

# Antenna Grouping Based Feedback Compression for FDD-Based Massive MIMO Systems

Byungju Lee, *Member, IEEE*, Junil Choi, *Member, IEEE*, Ji-Yun Seol, *Member, IEEE*,  
David J. Love, *Fellow, IEEE*, and Byonghyo Shim, *Senior Member, IEEE*

**Abstract**—Recent works on massive multiple-input multiple-output (MIMO) have shown that a potential breakthrough in capacity gains can be achieved by deploying a very large number of antennas at the base station. In order to achieve the performance that massive MIMO systems promise, accurate transmit-side channel state information (CSI) should be available at the base station. While transmit-side CSI can be obtained by employing channel reciprocity in time division duplexing (TDD) systems, explicit feedback of CSI from the user terminal to the base station is needed for frequency division duplexing (FDD) systems. In this paper, we propose an antenna grouping based feedback reduction technique for FDD-based massive MIMO systems. The proposed algorithm, dubbed antenna group beamforming (AGB), maps multiple correlated antenna elements to a single representative value using predesigned patterns. The proposed method modifies the feedback packet by introducing the concept of a *header* to select a suitable group pattern and a *payload* to quantize the reduced dimension channel vector. Simulation results show that the proposed method achieves significant feedback overhead reduction over conventional approach performing the vector quantization of whole channel vector under the same target sum rate requirement.

**Index Terms**—Massive multiple-input multiple-output, antenna group beamforming, feedback reduction, vector quantization, Grassmannian subspace packing.

## I. INTRODUCTION

MULTIPLE-INPUT multiple-output (MIMO) systems with large-scale transmit antenna arrays, often called massive MIMO, have been of great interest in recent years because of their potential to dramatically improve spectral efficiency of future wireless systems [3], [4]. By employing a

simple linear precoding in the downlink and receive filtering in the uplink, massive MIMO systems can control intra-cell interference and thermal noise [3]. Additionally, massive MIMO can improve the power efficiency by scaling down the transmit power of each terminal inversely proportional to the number of base station antennas for uplink [4].

Presently, standardization activity for massive MIMO has been initiated [5], [6], and there is on-going debate regarding the pros and cons of time division duplexing (TDD) and frequency division duplexing (FDD). In obtaining the channel state information (CSI), FDD requires the CSI to be fed back through the uplink [7] while no such procedure is required for TDD systems owing to the channel reciprocity [8]. In fact, under the assumption that RF chains are properly calibrated [9], the CSI of the downlink can be estimated using the pilot signal in the uplink. Due to this benefit, most of the massive MIMO works in the literature have focused on TDD [10] (possible exceptions are [11]–[19]). However, FDD dominates current cellular networks and offers many benefits over TDD (e.g., small latency, continuous channel estimation, backward compatibility), and it is important to identify and develop solutions for potential issues arising from FDD-based massive MIMO techniques.

One well-known problem of FDD system is that the amount of CSI feedback must scale linearly with the number of antennas to control the quantization error [20]–[23]. Therefore, it is not hard to convince oneself that the overhead of CSI feedback is a serious concern in the massive MIMO regime. Needless to say, a technique that efficiently reduces the feedback overhead while affecting minimal impact on system performance is crucial to the success of FDD-based massive MIMO systems.

In this paper, we provide a novel framework for FDD-based massive MIMO systems that achieves a reduction in the CSI feedback overhead by exploiting the spatial correlation among antennas. The proposed algorithm, henceforth dubbed antenna group beamforming (AGB), maps multiple correlated antenna elements to a single representative value using properly designed grouping patterns. When the antenna elements are correlated, the loss caused by grouping antenna elements is shown to be small, meaning that grouping of antenna elements with correlated channels is an effective means of reduced dimension channel vector generation. In fact, by allocating a small portion of the feedback resources to represent the grouping pattern, the number of bits required for channel vector quantization can be reduced substantially, resulting in a significant reduction in feedback overhead. In order to support the antenna grouping operation, the proposed AGB algorithm uses a new feedback

Manuscript received December 20, 2014; revised April 23, 2015 and June 17, 2015; accepted July 14, 2015. Date of publication July 24, 2015; date of current version September 3, 2015. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIP) (No. 2014R1A5A1011478 and 2014049151) and MSIP (Ministry of Science, ICT & Future Planning), Korea in the ICT R&D Program 2013 (No. 1291101110-130010100). This paper was presented in part at the International Conference on Communications (ICC), 2014 [1] and Vehicular Technology Conference (VTC), 2014 [2]. The associate editor coordinating the review of this paper and approving it for publication was M. Matthaiou.

B. Lee and B. Shim are with Institute of New Media and Communications, School of Electrical and Computer Engineering, Seoul National University, Seoul 151-742, Korea (e-mail: bjlee@islab.snu.ac.kr; bshim@snu.ac.kr).

J. Choi is with the University of Texas at Austin, Austin, TX 78712 USA (e-mail: junil.choi@utexas.edu).

J.-Y. Seol is with Samsung Electronics Co., Ltd., Suwon 443-742, Korea (e-mail: jiyun.seol@samsung.com).

D. J. Love is with School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: djlove@purdue.edu).

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

Digital Object Identifier 10.1109/TCOMM.2015.2460743

packet structure that divides the feedback resources into two parts: a *header* to indicate the antenna group pattern and a *payload* to indicate the codebook index of the reduced dimension channel vector. At the user terminal, a pair of group pattern and codeword minimizing the quantization distortion is chosen. Using the information delivered from the user terminal, the base station reconstructs the full-dimensional channel vector and then performs transmit beamforming.

In our analysis, we show that when the transmit antenna elements are correlated, the proposed AGB algorithm exhibits smaller quantization distortion than the conventional vector quantization employing a channel statistic-based codebook. This in turn implies that the number of quantization bits required to meet a certain level of the performance for the AGB algorithm is smaller than that of conventional vector quantization under the same level of quantization distortion. We also investigate an estimated required number of feedback bits to maintain a constant gap with respect to the system with perfect CSI. It is shown that the use of antenna grouping in correlated channels enables to considerably reduce the amount of feedback overhead. Moreover, due to the fact that dimension of the codeword being searched is reduced, and hence the proposed AGB brings additional benefits in search complexity over the conventional vector quantization. We confirm by simulation on realistic massive MIMO channels that the proposed AGB algorithm achieves up to 20%~70% savings in feedback information over the conventional vector quantization under the same target sum rate requirement.

The remainder of this paper is organized as follows. In Section II, we briefly review the system model and the conventional beamforming technique. In Section III, we provide a detailed description of the proposed AGB algorithm and subspace packing based grouping pattern generation scheme. We present the simulation results in Section IV and present conclusions in Section V.

**Notations:** Lower and upper boldface symbols are used to denote vectors and matrices, respectively. The superscripts  $(\cdot)^H$ ,  $(\cdot)^T$ , and  $(\cdot)^*$  denote Hermitian transpose, transpose, and conjugate, respectively.  $\|\mathbf{X}\|$  and  $\|\mathbf{X}\|_F$  are used as the two-norm and the Frobenius norm of a matrix  $\mathbf{X}$ , respectively.  $E[\cdot]$  denotes the expectation operation, and  $\mathcal{CN}(m, \sigma^2)$  indicates a complex Gaussian distribution with mean  $m$  and variance  $\sigma^2$ .  $\text{tr}(\cdot)$  is the trace operation, and  $\text{vec}(\mathbf{X})$  is the vectorization of matrix  $\mathbf{X}$ . Let  $\mathbf{X}_\Lambda \in \mathbb{C}^{|\Lambda| \times |\Lambda|}$  denote a submatrix of  $\mathbf{X}$  whose  $(i, j)$ -th entry is  $\mathbf{X}(\Lambda(i), \Lambda(j))$  for  $i, j = 1, \dots, |\Lambda|$  ( $\Lambda$  is the set of partial indices and  $|\Lambda|$  is the cardinality of  $\Lambda$ ).

## II. MIMO BEAMFORMING

### A. System Model and Conventional Beamforming

We consider a multiuser multiple-input single-output (MISO) downlink channel with  $N_t$  antennas at the base station and  $K$  user terminals each with a single antenna<sup>1</sup> (see Fig. 1). We assume spatially correlated and temporally correlated

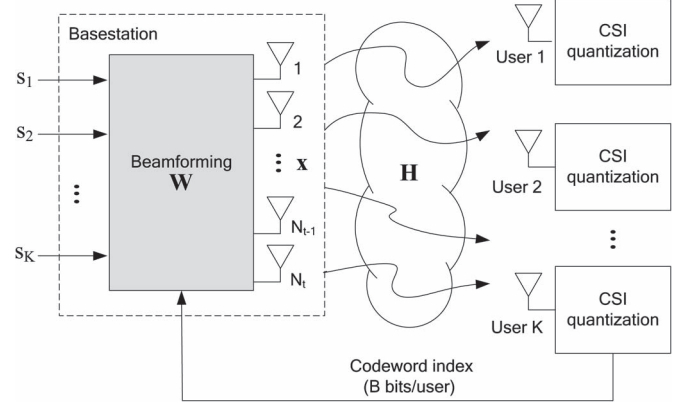


Fig. 1. CSI feedback in the multi-user downlink system.

block-fading channels where the channel vector  $\mathbf{h}_{i,\ell}$  follows the first-order Gauss-Markov model as

$$\mathbf{h}_{k,0} = \mathbf{R}_{t,k}^{\frac{1}{2}} \mathbf{g}_{k,0}$$

$$\mathbf{h}_{k,\ell} = \eta \mathbf{h}_{k,\ell-1} + \sqrt{1 - \eta^2} \mathbf{R}_{t,k}^{\frac{1}{2}} \mathbf{g}_{k,\ell}, \quad \ell \geq 1$$

where  $\mathbf{R}_{t,k} \in \mathbb{C}^{N_t \times N_t}$  is the transmit correlation matrix of the  $k$ -th user [24] and  $\mathbf{g}_{k,\ell} \in \mathbb{C}^{N_t}$  is the innovation process whose elements are independent and identically distributed according to  $\mathbf{g}_{k,\ell} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_t})$  and  $\eta$  is a temporal correlation coefficient ( $0 \leq \eta \leq 1$ ). We assume the block-fading channel has a coherence time of  $L$ , which means that the channel is static for  $L$  channel uses in each block and changes from block-to-block. In this setup, the received signal of the  $k$ -th user for the  $n$ -th channel use in the  $\ell$ -th fading block can be expressed as

$$y_{k,\ell}[n] = \mathbf{h}_{k,\ell}^H \mathbf{w}_{k,\ell} s_{k,\ell}[n] + \mathbf{h}_{k,\ell}^H \sum_{j \neq k} \mathbf{w}_{j,\ell} s_{j,\ell}[n] + z_{k,\ell}[n] \quad (1)$$

where  $\mathbf{h}_{k,\ell} \in \mathbb{C}^{N_t}$  is the channel vector from the base station antenna array to the  $k$ -th user,  $\mathbf{w}_{i,\ell} \in \mathbb{C}^{N_t}$  is the unit norm beamforming vector ( $\|\mathbf{w}_{i,\ell}\|^2 = 1$ ),  $s_{i,\ell}[n] \in \mathbb{C}$  is the message signal for the  $i$ -th user, and  $z_{k,\ell}[n] \sim \mathcal{CN}(0, 1)$  is normalized additive white Gaussian noise at the  $k$ -th user. Since the beamforming is performed separately per block, in the sequel we focus on the operation of a single block and drop the fading block index  $\ell$ . The matrix-vector form of (1) is expressed as

$$\mathbf{y}[n] = \mathbf{H} \mathbf{x}[n] + \mathbf{z}[n] \quad (2)$$

where  $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_K]^H \in \mathbb{C}^{K \times N_t}$  is the composite channel matrix,  $\mathbf{z}[n] = [z_1[n] z_2[n] \dots z_K[n]]^T \in \mathbb{C}^K$  is the complex Gaussian noise vector ( $\mathbf{z}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ ),  $\mathbf{x}[n]$  is the transmit vector normalized with the power constraint  $E(\|\mathbf{x}[n]\|^2) = P$ , and  $\mathbf{y}[n] = [y_1[n] y_2[n] \dots y_K[n]]^T$  is the vectorized received signal vector. In order to control the inter-user interference, beamforming is applied using  $\mathbf{x}[n] = \mathbf{W} \mathbf{s}[n]$  where  $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_K] \in \mathbb{C}^{N_t \times K}$  and  $\mathbf{s}[n] = [s_1[n] s_2[n] \dots s_K[n]]^T \in \mathbb{C}^K$  are the beamforming matrix and the message vector, respectively.

