CrossMark

A limited feedback scheme for massive MIMO systems based on principal component analysis

Tiankui Zhang[1*], Anmeng Ge[1], Norman C. Beaulieu[1], Zhirui Hu[2] and Jonathan Loo[3]

Abstract

Massive multiple-input multiple-output (MIMO) is becoming a key technology for future 5G cellular networks. Channel feedback for massive MIMO is challenging due to the substantially increased dimension of the channel matrix. This motivates us to explore a novel feedback reduction scheme based on the theory of principal component analysis (PCA). The proposed PCA-based feedback scheme exploits the spatial correlation characteristics of the massive MIMO channel models, since the transmit antennas are deployed compactly at the base station (BS). In the proposed scheme, the mobile station (MS) generates a compression matrix by operating PCA on the channel state information (CSI) over a long-term period, and utilizes the compression matrix to compress the spatially correlated high-dimensional CSI into a low-dimensional representation. Then, the compressed low-dimensional CSI is fed back to the BS in a short-term period. In order to recover the high-dimensional CSI at the BS, the compression matrix is refreshed and fed back from MS to BS at every long-term period. The information distortion of the proposed scheme is also investigated and a closed-form expression for an upper bound to the normalized information distortion is derived. The overhead analysis and numerical results show that the proposed scheme can offer a worthwhile tradeoff between the system capacity performance and implementation complexity including the feedback overhead and codebook search complexity.

Keywords: Massive MIMO, Limited feedback, Principal component analysis, Information distortion analysis

1 Introduction

The massive multiple-input multiple-output (MIMO) system which deploys large numbers of transmit antennas at the base station (BS) has been listed as one of the key techniques for fifth generation (5G) cellular networks [1]. The deployment of numerous antennas enables massive MIMO systems to achieve not only higher system capacity, but also higher spectrum and energy efficiency than conventional MIMO systems [2, 3].

The superior performance of the massive MIMO systems relies on the spatial multiplexing and the minor multi-user interference. As is the case for conventional MIMO systems, this in turn requires the BS to have perfect knowledge of the downlink channel state information

(CSI) [4]. In a time division duplexing (TDD) system, the channel reciprocity can be exploited to acquire the downlink CSI at the BS [5]. However, things become more challenging when the system operates in a frequency division duplexing (FDD) mode, where the channel reciprocity no longer holds. Therefore, a mobile station (MS) needs to feedback the downlink CSI through a rate-limited uplink channel. The authors in [6] drew the conclusion that the required feedback rate per user should be increased in proportion to the number of the transmit antennas for the sake of obtaining the full multiplexing gain. Therefore, feedback overhead turns into a key challenge in the massive MIMO systems.

The foundation of the works on feedback overhead reduction for MIMO systems is the correlation feature of MIMO channels. Limited feedback techniques for correlated MIMO channels were designed in [7–9]. A modified Grassmannian line packing codebook was proposed in [7], and the authors in [8, 9] rotated the codebook for i.i.d. channels with a unitary matrix to obtain the codebook for

*Correspondence: zhangtiankui@bupt.edu.cn

[1] Beijing Key Laboratory of Network System Architecture and Convergence, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, No. 10, Xitucheng Road, Haidian District, 100876 Beijing, China

Full list of author information is available at the end of the article

correlated MIMO channels. A systematic codebook was designed for quantized beamforming in [10], which was implemented by maps that can rotate and scale spherical caps on the Grassmannian manifold.

Furthermore, a codebook for uniform rectangular arrays (URA) for massive MIMO antennas was designed in [11, 12]. It was derived by the Kronecker product of two ULA codebooks. The authors in [13] proposed a feedback framework for FDD massive MIMO systems that divides the coverage area into sub-sectors, where each sub-sector is formed by a set of narrow beams that covers a pre-assigned area in azimuth and elevation. Non-coherent trellis-coded quantization and trellis-extended codebooks for massive MIMO systems were proposed in [14, 15], which exploited a Viterbi decoder for CSI quantization and a convolutional encoder for CSI reconstruction. A projection based feedback compression was utilized to project the high-dimensional channel space into a lower dimensional subspace [16]. However, [16] did not explain how to feedback the projection matrix.

The compressive sensing (CS)-based limited feedback schemes for massive MIMO were proposed to reduce the feedback overhead by exploiting the spatial correlation of CSI [17–20]. The authors in [17] introduced CS to massive MIMO for limited feedback. A unique insight was provided that strong spatial correlations are exhibited in massive closely-packed antenna arrays, so channel vectors can be represented in sparse form in the spatial-frequency domain. Subsequently, a compressed analog feedback strategy for spatially correlated massive MIMO channels was proposed in [18]. In contrast to the strategy in [18], the low-dimensional CSI was quantized with a codebook and the preferred index was fed back in [19] and [20].

The choice of orthogonal basis, which is intended for the sparse representation of the original signal, plays an important role in the recovery of the original high-dimensional signal at the BS. Two such kinds of orthogonal basis construction, the discrete cosine transform (DCT) and the Karhunen-Loeve transform (KLT), are usually employed [21]. If the channel correlation matrix is neither known at the MS nor the BS, the signal-independent DCT basis is a better option. On the one hand, because of its signal-independent nature, the utilization of the DCT basis does not require the MS to inform the BS of the channel correlation matrix. On the other hand, this makes the DCT basis incapable of tracking the real-time change of channel state, which has a negative effect on system capacity. In contrast to the DCT basis, the KLT basis can excellently adapt to CSI change. Therefore, when MS and BS both know the instantaneous channel correlation matrix, the KLT basis can provide the optimal sparse representation, which promises accurate recovery even if only a small number of measurements

are available. Unfortunately, the signal-dependent nature of the KLT basis requests the MS to feedback channel correlation matrix instantaneously [22]. This can hardly be implemented in practical systems because of the heavy feedback overhead.

In this case, principal component analysis (PCA) can offer a tradeoff between system capacity and practical implementation [23, 24]. Compared with a DCT basis, PCA can be more adaptive to the change of the channel state, since PCA is signal-dependent [23]. This guarantees PCA better system capacity than a DCT basis. Compared with a KLT basis, PCA only needs that the MS and the BS have knowledge of the channel correlation matrix in a long-term period. This makes PCA achieve feedback overhead reduction much better than a KLT basis. What is more, the most attractive characteristic of PCA is that it is effective for dimensionality reduction of high-dimensional data [24], whose elements are correlated. Inspired by this, PCA has great potential to be applied to the compression of high-dimensional CSI with strong spatial correlation to reduce feedback overhead in massive MIMO systems. To the best of our knowledge, there have not existed any works addressing a practical feedback scheme based on PCA.

This paper proposes a PCA-based feedback scheme for massive MIMO systems. In the proposed scheme, the MS utilizes a compression matrix, which is obtained by operating PCA on CSI observed over a long-term period, to compress spatially correlated high-dimensional CSI into low-dimensional representation. After quantizing the low-dimensional CSI with a random vector quantization (RVQ) codebook, the index of the preferred codeword is fed back to the BS in each short-term period. In order to track the channel changes and enable the BS to recover the high-dimensional CSI, it is necessary for the MS to refresh and feedback the compression matrix at every long-term period. Through the dimensionality reduction processing by PCA, feedback overhead and codebook search complexity can be reduced. The contributions of the paper are summarized as follows.

- A PCA-based feedback scheme for FDD massive MIMO systems is proposed. The operation procedures at the BS and the MS are divided into two types, which are long-term period operations and short-term period operations. In more detail, the exact operation procedures both at the BS and the MS, as well as the derivation of the compression matrix at the MS, are presented. The distortion of the proposed scheme is analyzed. An upper bound to the normalized distortion is derived.
- System performance comparisons of the PCA-based feedback scheme, the DCT-based CS scheme and the KLT-based CS scheme are presented. The feedback

Zhang *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:64

Page 3 of 12

overhead and the codebook search complexity are analyzed and the system capacity performance is simulated. Looking at the simulation results and the feedback overhead analysis comprehensively, we draw the conclusion that our proposed scheme can achieve a compromise between system capacity and implementation complexity (feedback overhead and codebook search complexity).

The remainder of this paper is organized as follows. In Section 2, the massive MIMO system model is described. Section 2 first reviews the PCA method itself in Subsection and then provides the details of the proposed scheme in Subsection . Moreover, distortion of the proposed scheme is analyzed in Subsection . The feedback overhead as well as codebook search complexity comparison and numerical results follow in Section 4 and Section 5, respectively. Finally, the conclusion of this paper is presented in Secton 6.

Notation: Throughout this paper, upper and lower case boldfaces are used to describe matrix $\mathbf{A}$ and vector $\mathbf{a}$, respectively. We denote the transpose and the conjugate transpose of matrix $\mathbf{A}$ or vector $\mathbf{a}$ by $\mathbf{A}^T (\mathbf{a}^T)$ and $\mathbf{A}^H (\mathbf{a}^H)$. In addiction, $\mathbf{A}^{-1}$ denotes the inverse of a square matrix.

2 System model

We consider a downlink massive MIMO system, where there is a single cell, in which the BS equipped with N_t antennas serves K single-antenna MSs.

