Robust Multi-Antenna Multi-User Precoding Based on Generalized Multi-Unitary Decomposition With Partial CSI Feedback

Wee Seng Chua, Chau Yuen, *Senior Member, IEEE*, Yong Liang Guan, *Member, IEEE*, and Francois Chin

***Abstract*—In this paper, we present a novel matrix decomposition method, called generalized multi-unitary decomposition (GMUD), and use it to achieve robust multi-antenna multi-user multiple-input–multiple-output (MIMO) precoding with limited and partial channel information feedback. GMUD transforms complex matrix {H} into $\mathbf{H} = \mathbf{P}_{\theta,r}\mathbf{R}_r\mathbf{Q}^H_{\theta,r}$, where $\mathbf{R}_r$ is a lower triangular matrix, and $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ are unitary matrices. A unique feature of GMUD is that it gives multiple solutions of the unitary matrices $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ based on two parameters, namely the *direction parameter* θ and *magnitude parameter* r. This property enables the GMUD precoder to steer the precoding vectors of different users to jointly optimize their performance. Compared with regularized-inverse precoding (an existing multi-user MIMO precoding technique that we have extended to handle multiple receive antennas), the advantages of GMUD precoding are that it requires only partial channel state information (CSI) from the users, and its performance is robust to CSI quantization.**

***Index Terms*—Linear precoding, multi-user multiple-input–multiple-output (MIMO) channels, partial feedback.**

I. Introduction

MULTIPLE-input–multiple-output (MIMO) schemes have become very important in current and next-generation broadband wireless networks. Both [1] and [2] show that the capacity of a point-to-point MIMO system with N_T transmit antennas and N_R receive antennas increases linearly as $\min(N_T, N_R)$ at high SNR. The additional antennas deployed in MIMO systems create virtual spatial sub-channels that allow multiple independent data streams to be transmitted simultaneously over the same frequency. These virtual spatial sub-channels provide a large linear increase in the data throughput and spectral efficiency. This linear increase in data throughput is also known as the *multiplexing gain*.

Subsequently, point-to-point MIMO channels are extended to MIMO broadcast channels (MIMO-BCs) where the multiple transmit antennas deployed at the base station are designed to serve multiple users simultaneously [3]–[6]. The sum-rate capacity of the Gaussian BC channel for the case of two users, each with a single receive antenna, was first obtained by Caire and Shamai [3]. Viswanath and Tse [4], Vishwanath *et al.* [5], and Yu and Cioffi [6] showed that the sum-rate capacity of MIMO-BCs grows linearly with the minimum number of transmit and receive antennas. Recently, Weingarten *et al.* provided a characterization of the entire capacity region [7].

When full channel state information (CSI) is available at the transmitter, precoding techniques such as zero-forcing precoding [8] and regularized-inverse precoding [9] can be used to precancel or pre-suppress multi-user interference at the transmitter in a downlink BC. For example, the regularized-inverse precoding proposed in [9] enables the base station to use N_T transmit antennas to simultaneously send N_T different data streams to N_T users equipped with single antennas. On the other hand, in many future communication systems such as 3GPP Long Term Evolution, the users may be equipped with more than one receive antenna. The additional receive antennas can be exploited to increase the number of data streams sent to each user [8] or to increase the quality of the CSI feedback [10]. In the latter case, the author proposed a composite-channel-based feedback technique that involves linearly combining the quantization vectors from an orthonormal basis.

In this paper, we focus on the non-codebook-based design of a robust multi-user MIMO precoding scheme in systems with limited channel feedback and multiple receive antennas per user. We first show that the regularized-inverse technique described in [9] can be augmented by applying receive-antenna selection at the precoder, but its performance is very sensitive to the CSI accuracy. If the feedback channel bandwidth is limited and the CSI has to be quantized, the performance of regularized-inverse precoding with antenna selection will deteriorate severely. To overcome this problem, we propose a new matrix decomposition technique, called generalized multi-unitary decomposition (GMUD), and use it to design a precoding scheme that makes use of multiple receive antennas to improve system performance without the need for full channel feedback information. Specifically, for N_R receive antennas per user, the proposed GMUD precoding matrix does not require the CSI of all $N_R \times N_T$ spatial channels but only the right eigenvectors and the singular values of the channel matrix, and they can be considerably quantized.

This paper is organized as follows. Section II describes the multi-antenna multi-user system model used and reviews

Manuscript received October 5, 2011; revised March 3, 2012, June 25, 2012 and September 30, 2012; accepted October 29, 2012. Date of publication November 15, 2012; date of current version February 12, 2013. This work was supported in part by the International Design Center under Grant IDG31100102 and Grant IDD11100101. The review of this paper was coordinated by Prof. W. A. Hamouda.

W. S. Chua and Y. L. Guan are with Nanyang Technological University, Singapore 639798.

C. Yuen is with Singapore University of Technology and Design, Singapore 138682 (e-mail: yuenchau@sutd.edu.sg).

F. Chin is with Institute for Infocomm Research, A*STAR, Singapore 138632.

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

Digital Object Identifier 10.1109/TVT.2012.2227857

related prior works. In Section III, we provide a detailed explanation of GMUD and prove its existence. We also provide a numerical example to illustrate its properties. In Section IV, we present the design of a GMUD precoding matrix for multi-user MIMO broadcast. The comparisons of the simulation results of the proposed scheme and other schemes are shown in Section V, where we also show that the GMUD precoding technique is robust to quantization errors. Finally, Section VI provides some concluding remarks.

II. Regularized-Inverse Precoding

A. Signal Model

Considering a base station with N_T transmit antennas that sends one data stream per user to a pool of K users, each with N_R receive antennas, the received signal of the kth user is given as

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{n}_k, \qquad k \in \mathcal{S}_K \tag{1}$$

where the italic subscript k represents the kth user; $\mathcal{S}_K$ is a set of indexes of the users being served; $\mathbf{y}_k$'s are $N_R \times 1$ column vectors representing the signal received by the kth user; $\mathbf{x} = [x_1 \ldots x_{N_T}]^T$ is a $N_T \times 1$ column vector representing the transmitted signal; $\mathbf{H}_k$ is a $N_R \times N_T$ matrix containing the channel information between the base station and the kth user, and is assumed to be known to both the transmitter and the receiver; and $\mathbf{n}_k$ is an additive white Gaussian noise (AWGN) vector such that $E[\mathbf{n}_k\mathbf{n}_k^H] = \sigma^2\mathbf{I}$ and that $[\cdot]^H$ denotes complex conjugate.

If one data stream is sent per user, the precoded signal $\mathbf{x}$ from the transmitter can be modeled as

$$\mathbf{x} = \sum_{k\in\mathcal{S}_K} \mathbf{g}_k u_k/\sqrt{\gamma} = \mathbf{G}\mathbf{u}/\sqrt{\gamma} \tag{2}$$

where $\mathbf{g}_k$ is the precoding vector for the kth user, $\mathbf{G}$ is a $N_T \times K$ precoding matrix; $\mathbf{u} = [u_1, \ldots, u_K]^T$ whose element u_k is the intended data for the kth user; and $\gamma = \|\mathbf{G}\mathbf{u}\|^2$ is used to normalize the power of the transmitted signal $\mathbf{x}$. The received signal hence becomes

$$\begin{aligned}\mathbf{y}_k &= \frac{\mathbf{H}_k\mathbf{G}\mathbf{u}}{\sqrt{\gamma}} + \mathbf{n}_k \\ &= \begin{bmatrix} \overline{h}_{1,1} & \cdots & \overline{h}_{1,k} \\ \vdots & & \vdots \\ \overline{h}_{i,1} & \ddots & \overline{h}_{i,k} \\ \vdots & & \vdots \\ \overline{h}_{N_R,1} & \cdots & \overline{h}_{N_R,k} \end{bmatrix} \begin{bmatrix} u_{1,1} \\ \vdots \\ u_{1,k} \end{bmatrix} + \mathbf{n}_k \end{aligned} \tag{3}$$

where $\overline{h}_{i,k}$ is the (i, k)th entry of the normalized equivalent channel matrix $\mathbf{H}_k\mathbf{G}/\sqrt{\gamma}$ for the kth user. The ith antenna of the kth user will receive

$$y_i = \sum_{j=1}^{K} \overline{h}_{i,j}u_j + n = \underbrace{\overline{h}_{i,k}u_k}_{\substack{\text{desired}\\ \text{signal}}} + \underbrace{\sum_{\substack{j=1\\ j\neq k}}^{K} \overline{h}_{i,j}u_j}_{\text{interference}} + \underbrace{n}_{\text{noise}} \tag{4}$$

and the signal-to-interference-plus-noise ratio (SINR) received by the ith antenna is

$$\mathrm{SINR}_i = \frac{\|\overline{h}_{i,k}\|^2}{\left\|\sum_{\substack{j=1\\ j\neq k}}^{K} \overline{h}_{i,j}u_j\right\|^2 + \sigma^2}. \tag{5}$$

In this paper, we considered an antenna selection diversity combining method, and the resultant SINR of the kth user is given as

$$\mathrm{SINR} = \max_{\forall i \in N_R} \mathrm{SINR}_i. \tag{6}$$

B. (Conventional) Regularized-Inverse Precoding

In a multi-user communication system with one receive antenna per user, the regularized-inverse precoding in [9] arranges the channel information into $K \times N_T$ matrix $\tilde{\mathbf{H}}$, such that the kth row of $\tilde{\mathbf{H}}$ represents the channel between the base station and the kth user. Precoding matrix $\mathbf{G}$ is

$$\mathbf{G}_{\text{reg-inv}} = \tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \alpha\mathbf{I})^{-1} \tag{7}$$

where the optimal value of $\alpha = K/\rho$ and $\rho = 1/\sigma^2$ are derived in [9] for uncoded system. Although this precoding matrix introduces some interference from the other users as $\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H(\tilde{\mathbf{H}}\tilde{\mathbf{H}}^H + \alpha\mathbf{I})^{-1} \neq \mathbf{I}$, it reduces the normalization constant γ significantly, which leads to a stronger signal with a higher SINR at the receiver. The amount of interference is determined by α. No matter how poorly conditioned $\tilde{\mathbf{H}}$ may be, the inverse in (7) can be improved by choosing appropriate α.