In generating the beamforming vectors, we consider zero-forcing beamforming (ZFBF) [25]–[29] where the right pseudo

<sup>1</sup>Note that the proposed method can be easily extended to a MIMO scenario by vectorizing the channel vector corresponding to each receive antenna. For simplicity, we consider the MISO setup for the rest of this paper.

inverse<sup>2</sup>  $\hat{\mathbf{W}}_{zf} = \hat{\mathbf{H}}^H (\hat{\mathbf{H}} \hat{\mathbf{H}}^H)^{-1}$  of the quantized channel matrix  $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \dots, \hat{\mathbf{h}}_K]^H$  is applied to the message vector  $\mathbf{s}[n]$  to alleviate the inter-user interference. In order to satisfy the transmit power constraint, the beamforming vector  $\hat{\mathbf{w}}_k$  should be normalized as

$$\hat{\mathbf{w}}_k = \frac{\hat{\mathbf{W}}_{zf}^k}{\|\hat{\mathbf{W}}_{zf}^k\|} \quad (3)$$

where  $\hat{\mathbf{W}}_{zf}^k$  is the  $k$ -th column of  $\hat{\mathbf{W}}_{zf}$ . Under the assumption that the base station allocates equal power for all users,<sup>3</sup> the achievable rate of the  $k$ -th user is

$$R_k = \log_2 \left( 1 + \frac{\frac{P}{K} |\mathbf{h}_k^H \hat{\mathbf{w}}_k|^2}{1 + \frac{P}{K} \sum_{j=1, j \neq k}^K |\mathbf{h}_k^H \hat{\mathbf{w}}_j|^2} \right) \quad (4)$$

and the corresponding sum rate becomes  $R_{\text{sum}} = \sum_{k=1}^K R_k$ .

### B. Conventional Limited Feedback

In order to feed back the CSI, a user quantizes its channel direction  $\mathbf{h}_k = \frac{\mathbf{h}_k}{\|\mathbf{h}_k\|}$  to a unit norm vector  $\hat{\mathbf{h}}_k$ . Specifically, the  $k$ -th user chooses the quantized vector (codeword)  $\hat{\mathbf{h}}_k$  from a pre-defined  $B$ -bit codebook set  $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_{2^B}\}$  that is closest to its channel direction<sup>4</sup>:

$$\hat{\mathbf{h}}_k = \arg \max_{\mathbf{c} \in \mathcal{C}} |\bar{\mathbf{h}}_k^H \mathbf{c}|^2. \quad (5)$$

Then, the index of the chosen codeword  $\hat{\mathbf{h}}_k$  is fed back to the base station.

In general, the number of bits needed to express the codeword should be scaled with the dimension of the channel vector to be quantized to control the distortion caused by the quantization process. In particular, when there is no spatial correlation among antenna elements and the random vector quantization (RVQ) codebook is used, the number of feedback bits per user should be scaled with the number of transmit antennas and SNR (in decibels) as [21], [33]

$$B_{\text{user}} = (N_t - 1) \log_2 P \approx \frac{N_t - 1}{3} P_{\text{dB}} \quad (6)$$

<sup>2</sup>The proposed AGB in this paper is aimed to generate  $\hat{\mathbf{h}}_k$  with reduced feedback overhead and can be applied to any precoding method including ZFBF.

<sup>3</sup>In this paper, we consider the equal power allocation scenario for simplicity. In order to maximize the sum rate, one might consider more deliberate power allocation strategies (e.g., waterfilling after the block diagonalization [30]).

<sup>4</sup>In practice, each user quantizes the estimated channel, which is obtained using the observations of the pilot signals. Once the channel information corresponding to the pilot signals are estimated, the channel information for the data tones are generated via proper interpolation among pilot channels. With an aim of reducing the pilot overhead of massive MIMO systems, various approaches have been proposed in recent years. In [31], an algorithm exploiting data tones for channel estimation has been proposed and also a pilot allocation strategy based on the sparsity of channel impulse response and compressive sensing (CS) principle [32]. Once the estimated channel information is obtained at the receiver, correlation matrix can be estimated using samples of instantaneous channel information (i.e.,  $\hat{\mathbf{R}} = E[\hat{\mathbf{h}}\hat{\mathbf{h}}^H]$ ).

to maintain a constant gap in terms of the sum rate from the system with perfect CSI. Hence, when the number of transmit antennas increases, the feedback overhead needs to be increased as well (e.g.,  $B_{\text{user}} = 210$  when  $N_t = 64$ ,  $P_{\text{dB}} = 10$ ), let alone the computational burden caused by the codebook selection. Therefore, a reduction in the number of channel vector dimensions would be beneficial in reducing the feedback overhead of the FDD-based massive MIMO systems.

## III. ANTENNA GROUPING BASED FEEDBACK REDUCTION

The key feature of the proposed scheme is to map multiple correlated antenna elements into a single representative value using grouping patterns. As a result, the channel vector dimension is reduced and a codeword is chosen from the codebook generated by the reduced dimension vector. When the channel is correlated (i.e., antenna elements in a group are similar), the loss caused by the grouping of antenna elements is negligible and the target performance can be achieved with smaller number of feedback bits than a conventional scheme requires. In this section, we explain the overall procedure of the proposed AGB algorithm and then discuss the pattern set design problem. We also analyze the quantization distortion caused by the proposed AGB technique and show that the quantization distortion indeed decreases with the transmit correlation coefficient.

### A. AGB Algorithm

As mentioned, the AGB algorithm reduces the dimension of the channel vector from  $N_t$  to  $N_g$  ( $N_t > N_g$ ) by mapping multiple correlated antenna elements to a single representative value (see Fig. 2). While a conventional scheme employs all feedback resources ( $B$  bits) to express the quantized channel vector, the proposed method uses a part of the feedback resources to quantize the (reduced dimension) channel vector and the rest to express the grouping pattern. Both base station and user terminal share a codebook of channel vector and grouping pattern matrices, and thus the receiver feeds back the index of these. In order to support this operation, we divide the feedback resources into two parts: a header ( $B_p$  bits) to indicate the antenna group pattern and a payload ( $B - B_p$  bits) to represent an index of the quantized channel vector (see Fig. 3). Since antenna grouping based quantization is performed separately for each user, in the sequel we focus on the operation of a single user and drop the user index  $k$ .

Suppose there are  $N_p = 2^{B_p}$  antenna group patterns, then  $N_p$  distinct reduced dimension channel vectors are generated. Each pattern converts an  $N_t$ -dimensional channel vector into an  $N_g$ -dimensional vector by multiplying the channel vector by a grouping matrix  $\mathbf{G}^{(i)} \in \mathbb{R}^{N_g \times N_t}$ . The reduced dimension channel vector  $\mathbf{h}_r^{(i)} \in \mathbb{C}^{N_g}$  of the group pattern  $i$  is

$$\mathbf{h}_r^{(i)} = \mathbf{G}^{(i)} \mathbf{h}, \quad i = 1, \dots, N_p. \quad (7)$$

In Fig. 4, we illustrate the concept of antenna group patterns. One simple way to generate the reduced dimension channel vector is to average the channel coefficients in an antenna group. For example, if  $\mathbf{h} = [h_1 \ h_2 \ h_3 \ h_4]^T$  and two adjacent

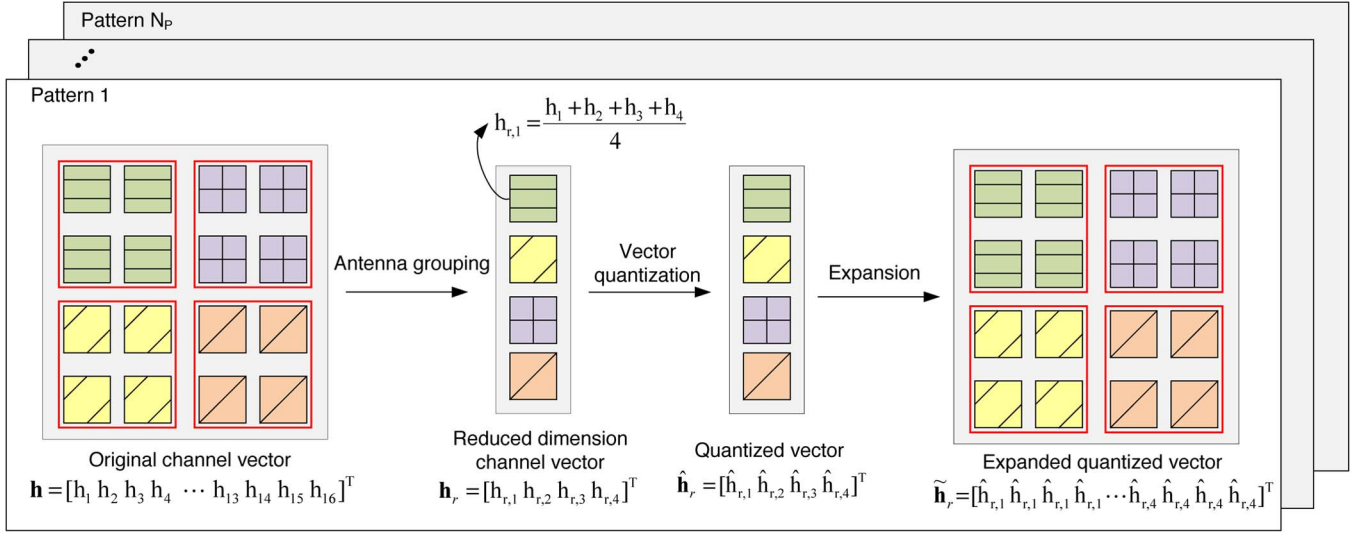


Fig. 2. Illustration of the AGB algorithm for  $N_t = 16$ ,  $N_g = 4$ . The reduced dimension channel vector  $\mathbf{h}_r$  is obtained by mapping antenna elements of a group as a representative value. Note that  $\hat{\mathbf{h}}_r$  is the quantized version of  $\mathbf{h}_r$  and  $\tilde{\mathbf{h}}_r^{(i)}$  is expanded version of  $\hat{\mathbf{h}}_r$ .

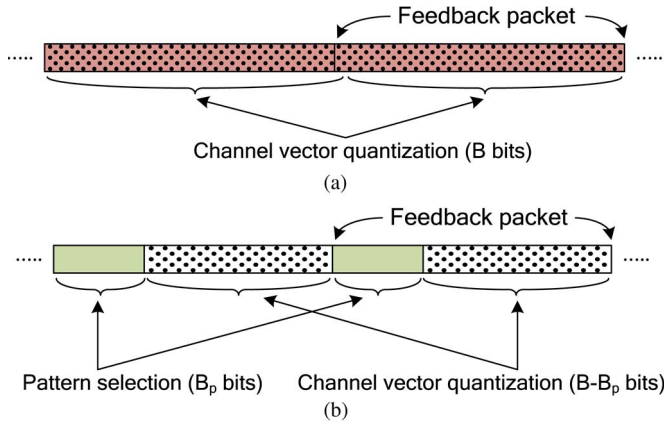


Fig. 3. Feedback packet structure: (a) conventional method and (b) proposed method.

channel coefficients (first and second, third and fourth) are grouped, the mapping matrix is

$$\mathbf{G}^{(i)} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad (8)$$

$$\text{and } \mathbf{h}_r^{(i)} = \mathbf{G}^{(i)} \mathbf{h} = \left[ \frac{h_1 + h_2}{2} \quad \frac{h_3 + h_4}{2} \right]^T.$$

Once the reduced dimension channel vector  $\mathbf{h}_r^{(i)}$  is obtained,  $\mathbf{h}_r^{(i)}$  is quantized by a  $B - B_p$  bit codebook  $\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_{2^{B-B_p}}\}$ . It is worth mentioning that a codebook designed for i.i.d channels is not a proper choice for correlated channels so that we use a channel statistic-based codebook for channel vector quantization [34] (see Section III-C for details). The codeword  $\hat{\mathbf{h}}_r^{(i)}$  maximizing the absolute inner product with  $\mathbf{h}_r^{(i)}$  is chosen as

$$\hat{\mathbf{h}}_r^{(i)} = \arg \max_{\mathbf{c} \in \mathcal{C}} \left| \tilde{\mathbf{h}}_r^{(i)H} \mathbf{c} \right|^2, \quad i = 1, \dots, N_p \quad (9)$$

where  $\tilde{\mathbf{h}}_r^{(i)} = \frac{\mathbf{h}_r^{(i)}}{\|\mathbf{h}_r^{(i)}\|}$  is the direction of the reduced dimension channel vector for the  $i$ -th pattern. This process is repeated for each group pattern and  $N_p$  candidate codewords  $\hat{\mathbf{h}}_r^{(i)}$ ,  $i = 1, \dots, N_p$ , are chosen in total.