2.1 Spatially correlated massive MIMO channel

A massive MIMO broadcast channel is modeled in this section. For simplicity, but without loss of generality, a large-scale uniform linear array (ULA) with an enormous number of antenna elements deployed compactly is assumed. The spatial correlations are exhibited in the massive MIMO channel model, because of the insufficient inter-element spacing. Additionally, a poor scattering environment may also contribute to the spatial correlation. Different from the previous works, which only consider either insufficient inter-element spacing or poor scattering environment, this paper combines the well-known Kronecker correlation model [25] with the geometrical one-ring model [26, 27], so as to describe the properties of the spatial correlation of the massive MIMO channel more precisely. Since the MS is equipped with a single antenna, the channel between the kth MS and BS is denoted by a $1 \times N_t$ row vector $\mathbf{h}_k$ $(k = 1, 2, \ldots, K)$. Based on the Kronecker correlation model, $\mathbf{h}_k$ can be modeled as

$$\mathbf{h}_k = \mathbf{h}_{\text{one-ring}} \mathbf{R}_{\text{Tx}}^{\frac{1}{2}}, \tag{1}$$

where $\mathbf{R}_{\text{Tx}}^{\frac{1}{2}}$ is the square root of the correlation matrix at the transmitter depicting the impact of insufficient

inter-element spacing and $\mathbf{h}_{\text{one-ring}}$ is derived from the one-ring model describing the spatial correlation caused by a scattering environment. Note that the correlation of the channel is time varying, due to the change of both the relative positions of scatterers and the correlation matrix at the transmitter.

In more detail, the uth row and the vth column entry of $\mathbf{R}_{\text{Tx}}$ (the correlation coefficient between the uth and the vth elements within the BS transmit antenna array) obeys the zeroth-order Bessel function of the first kind correlation model [18], that is

$$r_{uv} = J_0 \left(\frac{2\pi d_{uv}}{\lambda} \right), \tag{2}$$

where d_{uv} is the distance between the two antenna elements and λ denotes the carrier wavelength.

As to the one-ring model, we assume that each MS is surrounded by Q scatterers, which are uniformly distributed on a circle with the radius r, as shown in Fig. 1. The $\mathbf{h}_{\text{one-ring}}$ can be given as follows [27],

$$\mathbf{h}_{\text{one-ring}} \triangleq \frac{1}{\sqrt{Q}} \sum_{q=1}^{Q} \mathbf{h}_{kq}. \tag{3}$$

In (3), $\mathbf{h}_{kq}$ is the channel vector of the MS k over the qth scattering path, as given by

$$\mathbf{h}_{kq} \triangleq \left[\sqrt{\beta_{kq1}} e^{-j2\pi \frac{d_{kq1}+r}{\lambda}}, \ldots, \sqrt{\beta_{kqN_t}} e^{-j2\pi \frac{d_{kqN_t}+r}{\lambda}} \right] e^{j\varphi_{kq}}, \tag{4}$$

where d_{kqm} is the distance between the qth scatterer of the kth MS and the mth $(m = 1, 2 \ldots, N_t)$ BS antenna, while $d_{kqm} + r$ denotes the path length from the kth MS to the mth antenna via the qth path. Also, $e^{j\varphi_{kq}}$ represents the random common phase resulting from either the random perturbations of the MS location or the phase shift due to
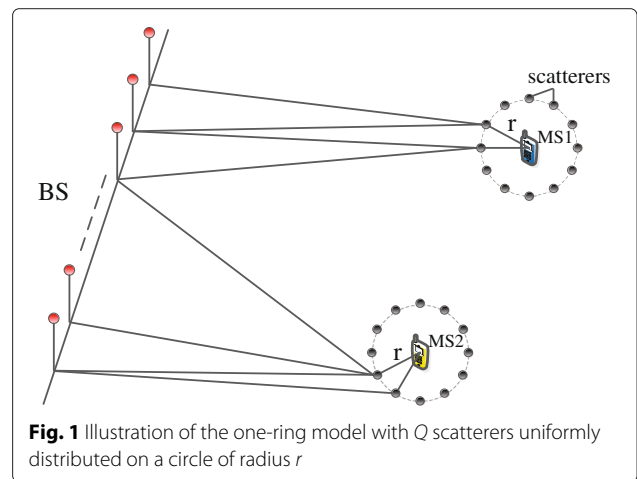

Fig. 1 Illustration of the one-ring model with Q scatterers uniformly distributed on a circle of radius r

Zhang *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:64

Page 4 of 12

the reflection of the scatterer, and β_{kqm} denotes the path loss of the qth scattering path, which is modeled by

$$\beta_{kqm} = \frac{\alpha}{(d_{kqm} + r)^{\gamma}}, \tag{5}$$

where α is a constant and γ is the path loss exponent.

2.2 Downlink signal model
In the downlink transmission, $s_k \in \mathbb{C}$ and $\mathbf{w}_k \in \mathbb{C}^{N_t \times 1}$ denote the transmit signal with power constraint $\mathbb{E}|s_k|^2 = 1$ and the column precoding vector intended for the kth MS, respectively. In this paper, zero-forcing precoding is adopted to eliminate multiuser interference [28]. Also, let n_k be additive Gaussian noise with zero mean and unit variance at the MS k. Then the received signal of the kth MS can be expressed as

$$y_k = \underbrace{\sqrt{\frac{P_t}{K}} \mathbf{h}_k \mathbf{w}_k s_k}_{\text{desired signal}} + \underbrace{\sum_{k' \neq k} \sqrt{\frac{P_t}{K}} \mathbf{h}_k \mathbf{w}_{k'} s_{k'} + n_k}_{\text{interfering signal and noise}}, \tag{6}$$

where P_t is the total transmit power of the BS. Equal power allocation is assumed with $\frac{P_t}{K}$ being the power distributed to each MS.

As seen in (6), y_k contains two main terms. The first term is the desired signal, while the other is the interfering signal and noise. From (6), we can derive the system capacity as

$$C = \sum_{k=1}^{K} \log_2 \left(1 + \frac{\frac{P_t}{K} |\mathbf{h}_k \mathbf{w}_k|^2}{\frac{P_t}{K} \sum_{k' \neq k} |\mathbf{h}_k \mathbf{w}_{k'}|^2 + 1} \right). \tag{7}$$

3 Feedback scheme for massive MIMO
3.1 Review of principal component analysis
We suppose that there are a data samples, each of which contains b characteristics. The b characteristics have complicated correlation relationships with each other, which makes it possible for dimensionality reduction with PCA. For convenience of description, let the $a \times b$ matrix $\mathbf{X}$ denote the original data containing the a data samples. The key point of PCA is how to derive a $b \times l$ ($l < b$) compression matrix $\bar{\boldsymbol{\Psi}}$, which is utilized to compress the high-dimensional data $a \times b$ $\mathbf{X}$ into a low-dimensional $a \times l$ $\bar{\mathbf{X}}$ as follows,

$$\bar{\mathbf{X}} = \mathbf{X} \bar{\boldsymbol{\Psi}}, \tag{8}$$

in which, $\bar{\boldsymbol{\Psi}}$ is composed of l-dominating eigenvectors, the so-called principal components, which are selected from all b eigenvectors of $\mathbf{X}$.

For the sake of determining which components are to be selected, the concept of contribution rate is introduced. Consider a descending ordering of the b eigenvalues $\lambda_1, \lambda_2 \ldots, \lambda_b$. Then, the contribution rate of the gth

eigenvalue λ_g is defined as $\frac{\lambda_g}{\sum_{g=1}^{b} \lambda_g}$, while the cumulative contribution rate of the top l eigenvalues can be expressed by $\frac{\sum_{g=1}^{l} \lambda_g}{\sum_{g=1}^{b} \lambda_g}$. Generally, when the cumulative contribution rate of the chosen l principal components exceeds a certain level, the information loss is acceptable.

Finally, the original $\mathbf{X}$ can be recovered from $\bar{\mathbf{X}}$ by

$$\hat{\mathbf{X}} = \bar{\mathbf{X}} \bar{\boldsymbol{\Psi}}^H. \tag{9}$$

3.2 Proposed PCA-based feedback scheme
A PCA-based feedback scheme for massive MIMO is proposed in this subsection. In the proposed scheme, different operations at the MS and the BS have different time periods, long-term period T_l and short-term period T_s. Every long-term period T_l contains several short-term periods T_s. In every T_s, the MS utilizes the compression matrix to compress high-dimensional CSI into low-dimensional representation. Then, the compressed low-dimensional CSI is quantized by the RVQ codebook and the index of the preferred codeword is fed back to the BS. Because of the signal-dependent nature of PCA, the compression matrix is derived by executing PCA on the CSI which is obtained through continuous channel estimation during a whole long-term period.