When the users have multiple receive antennas, the users may choose to feedback the channel of just one of the receive antennas; then, each user will appear to the transmitter as a single-antenna terminal. This is a simple extension of the regularized-inverse precoding from a single-antenna receiver to a multi-antenna receiver without adding to the burden of channel feedback, although it is suboptimal.

C. (New) Regularized-Inverse Precoding With Optimal Receive-Antenna Selection

When N_R receive antennas are available to each user for receiving one data stream per user, there is a total of N_R channels between the base station and one of the receive antenna. The equivalent communication channel between the base station and users can be arranged as a $K \times N_T$ matrix $\hat{\mathbf{H}}$, whose kth row represents the corresponding channel information between the N_T transmit antennas and user k. Since every user has a choice of N_R channel information from N_R antennas, there will be a total of $(N_R)^K$ different $\hat{\mathbf{H}}$. The regularized-inverse precoding matrix then becomes

$$\hat{\mathbf{G}}_{\text{reg-inv}} = \hat{\mathbf{H}}^H(\hat{\mathbf{H}}\hat{\mathbf{H}}^H + \alpha\mathbf{I})^{-1} \tag{8}$$

where $\hat{\mathbf{H}}$ is one of the $(N_R)^K$ different channel matrices with an optimal value of $\alpha = K\sigma^2$. The transmitted signal then becomes

$$\mathbf{x} = \frac{\hat{\mathbf{G}}_{\text{reg-inv}}}{\sqrt{\gamma}}\mathbf{u}. \tag{9}$$

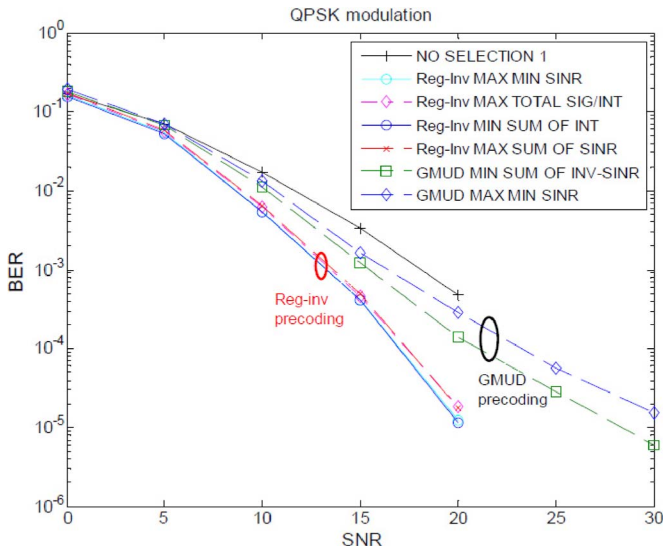

Fig. 1. Probability of bit error of a two-user MIMO QPSK system (a base station with two transmit antennas and two users with two receive antennas each) using different criteria to find the optimum precoding matrix (regularized-inverse precoding with antenna selection and GMUD precoding).

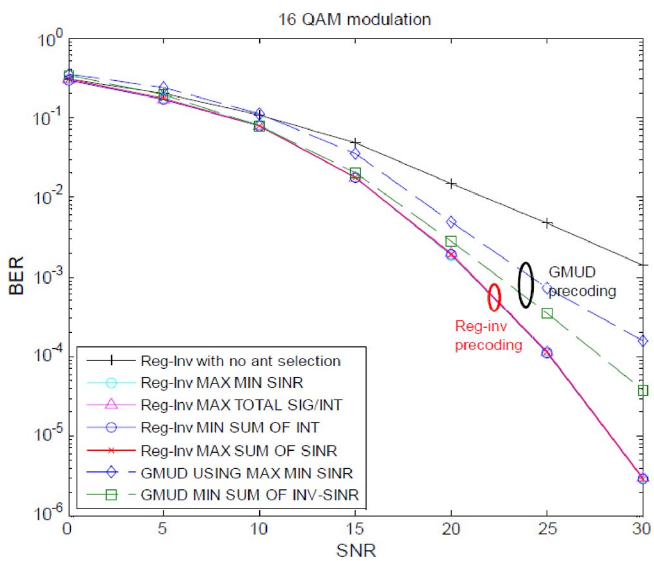

Fig. 2. Probability of bit error of a two-user MIMO 16-QAM system (a base station with two transmit antennas and two users with two receive antennas each) using different criteria to find the optimum precoding matrix (regularized-inverse precoding with antenna selection and GMUD precoding).

Clearly, the precoding matrix $\hat{\mathbf{G}}_{\text{reg}-\text{inv}}$ can be optimized by some form of antenna selection to optimize some performance criterion. This optimality criterion may be: 1) maximizing the minimum SINR of all users; 2) maximizing the sum of all users' signal power over the sum of all users' interference plus noise [11]; or 3) minimizing the sum of per user's inverse SINR [12]. Figs. 1 and 2 show that criterion 3, which was proposed in [12], outperforms the other two criteria; hence, we will use this criterion to optimize the precoding matrix in this paper.

The resultant precoding matrix is

$$\hat{\mathbf{G}}_{\text{reg}-\text{inv}} = \arg\min_{\hat{\mathbf{H}}} \sum_{m=1}^{K} \frac{\sum_{n=1,\, n\neq m}^{K} |\eta_{m,\,n}|^2 + \gamma/\rho}{|\eta_{m,\,m}|^2} \tag{10}$$

where

$$\hat{\mathbf{H}}\hat{\mathbf{G}}_{\text{reg}-\text{inv}} = \frac{\hat{\mathbf{H}}\hat{\mathbf{H}}^H(\hat{\mathbf{H}}\hat{\mathbf{H}}^H + \alpha\mathbf{I})^{-1}}{\sqrt{\gamma}} = \begin{bmatrix} \eta_{1,1} & \cdots & \eta_{1,K} \\ & \ddots & \\ \eta_{K,1} & & \eta_{K,K} \end{bmatrix}.$$

Effectively, the regularized-inverse precoding with antenna selection chooses the best-conditioned channel between the users and the transmit antennas to improve the bit-error-rate (BER) performance.

The received signal for user k is given by

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{n} = \mathbf{F}_k\mathbf{u} + \mathbf{n} \tag{11}$$

where $\mathbf{F}_k = \mathbf{H}_k\hat{\mathbf{G}}_{\text{reg}-\text{inv}}/\sqrt{\gamma}$ is the equivalent channel of user k. To recover the information symbol u_k, the user can either use a zero-forcing receiver or a MMSE receiver.[1] The MMSE-detected data vector is

$$\hat{\mathbf{x}}_k = \left(\mathbf{F}_k^H\mathbf{F}_k + \sigma^2\mathbf{I}\right)^{-1}\mathbf{y}_k \tag{12}$$

where σ^2 refers to the variance of the AWGN.

III. Generalized Multi-Unitary Decomposition

A complex channel matrix $\mathbf{H}$ can be transformed into various forms using different decomposition techniques such as the following.

1) *Singular value decomposition (SVD)* [14]: $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices, and $\mathbf{\Lambda}$ is a diagonal matrix containing the singular values of $\mathbf{H}$.
2) *Equal-diagonal QR decomposition* [15] *and geometric mean decomposition (GMD)* [16]: $\mathbf{H} = \mathbf{URV}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices, and $\mathbf{R}$ is an upper triangular matrix with diagonal elements containing the geometric mean of the positive singular values of $\mathbf{H}$.
3) *Generalized Triangular Decomposition (GTD)* [17]: $\mathbf{H} = \mathbf{URV}^H$, where $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices, and $\mathbf{R}$ is an upper triangular matrix with prescribed diagonal elements multiplicatively majorized by the singular values of $\mathbf{H}$.