Once  $N_p$  candidate codewords are obtained, we need to select the codeword that minimizes the distortion between  $\tilde{\mathbf{h}}$  and  $\hat{\mathbf{h}}_r^{(i)}$ . We note that the direct comparison between  $\tilde{\mathbf{h}}$  and  $\hat{\mathbf{h}}_r^{(i)}$  is not possible since the dimension of  $\hat{\mathbf{h}}_r^{(i)} \in \mathbb{C}^{N_g}$  is smaller than that of the original channel vector  $\mathbf{h} \in \mathbb{C}^{N_t}$ . In computing the distortion defined as  $D(\mathbf{h}, \tilde{\mathbf{h}}_r^{(i)}) = E[\|\mathbf{h}\|^2(1 - |\tilde{\mathbf{h}}_r^{(i)H} \mathbf{h}|^2)]$  caused by the grouping and quantization, therefore, we use  $\tilde{\mathbf{h}}_r^{(i)} \in \mathbb{C}^{N_t}$ , an expanded version of  $\hat{\mathbf{h}}_r^{(i)}$ . The expansion process, which essentially is done by copying each element in  $\hat{\mathbf{h}}_r^{(i)}$  to  $\frac{N_t}{N_g}$  elements in  $\tilde{\mathbf{h}}_r^{(i)}$ , is performed by multiplying an expansion matrix  $\mathbf{E}^{(i)} \in \mathbb{R}^{N_t \times N_g}$  to  $\hat{\mathbf{h}}_r^{(i)}$ . The expanded quantized vector  $\tilde{\mathbf{h}}_r^{(i)}$  is expressed as

$$\tilde{\mathbf{h}}_r^{(i)} = \mathbf{E}^{(i)} \hat{\mathbf{h}}_r^{(i)}, \quad i = 1, \dots, N_p \quad (10)$$

where  $\mathbf{E}^{(i)} = \kappa \mathbf{G}^{(i)T}$  and satisfies  $\mathbf{G}^{(i)} \mathbf{E}^{(i)} = \mathbf{I}_{N_g}$  ( $\kappa = \frac{N_t}{N_g}$ ). For example, for the grouping matrix in (8),  $\mathbf{E}^{(i)} = \kappa \mathbf{G}^{(i)T} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix}^T$  and the expanded quantized vector is

$$\tilde{\mathbf{h}}_r^{(i)} = \mathbf{E}^{(i)} \hat{\mathbf{h}}_r^{(i)} = \kappa \mathbf{G}^{(i)T} \hat{\mathbf{h}}_r^{(i)} = \left[ \hat{h}_{r,1}^{(i)} \hat{h}_{r,1}^{(i)} \hat{h}_{r,2}^{(i)} \hat{h}_{r,2}^{(i)} \right]^T. \quad (11)$$

The group pattern index  $i^*$  minimizing the distortion between  $\mathbf{h}$  and  $\tilde{\mathbf{h}}_r^{(i)}$  is

$$i^* = \arg \min_{i=1, \dots, N_p} D(\mathbf{h}, \tilde{\mathbf{h}}_r^{(i)}). \quad (12)$$

Once the pattern index  $i^*$  is obtained, this index and the corresponding codeword index are sent to the base station. After receiving the pattern index and codeword index of all user terminals, the base station decompresses the reduced dimension

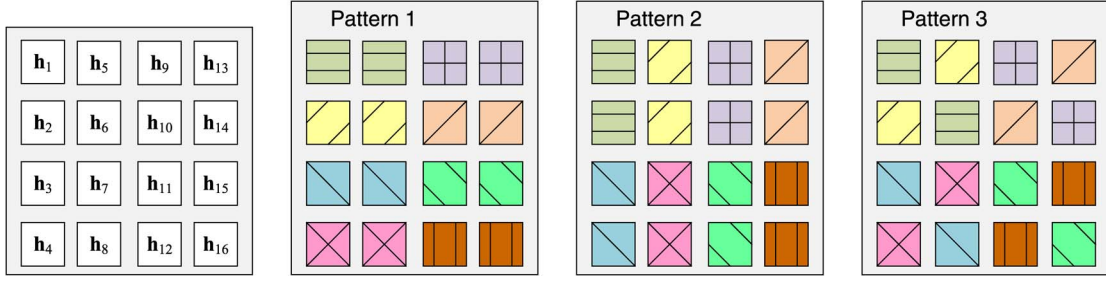


Fig. 4. Example of antenna group patterns ( $N_t = 16$ ,  $N_g = 8$ ,  $N_p = 3$ ). Antenna elements belonging to the same pattern are mapped to one representative value.

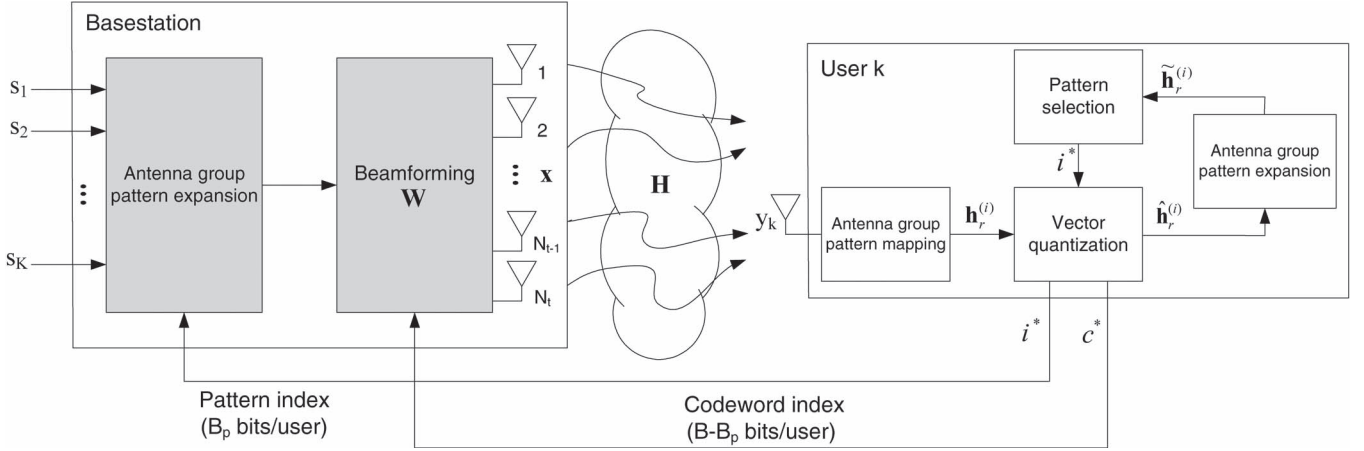


Fig. 5. Overall transceiver structure of the proposed AGB technique.

channel vector via the expansion ( $\hat{\mathbf{h}} = \mathbf{E}^{(i^*)} \hat{\mathbf{h}}_r^{(i^*)}$ ) and then performs the beamforming using the composite channel matrix  $\hat{\mathbf{H}}$ . A block diagram of the proposed AGB algorithm is depicted in Fig. 5.

### B. Antenna Group Pattern Generation

Since multiple correlated antenna elements are mapped to a single representative value, the AGB algorithm is sensitive to the choice of the antenna group pattern. Apparently, selecting the best pattern among all possible combinations would be an ideal option but it is undesirable since the number of patterns increases exponentially with the number of transmit antennas. Without doubt, a simple yet effective pattern design is crucial to the success of the AGB algorithm.

One easy and intuitive way to construct an antenna group pattern  $\mathbf{E}^{(i)}$  is to group highly correlated antenna elements together. Typically, adjacent antenna elements are highly correlated so that the grouping of nearby antenna elements would be a desirable option in practice (see the example in Fig. 4). Alternatively, one can consider Grassmannian subspace packing in the design of the antenna group patterns [35]. The main goal of Grassmannian subspace packing is, when the subspace distance metric and the number of feedback bits  $B$  are provided, to find a set of  $2^B$  subspaces in  $\mathcal{G}(N_t, m)$  that maximizes the minimum subspace distance between any pair of subspaces in the set.<sup>5</sup>

<sup>5</sup>  $\mathcal{G}(N_t, m)$  is the set of  $m$ -dimensional subspaces in  $\mathbb{C}^{N_t}$  (or  $\mathbb{R}^{N_t}$ ).

The chordal distance has been popularly used as a metric to measure the distance [36]. Our task of generating the pattern set is similar in spirit to the Grassmannian subspace packing based codebook generation in the sense that we construct a pattern set (containing  $2^{B_p}$  patterns) from all possible pattern candidates ( $\mathbf{E}^{(i)} \in \mathbb{R}^{N_t \times N_g}$ ) using a distance metric exploiting the spatial correlation among antenna elements.

In the first step of the pattern set design, we compute the quasi-correlation matrix norm  $\|\tilde{\mathbf{R}}_t^{(i)}\|_F$  to measure the spatial proximity of the antenna elements in the antenna group. The quasi-correlation matrix  $\tilde{\mathbf{R}}_t^{(i)}$ , defined as  $\tilde{\mathbf{R}}_t^{(i)} = \mathbf{R}_t^{1/2} \mathbf{E}^{(i)}$ , captures the actual influence of the pattern  $\mathbf{E}^{(i)}$  on the transmit correlation matrix  $\mathbf{R}_t$ . In general, a pattern generated by grouping closely spaced antenna elements tends to have a higher quasi-correlation matrix norm than that generated by grouping antenna elements apart. Thus, one can deduce that a pattern with a large-correlation matrix norm exhibits lower grouping loss than that with a small quasi-correlation matrix norm. For patterns with high quasi-correlation matrix norm, we perform subspace packing to generate  $2^{B_p}$  patterns (expansion matrices) maximizing the minimum distance metric between any pair of subspaces. In measuring the distance, we use the correlation matrix distance  $d_{\text{corr}}(\mathbf{A}, \mathbf{B})$  between two matrices  $\mathbf{A}$  and  $\mathbf{B}$  [37]

$$d_{\text{corr}}(\mathbf{A}, \mathbf{B}) = 1 - \frac{\text{tr}(\mathbf{A}^H \mathbf{B})}{\|\mathbf{A}\|_F \|\mathbf{B}\|_F}. \quad (13)$$

Note that  $d_{\text{corr}}(\mathbf{A}, \mathbf{B})$  measures the orthogonality between two correlation matrices  $\mathbf{A}$  and  $\mathbf{B}$ . When the correlation matrices



TABLE I  
SUMMARY OF THE ANTENNA GROUP PATTERN GENERATION

Initialization	$B_p$ : the number of bits for the pattern set $\mathcal{S}$ : the set of patterns to be selected
Main operation	<p>1) Initialize the index set <math>\Omega = \{1, \dots, N_{\max}\}</math> where</p> $N_{\max} = \frac{\prod_{n=0}^{N_g-1} \binom{N_t - n\kappa}{\kappa}}{N_g!}.$ <p><math>\kappa = \frac{N_t}{N_g}</math> is the number of elements in an antenna group.</p> <p>2) For each pattern <math>i \in \Omega</math>, calculate the Frobenius norm of the quasi-correlation matrix</p> $r_i = \ \tilde{\mathbf{R}}_t^{(i)}\ _F.$ <p>Without loss of generality, assume <math>r_1 \geq r_2 \geq \dots \geq r_{N_{\max}}</math>.</p> <p>3) Choose <math>J(\geq 2^{B_p})</math> patterns <math>\mathcal{T} = \{r_1, \dots, r_J\}</math>.</p> <p>4) Apply the subspace packing to <math>\mathcal{T}</math> to generate the pattern set <math>\mathcal{S}</math>.</p> <p>Construct <math>N_c = \binom{J}{2^{B_p}}</math> candidate sets <math>\{\mathcal{S}_k\}_{k=1}^{N_c}</math> where <math>\mathcal{S}_k = \{\tilde{\mathbf{R}}_t^{k,1}, \tilde{\mathbf{R}}_t^{k,2}, \dots, \tilde{\mathbf{R}}_t^{k,2^{B_p}}\}</math>.  <math>\tilde{\mathbf{R}}_t^{k,i}</math> is the <math>i</math>-th quasi-correlation matrix from the <math>k</math>-th candidate set.</p> <p>5) Calculate the minimum <math>d_{\text{corr}}</math> of a <math>\mathcal{S}_k</math></p> $d_{k,\min}(\mathcal{S}_k) = \min_{1 \leq m \leq n \leq 2^{B_p}} d_{\text{corr}}(\tilde{\mathbf{R}}_t^{k,m}, \tilde{\mathbf{R}}_t^{k,n}).$ <p>Decide the pattern set <math>\mathcal{S}</math></p> $\mathcal{S} = \arg \max_{k=1, \dots, N_c} d_{k,\min}(\mathcal{S}_k).$

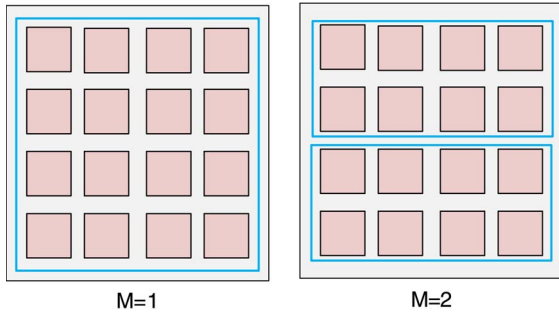


Fig. 6. Antenna group pattern of the partitioned sub-arrays when  $M = 1$  (no partition) and  $M = 2$ .

are equal up to a scaling factor,  $d_{\text{corr}}$  is minimized ( $d_{\text{corr}} = 0$ ). Whereas, when the inner product between the vectorized correlation matrices is zero (i.e.,  $\text{vec}(\mathbf{A})$  and  $\text{vec}(\mathbf{B})$  are orthogonal),  $d_{\text{corr}}$  is maximized ( $d_{\text{corr}} = 1$ ). In our numerical simulations, we show that the proposed subspace packing approach achieves a substantial gain over an approach using randomly selected patterns (see Section IV-B). We summarize the antenna group pattern generation procedures in Table I.