3.2.1 Compression matrix derivation
First of all, the detailed procedure for deriving the compression matrix in the nth long-term period $T_l^{(n)}$ with the PCA method is given in Table 1. We assume MS k can obtain S high-dimensional channel vectors $\left(\mathbf{h}_k^{(n,1)}, \mathbf{h}_k^{(n,2)} \ldots, \mathbf{h}_k^{(n,S)} \right)$ through ideal channel estimation in $T_l^{(n)}$. Here, the S channel vectors can be viewed as S data samples, each of which contains N_t characteristics. In order to compress the high-dimensional CSI (channel vectors), we choose M ($M \ll N_t$) dominating eigenvectors to compose the compression matrix $\bar{\mathbf{U}}^{(n)} \in \mathbb{C}^{N_t \times M}$.

The compression matrix obtained in the long-term period $T_l^{(n)}$ will be used by the MS to compress $1 \times N_t$ channel vectors into $1 \times M$ vectors, as well as by the BS to perform recovery in the period $T_l^{(n+1)}$.

3.2.2 MS operation
The main operations at the MS can be classified into two types: long-term period operations and short-term period operations.

In the long-term period $T_l^{(n)}$, the MS performs the continuous channel estimation to obtain S high-dimensional channel vectors, and derives the compression matrix $\bar{\mathbf{U}}^{(n)}$. Then, each column of the compression matrix is quantized by another RVQ codebook. After quantization, the compression matrix $\bar{\mathbf{U}}^{(n)}$ is fed back to the BS at the end of $T_l^{(n)}$.

Zhang *et al. EURASIP Journal on Advances in Signal Processing*   (2016) 2016:64

Page 5 of 12

Table 1 Procedure of deriving the compression matrix in each long-term period with the PCA method at the MS

Initialization	
S:	Number of the short-term period T_s in every long-term period T_l.
M:	Number of the dominating eigenvectors chosen to form the compression matrix.
Procedures	
1)	Channel estimation: perform channel estimation continuously in $T_l^{(n)}$ to obtain S high-dimensional channel vectors $\mathbf{h}_k^{(n,1)}, \mathbf{h}_k^{(n,2)} \dots, \mathbf{h}_k^{(n,S)}$.
2)	Computation of covariance matrix: $\tilde{\mathbf{H}}^{(n)} = \left[\left(\mathbf{h}_k^{(n,1)} \right)^H, \left(\mathbf{h}_k^{(n,2)} \right)^H \dots, \left(\mathbf{h}_k^{(n,S)} \right)^H \right]^H$ $\in \mathbb{C}^{S \times N_t}$ and compute its covariance matrix.
3)	Eigen-decomposition: perform eigen-decomposition on the above covariance matrix $\mathrm{Cov}\left(\tilde{\mathbf{H}}^{(n)}, \tilde{\mathbf{H}}^{(n)} \right) = \mathbf{U}^{(n)} \mathbf{D}^{(n)} \left(\mathbf{U}^{(n)} \right)^H$.
4)	Formation of compression matrix: choose M dominating eigenvectors to form compression matrix $\bar{\mathbf{U}}^{(n)} \in \mathbb{C}^{N_t \times M}$.

Operation in the sth ($s = 1, 2 \dots, S$) short-term period of the nth long-term period, $T_s^{(n,s)}$, is described as follows:

Step 1. Channel estimation is performed to obtain a $1 \times N_t$ channel vector $\mathbf{h}_k^{(n,s)}$.

Step 2. Multiply $\mathbf{h}_k^{(n,s)}$ by the compression matrix derived in the previous long-term period

$$\bar{\mathbf{h}}_k^{(n,s)} = \mathbf{h}_k^{(n,s)} \bar{\mathbf{U}}^{(n-1)}. \tag{10}$$

By this step, the original high-dimensional CSI ($1 \times N_t$) is compressed into a low-dimensional representation ($1 \times M$). The compression ratio is $\frac{M}{N_t}$.

Step 3. Quantize the low-dimensional CSI $\bar{\mathbf{h}}_k^{(n,s)}$ by RVQ codebook and obtain the index number of the codeword that best fits $\bar{\mathbf{h}}_k^{(n,s)}$, that is

$$j^{(n,s)} = \arg\max_j \left| \bar{\mathbf{h}}_k^{(n,s)} \mathbf{c}_j \right|, \tag{11}$$

where $\mathbf{c}_j$ is the jth codeword of the codebook. Compared with the quantizing high-dimensional CSI directly, the RVQ codebook used above can be designed to be much smaller. This not only reduces the feedback overhead, but also decreases the codebook search complexity.

Step 4. The index of the preferred codeword is fed back to the BS.

3.2.3 BS operation

Similarly, the main operation at the BS can also be classified into long-term period operation and short-term period operation. At the end of $T_l^{(n)}$, the BS receives the compression matrix $\bar{\mathbf{U}}^{(n)}$ to perform high-dimensional CSI recovery in the next long-term period $T_l^{(n+1)}$. Meanwhile, the short-term period operation follows the steps below:

Step 1. The codeword index $j^{(n,s)}$ is received in each short-term period;

Step 2. As the BS and the MS share the same codebook, it is easy for the BS to find the quantized low-dimensional CSI $\hat{\mathbf{h}}_k^{(n,s)}$ by letting $\hat{\mathbf{h}}_k^{(n,s)} = \mathbf{c}_{j^{(n,s)}}$;

Step 3. The high-dimensional CSI $\widehat{\mathbf{h}}_k^{(n,s)}$ can be recovered by

$$\widehat{\mathbf{h}}_k^{(n,s)} = \hat{\mathbf{h}}_k^{(n,s)} \left(\bar{\mathbf{U}}^{(n-1)} \right)^H, \tag{12}$$

where $\bar{\mathbf{U}}^{(n-1)}$ is derived from the period $T_l^{(n-1)}$.

3.3 Distortion analysis of proposed scheme

The distortion of our proposed scheme consists of three components, the distortion resulting from the PCA processing, the low-dimensional CSI $\bar{\mathbf{h}}$ quantization and the compression matrix $\bar{\mathbf{U}}$ quantization. To facilitate the distortion analysis below, different representations of CSI in different stages are enumerated in Table 2.

3.3.1 Distortion analysis of PCA

First, we consider the distortion caused by the PCA method itself, with no quantization errors resulting from low-dimensional CSI or the compression matrix taken into account. That is, we measure the mean square error between $\mathbf{h}$ and $\tilde{\mathbf{h}}$, where $\tilde{\mathbf{h}} = \mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H$. In this paper, $\mathbf{h}$ is a N_t dimensional row vector, which can be viewed as a point in the N_t dimensional space. Therefore, $\mathbf{h}$ can be

Table 2 Different representations of CSI in different stages

Representations	Implications
$\mathbf{h}$	Original high-dimensional CSI estimated by MS
$\bar{\mathbf{h}}$	Low-dimensional CSI after normalization and compression, that is $\bar{\mathbf{h}} = \frac{\mathbf{h}\bar{\mathbf{U}}}{\|\mathbf{h}\bar{\mathbf{U}}\|}$
$\hat{\mathbf{h}}$	Quantized low-dimensional CSI
$\widehat{\mathbf{h}}$	High-dimensional CSI recovered from $\hat{\mathbf{h}}$
$\tilde{\mathbf{h}}$	High-dimensional CSI recovered from $\bar{\mathbf{h}}$ (with no quantization error of low-dimensional CSI)

expressed by the linear combination of a set of orthogonal basis vectors, $\mathbf{u}_i$

$$\mathbf{h} = \sum_{i=1}^{N_t} \alpha_i \mathbf{u}_i, \qquad (13)$$

where $\mathbf{u}_i$ denotes the ith basis vector.

In the PCA method, the high-dimensional CSI $\mathbf{h}$ is compressed into low-dimensional (M-dimensional) $\bar{\mathbf{h}}$, the components of which are derived by projecting $\mathbf{h}$ onto the M dominating bais vectors. Given this, the reconstructed high-dimensional CSI $\tilde{\mathbf{h}}$ can be modeled as the combination of M dominating basis vectors and the other $N_t - M$ less dominating vectors,

$$\tilde{\mathbf{h}} = \sum_{i=1}^{M} z_i \mathbf{u}_i + \sum_{i=M+1}^{N_t} b_i \mathbf{u}_i. \qquad (14)$$

Therefore, the information distortion caused by the PCA itself, which is defined as the mean square error between $\mathbf{h}$ and $\tilde{\mathbf{h}}$, can be expressed by

$$J = \frac{1}{S} \sum_{n=1}^{S} \left\| \mathbf{h}_n - \tilde{\mathbf{h}}_n \right\|^2, \qquad (15)$$

where S denotes the number of short-term periods in a long-term period.

