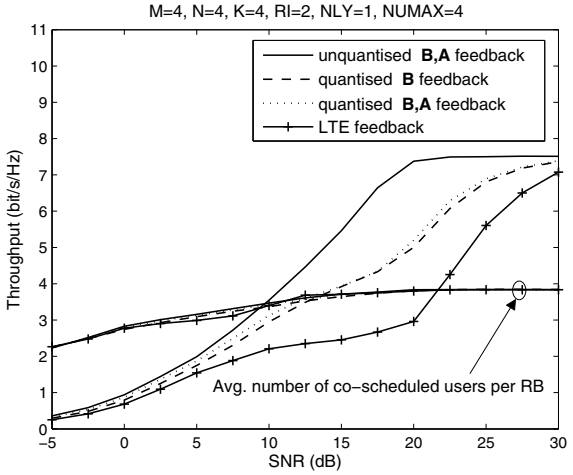

Fig. 6. Throughput comparison when the users report a rank-2 indicator (RI=2). 1 layer is allocated per user (NLY=1) and up to 4 users scheduled per RB (KMAX=4). The average number of users scheduled per RB is also plotted.

phase shift, which is irrelevant for the precoder calculation. It is well known that the LTE codebook-based feedback scheme is optimised primarily for single-user transmission and we can see that it performs very close to the unquantised feedback curve. We note in Fig. 2 that the addendum of information provided by the matrix $\mathbf{A}$ does not improve performance by much: the main reason, in this case, is that a single codeword is multiplexed on both spatial vectors signalled by a user such that the link adaptation needs to match their average channel gain. Hence, identifying the strongest eigenvector in this 2-dimensional subspace does not improve the precoder gain significantly. In both SU-MIMO parameter configurations the frequency selective feedback through the $\mathbf{B}$ matrix outperforms the LTE feedback scheme.

In Figs. 3 to 5 we consider a proper MU-MIMO transmission with a maximum of 2 users scheduled per RB and for RI=1,2,3, respectively. In all cases, the first stage of the CSI representation, by the matrix $\mathbf{B}$, is the most informative feedback by providing a much more accurate description of the MIMO channel structure than the LTE-type feedback and also by exploiting the channel frequency correlation. As we observed previously, the $\mathbf{B}$ feedback component provides the transmitter with interference rejection capability, whereas the $\mathbf{A}$ component helps maximising the channel gain for each user. However, when operating in MU-MIMO mode, the mismatch between the quantised directional channel feedback and the actual channel vectors causes residual interference between the scheduled users that cannot be predicted by either the transmitter or the receivers. An addition gain is provided by adding the $\mathbf{A}$ component, which is reflected in the small additional throughput we see in Figs. 4 to 6 when both feedback components are sent back.

Finally, in Fig. 6 we consider the case where all 4 users can be multiplexed together in the same RB, with RI=2. As in the previous configurations, the $\mathbf{B}$ component of the feedback offers a dramatic improvement in cell throughput compared to the LTE feedback.

From these results we note that the $\mathbf{B}$ component seem to have much greater impact in performance compared to the $\mathbf{A}$ component. This is in part motivated by the very coarse CSI representation provided by just 12 bits for a 4×4 system over a wide bandwidth, which impairs the accuracy of the refinement provided by the $\mathbf{A}$ component of the feedback.

VIII. Conclusion

In this paper we introduced a new method for encoding the MIMO CSI for frequency selective channels and limited feedback. The technique is based on the decomposition of the directional information of a MIMO channel in a two-coordinate system formed by the tangent space of a Grassmann manifold and the tangent space of a unitary group. The first component provides a non-redundant description of the subspace spanned by a set of channel eigenvectors, while the second component provides a non-redundant addendum of information to identify the strongest channel vector directions within that subspace. The first CSI feedback component, given by the $\mathbf{B}$ matrix, turns out to be the most important for MU-MIMO transmission because it conveys the information needed for to generate an interference-free transmission, while the second component, given by $\mathbf{A}$ matrix, improves the precoder gain for an individual user. One advantage of this feedback representation is that it allows to exploit the frequency correlation of selective MIMO channels. We showed that the two-stage feedback encoding and decoding are computationally inexpensive operations as they require each the equivalent complexity of one small SVD for either of the two components. We have run extensive simulations of an LTE downlink system and showed that the proposed CSI feedback description provides dramatic throughput improvement over the existing feedback scheme for MU-MIMO transmission.

Acknowledgement

The author would like to acknowledge the fruitful discussions with his colleagues at Toshiba Research Europe and the support of its directors.

References

[1] 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, 2003.