As a further means to lessen the computational burden of pattern generation process, we partition the antenna array into multiple sub-arrays and then apply Grassmannian subspace packing for each sub-array. In the partitioning process, we basically divide the antenna array such that the divided sub-arrays are close to the square matrix. Specifically, when the partitioning level is two, for the  $N_{t1} \times N_{t2}$  dimensional antenna array ( $N_t = N_{t1}N_{t2}$ ), we divide the axis with larger dimension. That is, if  $N_{t1} \geq N_{t2}$ , then the dimension of each sub-array becomes  $\frac{N_{t1}}{2} \times N_{t2}$  (see Fig. 6). When the partitioning level is larger than two, we perform the same procedure for each partitioned sub-array. In doing so, we achieve significant reduction in the number of antenna group pattern candidates. For example, if  $N_t = 16$ ,  $N_g = 8$ , and  $M = 1$  (no partition), then the total number of antenna group pattern candidates  $N_{\max}$  is about  $2 \times 10^6$  (see Table I). Whereas, if the antenna array is partitioned ( $M = 2$ ), then  $N_{\max}$  will be expressed as  $N_{\max} =$

$\prod_{i=1}^M N_{\max,i}$  where  $N_{\max,i}$  is the total number of candidates for the  $i$ -th sub-array. Since  $N_{\max,i} = 105$  in this case, we approximately have  $N_{\max} \approx 10^4$ . Finally, we note that since the pattern generation process is performed off the shelf, this process does not affect the real-time operation.

### C. Quantization Distortion Analysis

We now turn to the performance analysis of the AGB algorithm. In our analysis, we analyze the distortion  $D$  induced by the quantization of the channel direction vector  $\tilde{\mathbf{h}} = \frac{\mathbf{h}}{\|\mathbf{h}\|}$ , which is defined as

$$D = E \left[ \|\mathbf{h}\|^2 - |\mathbf{h}^H \tilde{\mathbf{h}}_r|^2 \right] \\ = E \left[ \|\mathbf{h}\|^2 \left( 1 - |\tilde{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right) \right] \quad (14)$$

where  $\tilde{\mathbf{h}}_r$  is the expanded version of the quantized vector  $\hat{\mathbf{h}}_r$  (see (10)).

In the evaluation of the distortion  $D$ , we use the quantization cell upper bound (QUB) [38]. As mentioned, since a codebook designed for the i.i.d channels is not the right choice for correlated channels, we employ a channel statistic-based codebook obtained by applying the transmit correlation matrix  $\mathbf{R}_t^{1/2}$  to the codebook generated from the Grassmannian line packing. Let  $\mathbf{f}_i \in \mathbb{C}^T$  be the  $i$ -th unit norm vector generated from the Grassmannian line packing, then the set of  $B$ -bit codewords for the channel statistic-based codebook is [34]

$$\mathcal{C} = \{\mathbf{c}_1, \dots, \mathbf{c}_{2^B}\} = \left\{ \frac{\mathbf{R}_t^{1/2} \mathbf{f}_1}{\|\mathbf{R}_t^{1/2} \mathbf{f}_1\|}, \dots, \frac{\mathbf{R}_t^{1/2} \mathbf{f}_{2^B}}{\|\mathbf{R}_t^{1/2} \mathbf{f}_{2^B}\|} \right\}. \quad (15)$$

When the channel statistic-based codebook is used, the normalized distortion  $\frac{D}{E[\|\mathbf{h}\|^2]} = 1 - |\tilde{\mathbf{h}}^H \mathbf{c}_i|^2$  between the channel direction vector  $\tilde{\mathbf{h}}$  and codebook vector  $\mathbf{c}_i$  can be upper bounded as [24]

$$\mathcal{R}_i \approx \left\{ \tilde{\mathbf{h}} : 1 - |\tilde{\mathbf{h}}^H \mathbf{c}_i|^2 \leq \delta \right\} \quad (16)$$

where  $\delta = \frac{\sigma_2^2}{\sigma_1^2} 2^{-\frac{B}{r-1}}$  ( $\sigma_i$  is the  $i$ -th largest singular value of the transmit correlation matrix  $\mathbf{R}_t \in \mathbb{C}^{r \times r}$ ) and  $B$  is the number of quantization bits.

In our analysis, we restrict our attention to the scenario where two antenna elements are mapped to a single representative value for mathematical tractability. Nevertheless, since the key factor affecting the quantization distortion is the transmit correlation coefficient (see (30)), our results can be readily applied to the general scenario where more than two antenna elements are grouped together. The minimal set of assumptions used for the analytical tractability are as follows:

- A-i) The channel vector  $\bar{\mathbf{h}}$  is partitioned into two subvectors  $\bar{\mathbf{h}}_A$  and  $\bar{\mathbf{h}}_B$ .  $\bar{\mathbf{h}}_A$  and  $\bar{\mathbf{h}}_B$  are composed of odd and even entries in  $\bar{\mathbf{h}}$  (i.e.,  $\bar{\mathbf{h}}_A = [\bar{\mathbf{h}}_1 \ \bar{\mathbf{h}}_3 \ \dots]^T$ ,  $\bar{\mathbf{h}}_B = [\bar{\mathbf{h}}_2 \ \bar{\mathbf{h}}_4 \ \dots]^T$ ). Thus,  $N_g = \frac{N_t}{2}$ .
- A-ii) The reduced dimension channel vector  $\mathbf{h}_r$  is designed such that  $\mathbf{h}_r = \bar{\mathbf{h}}_A$ . For example, if  $\bar{\mathbf{h}} = [\bar{\mathbf{h}}_1 \ \bar{\mathbf{h}}_2 \ \bar{\mathbf{h}}_3 \ \bar{\mathbf{h}}_4]^T$ , then the antenna grouping pattern is

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (17)$$

and hence  $\mathbf{h}_r = \bar{\mathbf{h}}_A = [\bar{\mathbf{h}}_1 \ \bar{\mathbf{h}}_3]^T$  ( $\bar{\mathbf{h}}_B = [\bar{\mathbf{h}}_2 \ \bar{\mathbf{h}}_4]^T$ ).

- A-iii) Antenna elements in a group are highly correlated. That is,  $E[|\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r|^2] \approx E[|\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r|^2]$  where  $\hat{\mathbf{h}}_r$  is the quantized vector of  $\mathbf{h}_r$  and generated from the channel statistic-based codebook. Note that this assumption justifies the use of antenna grouping pattern in (17).

It is worth mentioning that depending on the way of grouping elements of antenna array, there are several ways to generate subvectors  $\bar{\mathbf{h}}_A$  and  $\bar{\mathbf{h}}_B$ . Also, the group pattern used in (17) might be worse than the pattern generated by the subspace packing or proposed in (8). Thus, assumptions in A-i) and A-ii) would be clearly pessimistic, but makes our analysis tractable. The following proposition provides an approximate upper bound of the quantization distortion  $D$  under these assumptions.

**Proposition 1:** The quantization distortion  $D$  of the AGB algorithm under the channel statistic-based codebook satisfies

$$D \lesssim N_t \delta + \xi \sqrt{2N_t \delta} \quad (18)$$

where  $\delta = \frac{\sigma_2^2}{\sigma_1^2} 2^{-\frac{B-B_p}{N_g-1}}$  is an upper bound of the normalized distortion between  $\bar{\mathbf{h}}_A$  and  $\hat{\mathbf{h}}_r$  as defined in (16) ( $\sigma_i$  is the  $i$ -th largest singular value of  $\mathbf{R}_{t,A}$ )<sup>6</sup> and  $\xi$  is the correlation coefficient between two random variables  $\|\mathbf{h}\|^2$  and  $1 - |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2$ .

*Proof:* Using (14), we have

$$D = E \left[ \|\mathbf{h}\|^2 \left( 1 - |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right) \right] \quad (19)$$

$$= E \left[ \|\mathbf{h}\|^2 \right] \left( 1 - E \left[ |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right] \right) + \text{Cov} \left( \|\mathbf{h}\|^2, 1 - |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right) \quad (20)$$

$$= E \left[ \|\mathbf{h}\|^2 \right] \left( 1 - E \left[ |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right] \right) + \xi \sqrt{\text{Var}(\|\mathbf{h}\|^2)} \sqrt{\text{Var}(1 - |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2)} \quad (21)$$

$$\leq E \left[ \|\mathbf{h}\|^2 \right] \left( 1 - E \left[ |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right] \right) + \xi \sqrt{\text{Var}[\|\mathbf{h}\|^2]} \sqrt{1 - \left( E \left[ |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right] \right)^2} \quad (22)$$

where (20) is because  $\text{Cov}(X, Y) = E[XY] - E[X]E[Y]$ , (21) is because  $\text{Cov}(X, Y) = \xi \sqrt{\text{Var}(X)} \sqrt{\text{Var}(Y)}$  and (22) is because  $\text{Var}(1 - |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2) = E[|\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^4] - \left( E[|\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2] \right)^2 \leq 1 - \left( E[|\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2] \right)^2$ . The normalized distortion term  $1 - E[|\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2]$  in the right-hand side of (22) is approximately upper bounded as

$$\begin{aligned} & 1 - E \left[ |\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 \right] \\ & \stackrel{(a)}{=} 1 - E \left[ |\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r + \bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r|^2 \right] \\ & = 1 - E \left[ |\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r|^2 + |\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r|^2 + (\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r)^* (\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r) \right. \\ & \quad \left. + (\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r)^* (\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r) \right] \\ & = 1 - E \left[ |\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r|^2 \right] - E \left[ |\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r|^2 \right] \\ & \quad - 2E \left[ \text{Re} \left( (\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r)^* (\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r) \right) \right] \\ & \stackrel{(b)}{\approx} 1 - E \left[ |\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r|^2 \right] - E \left[ |\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r|^2 \right] \\ & \stackrel{(c)}{=} 1 - 2E \left[ |\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r|^2 \right] \\ & \stackrel{(d)}{\lesssim} 1 - 2(1 - \delta)E \left[ \|\bar{\mathbf{h}}_A\|^2 \right] \\ & \stackrel{(e)}{=} \delta \end{aligned} \quad (23)$$

where (a) is because  $|\bar{\mathbf{h}}^H \tilde{\mathbf{h}}_r|^2 = |\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r + \bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r|^2$ , (b) follows from  $E[\text{Re}((\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r)^* (\bar{\mathbf{h}}_B^H \hat{\mathbf{h}}_r))] \approx 0$  (see Appendix A), (c) follows from A-iii), and (d) follows from the QUB in (16). That is, by plugging  $\bar{\mathbf{h}} = \frac{\bar{\mathbf{h}}_A}{\|\bar{\mathbf{h}}_A\|}$ ,  $\mathcal{C} = \left\{ \frac{\mathbf{R}_{t,A}^{1/2} \mathbf{f}_1}{\|\mathbf{R}_{t,A}^{1/2} \mathbf{f}_1\|}, \dots, \frac{\mathbf{R}_{t,A}^{1/2} \mathbf{f}_{2^{B-B_p}}}{\|\mathbf{R}_{t,A}^{1/2} \mathbf{f}_{2^{B-B_p}}\|} \right\}$ , and  $\delta = \frac{\sigma_2^2}{\sigma_1^2} 2^{-\frac{B-B_p}{N_g-1}}$  into (16), we get  $E[|\bar{\mathbf{h}}_A^H \hat{\mathbf{h}}_r|^2] \geq (1 - \delta)E[\|\bar{\mathbf{h}}_A\|^2]$ .