Proposition. The PCA-caused information distortion J can be expressed by a linear sum of $N_t - M$ less dominating eigenvalues of the channel covariance matrix.

$$J = \sum_{i=M+1}^{N_t} \lambda_i.$$

$\square$

Proof. See Appendix A. $\square$

3.3.2 *Distortion analysis of quantization*
According to [6], to measure the quantization error, $\bar{\mathbf{h}}$ can be modeled as

$$\bar{\mathbf{h}} = \sqrt{1 - d^2}\hat{\mathbf{h}} + d\mathbf{e}, \qquad (16)$$

where $\hat{\mathbf{h}}$ is the quantization of $\bar{\mathbf{h}}$ and $\mathbf{e}$ is a unit norm vector isotropically distributed in the nullspace of $\hat{\mathbf{h}}$. Parameter d denotes the quantization error independent of $\mathbf{e}$, which satisfies $\mathbb{E}\left[d^2\right] \leq 2^{-\frac{B_1}{M-1}}$. Here, M represents the number of principal components and B_1 is the number of feedback bits of $\bar{\mathbf{h}}$. Similarly, $\bar{\mathbf{U}}$ can be modeled as

$$\bar{\mathbf{U}} = \sqrt{1 - D^2}\hat{\mathbf{U}} + D\mathbf{E}, \qquad (17)$$

where $\hat{\mathbf{U}}$ is the quantization of $\bar{\mathbf{U}}$ and $\mathbf{E}$ is composed of M unit norm vectors isotropically distributed in the

nullspace of $\hat{\mathbf{U}}$. Moreover, quantization error D is independent of $\hat{\mathbf{U}}$ satisfying

$$\mathbb{E}\left[D^2\right] \leq 2^{-\frac{B_2/M}{N_t-1}}, \qquad (18)$$

where B_2 denotes the number of feedback bits of the compression matrix.

Before analyzing the distortion between the original high-dimensional CSI $\mathbf{h}$ and the reconstructed $\widehat{\mathbf{h}}$, we first focus on how to express $\widehat{\mathbf{h}}$ in terms of $\mathbf{h}$.

Proposition. The reconstructed high-dimensional CSI $\widehat{\mathbf{h}}$ can be expressed in terms of $\mathbf{h}$ as

$$\widehat{\mathbf{h}} = \frac{\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H - D^2 \cdot \mathbf{h}\mathbf{I}_M - \left\|\mathbf{h}\bar{\mathbf{U}}\right\| \cdot d\sqrt{1-D^2} \cdot \mathbf{e}\hat{\mathbf{U}}^H}{\sqrt{1-d^2} \cdot \sqrt{1-D^2}}, \qquad (19)$$

where $\mathbf{I}_M = \begin{bmatrix} \mathbf{I}_{M \times M} & \mathbf{0}_{M \times (N_t - M)} \\ \mathbf{0}_{(N_t - M) \times M} & \mathbf{0}_{(N_t - M) \times (N_t - M)} \end{bmatrix}$. $\square$

Proof. See Appendix B. $\square$

Having derived an expression for $\widehat{\mathbf{h}}$ in terms of $\mathbf{h}$, our purpose is to analyze the distortion of the proposed scheme. We derive an upper bound to the normalized distortion (denoted by δ) between $\widehat{\mathbf{h}}$ and $\mathbf{h}$. Instead of calculating δ directly, we first calculate the normalized similarity (denoted by ρ) between $\widehat{\mathbf{h}}$ and $\mathbf{h}$. Then, δ can be conveniently obtained by $\delta = 1 - \rho$.

Before calculating the normalized similarity ρ between $\widehat{\mathbf{h}}$ and $\mathbf{h}$, it is insightful to look at the non-normalized similarity $\mathbb{E}\left[\left|\mathbf{h}\widehat{\mathbf{h}}^H\right|\right]$.

Proposition. A lower bound to the non-normalized similarity between $\widehat{\mathbf{h}}$ and $\mathbf{h}$ is given by

$$\mathbb{E}\left[\left|\mathbf{h}\widehat{\mathbf{h}}^H\right|\right] \geq A - M \cdot 2^{-\frac{B_2/M}{N_t-1}} - A \cdot 2^{-\frac{B_1}{M-1}}, \qquad (20)$$

where $A = \left(N_t - \sum_{i=M+1}^{N_t} \lambda_i\right)$. $\square$

Zhang *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:64

Page 7 of 12

Proof. Since we have derived an expression for $\widehat{\mathbf{h}}$ in terms of $\mathbf{h}$, as given in Proposition 2, then the non-normalized similarity can be expressed by

$$
\begin{aligned}
\mathbb{E}&\left[\left|\mathbf{h}\widehat{\mathbf{h}}^H\right|\right] \\
&= \mathbb{E}\left[\left|\mathbf{h}\left(\frac{\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H - D^2\cdot\mathbf{h}\mathbf{I}_M - \|\mathbf{h}\bar{\mathbf{U}}\|\cdot d\sqrt{1-D^2}\cdot\mathbf{e}\hat{\mathbf{U}}^H}{\sqrt{1-d^2}\cdot\sqrt{1-D^2}}\right)^H\right|\right] \quad (21)\\
&= \mathbb{E}\left[\left|\frac{\|\mathbf{h}\bar{\mathbf{U}}\|^2 - D^2\cdot\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H - \|\mathbf{h}\bar{\mathbf{U}}\|\cdot d\cdot\mathbf{h}(\bar{\mathbf{U}}-D\mathbf{E})\mathbf{e}^H}{\sqrt{1-d^2}\cdot\sqrt{1-D^2}}\right|\right].
\end{aligned}
$$

Now, because $d^2 < 1$ and $D^2 < 1$, $\sqrt{1-d^2}$ and $\sqrt{1-D^2}$ are also smaller than 1. Additionally, since the quantization of low-dimensional CSI and the compression matrix are independent, $\mathbb{E}\left[\mathbf{E}\mathbf{e}^H\right] = 0$. Therefore, the non-normalized similarity can be bounded by (22), shown at the top of the next page.

$$
\begin{aligned}
\mathbb{E}\left[\left|\mathbf{h}\widehat{\mathbf{h}}^H\right|\right] &\geq \mathbb{E}\left[\left|\|\mathbf{h}\bar{\mathbf{U}}\|^2 - D^2\cdot\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H - \|\mathbf{h}\bar{\mathbf{U}}\|\cdot d\cdot\mathbf{h}\left(\bar{\mathbf{U}}-D\mathbf{E}\right)\mathbf{e}^H\right|\right] \\
&\geq \mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2 - D^2\cdot\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H - \|\mathbf{h}\bar{\mathbf{U}}\|\cdot d\cdot\mathbf{h}\left(\bar{\mathbf{U}}-D\mathbf{E}\right)\mathbf{e}^H\right] \\
&= \mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2 - D^2\cdot\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H - \|\mathbf{h}\bar{\mathbf{U}}\|\cdot d\cdot\mathbf{h}\bar{\mathbf{U}}\mathbf{e}^H\right] \\
&= \mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2 - D^2\cdot\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H - \|\mathbf{h}\bar{\mathbf{U}}\|^2 d^2\right] \\
&= \mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2\right] - \mathbb{E}\left[D^2\right]\cdot\mathbb{E}\left[\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H\right] - \mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2\right]\cdot\mathbb{E}\left[d^2\right]
\end{aligned}
$$
$$(22)$$

Further, we take advantage of the equations

$$
\mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2\right] = N_t - J, \tag{23}
$$

$$
\mathbb{E}\left[\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H\right] = M, \tag{24}
$$

which are proved in Appendix C. Moreover, the upper boundary of $\mathbb{E}\left[D^2\right]$ and $\mathbb{E}\left[d^2\right]$ are $2^{-\frac{B_2/M}{N_t-1}}$ and $2^{-\frac{B_1}{M-1}}$, respectively. Consequently, we obtain (20). □
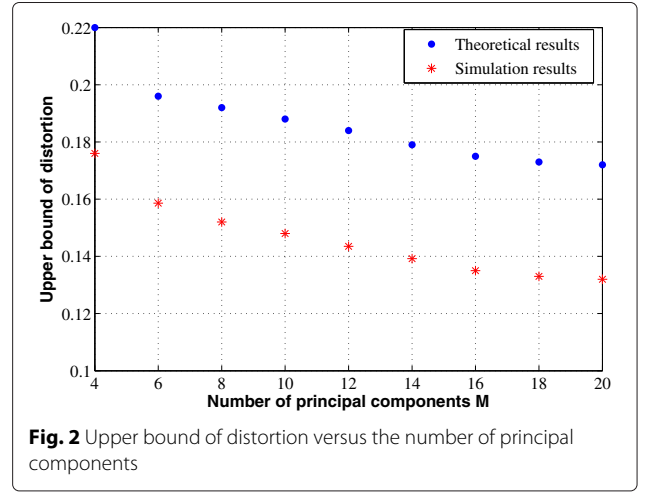
From (23), we can observe that when there is no distortion caused by PCA ($J = 0$), as well as no quantization error from low-dimensional CSI ($2^{-\frac{B_1}{M-1}} \to 0$) or compression matrix ($2^{-\frac{B_2/M}{N_t-1}} \to 0$), the maximum value of the lower bound reaches N_t. So the normalized similarity can be expressed as

$$
\rho \geq \frac{A - M\cdot 2^{-\frac{B_2/M}{N_t-1}} - A\cdot 2^{-\frac{B_1}{M-1}}}{N_t}. \tag{25}
$$

Finally, the upper bound of the normalized distortion of our proposed scheme can be obtained,

$$
\delta = 1 - \rho \leq 1 - \frac{A - M\cdot 2^{-\frac{B_2/M}{N_t-1}} - A\cdot 2^{-\frac{B_1}{M-1}}}{N_t}. \tag{26}
$$

According to the expression (26), we give the theoretical upper bound of the distortion and the simulated distortion in Fig. 2.