[2] P. Viswanath and D. N. C. Tse, "Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality," *IEEE Trans. Inf. Theory*, vol. 49, no. 8, pp. 1912–1921, 2003.

[3] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, 2003.

[4] W. Yu and J. M. Cioffi, "Sum capacity of Gaussian vector broadcast channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 1875–1892, 2004.

[5] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith, "Sum power iterative water-filling for multi-antenna Gaussian broadcast channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1570–1580, 2005.

[6] H. Weingarten, Y. Steinberg, and S. Shamai, "The capacity region of the Gaussian multiple-input multiple-output broadcast channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 3936–3964, 2006.

[7] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, 2004.

[8] T. Yoo and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 528–541, 2006.

[9] M. Stojnic, H. Vikalo, and B. Hassibi, "Rate maximization in multiantenna broadcast channels with linear preprocessing," in *Proc. 2004 IEEE Global Telecommunications Conf.*, vol. 6, pp. 3957–3961.

[10] C. B. Peel, B. M. Hochwald, and A. L. 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, 2005.

[11] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, 2008.

[12] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication—part II: perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537–544, 2005.

[13] J. H. Conway, R. H. Hardin, and N. J. A. Sloane, "Packing lines, planes, etc.: packings in Grassmannian space," *Experimental Mathematics*, vol. 5, no. 2, pp. 139–159, 1996.

[14] D. J. Love, J. Heath, R. W., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.

[15] A. Ashikhmin and R. Gopalan, "Grassmannian packings for efficient quantization in MIMO broadcast systems," in *Proc. 2007 IEEE Int. Symp. Inf. Theory*, pp. 1811–1815.

[16] I. S. Dhillon, R. W. Heath Jr., T. Strohmer, and J. A. Tropp, "Constructing packings in Grassmannian manifolds via alternating projection," *Experimental Mathematics*, vol. 17, no. 1, pp. 9–35, 2008.

[17] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5060, 2006.

[18] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran, "Multiuser MIMO achievable rates with downlink training and channel state feedback," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2845–2866, 2010.

[19] T. Pande, D. J. Love, and J. V. Krogmeier, "Reduced feedback MIMO-OFDM precoding and antenna selection," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2284–2293, 2007.

[20] H. Shirani-Mehr and G. Caire, "Channel state feedback schemes for multiuser MIMO-OFDM downlink," *IEEE Trans. Commun.*, vol. 57, no. 9, pp. 2713–2723, 2009.

[21] F. Boccardi, F. Tosato, and G. Caire, "Precoding schemes for the MIMO-GBC," in *Proc. 2006 Int Commun. Zurich Seminar*, pp. 10–13.

[22] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM J. Matrix Anal. Appl.*, vol. 20, no. 2, pp. 303–353, 1998.

[23] O. Henkel, "Sphere-packing bounds in the Grassmann and Stiefel manifolds," *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3445–3456, 2005.

[24] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd edition. The Johns Hopkins University Press, 1996.

[25] T. Chan, "An improved algorithm for computing the singular value decomposition," *ACM Trans. Math. Software*, vol. 8, no. 1, pp. 72–83, 1982.

[26] A. K. Jain, "A sinusoidal family of unitary transforms," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 4, pp. 356–365, 1979.

[27] Y. Linde, A. Buzo, and R. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, no. 1, pp. 84–95, 1980.

[28] S. Sesia, I. Toufik, and M. Baker, editors, *LTE, the UMTS Long Term Evolution*, 1st edition. Wiley, 2009.

[29] 3GPP TR 25.996, "Spatial channel model for MIMO simulations," 2009. Available: http://www.3gpp.org/ftp/Specs/archive/25_series/25.996/25996-900.zip

[30] P. Hoeher, S. Kaiser, and P. Robertson, "Two-dimensional pilot-symbol-aided channel estimation by Wiener filtering," in *Proc. 1997 IEEE Int. Acoust., Speech, Signal Process. Conf.*, vol. 3, pp. 1845–1848.

[31] 3GPP TR 36.211, "EUTRA physical channels and modulation," 2010. Available: http://www.3gpp.org/ftp/Specs/archive/36_series/36.211/36211-a00.zip

Filippo Tosato (S'99-M'04) received Laurea and Ph.D. degrees in telecommunications engineering from the University of Padova, Italy, in 2001 and 2004, respectively. He worked for Hewlett-Packard Research Laboratories, Bristol, UK, in summer 2001 and joined Philips Research Laboratories, Redhill, UK, in 2005, where he worked until 2008. Since then, he has been with the Telecommunications Research Lab of Toshiba Research Europe. His current research interests include communications theory and signal processing with a focus on multiple antenna systems and system analysis under realistic impairments.