<sup>6</sup>For example, if  $\mathbf{R}_t = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$  and  $A = \{1, 3\}$ , then

$$\mathbf{R}_{t,A} = \begin{bmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{bmatrix}.$$

Finally, (e) follows from  $E[\|\bar{\mathbf{h}}_A\|^2] = \frac{N_g}{N_t} = \frac{1}{2}$ . Plugging (23) into (22), we have

$$D \leq E[\|\mathbf{h}\|^2] \delta + \xi \sqrt{\text{Var}[\|\mathbf{h}\|^2]} \sqrt{1 - (1 - \delta)^2} \quad (24)$$

$$= N_t \delta + \xi \sqrt{N_t (1 - (1 - \delta)^2)} \quad (25)$$

$$\approx N_t \delta + \xi \sqrt{2N_t \delta} \quad (26)$$

where (25) is because  $E[\|\mathbf{h}\|^2] = N_t$  and  $\text{Var}[\|\mathbf{h}\|^2] = N_t$ , and (26) is because  $\delta(2 - \delta) \approx 2\delta$  where  $\delta \ll 2$ , which is the desired result.  $\square$

We note that the relationship between the quantization distortion  $D$  and the transmit antenna correlation is not clearly shown in (18). When a specific correlation model is used, however, we can observe the relationship between the two. For example, if the exponential correlation model is employed, the transmit correlation matrix  $\mathbf{R}_t$  is expressed as [39]

$$\mathbf{R}_t = \begin{bmatrix} 1 & \rho & \cdots & \rho^{N_t-1} \\ \rho^* & 1 & \cdots & \rho^{N_t-2} \\ \vdots & \vdots & \ddots & \vdots \\ (\rho^*)^{N_t-1} & (\rho^*)^{N_t-2} & \cdots & 1 \end{bmatrix} \quad (27)$$

where  $\rho = \alpha e^{j\theta}$  is the transmit correlation coefficient, and  $\alpha$  is the magnitude of correlation coefficient, and  $\theta$  is the phase of the coefficient. When the number of transmit antennas  $N_t$  is large, (non-ordered) singular value  $\mu_i$  of  $\mathbf{R}_t$  approximately behaves as [40]

$$\begin{aligned} \mu_i &\approx \sum_{k=-(N_t-1)}^{N_t-1} \rho^{|k|} e^{j \frac{2\pi i k}{N_t}} \\ &\approx \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos\left(\frac{2\pi i}{N_t}\right)}, \quad i = 1, \dots, N_t. \end{aligned} \quad (28)$$

Using the first and second largest singular values of<sup>7</sup> (28), we have

$$\frac{\sigma_2}{\sigma_1} \approx \frac{1 + \rho^2 - 2\rho}{1 + \rho^2 - 2\rho \cos\left(2\pi \frac{N_t-1}{N_t}\right)}. \quad (29)$$

Using this together with  $\delta = \frac{\sigma_2^2}{\sigma_1^2} 2^{-\frac{B-B_p}{N_g-1}}$  in Proposition 3.1, we have

$$\begin{aligned} D &\lesssim N_t \frac{(1 + \rho^2 - 2\rho)^2}{\left(1 + \rho^2 - 2\rho \cos\left(2\pi \frac{N_t-1}{N_t}\right)\right)^2} 2^{-\frac{B-B_p}{N_g-1}} \\ &\quad + \xi \sqrt{2N_t} \frac{1 + \rho^2 - 2\rho}{1 + \rho^2 - 2\rho \cos\left(2\pi \frac{N_t-1}{N_t}\right)} 2^{-\frac{B-B_p}{2(N_g-1)}}. \end{aligned} \quad (30)$$

In (30), we observe that the quantization distortion  $D$  decreases with the correlation coefficient  $\rho$ . Fig. 7 plots the normalized

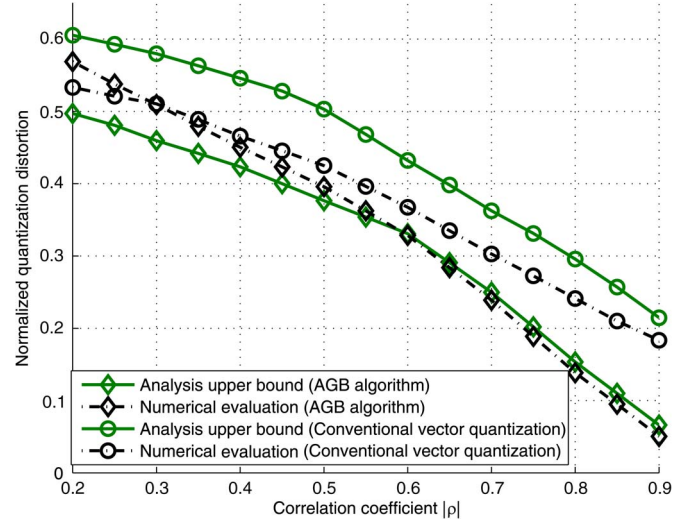


Fig. 7. Normalized quantization distortion as a function of the correlation coefficient ( $N_t = 16$ ,  $N_g = 8$ ,  $B = 16$ ,  $B_p = 8$ ).

quantization distortion  $\frac{D}{E[\|\mathbf{h}\|^2]}$  as a function of the correlation coefficient  $\rho$ . Note that  $\delta = \frac{\sigma_2^2}{\sigma_1^2} 2^{-\frac{B}{r-1}}$  is obtained from the assumption that all non-zero singular values except the dominant one (i.e.,  $\sigma_1$ ) are the same ( $\sigma_2 = \sigma_3 = \dots$ ). Note also that  $\mathcal{R}_i$  is tight in a regime where transmit antennas are highly correlated since  $\frac{\sigma_2}{\sigma_1}$  (i.e.,  $\delta$ ) decreases with the correlation coefficient. Readers are referred to [23], [24] for more details. We observe that if  $|\rho| > 0.3$ , the quantization distortion  $D$  of the AGB algorithm is better (smaller) than that of conventional vector quantization. We can also observe that the analysis matches well with the simulation results when the transmit antennas are highly correlated ( $|\rho| > 0.6$ ). However, when the magnitude of  $\rho$  is small, the assumption in A-iii) is violated so that the proposed bound is invalid.

The following proposition provides the upper bound of the sum rate gap  $\Delta R(P)$ .

**Proposition 2:** When an equal power allocation per user is applied, the sum rate gap (per user) between the ZFBF with perfect CSI and the proposed method satisfies

$$\Delta R(P) \lesssim \log_2 \left( 1 + P \frac{K-1}{K} (N_t \delta + \xi \sqrt{2N_t \delta}) \right) \quad (31)$$

where  $\Delta R(P)$  is the difference between the achievable rate achieved by (4) and  $\mathbf{w}_k = \frac{\mathbf{w}_{zf}^k}{\|\mathbf{w}_{zf}^k\|}$  where  $\mathbf{W}_{zf} = \mathbf{H}^H (\mathbf{H} \mathbf{H}^H)^{-1}$ .

*Proof:* Note that  $\Delta R(P)$  is given by  $\Delta R(P) = E[\log_2(1 + \frac{P}{K} |\mathbf{h}_k^H \mathbf{w}_k|^2)] - E[\log_2(1 + \frac{P}{1 + \frac{P}{K} \sum_{j=1, j \neq k}^K |\mathbf{h}_k^H \hat{\mathbf{w}}_j|^2})]$ . Using Jensen's inequality,  $\Delta R(P)$  can be upper bounded as [21]

$$\begin{aligned} \Delta R(P) &\leq E \left[ \log_2 \left( 1 + \frac{P}{K} \sum_{j=1, j \neq k}^K |\mathbf{h}_k^H \hat{\mathbf{w}}_j|^2 \right) \right] \\ &\leq \log_2 \left( 1 + \frac{P}{K} E \left[ \sum_{j=1, j \neq k}^K |\mathbf{h}_k^H \hat{\mathbf{w}}_j|^2 \right] \right). \end{aligned} \quad (32)$$

<sup>7</sup>Due to the symmetric property of  $\mu_i$  (i.e.,  $\mu_{N_t} > \mu_{N_t-1} = \mu_1 > \mu_{N_t-2} = \mu_2 > \dots$ ),  $\sigma_1 = \mu_{N_t}$  and  $\sigma_2 = \mu_{N_t-1} = \mu_1$ .



Using orthogonality between  $\hat{\mathbf{w}}_j$  and  $\tilde{\mathbf{h}}_{r,k}$ ,

$$\|\mathbf{h}_k\|^2 \geq |\mathbf{h}_k^H \hat{\mathbf{w}}_j|^2 + \|\mathbf{h}_k\|^2 |\tilde{\mathbf{h}}_k^H \tilde{\mathbf{h}}_{r,k}|^2, \quad (33)$$

then (32) becomes

$$\begin{aligned} \Delta R(P) &\leq \log_2 \left( 1 + \frac{P}{K} E \left[ \sum_{j=1, j \neq k}^K \|\mathbf{h}_k\|^2 \left( 1 - |\tilde{\mathbf{h}}_k^H \tilde{\mathbf{h}}_{r,k}|^2 \right) \right] \right) \\ &= \log_2 \left( 1 + P \frac{(K-1)}{K} D \right) \end{aligned} \quad (34)$$

where (34) is due to  $D = E[\|\mathbf{h}_k\|^2 (1 - |\tilde{\mathbf{h}}_k^H \tilde{\mathbf{h}}_{r,k}|^2)]$ . Using (18) and (34), we get the desired result.  $\square$

Next proposition specifies the number of feedback bits needed to maintain a constant rate gap from the system with perfect CSI.

*Proposition 3:* In order to maintain a rate gap (between the ZFBF with perfect CSI and the proposed method) within  $\log_2 \beta$  bps/Hz per user, it is sufficient to scale the number of bits per user according to

$$\begin{aligned} B \approx B_p + (N_g - 1) &\left[ \log_2 \left( \frac{(1 + \rho^2 - 2\rho)^2}{(1 + \rho^2 - 2\rho \cos(2\pi \frac{N_t-1}{N_t}))^2} \right) \right. \\ &\left. - 2 \log_2 \left( \frac{-\xi + \sqrt{\xi^2 + 4(\beta - 1) \frac{K}{P(K-1)}}}{2\sqrt{N_t}} \right) \right]. \end{aligned} \quad (35)$$

*Proof:* In order to maintain a rate loss of  $\Delta R(P) \leq \log_2 \beta$  bps/Hz per user, we set

$$\Delta R(P) \lesssim \log_2 \left( 1 + P \frac{K-1}{K} (N_t \delta + \xi \sqrt{2N_t \delta}) \right) \triangleq \log_2 \beta. \quad (36)$$

By inverting (36) and solving for  $B$ , we get the desired result.  $\square$

Fig. 8 plots the sum rate as a function of SNR when  $B$  in (35) is applied. We fix  $\beta = 2$  in order to maintain a SNR gap of 3 dB. We observe that by using a proper scaling of feedback bits, we can limit the rate loss within 3 dB (in fact around 2 dB), as desired.

#### D. Comments on Complexity

In this subsection, we discuss the complexity of the AGB algorithm and conventional vector quantization. Other than the codeword index search operation of the conventional approach, computations associated with pattern index selection, grouping and expansion process, pattern generation are additionally required for the proposed method. We first analyze the computational complexity of the pattern index selection, grouping process, and the expansion process, which are performed on the fly. Denoting the complexity associated with pattern index selection, grouping process, and the expansion process as

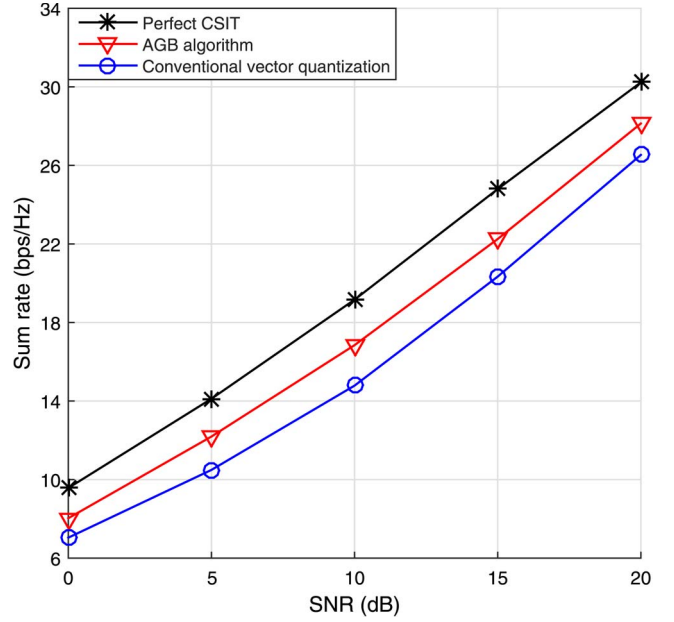


Fig. 8. Sum rate as a function of SNR when  $(N_t = 32, N_g = 16, K = 4, \rho = 0.9, \zeta = 0.05, \zeta B_p = 8)$ .