Fig. 2 Upper bound of distortion versus the number of principal components

4 Implementation complexity analysis

This section analyzes the feedback overhead and the codebook search complexity of the proposed scheme. For comparison, the existing CS-based schemes utilizing the KLT basis and the DCT basis are also taken into account.

The number of feedback bits per user increases linearly with the number of transmit antennas, as modeled by [6]

$$
B = (N_t - 1)\log_2\rho \approx \frac{N_t - 1}{3}\rho, \tag{27}
$$

where ρ denotes the received signal-to-noise ratio (SNR) in decibels at the MS.

Consider a long-term period containing S short-term periods. When the proposed scheme is adopted, the number of feedback bits can be represented by

$$
B_{\text{pro}} = S\frac{(M-1)\rho}{3} + M\frac{(N_t-1)\rho}{3}, \tag{28}
$$

where M is the number of the principal components, the first term denotes the number of feedback bits for quantizing low-dimensional CSI, and the second term is caused by the quantization of the compression matrix.

For the DCT-based CS scheme, there is no need for the MS to inform the BS of the channel correlation matrix, due to the signal-independent nature of the DCT basis. So the number of feedback bits for DCT-based CS is given by $B_{\text{DCT}} = S\cdot\frac{(M-1)\rho}{3}$.

In the case of the KLT-based CS scheme, the number of feedback bits dramatically increases. In order to feedback the complete correlation matrix in every short-term period, the second term of Eq. (28) needs to be modified to express the number of feedback bits of the KLT-based CS scheme, which is given by

$$B_{\text{KLT}} = S \cdot \frac{(M-1)\,\rho}{3} + S \cdot N_{\text{t}} \cdot \frac{(N_{\text{t}}-1)\,\rho}{3}. \tag{29}$$

As for the codebook search complexity, it is proportional to the number of conjugate multiplications when searching for the best codeword. So the search complexity of the proposed scheme, DCT-based CS and KLT-based scheme in a long-term period can be expressed as $S \cdot M \cdot 2^{\frac{(M-1)\rho}{3}} + M \cdot N_{\text{t}} \cdot 2^{\frac{(N_{\text{t}}-1)\rho}{3}}$, $S \cdot M \cdot 2^{\frac{(M-1)\rho}{3}}$ and $S \cdot M \cdot 2^{\frac{(M-1)\rho}{3}} + S \cdot N_{\text{t}}^2 \cdot 2^{\frac{(N_{\text{t}}-1)\rho}{3}}$, respectively.

Table 3 illustrates the comparison in detail. Based on the analysis above, we can observe that the number of feedback bits and the codebook search complexity of the proposed scheme falls in between the DCT-based and KLT-based CS schemes.

5   Simulation results

In this section, we present simulation results. A single cell scenario is considered, where the BS deploys a uniform linear array with $N_{\text{t}} = 128$ antennas serving $K = 6$ single-antenna MSs. Table 4 lists the detailed simulation parameters.

5.1   Feasibility validation

We verify whether PCA can be utilized to compress spatially correlated high-dimensional CSI into low-dimensional representation. To achieve this purpose, we simulate the eigenvalue distribution of the spatially correlated channels defined in (1). As shown in Fig. 3, the eigenvalue distribution of the spatially correlated channel is far from uniform. The eigenvalues are sorted by their contribution rate in descending order. The contribution rate of the biggest eigenvalue exceeds 50 %, while the fourth biggest eigenvalue only contributes 5.9 %. The cumulative contribution rate of the top four eigenvalues exceeds 95 %. We can conclude that the spatially correlated channel vectors can be expressed by several principal components with low information distortion.

Table 4 Simulation parameters

Parameters	Assumption
Antenna configuration of BS	ULA 0.5λ spaced
Feedback channel	Lossless & without delay
Carrier frequency	2.6 GHz
Bandwidth	10 MHz
Cell radius	200 m
Short-term period	1 ms
Long-term period	10 ms
Radius of scatterer ring r	10 m
Number of scatterers Q	10
Path loss exponent γ	2.5
Constant α in channel model	10^7

5.2   Evaluations of the proposed scheme

We show the simulation results of channel compression of the proposed scheme in Figs. 4 and 5. It is assumed that the SNR is 20 dB and there is no quantization error of the low-dimensional CSI.

Figure 4 shows the effect of the compression ratio $\left(\eta = \frac{M}{N_{\text{t}}}\right)$ on the system capacity. The comparison is among the proposed scheme, DCT-based CS and KLT-based CS. We can observe that whether in low or high compression ratio regimes, the KLT-based CS has the best performance, while the DCT-based CS performs the worst. To be emphasized, the best performance of the KLT-based CS is at a sacrifice of increased feedback overhead, as shown in Table 3. In this sense, the proposed scheme can offer a useful tradeoff. Additionally, as Fig. 4 shows, the proposed scheme performs much better than DCT-based CS in low compression ratio regimes.

Figure 5 illustrates the recovery performance of the high-dimensional CSI at the BS under the circumstances that the BS has perfect knowledge of the low-dimensional
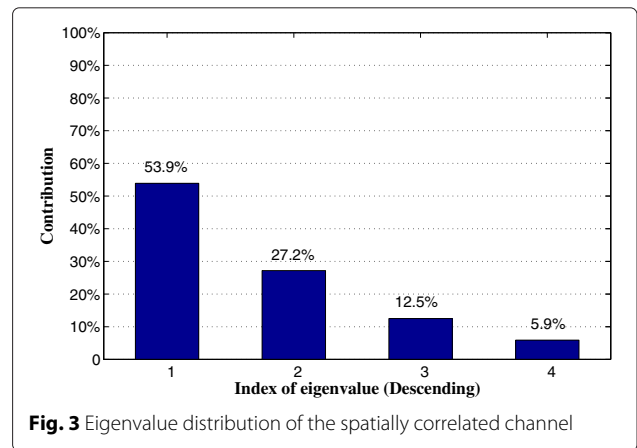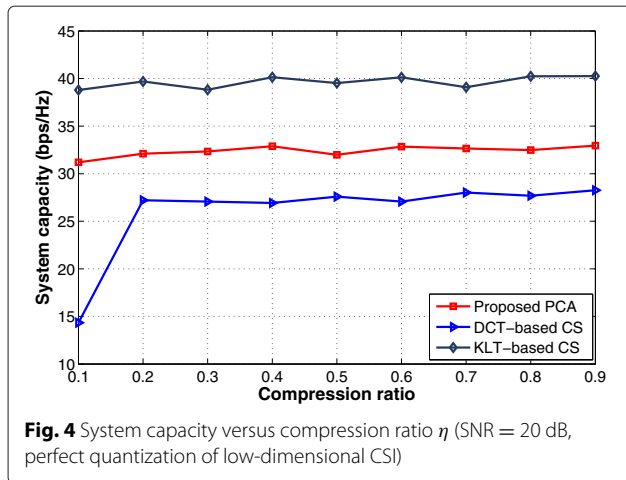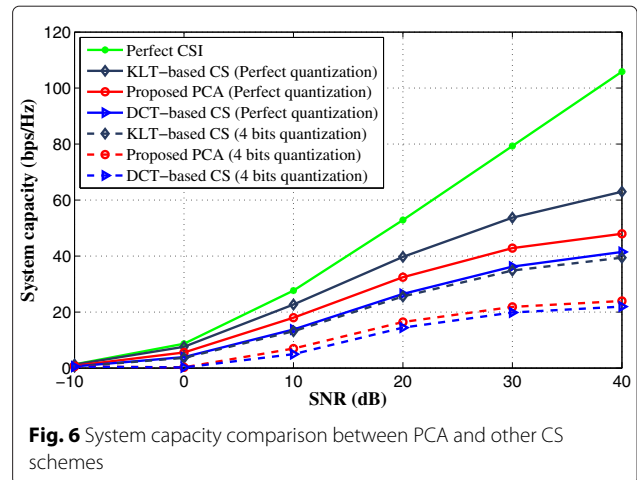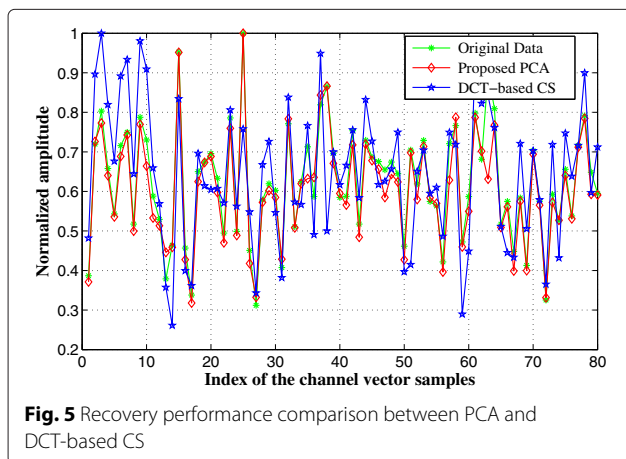
Table 3 Comparison of feedback overhead and codebook search complexity

Value Content Subject	Feedback bits	Search complexity
DCT-based CS	$S \cdot \frac{(M-1)\rho}{3}$	$S \cdot M \cdot 2^{\frac{(M-1)\rho}{3}}$
Proposed scheme	$S \cdot \frac{(M-1)\rho}{3} +$ $M \cdot \frac{(N_{\text{t}}-1)\rho}{3}$	$S \cdot M \cdot 2^{\frac{(M-1)\rho}{3}} +$ $M \cdot N_{\text{t}} \cdot 2^{\frac{(N_{\text{t}}-1)\rho}{3}}$
KLT-based CS	$S \cdot \frac{(M-1)\rho}{3} +$ $S \cdot N_{\text{t}} \cdot \frac{(N_{\text{t}}-1)\rho}{3}$	$S \cdot M \cdot 2^{\frac{(M-1)\rho}{3}} +$ $S \cdot N_{\text{t}}^2 \cdot 2^{\frac{(N_{\text{t}}-1)\rho}{3}}$

Fig. 3 Eigenvalue distribution of the spatially correlated channel

Zhang *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:64

Page 9 of 12

Fig. 4 System capacity versus compression ratio η (SNR = 20 dB, perfect quantization of low-dimensional CSI)

Fig. 6 System capacity comparison between PCA and other CS schemes

CSI without quantization. We take the proposed scheme and DCT-based CS for comparison. The $1 \times N_t$ original CSI is compressed into $1 \times M_{CS}$ low-dimensional information, where $M_{CS} = 20$ and the compression ratio is $\eta_{CS} = \frac{M_{CS}}{N_t} \approx 0.16$, while in the proposed scheme, the number of principal components is $M_{PCA} = 4$ with the compression ratio being $\eta_{PCA} = \frac{M_{PCA}}{N_t} \approx 0.03$.

As can be seen, the reconstructed high-dimensional CSI is considerably close to the original data when the PCA is utilized. But there still exists distortion because the PCA itself inevitably introduces information loss. However, the recovery performance gets poorer in the case of the DCT-based CS. The reason is that the proposed scheme takes advantage of the signal-dependent nature of PCA, which makes it possible for the compression matrix to change adaptively in every long-term period according to the variation of the original data.

Figure 6 shows a system capacity (defined as the sum of all the users' rates in the system) comparison. We choose four principal components to form the compression matrix, so the compression ratio of PCA is 0.03. For reference, we first consider the ideal situation, where

the BS can acquire perfect CSI with neither recovery distortion nor quantization error. As illustrated in Fig. 6, the best system capacity can be achieved only when the BS acquires perfect CSI. Meanwhile, the proposed scheme outperforms the existing DCT-based CS scheme whether there is quantization error resulting from low-dimensional CSI or not. But, it performs a little poorer than the KLT-based CS scheme.

When we utilize the RVQ codebook to quantize low-dimensional CSI, the system capacity decreases in both cases because the quantization error must be taken into account. Based on the results in Fig. 6 and the feedback overhead analysis in Subsection 3.3, we can draw the conclusion that our proposed scheme can offer a worthwhile design tradeoff between system capacity and feedback overhead.

6 Conclusions

In this paper, a PCA-based feedback scheme for massive MIMO was proposed. In the proposed scheme, two kinds of feedback information, the quantized low-dimensional CSI index and the compression matrix utilized to perform both compression and recovery, are fed back hierarchically. Moreover, we obtained a closed-form expression for an upper bound to the normalized information distortion. We analyzed the feedback overhead and codebook search complexity of the proposed scheme. Simulation results showed that without considering the low-dimensional CSI quantization, the proposed scheme outperforms the existing DCT-based CS scheme in terms of compression ratio. When a RVQ codebook is adopted to quantize the low-dimensional CSI, worthwhile system capacity and recovery performance can be achieved. Finally, we draw the conclusion that the proposed scheme can achieve a useful performance tradeoff between system capacity and feedback overhead, which gives it high potential to be implemented in practical massive MIMO systems.

Fig. 5 Recovery performance comparison between PCA and DCT-based CS

Zhang *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:64

Page 10 of 12

Appendix

A Proof of Proposition 1

The PCA-caused information distortion J is

$$
\begin{aligned}
J &= \frac{1}{S} \sum_{n=1}^{S} \left\| \mathbf{h}_n - \tilde{\mathbf{h}}_n \right\|^2 \\
&= \frac{1}{S} \sum_{n=1}^{S} \left\| \sum_{i=1}^{N_t} \alpha_{ni} \mathbf{u}_i - \sum_{i=1}^{M} z_{ni} \mathbf{u}_i - \sum_{i=M+1}^{N_t} b_i \mathbf{u}_i \right\|^2 \\
&= \frac{1}{S} \sum_{n=1}^{S} \left\| \sum_{i=1}^{M} (\alpha_{ni} - z_{ni}) \mathbf{u}_i + \sum_{i=M+1}^{N_t} (\alpha_{ni} - b_i) \mathbf{u}_i \right\|^2 \\
&= \frac{1}{S} \sum_{n=1}^{S} \left[\sum_{i=1}^{M} (\alpha_{ni} - z_{ni})^2 + \sum_{i=M+1}^{N_t} (\alpha_{ni} - b_i)^2 \right].
\end{aligned}
\tag{30}
$$

In order to minimize J, we take partial derivatives with respect to z_{ni} and b_i separately, as given by

$$
\frac{\partial J}{\partial z_{ni}} = \frac{1}{S} \sum_{n=1}^{S} \sum_{i=1}^{M} \left[-2 (\alpha_{ni} - z_{ni}) \right];
\tag{31}
$$

$$
\begin{aligned}
\frac{\partial J}{\partial b_i} &= \frac{1}{S} \sum_{n=1}^{S} \sum_{i=M+1}^{N_t} \left[-2 (\alpha_{ni} - b_i) \right] \\
&= -\frac{2}{S} \sum_{i=M+1}^{N_t} \left(\sum_{n=1}^{S} \alpha_{ni} - S \cdot b_i \right).
\end{aligned}
\tag{32}
$$

Letting $\frac{\partial J}{\partial z_{ni}} = 0$, we can obtain $z_{ni} = \alpha_{ni}$, that is

$$
z_{ni} = \mathbf{h}_n \mathbf{u}_i^H \ (i = 1, 2 \ldots, M; n = 1, 2 \ldots, S).
\tag{33}
$$

Similarly, when letting $\frac{\partial J}{\partial b_i} = 0$, we can also acquire

$$
b_i = \frac{1}{S} \sum_{n=1}^{S} \alpha_{ni} = \breve{\mathbf{h}} \mathbf{u}_i^H \ (i = M + 1, M + 2 \ldots, N_t),
\tag{34}
$$

where $\breve{\mathbf{h}}$ denotes the mean vector of all the S high-dimensional channel vectors estimated in a long-term period, as given by

$$
\breve{\mathbf{h}} = \frac{1}{S} \sum_{n=1}^{S} \mathbf{h}_n.
\tag{35}
$$

Substitute (33) and (34) into (14), and then $\tilde{\mathbf{h}}$ can be re-written as

$$
\tilde{\mathbf{h}}_n = \sum_{i=1}^{M} \mathbf{h}_n \mathbf{u}_i^H \mathbf{u}_i + \sum_{i=M+1}^{N_t} \breve{\mathbf{h}} \mathbf{u}_i^H \mathbf{u}_i.
\tag{36}
$$

Then,

$$
\begin{aligned}
\mathbf{h}_n &- \tilde{\mathbf{h}}_n \\
&= \sum_{i=1}^{N_t} \mathbf{h}_n \mathbf{u}_i^H \mathbf{u}_i - \left(\sum_{i=1}^{M} \mathbf{h}_n \mathbf{u}_i^H \mathbf{u}_i + \sum_{i=M+1}^{N_t} \breve{\mathbf{h}} \mathbf{u}_i^H \mathbf{u}_i \right) \\
&= \sum_{i=M+1}^{N_t} \mathbf{h}_n \mathbf{u}_i^H \mathbf{u}_i - \sum_{i=M+1}^{N_t} \breve{\mathbf{h}} \mathbf{u}_i^H \mathbf{u}_i \\
&= \sum_{i=M+1}^{N_t} \left[\left(\mathbf{h}_n - \breve{\mathbf{h}} \right) \mathbf{u}_i^H \right] \mathbf{u}_i.
\end{aligned}
\tag{37}
$$

Therefore, J can be re-expressed by

$$
\begin{aligned}
J &= \frac{1}{S} \sum_{n=1}^{S} \left\| \mathbf{h}_n - \tilde{\mathbf{h}}_n \right\|^2 \\
&= \frac{1}{S} \sum_{n=1}^{S} \sum_{i=M+1}^{N_t} \left(\mathbf{h}_n \mathbf{u}_i^H - \breve{\mathbf{h}} \mathbf{u}_i^H \right)^2 \\
&= \sum_{i=M+1}^{N_t} \frac{1}{S} \sum_{n=1}^{S} \left(\mathbf{h}_n \mathbf{u}_i^H - \breve{\mathbf{h}} \mathbf{u}_i^H \right)^2.
\end{aligned}
\tag{38}
$$

In (38), the expression $\frac{1}{S} \sum_{n=1}^{S} \left(\mathbf{h}_n \mathbf{u}_i^H - \breve{\mathbf{h}} \mathbf{u}_i^H \right)^2$ can be viewed as the covariance of $\mathbf{h}_n \mathbf{u}_i^H$. If we assume $\mathbf{C}_h$ to be the covariance matrix of $\mathbf{h}_n$, then J can be further given by

$$
J = \sum_{i=M+1}^{N_t} \mathbf{u}_i \mathbf{C}_h \mathbf{u}_i^H.
\tag{39}
$$

Our target is to minimize the PCA caused information distortion J, which can be solved by the Lagrange Multiplier (LM) method. After applying the LM, we can observe that the base vector must satisfy

$$
\mathbf{C}_h \mathbf{u}_i^H = \lambda_i \mathbf{u}_i^H.
\tag{40}
$$

Equation (40) indicates that base vector $\mathbf{u}_i$ should be chosen as the eigenvector of channel covariance matrix $\mathbf{C}_h$, and λ_i is the corresponding eigenvalue. As a result, the PCA-caused information distortion J is

$$
J = \sum_{i=M+1}^{N_t} \mathbf{u}_i \mathbf{C}_h \mathbf{u}_i^H = \sum_{i=M+1}^{N_t} \mathbf{u}_i \lambda_i \mathbf{u}_i^H = \sum_{i=M+1}^{N_t} \lambda_i.
\tag{41}
$$

B Proof of Proposition 2

As has been mentioned in Table 3, $\widehat{\mathbf{h}}$ is the reconstructed high-dimensional CSI recovered from $\hat{\mathbf{h}}$, which is given by

$$
\widehat{\mathbf{h}} = \left\| \mathbf{h} \bar{\mathbf{U}} \right\| \cdot \hat{\mathbf{h}} \hat{\mathbf{U}}^H,
\tag{42}
$$

where $\left\| \mathbf{h} \bar{\mathbf{U}} \right\|$ represents the modulus of low-dimensional CSI, since we have performed normalization in the very beginning, that is

Zhang *et al. EURASIP Journal on Advances in Signal Processing* (2016) 2016:64

Page 11 of 12

$$\bar{\mathbf{h}} = \frac{\mathbf{h}\bar{\mathbf{U}}}{\|\mathbf{h}\bar{\mathbf{U}}\|}. \tag{43}$$

Substituting (16), (17) and (43) into (42), we can rewrite $\widehat{\mathbf{h}}$ as

$$
\begin{aligned}
\widehat{\mathbf{h}} &= \|\mathbf{h}\bar{\mathbf{U}}\| \cdot \hat{\mathbf{h}}\hat{\mathbf{U}}^H \\
&= \|\mathbf{h}\bar{\mathbf{U}}\| \cdot \frac{\bar{\mathbf{h}} - d\mathbf{e}}{\sqrt{1-d^2}} \cdot \left(\frac{\bar{\mathbf{U}} - D\mathbf{E}}{\sqrt{1-D^2}}\right)^H \\
&= \|\mathbf{h}\bar{\mathbf{U}}\| \cdot \frac{\frac{\mathbf{h}\bar{\mathbf{U}}}{\|\mathbf{h}\bar{\mathbf{U}}\|} - d\mathbf{e}}{\sqrt{1-d^2}} \cdot \left(\frac{\bar{\mathbf{U}} - D\mathbf{E}}{\sqrt{1-D^2}}\right)^H \\
&= \frac{\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H - \mathbf{h}\bar{\mathbf{U}}(D\mathbf{E})^H - \|\mathbf{h}\bar{\mathbf{U}}\| \cdot d\mathbf{e}(\bar{\mathbf{U}} - D\mathbf{E})^H}{\sqrt{1-d^2} \cdot \sqrt{1-D^2}}.
\end{aligned}
\tag{44}
$$

Moreover, because of the independence between $\hat{\mathbf{U}}$ and $\mathbf{E}$, when multiplying $\bar{\mathbf{U}}$ in (17) by $\mathbf{E}^H$, one obtains

$$
\begin{aligned}
\bar{\mathbf{U}}\mathbf{E}^H &= \sqrt{1-D^2}\hat{\mathbf{U}}\mathbf{E}^H + D\mathbf{E}\mathbf{E}^H = D\mathbf{E}\mathbf{E}^H \\
&= D\begin{bmatrix} \mathbf{I}_{M\times M} & \mathbf{0}_{M\times(N_t-M)} \\ \mathbf{0}_{(N_t-M)\times M} & \mathbf{0}_{(N_t-M)\times(N_t-M)} \end{bmatrix},
\end{aligned}
\tag{45}
$$

where $\mathbf{I}_{M\times M}$ represents the $M \times M$ identity matrix; $\mathbf{0}_{M\times(N_t-M)}$, $\mathbf{0}_{(N_t-M)\times M}$ and $\mathbf{0}_{(N_t-M)\times(N_t-M)}$ denotes the $M \times (N_t - M)$, $(N_t - M) \times M$ and $(N_t - M) \times (N_t - M)$ zero matrix respectively. When assuming $\mathbf{I}_M = \begin{bmatrix} \mathbf{I}_{M\times M} & \mathbf{0}_{M\times(N_t-M)} \\ \mathbf{0}_{(N_t-M)\times M} & \mathbf{0}_{(N_t-M)\times(N_t-M)} \end{bmatrix}$, the following expression results,

$$\bar{\mathbf{U}}\mathbf{E}^H = D \cdot \mathbf{I}_M. \tag{46}$$

Substitute (46) into (44), then (44) can be rewritten as

$$
\begin{aligned}
\widehat{\mathbf{h}} &= \frac{\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H - D\mathbf{h}\bar{\mathbf{U}}\mathbf{E}^H - \|\mathbf{h}\bar{\mathbf{U}}\| \cdot d\mathbf{e}(\bar{\mathbf{U}} - D\mathbf{E})^H}{\sqrt{1-d^2} \cdot \sqrt{1-D^2}} \\
&= \frac{\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H - D^2 \cdot \mathbf{h}\mathbf{I}_M - \|\mathbf{h}\bar{\mathbf{U}}\| \cdot d\sqrt{1-D^2} \cdot \mathbf{e}\hat{\mathbf{U}}^H}{\sqrt{1-d^2} \cdot \sqrt{1-D^2}}.
\end{aligned}
\tag{47}
$$

C   Proof of Eqs. (23) and (24)

First, we focus on Eq. (23). Assume $\mathbf{U} = \begin{bmatrix} \bar{\mathbf{U}} \vdots \Delta\mathbf{U} \end{bmatrix}$, where $\bar{\mathbf{U}}$ is composed of the M dominating eigenvectors, while $\Delta\mathbf{U}$ is composed of the less dominating $N_t - M$ eigenvectors. In the proposed scheme, we only choose M dominating eigenvectors to compose compression matrix $\bar{\mathbf{U}}$, which is to be utilized to compress high-dimensional CSI into low-dimensional representation. Particularly, if choosing

all of the N_t eigenvectors, that is $\bar{\mathbf{U}} = \mathbf{U}$, the distortion disappears, as given by

$$\mathbf{h} = \mathbf{h}\mathbf{U}\mathbf{U}^H. \tag{48}$$

Substituting $\mathbf{U} = \begin{bmatrix} \bar{\mathbf{U}} \vdots \Delta\mathbf{U} \end{bmatrix}$ into (48), and we can obtain

$$
\begin{aligned}
\mathbf{h} &= \mathbf{h}\begin{bmatrix} \bar{\mathbf{U}} \vdots \Delta\mathbf{U} \end{bmatrix}\begin{bmatrix} \bar{\mathbf{U}} \vdots \Delta\mathbf{U} \end{bmatrix}^H \\
&= \mathbf{h}\left(\bar{\mathbf{U}}\bar{\mathbf{U}}^H + \Delta\mathbf{U}\Delta\mathbf{U}^H\right).
\end{aligned}
\tag{49}
$$

As mentioned above, the choice of M dominating eigenvectors inevitably leads to information distortion. According to (49), the distortion caused by PCA can be expressed by

$$J = \mathbb{E}\left[\|\mathbf{h} - \mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H\|^2\right] = \mathbb{E}\left[\|\mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\|^2\right]. \tag{50}$$

Meanwhile,

$$
\begin{aligned}
E\left[\|\mathbf{h}\|^2\right] &= E\left[\|\mathbf{h}\left(\bar{\mathbf{U}}\bar{\mathbf{U}}^H + \Delta\mathbf{U}\Delta\mathbf{U}^H\right)\|^2\right] \\
&= E\left[\left(\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H + \mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\right)\left(\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H + \mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\right)^H\right] \\
&= E\left[\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H\bar{\mathbf{U}}\bar{\mathbf{U}}^H\mathbf{h}^H + \|\mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\|^2 + \right. \\
&\quad \left. \mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H\Delta\mathbf{U}\Delta\mathbf{U}^H\mathbf{h}^H + \mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\bar{\mathbf{U}}\bar{\mathbf{U}}^H\mathbf{h}^H\right].
\end{aligned}
\tag{51}
$$

Because of the orthogonality between $\bar{\mathbf{U}}$ and $\Delta\mathbf{U}$, one has

$$\bar{\mathbf{U}}^H\Delta\mathbf{U} = \mathbf{0}_{M\times(N_t-M)}; \tag{52}$$

$$\Delta\mathbf{U}^H\bar{\mathbf{U}} = \mathbf{0}_{(N_t-M)\times M}. \tag{53}$$

Therefore, Eq. (51) can be rewritten as

$$
\begin{aligned}
\mathbb{E}\left[\|\mathbf{h}\|^2\right] &= \mathbb{E}\left[\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H\bar{\mathbf{U}}\bar{\mathbf{U}}^H\mathbf{h}^H + \|\mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\|^2 + \right. \\
&\quad \left. \underbrace{\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H\Delta\mathbf{U}\Delta\mathbf{U}^H\mathbf{h}^H}_{0} + \underbrace{\mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\bar{\mathbf{U}}\bar{\mathbf{U}}^H\mathbf{h}^H}_{0}\right] \\
&= \mathbb{E}\left[\mathbf{h}\bar{\mathbf{U}}\bar{\mathbf{U}}^H\bar{\mathbf{U}}\bar{\mathbf{U}}^H\mathbf{h}^H + \|\mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\|^2\right] \\
&= \mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2 + \|\mathbf{h}\Delta\mathbf{U}\Delta\mathbf{U}^H\|^2\right] \\
&= \mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2\right] + J.
\end{aligned}
\tag{54}
$$

Since each element of channel vector $\mathbf{h}$ obeys the Gaussian distribution with unit variance, then $\mathbb{E}\left[\|\mathbf{h}\|^2\right] = N_t$. So we can easily obtain that $\mathbb{E}\left[\|\mathbf{h}\bar{\mathbf{U}}\|^2\right] = N_t - J$.

As for Eq. (24), $\mathbb{E}\left[\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H\right]$ can be expressed by

$$\mathbb{E}\left[\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H\right] = \mathbb{E}\left[\left|\mathbf{h}_{(1)}\right|^2 + \left|\mathbf{h}_{(2)}\right|^2 + \ldots + \left|\mathbf{h}_{(M)}\right|^2\right], \tag{55}$$

Zhang et al. EURASIP Journal on Advances in Signal Processing (2016) 2016:64

Page 12 of 12

where $\mathbf{h}_{(m)}$ ($m = 1, 2, \ldots M$) denotes the mth element of $\mathbf{h}$. As mentioned above, each element of channel vector $\mathbf{h}$ obeys the Gaussian distribution with unit variance. Therefore,

$$\mathbb{E}\left[\mathbf{h}\mathbf{I}_M{}^H\mathbf{h}^H\right] = M. \tag{56}$$

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Beijing Key Laboratory of Network System Architecture and Convergence, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, No. 10, Xitucheng Road, Haidian District, 100876 Beijing, China. [2]School of Communication Engineering, Hangzhou Dianzi University, 310018 Hangzhou, China. [3]School of Science and Technology, Middlesex University, NW4 4BT, London, UK.

References
1. F Boccardi, RW Heath Jr., A Lozano, TL Marzetta, P Popovski, Five disruptive technology directions for 5G. IEEE Comm. Mag. **52**(2), 74–80 (2014)
2. F Rusek, D Persson, BK Lau, EG Larsson, TL Marzetta, O Edfors, F Tufvesson, Scaling up MIMO: Opportunities and challenges with very large arrays. IEEE Sig. Proc. Mag. **30**(1), 40–60 (2013)
3. X Su, J Zeng, L-P Rong, Y-J Kuang, Investigation on key technologies in large-scale MIMO. J. Comput. Sci. Technol. **28**(3), 412–419 (2013)
4. X Rao, VKN Lau, Interference alignment with partial CSI feedback in MIMO cellular networks. IEEE Trans. Sig. Proc. **62**(8), 2100–2110 (2014)
5. J Hoydis, S ten Brink, M Debbah, Massive MIMO in the UL/DL of cellular networks: how many antennas do we need. IEEE Jour.Select. Areas in Comm. **31**(2), 160–171 (2013)
6. N Jindal, MIMO broadcast channels with finite-rate feedback. IEEE Trans. Info. Th. **52**(11), 5045–5060 (2006)
7. DJ Love, RW Heath Jr., Limited feedback diversity techniques for correlated channels. IEEE Trans. Vehicular Technol. **55**(2), 718–722 (2006)
8. P Xia, GB Giannakis, Design and analysis of transmit beamforming based on limited-rate feedback. IEEE Trans. Sig. Proc. **54**(5), 1853–1863 (2006)
9. J Choi, V Raghavan, DJ Love, *Limited feedback design for the spatially correlated multi-antenna broadcast channel,* 2013 IEEE Global Communications Conference (GLOBECOM), vol. 1, (Atlanta, GA, 2013), pp. 3481–3486
10. V Raghavan, RW Heath Jr., AM Sayeed, Systematic codebook designs for quantized beamforming in correlated MIMO channels. IEEE J. Select. Areas Commun. **25**(7), 1298–1310 (2007)
11. J Li, X Su, Z Zeng, Y Zhao, S Yu, L Xiao, X Xu, *Codebook design for uniform rectangular arrays of massive antennas,* Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th, (Dresden, 2013), pp. 1–5
12. X Su, J Zeng, J Li, L Rong, L Liu, X Xu, J Wang, *International Journal of Antennas and Propagation*, vol. 2013. (Hindawi Publishing Corporation, New York, US, 2013)
13. D Ying, FW Vook, T Thomas, DJ Love, *Sub-sector-based codebook feedback for massive MIMO with 2D antenna arrays,* 2014 IEEE Global Communications Conference, (Austin, TX, 2014), pp. 3702–3707
14. J Choi, Z Chance, DJ Love, U Madhow, Noncoherent trellis coded quantization: a practical limited feedback technique for massive MIMO systems. IEEE Trans. Commun. **61**(12), 5016–5029 (2013)
15. J Choi, DJ Love, T Kim, Trellis-extended codebooks and successive phase adjustment: a path from LTE-advanced to FDD massive MIMO systems. IEEE Trans. Wireless Commun. **14**(4), 2007–2016 (2015)
16. Y Han, S Wonjae, L Jungwoo, *Projection based feedback compression for FDD massive MIMO systems,* 2014 IEEE Globecom Workshops (GC Wkshps), (Austin, TX, 2014), pp. 364–369
17. PH Kuo, HT Kung, PA Ting, *Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays,* 2012 IEEE Wireless Communications and Networking Conference (WCNC), (Shanghai, 2012), pp. 492–497
18. J Lee, SH Lee, *A Compressed Analog Feedback Strategy for Spatially Correlated Massive MIMO Systems,* Vehicular Technology Conference (VTC Fall), 2012 IEEE, (Quebec City, QC, 2012), pp. 1–6
19. P Cao, E Jorswieck, *DCT and VQ based limited feedback in spatially-correlated massive MIMO systems,* 2014 IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM), (A Coruna, 2014), pp. 285–288
20. MS Sim, CB Chae, *Compressed channel feedback for correlated massive MIMO systems,* 2014 IEEE Globecom Workshops (GC Wkshps), (Austin, TX, 2014), pp. 360–364
21. W Lu, X Tan, Q Liu, Y Liu, D Wang, Compressive Channel Feedback Schemes Based on Redundant Dictionary in MIMO Communication Systems. Wireless Personal Communications. **82**(4), 2215–2229 (2015)
22. MS Sim, C-B Chae, in *Proc. IEEE Globecom Workshops (GC Wkshps)*. Compressed channel feedback for correlated massive MIMO systems, (Austin, TX, 2014), pp. 327–332
23. JE Fowler, Compressive-projection principal component analysis. IEEE Trans. Image Process. **18**, 2230–2242 (2009)
24. LI Smith, *A tutorial on principal components analysis. Technical report.* (Cornell University, USA, 2002)
25. D Shiu, G Foschini, M Gans, J Kahn, Fading correlation and its effect on the capacity of multi-element antenna systems. IEEE Trans. Comm. **48**(3), 502–513 (2000)
26. J Nam, J-Y Ahn, A Adhikary, G Caire, *Joint spatial division and multiplexing: realizing massive MIMO gains with limited channel state information,* 46th Annual Conference on Information Sciences and Systems(CISS 2012). (Princeton, NJ, USA, 2012)
27. H Yin, D Gesbert, L Cottatellucci, Dealing with interference in distributed large-scale MIMO systems: a statistical approach. IEEE Jour. Select. Topics Sig. Proc. **8**(5), 942–953 (2014)
28. A Wiesel, YC Eldar, S Shamai, Zero-forcing precoding and generalized inverses. IEEE Trans. on Sig. Proc. **56**(9), 4409–4418 (2008)