In this paper, we propose a new matrix decomposition technique called GMUD that gives the transformation $\mathbf{H} = \mathbf{P}_{\theta,\,r}\mathbf{R}_r\mathbf{Q}_{\theta,\,r}^H$, where $\mathbf{R}_r$ can be a lower triangular matrix or a special matrix with a prescribed value called "*magnitude parameter*" r in the leftmost entry and zeros at the remaining entries in the first row. $\mathbf{P}_{\theta,\,r}$ and $\mathbf{Q}_{\theta,\,r}$ are different unitary matrices generated based on r and "*direction parameter*" θ.

A. Derivation of R_r in GMUD

Consider rank-L complex $\mathbf{H}$ with singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. Its GMUD gives $\mathbf{H} = \mathbf{P}_{\theta,\,r}\mathbf{R}_r\mathbf{Q}_{\theta,\,r}^H$, where $\mathbf{P}_{\theta,\,r}$ and

[1]MMSE receiver or Wiener filter has been shown to be optimum for a MIMO system with any precoding matrix in [11].

$\mathbf{Q}_{\theta,r}$ are unitary matrices, and $\mathbf{R}_r$ is a $L \times L$ matrix in the form of

$$\mathbf{R}_r = \begin{bmatrix} r & 0 & \cdots & 0 \\ z_{21} & z_{22} & \cdots & z_{2L} \\ \vdots & \vdots & \vdots & \vdots \\ z_{L1} & z_{L2} & \cdots & z_{LL} \end{bmatrix}, \quad \text{where } \lambda_L < r < \lambda_1. \tag{13}$$

In the first row of $\mathbf{R}_r$, only the leftmost entry is a nonzero positive element. This element r can take any value between the smallest and largest singular values of $\mathbf{H}$ (corresponding to λ_L and λ_1 in (13), respectively). The values of the remaining elements $z_{21}, \ldots, z_{LL}$ in $\mathbf{R}_r$ are calculated based on r and the singular values (details will be shown later). When the rank of $\mathbf{H}$ is greater than 2, $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ are generally not unique under this definition of $\mathbf{R}_r$, i.e., there are multiple solutions of $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$, given a value of r. However, in the special case where $\mathbf{R}_r$ is defined as a lower triangular matrix with prescribed diagonal elements, $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ will both be unique. Such $\mathbf{R}_r$ takes the following form:

$$\mathbf{R}_r = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ z_{21} & r_2 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ z_{L1} & z_{L2} & \cdots & r_L \end{bmatrix} \tag{14}$$

where the values of the diagonal entries must satisfy Weyl's multiplicative majorization conditions [17]. In this case, GTD turns out to be a special case of GMUD.

The complex channel $\mathbf{H}$ is first transformed into $\mathbf{H} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H$ using SVD, where $\boldsymbol{\Lambda} \in \mathbb{R}^{L \times L}$ is a diagonal matrix containing all the singular values λ_i in descending order, and both $\mathbf{U}$ and $\mathbf{V}$ are orthonormal matrices. For GMUD, its matrix $\mathbf{R}_r$ with prescribed r can be arranged in the form of $\mathbf{R}_r = \mathbf{W}^H\boldsymbol{\Lambda}\mathbf{X}$, where $\boldsymbol{\Lambda}$ is the same SVD diagonal matrix of $\mathbf{H}$, whereas $\mathbf{W}$ and $\mathbf{X}$ are unitary matrices defined by the following Givens rotations [18], [19]

$$\prod_{i=1}^{L} \prod_{l=i+1}^{L} \mathbf{F}_{li}^T(\psi_{li}) \tag{15}$$

where

$$\mathbf{F}_{li}(\psi_{li}) = \begin{bmatrix} \mathbf{I}_{i-1} & 0 & 0 & 0 & 0 \\ 0 & \cos(\psi_{li}) & 0 & \sin(\psi_{li}) & 0 \\ 0 & 0 & I_{l-i-1} & 0 & 0 \\ 0 & -\sin(\psi_{li}) & 0 & \cos(\psi_{li}) & 0 \\ 0 & 0 & 0 & 0 & \mathbf{I}_{L-1} \end{bmatrix}$$

and the product $\mathbf{F}_{li}(\psi_{li})\mathbf{x}$ represents a counterclockwise rotation of vector $\mathbf{x}$ in the (l, i) plane of ψ_{li} radians. $\mathbf{I}_{i-1}$ and $\mathbf{I}_{L-1}$ in (15) are identity matrices with $i-1$ and $L-1$ dimensions, respectively.

From this point onward, for simplicity of illustration, we consider $L = 2$. This gives

$$\mathbf{R}_r = \begin{bmatrix} r & 0 \\ z_1 & z_2 \end{bmatrix} \triangleq \mathbf{W}^H\boldsymbol{\Lambda}\mathbf{X} \tag{16}$$

where $\boldsymbol{\Lambda}$ is the SVD diagonal matrix of $\mathbf{H}$, and based on (15), we can write $\mathbf{W}$ and $\mathbf{X}$ as

$$\mathbf{W} = \begin{bmatrix} a^* & b^* \\ -b^* & a^* \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \tag{17}$$

where * denotes complex conjugates $|b| = \sqrt{1-a^2}$ and $|s| = \sqrt{1-c^2}$. As a result, r, z_1, and z_2 can be expanded in terms of a, b, c, s, λ_1, and λ_2, i.e.,

$$\begin{aligned} \mathbf{R}_r &= \mathbf{W}^H\boldsymbol{\Lambda}\mathbf{X} \\ \begin{bmatrix} r & 0 \\ z_1 & z_2 \end{bmatrix} &= \begin{bmatrix} a^* & b^* \\ -b^* & a^* \end{bmatrix}^H \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \\ \begin{bmatrix} r & 0 \\ z_1 & z_2 \end{bmatrix} &= \begin{bmatrix} ac\lambda_1 + bs\lambda_2 & as\lambda_1 - bc\lambda_2 \\ bc\lambda_1 - as\lambda_2 & bs\lambda_1 + ac\lambda_2 \end{bmatrix}. \end{aligned} \tag{18}$$

By prescribing a value of r, the values of a and c can be obtained from the first row of (18) after substituting b and s from (17); then, $\mathbf{W}$ and $\mathbf{X}$ can be determined from the following:

$$ac\lambda_1 + bs\lambda_2 = r, \quad as\lambda_1 - bc\lambda_2 = 0. \tag{19}$$

The values of z_1 and z_2 can be found from the following after a and c are determined:

$$z_1 = bc\lambda_1 - as\lambda_2, \quad z_2 = bs\lambda_1 + ac\lambda_2. \tag{20}$$

Next, by substituting $\boldsymbol{\Lambda} = \mathbf{W}\mathbf{R}_r\mathbf{X}^H$ into the SVD for $\mathbf{H}$, we have

$$\begin{aligned} \mathbf{H} &= \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^H = \mathbf{U}\mathbf{W}\mathbf{R}_r\mathbf{X}^H\mathbf{V}^H \\ &= (\mathbf{U}\mathbf{W})\mathbf{R}_r(\mathbf{V}\mathbf{X})^H = \mathbf{P}_r\mathbf{R}_r\mathbf{Q}_r^H. \end{aligned} \tag{21}$$

Since $\mathbf{U}$, $\mathbf{W}$, $\mathbf{V}$, and $\mathbf{X}$ are unitary, $\mathbf{P}_r$ and $\mathbf{Q}_r$ are also unitary.

B. Generating Multiple Unitary Matrices P_r and Q_r

To generate multiple $\mathbf{P}_r$ and $\mathbf{Q}_r$ matrices, we introduce a phase rotation matrix $\mathbf{M}_\theta$ to the SVD of $\mathbf{H}$ in (21). $\mathbf{M}_\theta$ is a diagonal matrix whose elements have unity gain, as shown in the following:

$$\mathbf{M}_\theta = \begin{bmatrix} e^{j\theta_1} & 0 \\ 0 & e^{j\theta_2} \end{bmatrix} \tag{22}$$

where θ_1 and θ_2 are direction parameters of any value from 0 to 2π. Since $\mathbf{M}_\theta\boldsymbol{\Lambda}\mathbf{M}_\theta^H = \boldsymbol{\Lambda}$, the inclusion of $\mathbf{M}_\theta$ and $\mathbf{M}_\theta^H$ into (21) does not change the value of $\mathbf{R}_r$, and (21) becomes

$$\begin{aligned} \mathbf{H} &= \mathbf{U}\mathbf{M}_\theta\boldsymbol{\Lambda}\mathbf{M}_\theta^H\mathbf{V}^H \\ &= \mathbf{U}\mathbf{M}_\theta(\mathbf{W}\mathbf{R}_r\mathbf{X}^H)\mathbf{M}_\theta^H\mathbf{V}^H \\ &= (\mathbf{U}\mathbf{M}_\theta\mathbf{W})\mathbf{R}_r(\mathbf{V}\mathbf{M}_\theta\mathbf{X})^H \\ &= \mathbf{P}_{\theta,r}\mathbf{R}_r\mathbf{Q}_{\theta,r}^H. \end{aligned} \tag{23}$$

From (23), it is clear that the values of $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ change with the values of θ_1 and θ_2 in $\mathbf{M}_\theta$ and r in $\mathbf{R}_r$. Moreover, since $\mathbf{U}$, $\mathbf{W}$, $\mathbf{V}$, $\mathbf{X}$, $\mathbf{M}_\theta$, and $\mathbf{M}_\theta^H$ are unitary matrices, $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ are also unitary matrices. If $\mathbf{H}$ is not a square matrix,

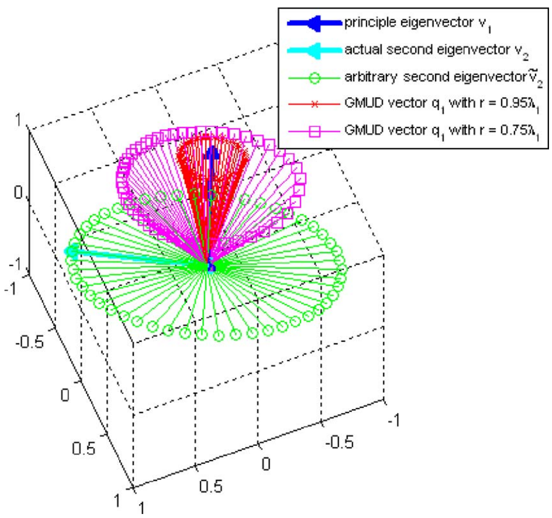

Fig. 3. Principal and second eigenvectors of SVD, arbitrary vectors generated from the principal eigenvector, and the first column vectors of the unitary matrix $\mathbf{Q}_{\theta,r}$ of GMUD with different $\mathbf{M}_i$. The (1, 1) element of the first column vector of $\mathbf{Q}_{\theta,r}$ matrix is made to be real by lumping the phase into the $\mathbf{R}$ matrices. The three axes are used to represent the real part of the (1, 1) element, and the real and imaginary parts of the (2, 1) element of the first column vector of the $\mathbf{Q}_{\theta,r}$ matrix. With the same r, different θ_{i1} or θ_{i2} produces different vectors in $\mathbf{Q}_{\theta,r}$, which form a cone whose center is the principal eigenvector of SVD.

the dimensions of matrices $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ will be different. For example, let the number of rows and columns of $\mathbf{H}$ be denoted as h_r and h_c, respectively; then, $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ will be $h_r \times h_r$ and $h_c \times h_c$, respectively [20].

GMUD is a generalized decomposition method that includes SVD, GMD, and GTD as special cases with specific $\mathbf{R}_r$ and $\mathbf{M}_\theta$ matrices. In the special case where θ in $\mathbf{M}$ is zero, GMUD reduces to GTD. Both GTD and GMUD allow the user to prescribe the diagonal values of the $\mathbf{R}_r$ matrices, but GTD only produces a pair of unitary matrices for a given channel $\mathbf{H}$, whereas GMUD, with different $\mathbf{M}_\theta$ matrices, produces multiple pairs of unitary matrices for a given channel $\mathbf{H}$. We will show later that this is useful for multi-user MIMO precoding because each orthonormal vector in the unitary matrices can represent one precoding vector for a user, and the GMUD precoding vectors for different users can be steered in some optimal manner.

C. Geometrical Interpretation of GMUD

Fig. 3 shows the characteristics of multiple $\mathbf{P}_{\theta,r}$'s and $\mathbf{Q}_{\theta,r}$'s plotted on a uniform 3-D sphere with equal axes, where it shows the principal eigenvector and the possible first column vector of the matrix $\mathbf{Q}_{\theta,r}$, obtained using θ_1 ranging from 0 to 2π. As shown in Fig. 3, they form a cone surrounding the principal eigenvector. The value of $r(\lambda_R < r < \lambda_1)$ varies inversely with the radius of the cone, i.e., the closer the value of r to the largest singular value λ_1, the smaller the cone radius. In Fig. 3, the lines with square markers represent GMUD vectors with a smaller $r = 0.75\lambda_1$, whereas the lines with cross markers represent GMUD vectors with larger $r = 0.95\lambda_1$. When r takes its largest value at λ_1, the first column vector of $\mathbf{Q}_{\theta,r}$ becomes the principal eigenvector.

With the use of $\mathbf{M}_\theta$, the exact "phase" of the second eigenvector becomes unimportant. This is because under GMUD decomposition, both $\mathbf{M}_\theta$ and the second eigenvector will generate the same cone in Fig. 3. This property leads to further reduction of the amount of required channel feedback (as only the principal eigenvector is required) in GMUD precoding systems, which is shown later.

IV. Precoding Based on Generalized Multi-Unitary Decomposition

Considering the scenario where the base station uses N_T transmit antennas to send K data streams to K users, we have one data stream per user, each with N_R receive antennas. The received signal per user is given in (1), and each user channel $\mathbf{H}_k$ can be decomposed using GMUD into

$$\mathbf{H}_k = \mathbf{P}_{\theta_k,r_k}\mathbf{R}_{r_k}(\mathbf{Q}_{\theta_k,r_k})^H \tag{24}$$

where multiple pairs of $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ matrices can be generated using (23).

The precoding matrix $\mathbf{G}$ at the transmitter is formed by a collection of precoding vectors, i.e., $\mathbf{g}_k$ for the kth user, as follows:

$$\mathbf{G} = [\mathbf{g}_1 \quad \cdots \quad \mathbf{g}_k \quad \cdots \quad \mathbf{g}_K]. \tag{25}$$

In this paper, we assign the first column vector of $\mathbf{Q}_{\theta_k,r_k}$, i.e., $\mathbf{q}_{1,\theta_k,r_k}$, as the precoding vector $\mathbf{g}_k$. The availability of multiple $\mathbf{Q}_{\theta_k,r_k}$ generated from GMUD is important because the base station can steer the precoding vectors of different users to optimize certain performance criteria. By changing the "*magnitude parameter*" r_k and *"direction parameter"* θ_k, different beamforming vectors "pointing" at different directions with different amplitudes can be obtained.

From (23), for user k, we have

$$\begin{aligned}\mathbf{Q}_{\theta_k,r_k} &= \mathbf{V}_k\mathbf{M}_{\theta_k}\mathbf{X}\\ &= \Big[\underbrace{\mathbf{q}_{1,\theta_k,r_k}}_{\mathbf{g}_k} \quad \cdots \quad \mathbf{q}_{K,\theta_k,r_k}\Big]\end{aligned} \tag{26}$$

where $\mathbf{q}_{1,\theta_k,r_k}$–$\mathbf{q}_{K,\theta_k,r_k}$ are the column vectors of $\mathbf{Q}_{\theta_k,r_k}$, and the first subscript denotes the column position.

In the case of $K = 2$, we denote the two users as user k and user l; the precoding matrix $\mathbf{G}$ used at the transmitter will have two vectors

$$\mathbf{G} = [\mathbf{g}_k \quad \mathbf{g}_l] = [\mathbf{q}_{1,\theta_k,r_k} \quad \mathbf{q}_{1,\theta_l,r_l}] \tag{27}$$

where $\mathbf{g}_k$ and $\mathbf{g}_l$ are the first column vectors of $\mathbf{Q}_{\theta,r}$ of user k and l, respectively. As explained in Section II-C, $\mathbf{G}$ will be optimized to minimize the sum of per user's inverse SINR.

From (23), for the user k

$$\begin{aligned}\mathbf{Q}_{\theta_k,r_k} &= \mathbf{V}\mathbf{M}_{\theta_k}\mathbf{X}\\ &= \begin{bmatrix} v_{11} & v_{12}\\ v_{21} & v_{22}\end{bmatrix}\begin{bmatrix} e^{j\theta_k} & 0\\ 0 & e^{j0}\end{bmatrix}\begin{bmatrix} c & s\\ -s & c\end{bmatrix}\\ &= \begin{bmatrix} v_{11}e^{j\theta_k}c - v_{12}s & v_{11}e^{j\theta_k}s + v_{12}c\\ v_{21}e^{j\theta_k}c - v_{22}s & v_{21}e^{j\theta_k}s + v_{22}c\end{bmatrix}\\ &= \Big[\underbrace{\mathbf{q}_{1,\theta_k,r_k}}_{\mathbf{g}_k} \quad \mathbf{q}_{2,\theta_k,r_k}\Big]\end{aligned} \tag{28}$$

the phase of the last eigenvector is not important to GMUD precoding as explained in Section III-C, hence for the 2×2 case, only the principal eigenvector $\mathbf{v}_{k,1}$ is needed. The transmitter at the base station then makes use of $\mathbf{v}_{k,1}$ to generate an arbitrary second eigenvector $\tilde{\mathbf{v}}_{k,2}$, which is orthogonal to $\mathbf{v}_{k,1}$, as it is not important in the optimization.

The precoding matrix vectors for the users k and l are, respectively

$$\mathbf{g}_k = \mathbf{q}_{1,\theta_k,r_k}, \quad \mathbf{g}_l = \mathbf{q}_{1,\theta_l,r_l}. \tag{29}$$

Combining (1), (24), and (27), the signal received by the user k becomes

$$\begin{aligned}\mathbf{y}_k &= \mathbf{P}_{\theta_k,r_k}\mathbf{R}_{r_k}(\mathbf{Q}_{\theta_k,r_k})^H \frac{\mathbf{Gu}}{\sqrt{\gamma}} + \mathbf{n}_k \\ &= \frac{1}{\sqrt{\gamma}}\mathbf{P}_{\theta_k,r_k}\mathbf{R}_{r_k}\begin{bmatrix}\mathbf{q}_{1,\theta_k,r_k}^H \\ \mathbf{q}_{2,\theta_k,r_k}^H\end{bmatrix}[\mathbf{g}_k \quad \mathbf{g}_l]\begin{bmatrix}u_k \\ u_l\end{bmatrix} + \mathbf{n}_k \\ &= \frac{1}{\sqrt{\gamma}}\mathbf{P}_{\theta_k,r_k}\mathbf{R}_{r_k}\begin{bmatrix}\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_k u_k + \mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_l u_l \\ \mathbf{q}_{2,\theta_k,r_k}^H\mathbf{g}_k u_k + \mathbf{q}_{2,\theta_k,r_k}^H\mathbf{g}_l u_l\end{bmatrix} + \mathbf{n}_k\end{aligned} \tag{30}$$

where $\gamma = \|\mathbf{Gu}\|^2$ is used to normalize the power of the transmitted signal.

Since $\mathbf{R}_{r_k}$ has a special form shown in (13) and (14), and $\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_k = 1$ and $\mathbf{q}_{2,\theta_k,r_k}^H\mathbf{g}_k = 0$, (30) can be reduced to

$$\begin{aligned}\mathbf{y}_k &= \frac{1}{\sqrt{\gamma}}\mathbf{P}_{\theta_k,r_k}\begin{bmatrix} r_k\left[\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_k u_k + \mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_l u_l\right] \\ \varepsilon\end{bmatrix} + \mathbf{n}_k \\ &= \frac{1}{\sqrt{\gamma}}\mathbf{P}_{\theta_k,r_k}\begin{bmatrix} r_k\left[u_k + \mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_l u_l\right] \\ \varepsilon\end{bmatrix} + \mathbf{n}_k\end{aligned} \tag{31}$$

where $\varepsilon = z_{21,k}u_k + (z_{21,k}\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_l + z_{22,k}\mathbf{q}_{2,\theta_k,r_k}^H\mathbf{g}_k)u_l$, $z_{21,k}$, and $z_{22,k}$ are the elements of $\mathbf{R}_{r_k}$ from (18)–(20), respectively. To detect the information symbol u_k, the user k can either use a zero-forcing receiver or an MMSE receiver. In this paper, we are using an MMSE receiver as it has been shown to be optimum for a MIMO system for any precoding matrix [11].

In (31), the useful signal and interference terms are $r_k u_k/\sqrt{\gamma}$ and $r_k\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{g}_l u_l/\sqrt{\gamma}$, respectively. To reduce the BER, the interference terms can be minimized by making $\mathbf{q}_{1,\theta_k,r_k}$ and $\mathbf{g}_l(\mathbf{g}_l = \mathbf{q}_{1,\theta_l,r_l})$ as orthogonal to each other as possible while taking the noise term $\mathbf{n}_k$ and normalization constant γ into account. We apply the same optimality criteria used in regularized-inverse precoding to optimize $\mathbf{G}$ as follows:

$$\begin{aligned}\mathbf{G} &= \underset{r_1\cdots r_K,\theta_1\cdots\theta_K}{\arg\min}\sum_{k=1}^{K}(\mathrm{SINR}_k)^{-1} \\ &= \underset{r_1\cdots r_K,\theta_1\cdots\theta_K}{\arg\min}\sum_{k=1}^{K}\left(\frac{|r_k|^2}{|r_k|^2\sum\limits_{\substack{j=1\\j\neq k}}^{K}\left|\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{q}_{1,\theta_j,r_j}\right|^2 + \sigma^2\gamma}\right)^{-1} \\ &= \underset{r_1\cdots r_K,\theta_1\cdots\theta_K}{\arg\min}\sum_{\substack{j=1\\j\neq k}}^{K}\left|\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{q}_{1,\theta_j,r_j}\right|^2 + \frac{\sigma^2\gamma}{|r_k|^2}.\end{aligned} \tag{32}$$

For the case of two users, the given expression reduces to

$$\begin{aligned}\mathbf{G} = \underset{r_k,r_l,\theta_k,\theta_l}{\arg\min}&\Big(\left|\mathbf{q}_{1,\theta_k,r_k}^H\mathbf{q}_{1,\theta_l,r_l}\right|^2 + \sigma^2\gamma/|r_k|^2 \\ &+ \left|\mathbf{q}_{1,\theta_l,r_l}^H\mathbf{q}_{1,\theta_k,r_k}\right|^2 + \sigma^2\gamma/|r_l|^2\Big) \\ \text{s.t.}\quad &\lambda_{1,k} > r_k > \lambda_{L,k},\ -\pi \geq \theta_i \geq \pi, \\ &\lambda_{1,l} > r_l > \lambda_{L,l},\ -\pi \geq \theta_i \geq \pi\end{aligned} \tag{33}$$

where $(\lambda_{1,k}, \lambda_{L,k})$ and $(\lambda_{1,l}, \lambda_{L,l})$ are the largest and smallest singular values of $\mathbf{H}_k$ and $\mathbf{H}_l$, respectively.

The magnitude parameters r_k and r_l determine the channel gains received by users k and l, respectively; however a larger r results in a smaller cone and less direction steering possibilities (see Fig. 3).

Since $\lambda_{R,k} < r_k < \lambda_{1,k}$ [see (13)], the optimization range of r_k for the user k depends on the spread of the channel singular values $(\lambda_{1,k} - \lambda_{L,k})$, which may be different for different users, leading to unequal performance among the users. To address this problem, we apply suitable power loading weights β_k and β_l, with values between 0 and 1, to scale $\mathbf{g}_k$ and $\mathbf{g}_l$, respectively. The final precoding matrix $\mathbf{G}$ hence becomes

$$\mathbf{G} = [\beta_k\mathbf{g}_k \quad \beta_l\mathbf{g}_l] \tag{34}$$

and the corresponding cost function becomes

$$\begin{aligned}\mathbf{G} = \underset{r_k,r_l,\theta_k,\theta_l,\beta_k,\beta_l}{\arg\min}&\left(\frac{\beta_l^2\left|\mathbf{q}_{1,\theta_k,r_k}^{\mathbf{H}}\mathbf{q}_{1,\theta_l,r_l}\right|^2}{\beta_k^2} + \frac{\sigma^2\gamma}{\beta_k^2|r_k|^2}\right. \\ &\left.+\frac{\beta_k^2\left|\mathbf{q}_{1,\theta_l,r_l}^{\mathbf{H}}\mathbf{q}_{1,\theta_k,r_k}\right|^2}{\beta_l^2} + \frac{\sigma^2\gamma}{\beta_l^2|r_l|^2}\right) \\ \text{s.t. } \lambda_{1,k} > r_k > \lambda_{L,k},\ &-\pi \geq \theta_i \geq \pi,\ \lambda_{1,l} > r_l > \lambda_{L,l}, \\ -\pi \geq \theta_i \geq \pi,\ &1 \geq \beta_k \geq 0,\ 1 \geq \beta_l \geq 0.\end{aligned} \tag{35}$$

The expected receiver output for user k can be written as

$$\eta_k = \left[\beta_k r_k \quad \beta_l r_k\mathbf{q}_{1,r_k,\theta_k}^H\mathbf{q}_{1,r_l,\theta_l}\right]\mathbf{u} + \sqrt{\gamma}n_k \tag{36}$$

and the normalized ouput is given as

$$\begin{aligned}\overline{\eta}_k &= \left[1 \quad \tfrac{\beta_l\mathbf{q}_k^*\mathbf{q}_l}{\beta_k}\right]\mathbf{u} + \frac{\sqrt{\gamma}}{\beta_k r_k}n_k \\ &= \upsilon\mathbf{u} + \frac{\sqrt{\gamma}}{\beta_k r_k}n_k.\end{aligned} \tag{37}$$

The expected values of $\overline{\eta}_k$ becomes

$$\begin{aligned}E\left(\overline{\eta}_k^2\right) &= E\left(\|\upsilon\mathbf{u}\|^2\right) + E\left(\left|\frac{\sqrt{\gamma}}{r_k}n_k\right|^2\right) \\ &= E(\upsilon\mathbf{u}\mathbf{u}^H\upsilon^H) + \frac{\gamma}{|\beta_k r_k|^2}\sigma^2 \\ &= \upsilon\upsilon^H + \frac{\gamma}{|\beta_k r_k|^2}\sigma^2 \\ &= \begin{bmatrix}1 & \frac{\beta_l}{\beta_k}\mathbf{q}_k^H\mathbf{q}_l\end{bmatrix}\begin{bmatrix}1 \\ \frac{\beta_l}{\beta_k}\mathbf{q}_l^H\mathbf{q}_k\end{bmatrix} + \frac{\gamma}{|\beta_k r_k|^2}\sigma^2 \\ &= 1 + \left|\frac{\beta_l}{\beta_k}\right|^2\left\|\mathbf{q}_k^H\mathbf{q}_l\right\|^2 + \frac{\gamma}{|\beta_k r_k|^2}\sigma^2\end{aligned} \tag{38}$$

and the resultant normalized SINR becomes

$$\mathrm{SINR}_k = \frac{1}{\left|\frac{\beta_l}{\beta_k}\right|^2 \left\|\mathbf{q}_k^H \mathbf{q}_l\right\|^2 + \frac{\gamma}{|\beta_k r_k|^2}\sigma^2} = \frac{|\beta_k|^2}{|\beta_l|^2 \left\|\mathbf{q}_k^H \mathbf{q}_l\right\|^2 + \frac{\gamma}{|r_k|^2}\sigma^2}. \quad (39)$$

Note that, in the given optimization process, the SVD left eigenvectors $\mathbf{U}$ in (23) are not needed. This shows that the proposed GMUD precoding only requires *partial CSI feedback*, in contrast with regularized-inverse precoding, which requires *full* CSI feedback of the $\mathbf{U}$, $\mathbf{V}$, and singular values. For GMUD precoding with two transmit antennas at the base station and two receive antennas per user, the feedback channel can be modeled as

$$\mathbf{H} = \mathbf{U\Lambda V}^H = \mathbf{U}\begin{bmatrix}\delta_1 & 0\\ 0 & \delta_2\end{bmatrix}[\boldsymbol{\nu}_1 \quad \boldsymbol{\nu}_2]^H. \quad (40)$$

The users only send back the singular values (denoted as δ_1 and δ_2) and the principal eigenvector (denoted as $\boldsymbol{\nu}_1$). The second eigenvector $\boldsymbol{\nu}_2$ is not fed back, but since $\boldsymbol{\nu}_2$ is normal to $\boldsymbol{\nu}_1$, the base station can generate an arbitrary second eigenvector $\tilde{\boldsymbol{\nu}}_2$ based on the principal eigenvector $\boldsymbol{\nu}_1$ using Givens rotation, as shown in the following.

The principal eigenvector $\boldsymbol{\nu}_1$ can be characterized as

$$\boldsymbol{\nu}_1 = [\cos\phi_1 \quad -\sin\phi_1 \exp(j\phi_2)]^T. \quad (41)$$

An arbitrary second eigenvector vector $\tilde{\boldsymbol{\nu}}_2$, which is normal to $\boldsymbol{\nu}_1$, can be formed as

$$\tilde{\boldsymbol{\nu}}_2 = \exp(j\varphi)[\sin\phi_1 \quad \cos\phi_1 \exp(j\phi_2)]^T \quad (42)$$

where φ can take any value from 0 to 2π.[2] Finally, the arbitrary matrix $\tilde{\mathbf{V}}$ is formed as follows:

$$\tilde{\mathbf{V}} = [\boldsymbol{\nu}_1 \quad \tilde{\boldsymbol{\nu}}_2]. \quad (43)$$

The steering matrix $\mathbf{M}_\theta$ will still find the optimum first column vector $\mathbf{q}_{1,\theta,r}$ used to form the precoding matrix $\mathbf{G}$ in (32) and (35), although a different arbitrary matrix $\tilde{\mathbf{V}}$ is used in place of the actual $\mathbf{V}$ at the precoder.

The geometrical relationship between $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$, and $\tilde{\boldsymbol{\nu}}_2$ is shown in Fig. 3. The arbitrary second eigenvectors, including the actual second eigenvector, form a plane normal to the principal eigenvector of SVD. For a general $n \times n$ case, by using the definition of the Givens rotations in [19], we need $n-1$ eigenvectors, and the last arbitrary eigenvector can be calculated from the $n-1$ eigenvectors.

V. Simulation Results

With finite feedback channel resource, it is desirable to use a precoding scheme that requires less CSI and is robust to CSI quantization. Here, the BER performance and CSI resource requirement of a MIMO system using the proposed regularized-inverse precoding with antenna selection (see Section II-B and C), and the proposed GMUD precoding with power loading (see Section IV) are compared with the SVD precoding scheme. The MIMO base station is equipped with either two or three transmit antennas. It serves each user with one data stream. All users have two or three receive antennas each. All users are allocated equal data rates, and their signals are transmitted simultaneously. All the simulation results (which includes finding the optimized $\mathbf{G}$ in this paper) are obtained using Matlab simulation software and its optimization toolbox with the fmincon solver.

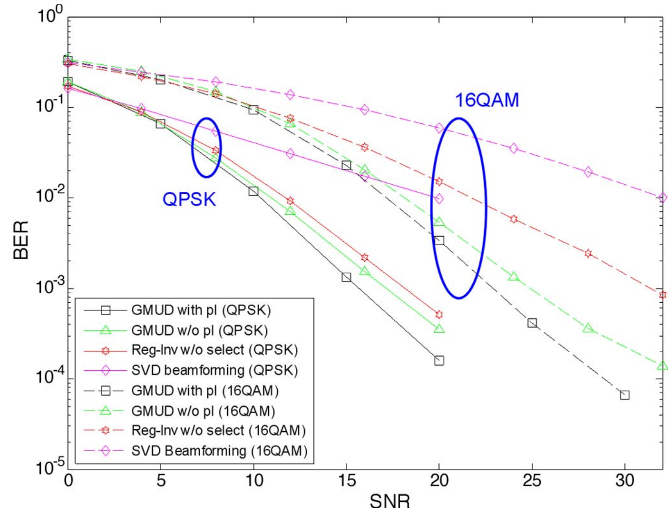

Fig. 4. Probability of bit error of a two-user MIMO system (a base station with two transmit antennas and two users with two receive antennas each) applying various precoding schemes with an MMSE receiver. "pl" denotes power loading.

A. GMUD Precoding With Power Loading Versus Without Power Loading

We first show the impact of power loading in (34) on the precoding optimization in (35). These weights compensate for the difference between the channel gains of different users, which are determined by the magnitude parameter r in (16).

Fig. 4 compares the BER of the various precoding schemes in a Rayleigh flat-fading downlink channel with an equal SNR. It shows that the proposed GMUD precoding with power loading [see (35)] performs better than without power loading [see (33)]. Furthermore, for 16-quadratic-amplitude modulation (QAM) transmission at BER $= 10^{-3}$, GMUD precoding with power loading has a performance gain of 8.8 dB over the conventional regularized-inverse precoding and much better than the SVD beamforming. The SVD beamforming precoder is generated by using the principal eigenvectors of the users with optimized power loading. Similar observations can be made for the QPSK results. Therefore, in the following, we will only consider GMUD with power loading.

In Fig. 5, we compare the BER of the GMUD and regularized-inverse precoding in a three-user MIMO system with three transmit antennas at the base station and three receive antennas per user. All the users are allocated equal rates, and the information is transmitted simultaneously from the three transmit antenna at the base station.

[2]The value of φ of the exact second eigenvector is either 0 or π

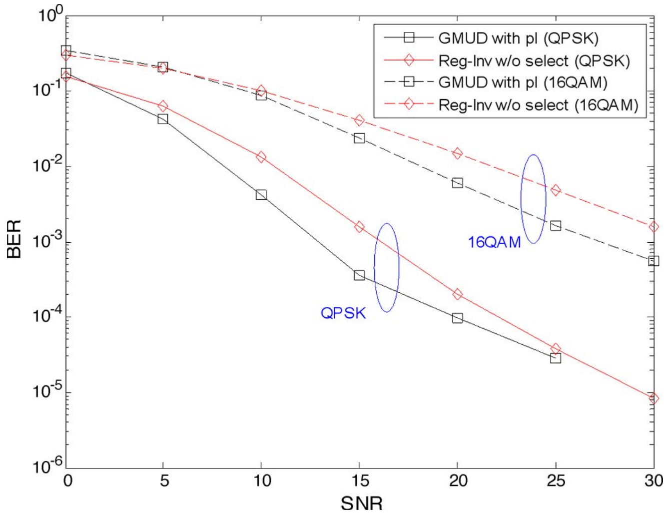

Fig. 5. Probability of bit error of a three-user MIMO system (a base station with three transmit antennas and three users with three receive antennas each) applying various precoding schemes with an MMSE receiver.

B. GMUD Precoding Versus Regularized-Inverse Precoding With Antenna Selection

Here, we compare the two proposed schemes, i.e., regularized-inverse precoding with optimal antenna selection and GMUD precoding, with CSI quantization. To quantify the amount of CSI feedback across different precoding schemes, we view the channel coefficient between a transmit antenna and a receive antenna as an element in the channel matrix $\mathbf{H}$. For the case of regularized-inverse precoding, the channel $\mathbf{H}$ can be represented in the form of

$$\mathbf{H} = \begin{bmatrix} \mathbf{h_1} \\ \vdots \\ \mathbf{h_{N_T}} \end{bmatrix} = \begin{bmatrix} \delta_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \delta_{N_T} \end{bmatrix} \begin{bmatrix} \boldsymbol{\nu}_1^{\mathbf{T}} \\ \vdots \\ \boldsymbol{\nu}_{N_T}^{\mathbf{T}} \end{bmatrix} \quad (44)$$

where δ are normalization constants (norm of the vectors) for the respective $\mathbf{h}$, and $\boldsymbol{\nu}$ are the resultant unit vectors, respectively. To feedback $\mathbf{H}$, the users need to send back all the unit vectors and their normalization constants. However, the normalized $\mathbf{H}$ is not a unitary matrix. Alternatively, the users can feedback the singular values and the right and left eigenvectors after performing SVD to channel $\mathbf{H}$. For the case of GMUD precoding, the users only send back the singular values and the right eigenvectors, and not the left eigenvectors. The feedback can be further reduced to the singular values and the principal eigenvector when the channel dimension is 2 by 2, as explained in Section IV.

Considering each of the real and imaginary parts of δ and every element of $\boldsymbol{\nu}$ as a scalar element, the total number of scalar CSI elements required to be fed back by different precoding schemes is compared in Table I. It shows that the GMUD precoding needs to feed back less scalar elements than the regularized-inverse precoding with antenna selection. This saving will be significant when the number of transmit and receive antennas are large.

To facilitate a fair BER comparison of different precoding schemes under the effect of CSI quantization, we allocate

TABLE I
ALLOCATION OF $12N$ BITS FOR CSI FEEDBACK FOR DIFFERENT PRECODING SCHEMES (FOR TWO USERS)

Precoding scheme \ Feedback Info	Total no. of scalar elements for all δ (quantization bits allocated per δ)	Total no. of scalar elements for all vectors $\boldsymbol{\nu}$ (quantization bits allocated per scalar element of $\boldsymbol{\nu}$)
Reg-Inv with antenna selection	2 ($2N$ bits) [0 , 4]*	8 (N bits) [-1, 1]*
GMUD precoding	2 ($2N$ bits) [0, 5]*	4 ($2N$ bits) [-1, 1]*

* [*a*, *b*] indicates the range of numerical values

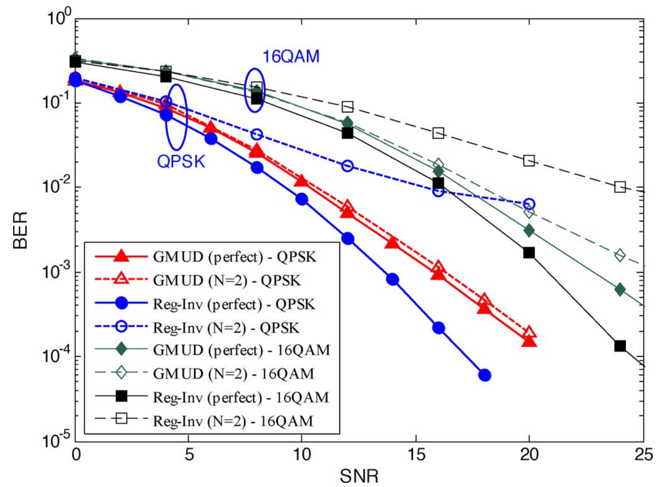

Fig. 6. Probability of bit error of a two-user MIMO system with quantized CSI for precoding and an MMSE receiver for detection (a base station with two transmit antennas and two users with two receive antennas each).

the same amount of feedback bits to all precoding schemes. Specifically, we use a total of $12N$ bits to quantize all the necessary CSI elements, where N is an integer. For regularized-inverse precoding with two-antenna selection, $12N$ bits in total correspond to $2N$ and N bits for each δ and each scalar element of $\boldsymbol{\nu}$, respectively. For GMUD precoding, $12N$ bits in total correspond to $2N$ bits for each CSI element.

The Rayleigh flat-fading BER performance (for a two-user scenario) of the two proposed precoding schemes with perfect feedback and quantized feedback are shown in Fig. 6 for both QPSK and 16-QAM transmissions. The required SNR in decibels to achieve BER $= 10^{-3}$ by the two precoding schemes with different amount of CSI quantization feedback are shown in Table II. For the case of perfect CSI, it is noted that the regularized-inverse precoding with antenna selection scheme performs better than the GMUD scheme. However, with increasing CSI quantization (quantified by decreasing the total CSI feedback bits from ∞ to 48, 36, and 24 bits), the performance of regularized-inverse precoding deteriorates rapidly. For example, Table II shows that the performance of two-user regularized-inverse 16-QAM precoding with 48 feedback bits $(N = 4)$ used requires an SNR of 22.4 dB at BER $= 10^{-3}$, i.e., an increase of 1.55 dB from the perfect CSI case. This SNR penalty increases to 5.05 dB when 36 bits $(N = 3)$ is

TABLE II
REQUIRED SNR TO ACHIEVE BER $= 10^{-3}$ OF DIFFERENT PRECODING SCHEMES WITH QUANTIZED CSI (USING THE MMSE RECEIVER)

Precoding \ Feedback Info		Required SNR to achieve BER $= 10^{-3}$ using QPSK (16QAM) transmission			
		$N \to \infty$ {un-quantized}	$N=4$	$N=3$	$N=2$
Reg-Inv with antenna selection	2 users	13.65dB (20.85dB)	14.25dB (22.40dB)	16.10dB (25.90dB)	error floor at BER≈0.05
	3 users	10.65dB (17.25)	10.75dB (19.30dB)	13.20dB (24.80dB)	error floor at BER≈0.05
GMUD precoding	2 users	15.81dB (22.79dB)	15.82dB (22.80dB)	15.83dB (23.33dB)	16.25dB (25.70dB)
	3 users	12.90dB (26.95dB)	13.90dB (27.20dB)	14.35dB (29.95dB)	17.50dB (30.30dB)

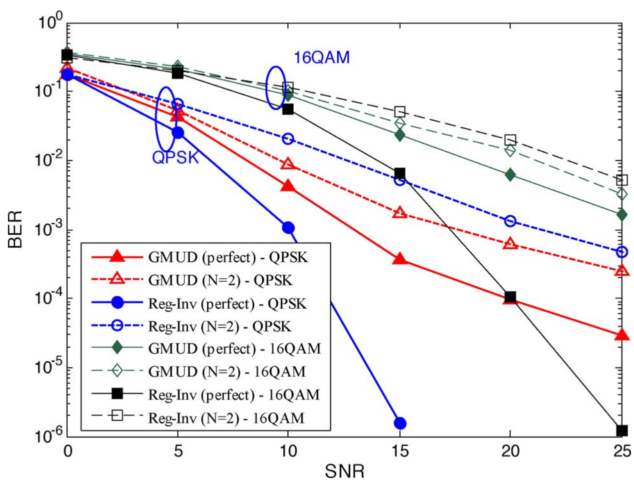

Fig. 7. Probability of bit error of a three-user MIMO system with quantized CSI for precoding and an MMSE receiver for detection (a base station with three transmit antennas and three users with three receive antennas each).

used. Eventually, a very high error floor occurs when 24 bits (or $N = 2$) are used. In sharp contrast, the GMUD precoding performance does not vary much with CSI quantization. Even when only two bits per CSI element are used (corresponding to 24 total quantization bits), the required SNR to attain BER $= 10^{-3}$ only increases by 0.44 dB compared with the unquantized CSI case. All in all, they show that GMUD precoding is much more insensitive to CSI quantization.

Next, the quantized CSI simulations are extended to a three-user MIMO system with three transmit antennas at the base station and three receive antennas per user in Fig. 7, using the same simulation parameters as in Fig. 6. All the users are allocated equal rates, and information is transmitted simultaneously from all three transmit antennas at the base station. Once again, the results show that the GMUD precoding system performance is highly robust, whereas the regularized-inverse precoding performance is highly sensitive to CSI quantization.

Similar results can be observed for a four-user MIMO system with two transmit antennas at the base station and three receive antennas per user in Fig. 8.

The remarkable robustness of GMUD precoding to CSI quantization can be explained as follows. Recall that GMUD

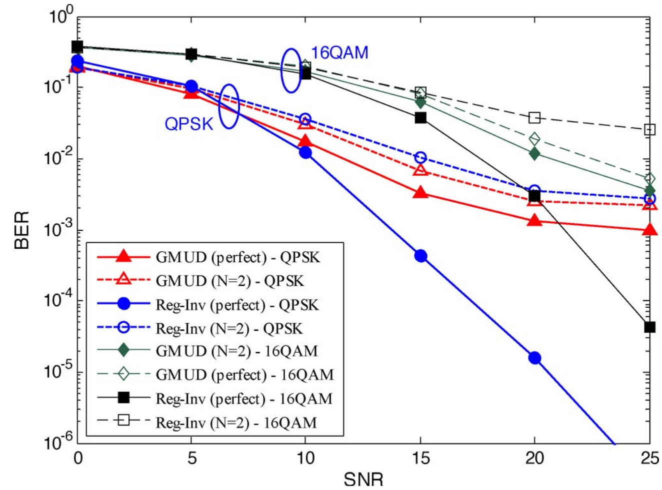

Fig. 8. Probability of bit error of a four-user MIMO system with quantized CSI for precoding and an MMSE receiver for detection (a base station with four transmit antennas and four users with two receive antennas each).

precoding uses the singular values and the first eigenvector of the MIMO channel per user. The singular values define the range of r in (14), which is subsequently used for the precoding matrix optimization in (33) or (35). When these singular values become less accurate due to quantization, it affects the search range of r but does not directly affect the optimization outcome. On the other hand, when the first eigenvector of a user becomes more quantized, it shifts the center of the search space for that user (the cones in Fig. 3). This may result in a smaller overall optimization search space for all users but again may not lead to incorrect optimization outcomes. As a result, the GMUD precoder in (33) or (35) may still be able to find near-optimal precoding vectors for all users. Moreover, as GMUD precoding requires fewer CSI elements to be fed back, more quantization bits per CSI element are available to GMUD precoding.

Regularized-inverse precoding outperforms GMUD precoding where there are large amounts of feedback. However, when the feedback becomes limited, GMUD precoding shows its robustness and lack of sensitivity to CSI quantization. In the scenario where the base station can afford higher complexity and the feedback bandwidth is limited, GMUD precoding is the preferred solution over regularized-inverse precoding.

VI. CONCLUSION

In this paper, we presented two precoding techniques for MIMO-BCs with multiple receive antennas per user. The first precoding technique performs receive antenna selection based on regularized-inverse precoding. The second technique is based on a new MIMO channel decomposition method called GMUD. GMUD decomposes complex matrix $\mathbf{H}$ into $\mathbf{H} = \mathbf{P}_{\theta,r}\mathbf{R}_r\mathbf{Q}_{\theta,r}^H$, where $\mathbf{R}_r$ is a special matrix whose first row contains a nonzero value only at the leftmost position, and $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ are a pair of unitary matrices. A unique feature of GMUD is that it gives multiple solutions of the unitary $\mathbf{P}_{\theta,r}$ and $\mathbf{Q}_{\theta,r}$ matrices, which present additional flexibility for precoder optimization.

We show that, although the regularized-inverse precoding with antenna selection performs well, it requires full CSI feedback and is very sensitive to CSI quantization. On the other hand, the GMUD precoding requires only partial CSI feedback (minus left eigenvectors of the user channels) and is robust against CSI quantization. Simulation results show, that with 16-QAM modulation and MMSE detection, the required E_b/N_o at $\mathrm{BER} = 10^{-3}$ for two-user and three-user GMUD precoding systems varies by less than 3 and 4 dB, respectively, when the CSI varies from unquantized to 4-bit quantization per scalar channel coefficient.

REFERENCES

[1] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *J. Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[2] E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[3] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.

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

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

[6] W. Yu and J. Cioffi, "Sum capacity of a Gaussian vector broadcast channel," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2002, p. 498.

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

[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, Mar. 2006.

[9] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multi-antenna multi-user communication—Part I: Channel inversion and regularization," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 195–202, Jan. 2005.

[10] N. Jindal, "A feedback reduction technique for MIMO broadcast channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2006, pp. 2699–2703.

[11] C. Yuen and B. M. Hochwald, "Achieving near-capacity at Low SNR on a multiple-antenna multiple-user channel," *IEEE Trans. Commun.*, vol. 57, no. 1, pp. 69–74, Jan. 2009.

[12] M. Morelli and L. Sanguinetti, "A novel prefiltering technique for downlink transmissions in TDD MC-CDMA systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2064–2069, Sep. 2005.

[13] D. P. Palomar, J. M. Cioffi, and M. A. Lagunas, "Joint Tx-Rx beamforming design for multicarrier MIMO channels: A unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sep. 2003.

[14] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[15] J.-K. Zhang, A. Kavcic, and K. M. Wong, "Equal-diagonal QR decomposition and its application to precoder design for successive-cancellation detection," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 154–172, Jan. 2005.

[16] Y. Jiang, W. W. Hager, and J. Li, "The geometric mean decomposition," *Linear Algebra Appl.*, vol. 396, pp. 373–384, Feb. 2005.

[17] Y. Jiang, W. W. Hager, and J. Li, "The generalized triangular decomposition," *Math. Comput.*, vol. 77, no. 262, pp. 1037–1056, Jun. 2008.

[18] W. H. Chin and C. Yuen, "Design of differential quantization for low bitrate channel state information feedback in MIMO-OFDM systems," in *Proc. IEEE VTC*, 2008, pp. 827–831.

[19] C. Yuen, S. Sun, and M. M. S. Ho, "Beamforming matrix quantization with variable feedback rate," in *Proc. IEEE PIMRC*, 2008, pp. 1–5.

[20] W. S. Chua, C. Yuen, Y. L. Guan, and F. Chin, "Multi-user SISO precoding based on generalized multi-unitary decomposition for single carrier transmission in frequency selective channel," in *Proc. IEEE WCNC*, Apr. 2009, pp. 1–5.

Wee Seng Chua received the B.Eng. degree from Nanyang Technological University (NTU), Singapore, in 2005. He is currently working toward the Ph.D. degree with NTU.

His research interests include multiuser multiple-input–multiple-ouput precoding and perturbation.

Chau Yuen (SM'12) received the B.Eng. and Ph.D. degrees from Nanyang Technological University, Singapore, in 2000 and 2004, respectively.

In 2005, he was a Postdoctoral Fellow with Lucent Technologies Bell Labs, Murray Hill, NJ. In 2008, he was a Visiting Assistant Professor with Hong Kong Polytechnic University, Kowloon, Hong Kong. From 2006 to 2010, he was with the Institute for Infocomm Research, A*STAR, Singapore, as a Senior Research Engineer. Since June 2010, he has been with Singapore University of Technology and Design, Singapore, as an Assistant Professor. He is the author of more than 100 research papers in international journals or conference proceedings. His current research interests include green communications, cognitive networks, smart grids, network coding, and distributed storage.

Dr. Yuen serves as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He received the Seventh IEEE Asia-Pacific Outstanding Young Researcher Award from the IEEE Communication Society.

Yong Liang Guan (M'99) received the Bachelor of Engineering degree (with first-class honors) from National University of Singapore, Singapore, and the Ph.D. degree from Imperial College of London, London, U.K.

He is currently an Associate Professor and the Head of the Communication Engineering Division with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore. His research interests include modulation, coding, and signal processing for communication systems and information security systems.

Dr. Guan is an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.

Francois Chin received the Ph.D. degree in electrical engineering from National University of Singapore, Singapore.

Since 1995, he has been with the Institute for Infocomm Research, A*STAR, Singapore, where he is currently a Program Manager leading a research and development team in millimeter-wave (mmWave) wireless chipset development and future broadband wireless access technologies. He is actively contributing to various standardization activities/efforts, including Third-Generation Partnership Project Long-Term Evolution, the WiGig Alliance for mmWave Wireless Personal Area Networks (WPAN), the IEEE 802.11ad Enhancements for Very High Throughput in the mmWave Band, the IEEE 802.11cmmw Enhancements for Very High Throughput in the China mmWave Band, the WiMedia Alliance for Ultrawideband (UWB) WPAN, the IEEE 802.15.3a UWB Alternate PHY for High Rate, the IEEE 802.15.4a UWB Alternate PHY for Low-Rate Low-power WPAN, and the IEEE 802.22 Wireless Regional Area Networks. He is the author of over 180 papers in well-known international journals and technical conferences and over 60 technical documents in wireless-communications-related standards. He has 30 pending patents.

Dr. Chin was a General Cochair for the 2010 IEEE Dynamic Spectrum Access Networks Conference and the International Conference on Software Testing, Verification and Validation (ICST) CrownCom 2008 Conference. He was a member of the Radio Standard Technical Committee, Infocomm Development Authority (Singapore). He received the Singapore IT Federation 2012 Special Mention Award (Emerging Technologies category) for his next-generation mmWave-based Wi-Fi chipset; the Institution Engineering of Singapore Prestigious Engineering Achievement Awards, in 2005 and 2007, for his contributions to the IEEE 802.15.4b and IEEE 802.22 Standards, respectively; and the 50th IEEE Vehicular Technology and Communications Conference Best Paper Award.