$C_p$ ,  $C_g$ ,  $C_e$ , respectively, then the number of required floating-point operations (flops) for each step is as follows [41]:

- $C_p$  requires  $4N_t$  flops for computing the distortion  $D$  in (12).
- $C_g$  requires  $(2N_t - 1)N_g$  flops for the matrix multiplication in (7).
- $C_e$  requires  $(2N_g - 1)N_t$  flops for the matrix multiplication in (10).

Note that these operations need to be computed for  $N_p$  times. We next measure the computational complexity of the pattern generation process. The number of required flops for computing the quasi-correlation matrix norm  $\|\tilde{\mathbf{R}}_t^{(i)}\|_F$  and the correlation matrix distance  $d_{\text{corr}}$  can be obtained as  $2N_t N_g$  and  $2N_g^2 N_t + 4N_t N_g$ , respectively. Then, according to Table I, the total computational complexity becomes  $2N_{\max} N_t N_g + \binom{J}{N_p} \binom{N_p}{2} (2N_g^2 N_t + 4N_t N_g)$ . Note that  $N_{\max}$  can be reduced significantly by applying the partitioning approach discussed in Section III-B. As mentioned, the pattern generation process does not affect the real-time operation since this process is performed off the shelf. In contrast to the operations we just described, the codeword search complexity is quantified by  $O(\cdot)$  notation. Note that codeword search complexity grows exponentially with the dimension of the vector to be quantized and the codeword search complexity for the conventional approach and proposed method is given by  $O(N_t 2^B)$  and  $N_p O(N_g 2^{B-B_p})$ , respectively. Note also that the complexity of additional operations (i.e.,  $N_p(C_p + C_g + C_e)$ ) is much smaller than that of the codeword search complexity  $N_p O(N_g 2^{B-B_p})$ . Overall, the proposed method brings additional benefits in search complexity over the conventional approach due to the fact that dimension of the codeword being searched is reduced.

#### IV. SIMULATION RESULTS AND DISCUSSIONS

##### A. Simulation Setup

In this section, we compare the sum rate performance of the conventional vector quantization using the channel statistic-based codebook [34] and the proposed AGB algorithm. While all the feedback resources ( $B$ -bit) are used to quantize the channel vector  $\mathbf{h}_k$  in the conventional vector quantization approach,  $B$ -bit feedback resource is divided into  $B_q$  (channel vector quantization) and  $B_p$  (pattern selection) in the proposed method. To express the feedback allocation, we use the notation  $B = (B_q, B_p)$  in the sequel. As a pattern set, we use the combination of patterns for each sub-array. Let  $B_{p,\text{sub}}$  be the number of pattern bits of each sub-array ( $B_{p,\text{sub}} = \frac{B_p}{M}$ ), then  $2^{B_{p,\text{sub}}}$  patterns are generated by applying the proposed subspace packing approach. As a transmit antenna model, we consider the exponential correlation model in (27) [42] and two-dimensional uniform planar array (UPA) model [43]. We use Jakes' model [44] for the temporal correlation coefficient  $\eta = J_0(2\pi f_D \tau)$  where  $J_0(\cdot)$  is the 0-th order Bessel function of the first kind,  $f_D = v f_c / c$  denotes the maximum Doppler frequency, and  $\tau = 5$  ms is the channel instantiation interval. With the user speed  $v = 3$  km/h, the carrier frequency  $f_c = 2.5$  GHz, and the speed of light  $c = 3 \times 10^8$  m/s, the temporal correlation coefficient becomes  $\eta = 0.9881$ . If we set the coherence time to 5 ms and use the frame structure of 3GPP LTE FDD systems [5], each fading block consists of  $L \approx 10$  static channel uses.

##### B. Simulation Results

We first consider the exponential channel model

$$r_{ij} = \begin{cases} \rho_k^{|j-i|} & i \leq j \\ \left(\rho_k^{|j-i|}\right)^H & i > j \end{cases} \quad (37)$$

where  $r_{ij}$  is the  $(i, j)$ -th element of  $\mathbf{R}_{t,k}$  and  $\rho_k = \alpha e^{j\theta_k}$  is a transmit correlation coefficient for the  $k$ -th user where  $\alpha$  is the magnitude of correlation coefficient and  $\theta_k$  is the phase of the  $k$ -th user. Note that the phase of each user is randomly generated from  $-\pi$  to  $\pi$  and independent among each user. Note also that all users have the same transmit correlation coefficient  $|\rho_k| = \alpha$  since  $\alpha$  is determined by the antenna spacing at the base station.

In order to observe the effectiveness of the subspace packing approach discussed in Section III-B, we compare the proposed approach to the random pattern generation and grouping of adjacent antenna elements. In our simulations, we set  $B_q = 16$ ,  $N_t = 16$ ,  $N_g = 8$ ,  $M = 2$ ,  $K = 1$  and measure the sum rate as a function of the number of pattern bits  $B_p$ . To set the same level of feedback, we set  $B = B_q + B_p$  bit for the conventional vector quantization. Overall, we observe from Fig. 9 that the subspace packing approach provides a considerable sum rate gain over the approach using randomly generated patterns, AGB with grouping of adjacent antenna elements as well as the conventional vector quantization technique. For example, to achieve 9 bps/hz, AGB with subspace packing requires  $B = 18$  bits while AGB with grouping of adjacent antenna

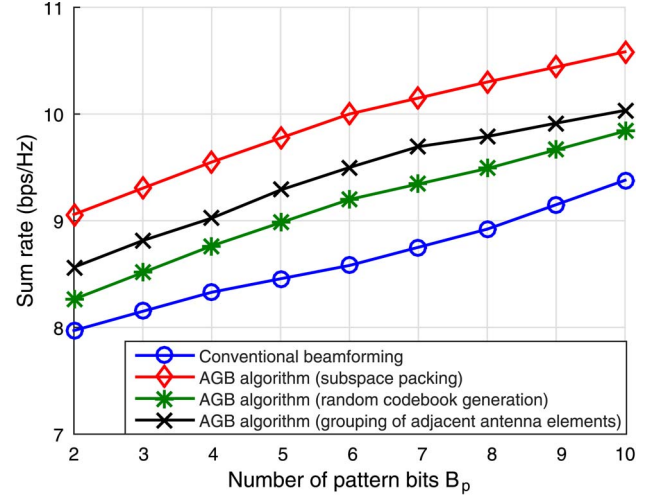


Fig. 9. Sum rate as a function of the number of pattern bits ( $N_t = 16$ ,  $M = 2$ ,  $K = 1$ ,  $N_g = 8$ ,  $B_q = 16$ ,  $B = B_q + B_p$ ),  $\text{SNR} = 10$  dB.

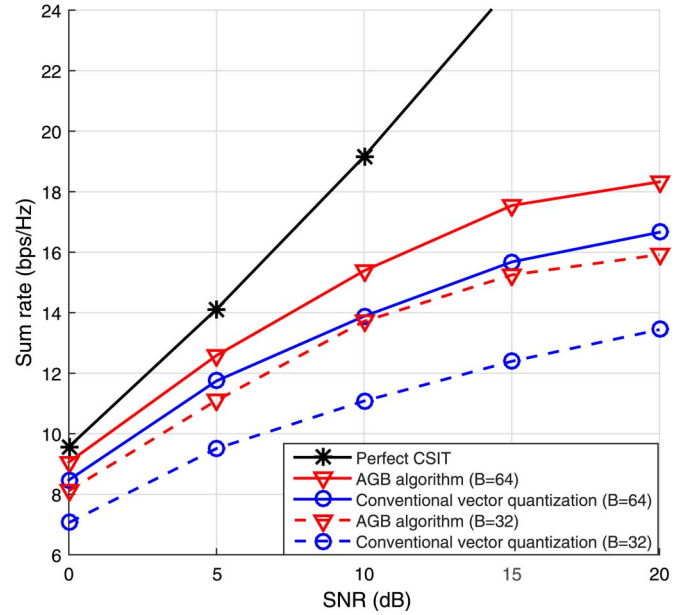


Fig. 10. Sum rate as a function of SNR when  $B = (B_q, 8)$  ( $N_t = 32$ ,  $M = 4$ ,  $K = 4$ ,  $N_g = 16$ ,  $\alpha = 0.8$ ).

elements, AGB with random patterns and conventional vector quantization require 20, 21, and 24 bits, respectively.

We next measure the sum rate as a function of SNR. In this case, we set  $N_t = 32$ ,  $M = 4$ ,  $K = 4$ ,  $\alpha = 0.8$  and investigate the performance for two scenarios ( $B = N_t$  and  $2N_t$ ). In addition, we plot the system with perfect CSIT as an upper bound. As shown in Fig. 10, the AGB algorithm achieves significant gain over the conventional vector quantization technique, bringing in more than 3 dB gain at mid SNR regime. In particular, with  $B = 2N_t$ , performance loss of the AGB algorithm over the perfect CSIT system is within about 2 dB until  $\text{SNR} = 7$  dB while others suffer from more than 5 dB loss compared to the perfect CSIT system.

In Fig. 11, we plot the sum rate as a function of the number of feedback bits. In this case, we set  $K = 4$ ,  $\alpha = 0.8$ ,  $N_g = 16$  and compare the performance of the AGB algorithm when  $N_t = 32$

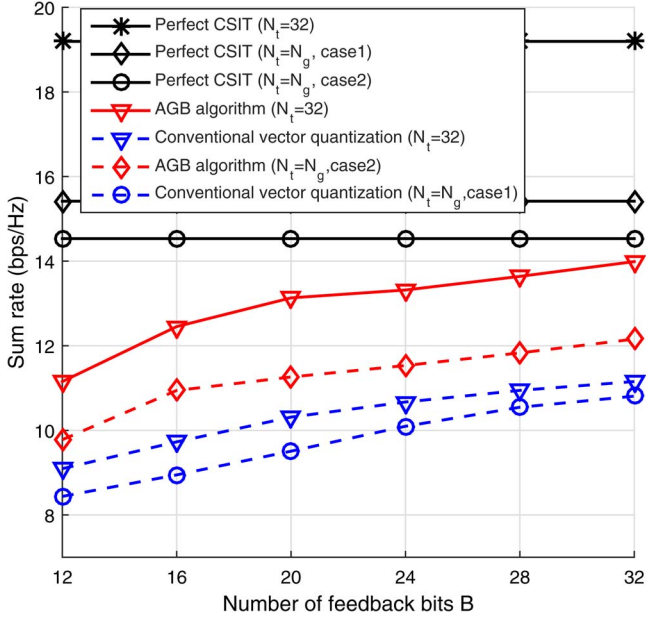


Fig. 11. Sum rate as a function of the number of feedback bits  $B$  when SNR = 10 dB and  $B = (Bq, 8)$  ( $N_t = 32$ ,  $M = 4$ ,  $K = 4$ ,  $N_g = 16$ ,  $\alpha = 0.8$ ).

with the following two scenarios; 1) a system having a reduced number of transmit antennas ( $N_t = N_g$ ) and 2) a system where one antenna per group is selected and all remaining antennas per group is shut down ( $N_t = N_g$ ). Interestingly, by taking advantage of high correlation among the antennas in a group, we observe that the performance of case 2) is better than that of the conventional system and case 1). Nevertheless, due to the number of active antennas, the proposed algorithm when  $N_t = 32$  still achieves significant feedback overhead reduction over the case 2). We observe that the proposed AGB algorithm with  $N_t = 32$  requires smaller number of bits to achieve the same level of performance. For example, the proposed approach achieves significant gain over the conventional vector quantization techniques, resulting in more than 60% feedback overhead reduction.

In Fig. 12, we consider the two-dimensional UPA ( $N_V \times N_H$  array) model, which is more realistic antenna model for massive MIMO scenarios. The UPA model can be obtained by the Kronecker product of the vertical correlation matrix  $\mathbf{R}_V \in \mathbb{C}^{N_V \times N_V}$  and the horizontal correlation matrix  $\mathbf{R}_{H,k} \in \mathbb{C}^{N_H \times N_H}$ . The resulting transmit correlation matrix of the UPA model is expressed as  $\mathbf{R}_{t,k} = \mathbf{R}_V \otimes \mathbf{R}_{H,k}$  where  $\otimes$  is the Kronecker product operator and each of the spatial correlation matrices is defined by

$$[\mathbf{R}_{q,k}]_{m,p} = \frac{\gamma}{2\Delta_q} \int_{-\Delta_q + \phi_{q,k}}^{\Delta_q + \phi_{q,k}} e^{-j2\pi D(m-p)\sin(\alpha)} d\alpha \quad (38)$$

where  $q \in \{H, V\}$ ,  $\gamma$  denotes propagation path loss between the transmitter and the receiver,  $\Delta_q$  is the angular spread,  $D$  is the antenna elements spacing, and  $\phi_{q,k}$  is the angle of arrival (AoA) for the  $k$ -th user. We summarize the simulation parameters for UPA model in Table II. In Fig. 12, we plot the sum rate as a function of SNR and the number of feedback bits for  $N_t = 32, 64$  and  $K = 2$ . For the UPA model, we set  $N_V = 4$ ,  $N_H = 8$

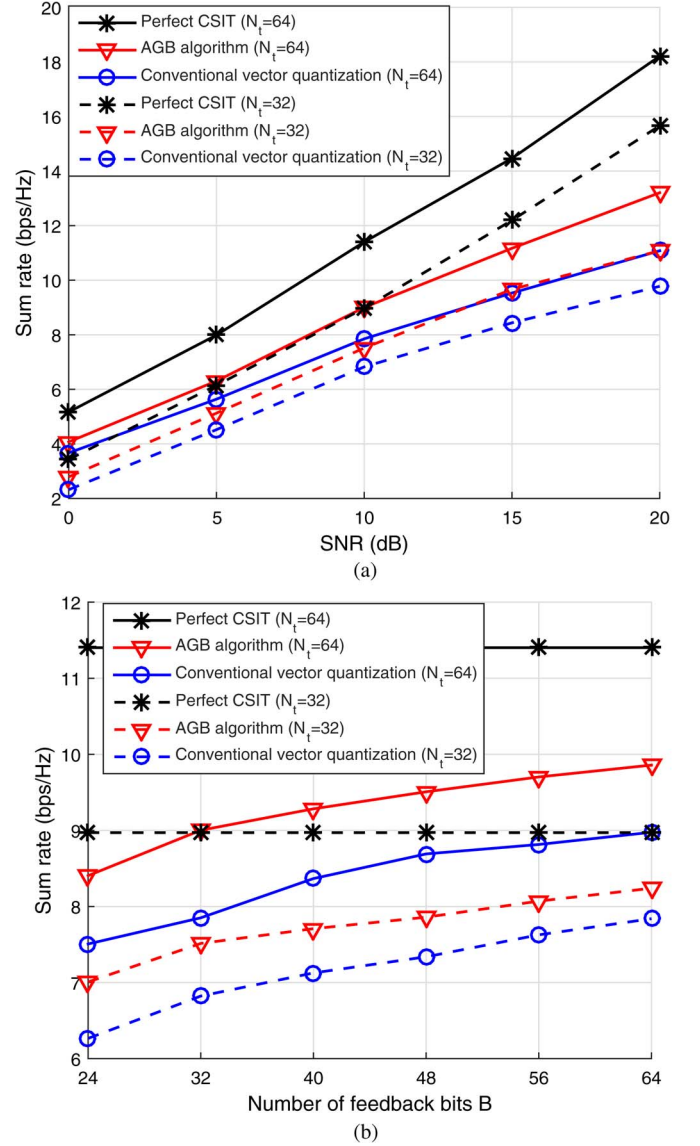


Fig. 12. Sum rate as a function of (a) SNR when  $B = (24, 8)$  for  $N_t = 32$  and  $B = (48, 16)$  for  $N_t = 64$  and (b) number of feedback bits  $B$  for UPA correlation model when SNR = 10 dB,  $B = (Bq, 8)$  for  $N_t = 32$ , and  $B = (Bq, 16)$  for  $N_t = 64$ .

for  $N_t = 32$ ,  $M = 4$  and  $N_V = 8$ ,  $N_H = 8$  for  $N_t = 64$ ,  $M = 8$ , respectively. We observe from Fig. 12(a) that the proposed approach achieves better sum rate than the conventional scheme produces, in particular for high SNR regime. We also observe from Fig. 12(b) that the AGB algorithm outperforms the conventional vector quantization technique with a large margin, resulting in more than 50% feedback overhead reduction.

So far, we have assumed that the receiver has knowledge of full CSI. In Fig. 13, we investigate the performance of the AGB algorithm when the estimated CSI is employed. Since the mismatch between the actual CSI and the estimated CSI is unavoidable in a real communication, and this might result in degradation performance, it is of importance to investigate the effect of channel estimation error. In our simulation, we use an additive channel estimation model where  $\mathbf{h}_{k,\text{est}} = \mathbf{h}_k + \mathbf{h}_{k,\text{err}}$  where  $\mathbf{h}_{k,\text{est}}$ ,  $\mathbf{h}_k$ , and  $\mathbf{h}_{k,\text{err}}$  represent the estimated channel

TABLE II  
SIMULATION PARAMETERS FOR UPA MODEL

Variables	Simulation parameters
Antenna elements spacing	$D = 0.5$
Propagation path loss	$\gamma = (1 + (\frac{s}{r})^{\alpha_{pl}})^{-1}$
Path loss exponent	$\alpha_{pl} = 3$
Angular spread (vertical)	$\Delta_V = \frac{1}{2} (\arctan(\frac{s+r}{u}) - \arctan(\frac{s-r}{u}))$
Angle of arrival (vertical)	$\phi_V = \frac{1}{2} (\arctan(\frac{s+r}{u}) + \arctan(\frac{s-r}{u}))$
Angular spread (horizontal)	$\Delta_H = \arctan(\frac{r}{s})$
Angle of arrival (horizontal)	$\phi_{H,k} \in (-\pi, \pi]$
Elevation of the transmit antenna	$u = 60m$
Radius of the scattering ring for the receiver	$r = 30m$
Distance from the transmitter	$s = 50m$

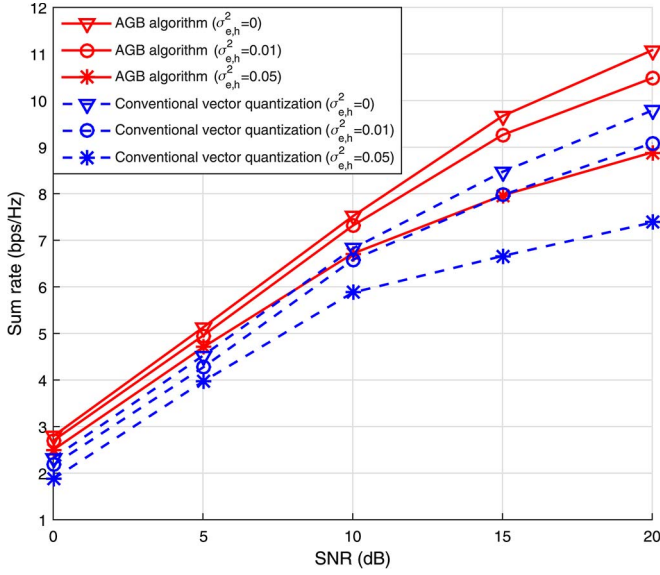


Fig. 13. Sum rate performance with various  $\sigma_{e,h}^2$  for UPA correlation model when  $B = (24, 8)$  ( $N_t = 32$ ,  $M = 4$ ,  $K = 4$ ,  $N_g = 16$ ).

vector, the original channel vector and the estimated error vector, respectively. We assume that  $\mathbf{h}_{k,\text{err}}$  is uncorrelated with  $\mathbf{h}_{k,\text{est}}$ , and  $\mathbf{h}_{k,\text{err}}$  has i.i.d elements with zero mean and the estimation error variance  $\sigma_{e,h}^2$ . We observe from Fig. 13 that the AGB algorithm is more robust to the estimation errors than the conventional vector quantization. For example, the sum rate gain at 7 bps/Hz of the AGB algorithm is about 5 dB over the conventional vector quantization when  $\sigma_{e,h}^2 = 0.05$ , while the gain is around 3 dB for  $\sigma_{e,h}^2 = 0.01$ .

Finally, in Fig. 14, we plot the sum rate as a function of  $\alpha$  for system with  $N_t = 32, 64$  and  $K = 4$ . In the AGB algorithm, we assign one bit per antenna elements on average. Specifically, for  $N_t = 32$ , we set  $B = (24, 8)$  and for  $N_t = 64$ , we set  $B = (48, 16)$ , respectively. In this case,  $N_g = 16$  for  $N_t = 32$  and  $N_g = 32$  for  $N_t = 64$ , respectively. It is worth mentioning that correlated fading tends to decrease the size of space that channel vectors span and hence is beneficial to reduce the quantization distortion by employing a user dependent channel statistic-based codebook [45]. As a result, the sum rate of multiuser MIMO systems increases with the transmit correlation coefficient. As shown in Fig. 14, when the transmit correlation coefficient  $\alpha$  increases, the antenna grouping operation becomes effective and thus the sum rate of

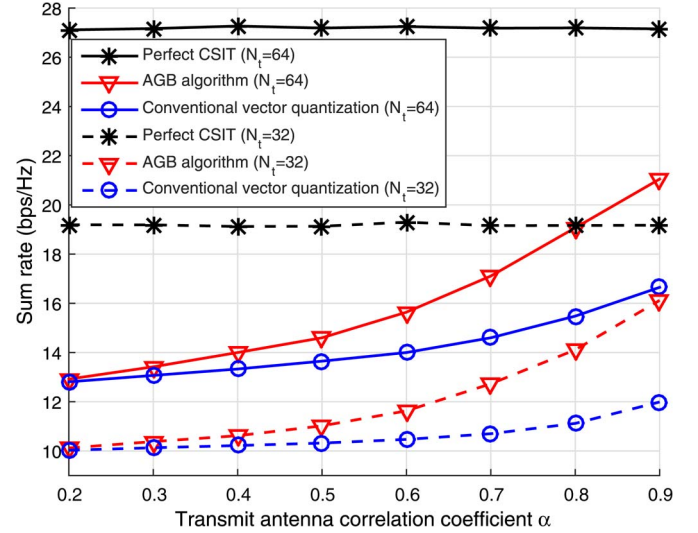


Fig. 14. Sum rate as a function of the transmit antenna correlation coefficient  $\alpha$  when SNR = 10 dB.

the AGB algorithm improves drastically. For example, when  $\alpha = 0.8$ , the sum rate gains of the AGB algorithm over the conventional vector quantization technique is 30% for  $N_t = 32$  and 25% for  $N_t = 64$ , respectively.

## V. CONCLUSION

In this paper, we proposed an efficient feedback reduction algorithm for FDD-based massive MIMO systems. Our work is motivated by the observation that the CSI feedback overhead must scale linearly with the number of transmit antennas so that conventional vector quantization approach performing the quantization of the whole channel vector is not an appropriate option for the massive MIMO regime. The key feature of the antenna group beamforming (AGB) algorithm is to control relentless growth of the CSI feedback information in the massive MIMO regime by mapping multiple correlated antenna elements into a single representative value with grouping patterns and then choosing the codeword from the codebook generated from the reduced dimension channel vector. It has been shown by distortion analysis and simulation results that the proposed AGB algorithm is effective in achieving a substantial reduction in the feedback overhead in the realistic massive MIMO channels.



Although our study in this work focused on the single-cell scenario, we expect that the effectiveness of the proposed method can be readily extended to multi-cell scenario. In fact, in the multi-cell scenario, more aggressive feedback compression is required since the channel information of the interfering cells as well as the desired cell may be needed at the base station to properly control inter-cell interference. In this scenario, the proposed AGB algorithm can be used as an effective means to achieve reduction in the feedback information. Also, investigation of nonlinear transmitter techniques with user scheduling [46] would be interesting direction to be investigated. Finally, we note that the proposed method can be nicely integrated into the dual codebooks structure in LTE-Advanced [47], [48] by feeding back the pattern index for long-term basis and the codebook index for short-term basis. Since the main target of the massive MIMO system is slowly varying or static channels, dual codebook based AGB algorithm will bring further reduction in feedback overhead.

#### APPENDIX A DERIVATION OF (23)

Denoting  $(\hat{\mathbf{h}}_A^H \hat{\mathbf{h}}_r)^*$  and  $\hat{\mathbf{h}}_B^H \hat{\mathbf{h}}_r$  as  $r_1(\cos \theta_1 + j \sin \theta_1)$  and  $r_2(\cos \theta_2 + j \sin \theta_2)$ ,  $\text{Re}[(\hat{\mathbf{h}}_A^H \hat{\mathbf{h}}_r)^*(\hat{\mathbf{h}}_B^H \hat{\mathbf{h}}_r)]$  becomes  $r_1 r_2 (\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2)$ . Under the assumption that sufficient number of bits is used and hence the distortion between  $\hat{\mathbf{h}}_A$  and  $\hat{\mathbf{h}}_r$  is small (i.e.,  $\theta_1 \approx 0$ ,  $r_1 \approx 1$ ),  $E[\text{Re}[(\hat{\mathbf{h}}_A^H \hat{\mathbf{h}}_r)^*(\hat{\mathbf{h}}_B^H \hat{\mathbf{h}}_r)]]$  is expressed as

$$\begin{aligned} E[\text{Re}[(\hat{\mathbf{h}}_A^H \hat{\mathbf{h}}_r)^*(\hat{\mathbf{h}}_B^H \hat{\mathbf{h}}_r)]] &= E[r_1 r_2 (\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2)] \\ &= E[r_1 r_2] E[\cos \theta_1 \cos \theta_2 - \sin \theta_1 \sin \theta_2] \\ &= E[r_1 r_2] E[\cos(\theta_1 + \theta_2)] \\ &\approx E[r_2] E[\cos \theta_2] \quad (39) \\ &= 0 \quad (40) \end{aligned}$$

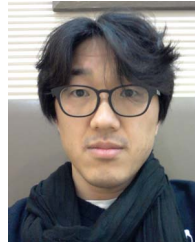
where (39) follows from the fact that  $\theta_1 \approx 0$ ,  $r_1 \approx 1$ , and (40) is because  $E[\cos \theta_2] = 0$  since  $\theta_2$  is uniformly distributed between  $-\pi$  and  $\pi$ .

#### REFERENCES

- [1] B. Lee, J. Choi, D. Love, J. Seol, and B. Shim, "Antenna grouping based feedback reduction technique for FDD-based massive MIMO systems," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 4477–4482.
- [2] B. Lee and B. Shim, "An efficient feedback compression for large-scale MIMO systems," in *Proc. IEEE Veh. Technol. Conf. Spring*, May 2014, pp. 1–5.
- [3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.
- [4] J. Hoydis, S. T. Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 160–171, Feb. 2013.
- [5] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation*, 3GPP TS 36.211 v11.0.0, 2012.
- [6] Y. Nam *et al.*, "Full-Dimension MIMO (FD-MIMO) for next generation cellular technology," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 172–179, Jun. 2013.
- [7] D. J. Love *et al.*, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, Oct. 2008.
- [8] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Trans. Wireless Commun.*, vol. 10, no. 8, pp. 2640–2651, Aug. 2011.
- [9] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7112–7139, Nov. 2014.
- [10] C. Shepard *et al.*, "Argos: Practical many-antenna base stations," in *Proc. ACM Mobicom Netw. Conf.*, Aug. 2012, pp. 53–64.
- [11] J. Choi, Z. Chance, D. J. Love, and U. Madhow, "Noncoherent trellis coded quantization: A practical limited feedback technique for massive MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 5016–5029, Dec. 2013.
- [12] J. Nam, J. Ahn, A. Adhikary, and G. Caire, "Joint spatial division and multiplexing: Opportunistic beamforming, user grouping and simplified downlink scheduling," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 876–890, Oct. 2014.
- [13] P. Kuo, H. T. Kung, and P. Ting, "Compressive sensing based channel feedback protocols for spatially correlated massive antenna arrays," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2012, pp. 492–497.
- [14] C. K. Au-Yeung, D. J. Love, and S. Sanayei, "Trellis coded line packing: Large dimensional beamforming vector quantization and feedback transmission," *IEEE Trans. Commun.*, vol. 10, no. 6, pp. 1844–1853, Jun. 2011.
- [15] K. T. Truong and R. W. Heath, Jr., "Effects of channel aging in massive MIMO systems," *J. Commun. Netw.*, vol. 15, no. 4, pp. 338–351, Aug. 2013.
- [16] D. Ying, F. W. Vook, T. A. Thomas, D. J. Love, and A. Ghosh, "Kronecker product correlation model and limited feedback codebook design in a 3D channel model," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 5865–5870.
- [17] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.
- [18] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 787–801, Oct. 2014.
- [19] M. Joham, P. M. Castro, W. Utschick, and L. Castedo, "Robust precoding with limited feedback design based on precoding MSE for MU-MISO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 3101–3111, Jun. 2012.
- [20] W. Santipach and M. L. Honig, "Asymptotic performance of MIMO wireless channels with limited feedback," in *Proc. IEEE Mil. Commun. Conf.*, Oct. 2003, pp. 141–146.
- [21] N. Jindal, "MIMO broadcast channels with finite rate feedback," *IEEE Trans. Inf. Theory*, vol. 52, no. 11, pp. 5045–5059, Nov. 2006.
- [22] C. K. Au-Yeung and D. J. Love, "On the performance of random vector quantization limited feedback beamforming in a MISO system," *IEEE Trans. Wireless Commun.*, vol. 6, no. 2, pp. 458–462, Feb. 2007.
- [23] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, "On beamforming with finite rate feedback in multiple-antenna systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2562–2579, Oct. 2003.
- [24] B. Clerckx, G. Kim, and S. J. Kim, "MU-MIMO with channel statistics based codebooks in spatially correlated channels," in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2008, pp. 1–5.
- [25] G. Caire and S. Shamai, "On the achievable throughput of a multi-antenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691–1706, Jul. 2003.
- [26] 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.
- [27] J. Park, B. Lee, and B. Shim, "A MMSE vector precoding with block diagonalization for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 60, no. 2, pp. 569–577, Feb. 2012.
- [28] K. Huang, J. G. Andrews, and R. W. Heath, Jr., "Performance of orthogonal beamforming for SDMA with limited feedback," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 152–164, Jan. 2009.
- [29] N. Ravindran and N. Jindal, "Limited feedback-based block diagonalization for the MIMO broadcast channel," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1473–1482, Oct. 2008.
- [30] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.
- [31] S. Park, B. Shim, and J. Choi, "Iterative channel estimation using virtual pilot signals for MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3032–3045, Jun. 2015.



- [32] J. Choi and B. Shim, "Compressive sensing based pilot reduction technique for massive MIMO systems," in *Proc. IEEE Inf. Theory Appl. Workshop*, San Diego, CA, USA, Feb. 2015, pp. 1–4.
- [33] P. Ding, D. J. Love, and M. D. Zoltowski, "Multiple antenna broadcast channels with shape feedback and limited feedback," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3417–3428, Jul. 2007.
- [34] D. J. Love and R. W. Heath, Jr., "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 718–722, Mar. 2006.
- [35] D. J. Love and R. W. Heath, Jr., "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Trans. Inf. Theory*, vol. 51, no. 8, pp. 2967–2976, Aug. 2005.
- [36] D. J. Love, R. W. Heath, Jr., and T. Strohmer, "Grassmannian beamforming for multiple-input multiple-output wireless systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2735–2747, Oct. 2003.
- [37] M. Herdin, N. Czink, H. Ozelcik, and E. Bonek, "Correlation matrix distance, a meaningful measure for evaluation of non-stationary MIMO channels," in *Proc. IEEE Veh. Technol. Conf. Spring*, Jun. 2005, pp. 136–140.
- [38] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and users selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, Sep. 2007.
- [39] C. Martin and B. Ottersten, "Asymptotic eigenvalue distributions and capacity for MIMO channels under correlated fading," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1350–1359, Jul. 2004.
- [40] R. Gray, "On the asymptotic eigenvalue distribution of Toeplitz matrices," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 6, pp. 725–730, Nov. 1972.
- [41] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1989.
- [42] J. Choi and D. J. Love, "Bounds on eigenvalues of a spatial correlation matrix," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1391–1394, Aug. 2014.
- [43] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing: The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.
- [44] J. G. Proakis, *Digital Communication*, 4th ed. New York, NY, USA: McGraw-Hill, 2000.
- [45] B. Clerckx, G. Kim, and S. J. Kim, "Correlated fading in broadcast MIMO channels: Curse or blessing?" in *Proc. IEEE Global Telecommun. Conf.*, Dec. 2008, pp. 1–5.
- [46] B. Lee and B. Shim, "A vector perturbation with user selection for multiuser MIMO downlink," *IEEE Trans. Commun.*, vol. 60, no. 11, pp. 3322–3331, Nov. 2012.
- [47] "Way Forward on 8Tx Codebook for Rel.10 DL MIMO," 3GPP TSG RAN WG1 #62, R1-105011, Aug. 2010.
- [48] C. Lim, T. Yoo, B. Clerckx, B. Lee, and B. Shim, "Recent trend of multiuser MIMO in LTE-advanced," *IEEE Commun. Mag.*, vol. 51, no. 3, pp. 127–135, Mar. 2013.



mobile communications since 2011. His current fields of interest include research/development of next generation mobile communication system and advanced PHY algorithms.



**David J. Love** (F'15) received the B.S. (with highest honors), M.S.E., and Ph.D. degrees in electrical engineering from The University of Texas at Austin, Austin, TX, USA, in 2000, 2002, and 2004, respectively. During the summers of 2000 and 2002, he was with Texas Instruments, Dallas, TX, USA. Since August 2004, he has been with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, where he is now a Professor and recognized as a University Faculty Scholar. He has served as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING, and a Guest Editor for special issues of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS and the *EURASIP Journal on Wireless Communications and Networking*. He is recognized as a Thomson Reuters Highly Cited Researcher and holds 23 issued U.S. patents.

Dr. Love is a Fellow of the Royal Statistical Society, and he has been inducted into Tau Beta Pi and Eta Kappa Nu. Along with co-authors, he was awarded the 2009 IEEE Transactions on Vehicular Technology Jack Neubauer Memorial Award for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY in that year and multiple Globecom best paper awards. He was the recipient of the Fall 2010 Purdue HKN Outstanding Teacher Award, Fall 2013 Purdue ECE Graduate Student Association Outstanding Faculty Award, and Spring 2015 Purdue HKN Outstanding Professor Award.



**Byungju Lee** (M'15) received the B.S. and Ph.D. degrees from the School of Information and Communication, Korea University, Seoul, Korea, in 2008 and 2014, respectively. He was a Visiting Scholar with Purdue University, West Lafayette, IN, USA, in 2013. He is presently a Postdoctoral Fellow at Seoul National University. His research interests include information theory and signal processing for wireless communications.



**Junil Choi** (M'15) received the B.S. (with honors) and M.S. degrees in electrical engineering from Seoul National University, Seoul, Korea, in 2005 and 2007, respectively, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2015. He is now a Post-Doctorate Fellow at The University of Texas at Austin, Austin, TX, USA. From 2007 to 2011, he was a Member of Technical Staff at Samsung Advanced Institute of Technology (SAIT) and Samsung Electronics Co. Ltd. in Korea, where

he contributed to advanced codebook and feedback framework designs for the 3GPP LTE/LTE-Advanced and IEEE 802.16 m standards. His research interests are in the design and analysis of massive MIMO, distributed communication, and vehicular communication systems.



**Byonghyo Shim** (SM'09) received the B.S. and M.S. degrees in control and instrumentation engineering from Seoul National University, Seoul, Korea, in 1995 and 1997, respectively. He received the M.S. degree in mathematics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana-Champaign (UIUC), Urbana, IL, USA, in 2004 and 2005, respectively.

From 1997 and 2000, he was with the Department of Electronics Engineering, Korean Air Force Academy as an Officer (First Lieutenant) and an Academic Full-Time Instructor. From 2005 to 2007, he was with Qualcomm Inc., San Diego, CA, USA, as a Staff Engineer. From 2007 to 2014, he was with the School of Information and Communication, Korea University, Seoul, Korea, as an Associate Professor. Since September 2014, he has been with the Department of Electrical and Computer Engineering, Seoul National University, Seoul, Korea, where he is presently an Associate Professor. His research interests include wireless communications, statistical signal processing, estimation and detection, compressive sensing, and information theory.

Dr. Shim was the recipient of the 2005 M. E. Van Valkenburg Research Award from the Electrical and Computer Engineering Department of the University of Illinois and 2010 Hadong Young Engineer Award from IEIE. He is currently an Associate Editor of the IEEE WIRELESS COMMUNICATIONS LETTERS, JOURNAL OF COMMUNICATIONS AND NETWORKS, and a Guest Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS.