Massive MIMO Channel Estimation via Compressed and Quantized Feedback

Mingjie Shao
Department of Electronic Engineering
The Chinese University of Hong Kong
Hong Kong SAR of China
mingjieshao@cuhk.edu.hk

Xiao Fu
School of Electrical Engineering and Computer Science
Oregon State University
Corvallis, Oregon, USA
xiao.fu@oregonstate.edu

***Abstract*—This paper focuses on downlink channel state information (CSI) acquisition. A frequency division duplex (FDD) of massive MIMO system is considered. In such systems, the base station (BS) obtains the downlink CSI from the mobile users' feedback. A key consideration is to reduce the feedback overhead while ensuring that the BS accurately recovers the downlink CSI. Existing approaches often resort to dictionary-based or tensor/matrix decomposition techniques, which either exhibit unsatisfactory accuracy or induce heavy computational load at the mobile end. To circumvent these challenges, this work formulates the limited channel feedback problem as a quantized and compressed matrix recovery problem. The formulation presents a computationally challenging maximum likelihood estimation (MLE) problem. An ADMM algorithm leveraging existing harmonic retrieval tools is proposed to effectively tackle the optimization problem. Simulations show that the proposed method attains promising channel estimation accuracy, using a much smaller amount of feedback bits relative to existing methods.**

***Index Terms*—channel state information, matrix/tensor recovery, quantization, compression**

I. Introduction

In multiple-input-multiple-output (MIMO) communication systems, downlink channel state information (CSI) at the base station (BS) is vital for many tasks including multiplexing, beamforming, and resource allocation [1], [2]. In the *massive* MIMO systems, the number of antennas at the base station/mobile users can easily reach tens or hundreds—i.e., the downlink channel matrix can be of very high dimension [1], [3]. Under such circumstances, acquiring accurate downlink CSI at the BS becomes a challenging task. Particularly, in frequency division duplex (FDD) systems, downlink CSI on the BS side is usually acquired from mobile users' feedback [4], [5], as channel reciprocity does not hold in FDD systems. However, transmitting back the full channel matrix to the BS is not viable, as it incurs high feedback overhead; see discussions in [6], [7].

Central to the implementation of an FDD massive MIMO system is a downlink CSI acquisition strategy using limited user feedback. Quantization and compression are most widely considered techniques for this purpose. A classic way is to apply vector quantization (VQ) on the (Rayleigh) channel matrix; see, e.g., [5]. Both mobile user and BS maintain a common VQ dictionary. The channel matrix is quantized as a codeword in the dictionary—and only the codeword is fed back to the BS [8], [9]. For angular channel models that is often used in massive MIMO systems, one can exploit the fact that the channel is sparse in the spatial angular domain [1], [6], [10], [11]. By applying grid discretization on the spatial angle, the channel estimation problem can be formulated as a compressed sensing (CS) problem [6], [10]. These approaches often exhibit unsatisfactory recovery accuracy, as neither the codewords nor the discretized spatial domain could fully capture the downlink CSI. Another line of work estimates the key parameters of the downlink channel at the mobile end, leveraging the parsimonious angular parametrization using tensor/matrix decomposition and harmonic retrieval techniques; see, e.g., [7], [12]–[14]. However, these approaches render heavy computational loads on the mobile end, which is undesired for various reasons, e.g., energy consumption considerations and hardware limitations.

In this paper, we revisit the limited feedback-based CSI acquisition problem in FDD massive MIMO systems. Our proposed scheme is as follows: the mobile user first compresses the channel matrix using a random compression matrix, and then quantizes the compressed channel matrix with random dithering. The quantized and compressed channel matrix is then transmitted back to the BS. This way, no complex computation is involved at the user end, and the amount of the feedback bits is under control. On the BS side, we propose to recover the downlink channel matrix using maximum-likelihood estimation (MLE)—which is in accordance with our random compression/quantization strategy. The MLE problem presents an optimization challenge due to its array manifold-enforcing nonconvex constraints. We propose an alternating direction method of multipliers (ADMM) to tackle the MLE problem. By a careful variable-splitting design, each ADMM subproblem can be efficiently tackled, leveraging existing effective harmonic retrieval solvers and quantzied compressive sensing tools. We showcase the effectiveness of the proposed approach through numerical simulations.

The work of X. Fu was supported by the National Science Foundation (NSF) under MLWiNS CNS-2003082. His work was also supported by a gift from Intel through the MLWiNS program.

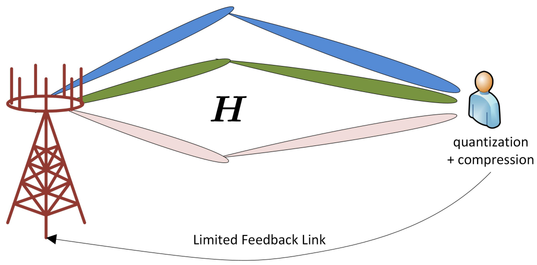


Fig. 1: Quantized and compressed channel CSI feedback

II. Problem Statement

A. Channel Model

We consider a frequency division duplex (FDD) massive MIMO system, where BS and each mobile user have N transmit antennas and M receive antennas, respectively. Suppose that both the BS and the user are equipped with uniform linear arrays. The downlink channel from BS to a user can be characterized by the following multipath model [1]:

$$\boldsymbol{H} = \sum_{k=1}^{K} \beta_k \boldsymbol{a}_r(\theta_k)\boldsymbol{a}_t(\phi_k)^H = \boldsymbol{A}_r(\boldsymbol{\theta})\mathrm{Diag}(\boldsymbol{\beta})\boldsymbol{A}_t(\boldsymbol{\phi})^H, \quad (1)$$

where K is the number of paths, $\beta_k \in \mathbb{C}$ is the path loss of path k, and $\boldsymbol{a}_r$ and $\boldsymbol{a}_t$ are the steering vectors at the user and BS sides, respectively, which are defined as

$$\boldsymbol{a}_r(\theta) = [1, e^{-\mathrm{j}\frac{2\pi d}{\lambda}\sin(\theta)}, \dots, e^{-\mathrm{j}\frac{2\pi d}{\lambda}\sin(\theta)(M-1)}]^\top,$$
$$\boldsymbol{a}_t(\phi) = [1, e^{-\mathrm{j}\frac{2\pi d}{\lambda}\sin(\phi)}, \dots, e^{-\mathrm{j}\frac{2\pi d}{\lambda}\sin(\phi)(N-1)}]^\top,$$

in which d is inter-antenna spacing, λ is the carrier wavelength, θ_k and ϕ_k are the angle of arrival at the user and angle of departure at the BS, respectively. In (1), we have used the notation $\boldsymbol{\theta} = [\theta_1, \dots, \theta_K]^\top$, $\boldsymbol{\phi} = [\phi_1, \dots, \phi_K]^\top$, and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^\top$; in addition, $\mathrm{Diag}(\boldsymbol{x})$ forms a diagonal matrix with $\boldsymbol{x}$ being the diagonal elements.

The BS needs the downlink CSI, i.e., $\boldsymbol{H}$, to perform many tasks, such as beamforming and resource allocation [1], [2], [4]. In FDD systems, the mobile user can acquire $\boldsymbol{H}$ in the training phase and send it back to the BS via feedback channel; see Fig. 1 for an illustration. However, $\boldsymbol{H}$ is a high-dimensional matrix in massive MIMO systems, which may incur high overhead if the feedback is done naïvely.

B. Existing Approaches

From (1), it is seen that $\boldsymbol{H}$ can be represented by K paths, which exhibits sparsity in the spatial angular domain. As a result, the works in [6], [10] proposed to feed back a compressed version of $\boldsymbol{H}$, and let the BS recover it via a compressive sensing(CS)-based formulation. For example, the work in [6] approximated $\boldsymbol{H}$ via spatial discretization, i.e.,

$$\boldsymbol{H} \approx \boldsymbol{D}_r\boldsymbol{G}\boldsymbol{D}_t^H, \quad (2)$$

where

$$\boldsymbol{D}_r = [\boldsymbol{a}_r(\theta_1), \dots, \boldsymbol{a}_r(\theta_L)], \ \ \boldsymbol{D}_t = [\boldsymbol{a}_t(\phi_1), \dots, \boldsymbol{a}_t(\phi_L)]$$

are the steering vector dictionaries formed by discretizing angle $[0, \pi]$ using L grids, and $\boldsymbol{G}$ is a sparse matrix whose (i, j)th entry being nozero "activates" the ith and jth columns of $\boldsymbol{D}_r$ and $\boldsymbol{D}_t$, respectively. The BS receives a compressed version of $\boldsymbol{H}$ and uses a CS-based formulation to recover the sparse $\boldsymbol{G}$. A caveat lies in the tradeoff between approximation accuracy and model complexity. Increasing L can improve the approximation accuracy in (2), but the dimension of $\boldsymbol{G}$ increases quadratically with L. Thus, this approach often exhibits unsatisfactory CSI estimation accuracy if the number of feedbac bits is limited, as will be seen in the simulations.

Another line of work exploited the fact that $\boldsymbol{H}$ resides on the Vandermonde array manifolds parameterized by $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi})$, and to use tensor/matrix methods and harmonic retrieval algorithms to estimate the parameters at the user side [7], [12]–[14]. Then, the mobile user sends the parameters to the BS. These methods usually show better recovery performance, as the angular domain is not discretized. But the computational burden on the user side is often quite high, as estimating $(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\phi})$ from $\boldsymbol{H}$ *per se* is a highly nontrivial optimization problem. Note that from a system design viewpoint, letting mobile users carry out heavy computations is not desired, as the mobile devices only have limited energy supplies and the hardware is less powerful compared to the BS.

C. Quantized and Compressed Feedback

In this work, we propose a CSI estimation method via exploiting the manifold structure of $\boldsymbol{H}$ as in [7], [14], but our method shifts the computational burden to the BS.

To proceed, let

$$\boldsymbol{h} = \mathrm{vec}(\tilde{\boldsymbol{H}}), \quad \tilde{\boldsymbol{H}} = \begin{bmatrix} \Re(\boldsymbol{H}) \\ \Im(\boldsymbol{H}) \end{bmatrix} \in \mathbb{R}^{2M \times N}. \quad (3)$$

We first compress $\boldsymbol{h}$ by a random matrix $\boldsymbol{A} \in \mathbb{R}^{R \times 2MN}$ with $R \ll 2MN$, whose elements follow $a_{i,j} \sim \mathcal{N}(0, 1)$. The compressed channel measurements are expressed as follows:

$$\boldsymbol{r} = \boldsymbol{A}\boldsymbol{h}. \quad (4)$$

Then, we apply quantization with random dithering to $\boldsymbol{r}$ and get

$$\boldsymbol{y} = \mathcal{Q}(\boldsymbol{r} + \boldsymbol{v}), \quad (5)$$

where $\mathcal{Q}$ is an elementwise uniform quantizer with $2D$ quantization levels $\{\pm b, \pm 3b, \dots, \pm(2D-1)b\}$, $2b$ is the quantization interval, and $\boldsymbol{v} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ is artificially added noise. For signal quantization, it is well-known that adding a certain amount of artificial noise $\boldsymbol{v}$ may have a "dithering" effect and enhances the recovery performance. This is because signal correlation together with the noise perturbation help retain information (e.g., whether or not the unquantized signal value is close to the bin boundary) lost in the quantization process; see [15]. The mobile user feeds back $\boldsymbol{y}$, rather than $\boldsymbol{H}$ itself, to the BS. Assuming that $\boldsymbol{y}$ is received by the BS without losses, the BS recovers $\boldsymbol{H}$ from $\boldsymbol{y}$, given the knowledge of $\boldsymbol{A}$ and $\mathcal{Q}$.

III. Proposee CSI Recovery Algorithm

We adopt maximum-likelihood estimation (MLE) at the BS to recover $\boldsymbol{H}$. Under (3)-(5), the ML estimator is given by

$$\begin{aligned}\min_{\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\varphi}} & -\sum_{i=1}^{R}\log\left[\Phi\Big(\frac{\bar{b}_i-\boldsymbol{a}_i^\top\boldsymbol{h}}{\sigma}\Big)-\Phi\Big(\frac{\underline{b}_i-\boldsymbol{a}_i^\top\boldsymbol{h}}{\sigma}\Big)\right] \\ \text{s.t. } & \Omega(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\phi})=\boldsymbol{H},\end{aligned} \tag{6}$$

where $\Omega(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\phi})$ represents the model in (1) of the channel matrix $\boldsymbol{H}$, Φ is the cumulative distribution function (CDF) of standard Gaussian distribution, and $\boldsymbol{a}_i$ is the ith row of $\boldsymbol{A}$. The estimator finds the model parameters that maximize the likelihood of y_i residing in the quantization interval $[\underline{b}_i,\bar{b}_i]$. Problem (6) can be regarded as a structured matrix recovery problem, where the latent factors of $\boldsymbol{H}$ are generated over array manifolds. We see that the manifold constraint poses great challenge in algorithm design.

Problem 6 is a nontrivial manifold-constrained nonconvex optimization problem. We propose an ADMM algorithm [16] to tackle (6). The augmented Lagrangian function associated to problem (6) can be expressed as

$$\begin{aligned}\mathcal{L}_{\boldsymbol{\Lambda}}(\boldsymbol{\Gamma},\boldsymbol{U})=&-\sum_{i=1}^{R}\log\left[\Phi\Big(\frac{\bar{b}_i-\boldsymbol{a}_i^\top\boldsymbol{h}}{\sigma}\Big)-\Phi\Big(\frac{\underline{b}_i-\boldsymbol{a}_i^\top\boldsymbol{h}}{\sigma}\Big)\right] \\ &+\langle\boldsymbol{\Lambda},\boldsymbol{H}-\boldsymbol{U}\rangle+\frac{\rho}{2}\|\boldsymbol{H}-\boldsymbol{U}\|_F^2,\end{aligned} \tag{7}$$

where $\boldsymbol{\Gamma}=(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\phi})$, $\boldsymbol{U}=\Omega(\boldsymbol{\Gamma})$, $\boldsymbol{\Lambda}$ is the Lagrangian dual variable, $\rho>0$ is a pre-specified stepsize parameter. At the ℓth iteration, the ADMM algorithm has the following updates:

$$\boldsymbol{\Gamma}^{\ell+1}=\arg\min_{\boldsymbol{\Gamma}}\frac{1}{2}\|\boldsymbol{U}-(\boldsymbol{H}^\ell+\frac{1}{\rho}\boldsymbol{\Lambda}^\ell)\|_F^2, \quad \text{s.t. } \Omega(\boldsymbol{\Gamma})=\boldsymbol{U} \tag{8a}$$

$$\begin{aligned}\boldsymbol{H}^{\ell+1}=\arg\min_{\boldsymbol{H}}&-\sum_{i=1}^{R}\log\left[\Phi\Big(\frac{\bar{b}_i-\boldsymbol{a}_i^T\boldsymbol{h}}{\sigma}\Big)-\Phi\Big(\frac{\underline{b}_i-\boldsymbol{a}_i^T\boldsymbol{h}}{\sigma}\Big)\right] \\ &+\langle\boldsymbol{\Lambda}^\ell,\boldsymbol{H}-\boldsymbol{U}^{\ell+1}\rangle+\frac{\rho}{2}\|\boldsymbol{H}-\boldsymbol{U}^{\ell+1}\|_F^2,\end{aligned} \tag{8b}$$

$$\boldsymbol{\Lambda}^{\ell+1}=\boldsymbol{\Lambda}^\ell+\rho(\boldsymbol{H}^{\ell+1}-\boldsymbol{U}^{\ell+1}). \tag{8c}$$

Note Problems (8a) and (8b) are still nontrivial optimization problems—yet each of them admits efficient off-the-shelf solvers.

A. Harmonic Retrieval for Problem (8a)

Problem (8a) is known as a 2D *harmonic retrieval* (HR) problem in the literature, which has been widely studied [17]–[19]. We adopt the RELAX method [18] to find a solution, which mainly uses 2D FFT. We outline how RELAX is applied to problem (8a). Note the fact that

$$\text{vec}(\boldsymbol{A}_r(\boldsymbol{\theta})\text{Diag}(\boldsymbol{\beta})\boldsymbol{A}_t(\boldsymbol{\phi})^H)=\sum_{k=1}^{K}\boldsymbol{a}_t^*(\phi_k)\otimes\boldsymbol{a}_r(\theta_k)\beta_k,$$

where $\otimes$ is the Kronecker product and * denotes the complex conjugate. Problem (8a) can be rewritten as

$$\min_{\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\phi}}\left\|\boldsymbol{\xi}-\sum_{k=1}^{K}\boldsymbol{a}_t^*(\phi_k)\otimes\boldsymbol{a}_r(\theta_k)\beta_k\right\|_2^2,$$

where $\boldsymbol{\xi}=\text{vec}(\boldsymbol{H}+\frac{1}{\rho}\boldsymbol{\Lambda})$.

RELAX updates $(\boldsymbol{\theta},\boldsymbol{\beta},\boldsymbol{\phi})$ in a coordinate-wise fashion. In each step, given $\{\theta_j,\phi_j,\beta_j\}_{j\neq k}$, RELAX updates $(\theta_k,\phi_k,\beta_k)$. Define

$$\boldsymbol{\xi}_k=\boldsymbol{\xi}-\sum_{j\neq k}\boldsymbol{a}_t^*(\phi_j)\otimes\boldsymbol{a}_r(\theta_j)\beta_j.$$

Then, the optimization problem with respect to $(\theta_k,\phi_k,\beta_k)$ is given by

$$\min_{\theta_k,\phi_k,\beta_k}\|\boldsymbol{\xi}_k-\boldsymbol{a}_t^*(\phi_k)\otimes\boldsymbol{a}_r(\theta_k)\beta_k\|_2^2. \tag{9}$$

The optimal β_k to (9) given other parameters is $\hat{\beta}_k=\frac{(\boldsymbol{a}_t(\phi_k)\otimes\boldsymbol{a}_r(\theta_k))^H\boldsymbol{\xi}_k}{MN}$. By substituting $\hat{\beta}_k$ back into (9), we get

$$(\theta_k,\phi_k)=\arg\max_{\theta_k,\phi_k}|(\boldsymbol{a}_t^*(\phi_k)\otimes\boldsymbol{a}_r(\theta_k))^H\boldsymbol{\xi}_k|^2,$$

which can be solved by 2D FFT; see [18] for more details. As seen, the subproblem with respect to $(\theta_k,\phi_k,\beta_k)$ for each k can be efficiently solved. The RELAX algorithm is shown in Algorithm 1.

Algorithm 1 RELAX for 2D Harmonic Retrieval [18]

```
Input: β = 0
for i = 1, ..., K do
    repeat
        for k = 1, ..., i do
            ξ_k = ξ − Σ_{j≠k}^{i} a_t^*(φ_j) ⊗ a_r(θ_j) β_j
            solve max_{θ_k,φ_k,β_k} ‖ξ_k − a_t^*(φ_k) ⊗ a_r(θ_k) β_k‖_2^2
        end for
    until {θ_k, φ_k, β_k}_{k=1}^{i} converges
end for
```

B. EM Algorithm for Problem (8b)

Problem (8b) is an unconstrained convex optimization problem, which can be solved by proximal gradient method [20] or expectation-maximization (EM) algorithm [21], [22]. We apply the EM algorithm to tackle (8b). The EM algorithm relies on the following variational inequality

$$-\log\left[\Phi\Big(\frac{\bar{b}_i-\boldsymbol{a}_i^T\boldsymbol{h}}{\sigma}\Big)-\Phi\Big(\frac{\underline{b}_i-\boldsymbol{a}_i^T\boldsymbol{h}}{\sigma}\Big)\right]\leq\frac{1}{2\sigma^2}|z_i-\boldsymbol{a}_i^T\boldsymbol{h}|^2+c,$$

where

$$z_i=\boldsymbol{a}_i^\top\bar{\boldsymbol{h}}+\sigma\frac{\phi(\frac{\underline{b}_i-\boldsymbol{a}_i^T\bar{\boldsymbol{h}}}{\sigma})-\phi(\frac{\bar{b}_i-\boldsymbol{a}_i^T\bar{\boldsymbol{h}}}{\sigma})}{\Phi(\frac{\bar{b}_i-\boldsymbol{a}_i^T\bar{\boldsymbol{h}}}{\sigma})-\Phi(\frac{\underline{b}_i-\boldsymbol{a}_i^T\bar{\boldsymbol{h}}}{\sigma})}$$

for some $\bar{\boldsymbol{h}}$, and equality holds when $\bar{\boldsymbol{h}}=\boldsymbol{h}$; ϕ is the probability density function of standard Gaussian distribution and c is a constant irrelevant to $\boldsymbol{h}$. Let t be the EM iteration index.

For solving (8b), each EM iteration involves the following two iterative steps:

$$\text{E-step}: z_i^{t+1} = \boldsymbol{a}_i^\top \boldsymbol{h}^t + \sigma \frac{\phi(\frac{b_i - \boldsymbol{a}_i^T \boldsymbol{h}^t}{\sigma}) - \phi(\frac{\bar{b}_i - \boldsymbol{a}_i^T \boldsymbol{h}^t}{\sigma})}{\Phi(\frac{\bar{b}_i - \boldsymbol{a}_i^T \boldsymbol{h}^t}{\sigma}) - \Phi(\frac{b_i - \boldsymbol{a}_i^T \boldsymbol{h}^t}{\sigma})},$$

$$\begin{aligned}\text{M-step}: \boldsymbol{h}^{t+1} &= \arg\min_{\boldsymbol{h}} \frac{1}{2\sigma^2}\|\boldsymbol{A}\boldsymbol{h} - \boldsymbol{z}^{t+1}\|^2 \\ &+ \langle \boldsymbol{\Lambda}^\ell, \boldsymbol{H} - \boldsymbol{U}^{\ell+1}\rangle + \frac{\rho}{2}\|\boldsymbol{H} - \boldsymbol{U}^{\ell+1}\|_F^2 \\ &= (\frac{1}{2\sigma^2}\boldsymbol{A}^\top \boldsymbol{A} + \frac{\rho}{2}\boldsymbol{I})^{-1}(\frac{1}{\sigma^2}\boldsymbol{A}^\top \boldsymbol{z}^{k+1} - \boldsymbol{\lambda}^k + \rho \boldsymbol{u}^{\ell+1}),\end{aligned} \tag{10}$$

where $\boldsymbol{\lambda} = \text{vec}(\boldsymbol{\Lambda})$, $\boldsymbol{u} = \text{vec}(\boldsymbol{U})$. When the EM algorithm converges, e.g., $\|\boldsymbol{h}^{t+1} - \boldsymbol{h}^t\|_2 \leq \epsilon$ for some $\epsilon > 0$, we output $\boldsymbol{h}^{\ell+1} = \boldsymbol{h}^{t+1}$.

To summarize, the overall channel feedback and recovery scheme is shown in Algorithm 2. One can see that most of computations are carried out at the BS end, which is desired.

Algorithm 2 Channel Feedback and Recovery Scheme

1: **input** channel $\boldsymbol{H}$
2: **mobile user**:
3: compress channel by $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{h}$,
4: quantize channel by $\boldsymbol{y} = \mathcal{Q}(\boldsymbol{r} + \boldsymbol{v})$,
5: send $\boldsymbol{y}$ to the BS.
6: **BS**:
7: initializations of $\boldsymbol{H}^0$ and $\boldsymbol{\Lambda}^0$
8: **repeat**
9: solve (8a) by Algorithm 1 to get $\boldsymbol{\Gamma}^{\ell+1}$,
10: solve (8b) by EM (10) to get $\boldsymbol{H}^{\ell+1}$,
11: update $\Lambda^{\ell+1}$ by (8c),
12: **until** some convergence criterion is satisfied.

IV. Simulation

First, we consider the case where the quantizer $\mathcal{Q}$ only uses one bit, i.e., binary quantization. We consider the following benchmark algorithms: 1) CS formulation solved by proximal gradient [6], termed as "CS-PG" and 2) CS formulation solved by EM and matching pursuit [10], termed as "EM-MP". For these methods, we discretzie $[0, \pi]$ into 256 equi-spaced intervals. We need to mention that CS-PG [6] and EM-MP [10] were not proposed under the same feedback scheme in Algorithm 2; we repurpose them to handle the feedback scheme in Algorithm 2 for the sake of performance comparison.

The MIMO channel size is $(M, N) = (16, 16)$. The number of paths $K = 4$. The noise power is 30dB. Fig. 2 shows the normalized mean squared error (NMSE) performance under different number of compression measurements R. The NMSE is defined as

$$\text{NMSE} = \left\| \frac{\hat{\boldsymbol{H}}}{\|\hat{\boldsymbol{H}}\|_F} - \frac{\boldsymbol{H}}{\|\boldsymbol{H}\|_F} \right\|_F^2,$$

where $\hat{\boldsymbol{H}}$ is the output of considered algorithms.

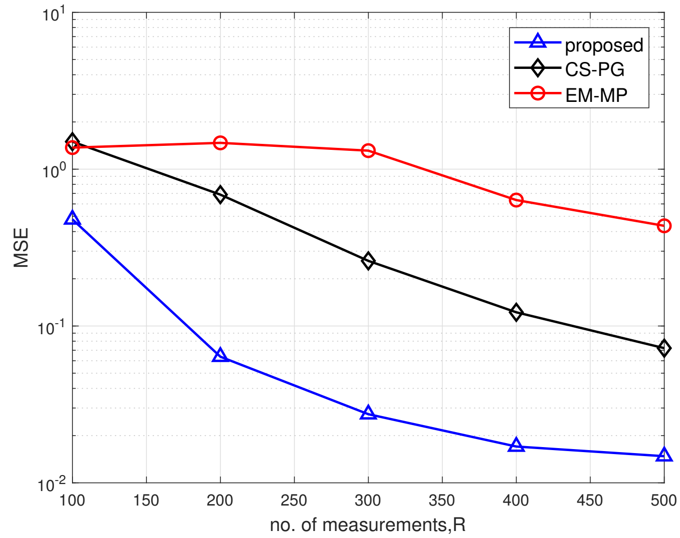


Fig. 2: NMSE performance of considered algorithms

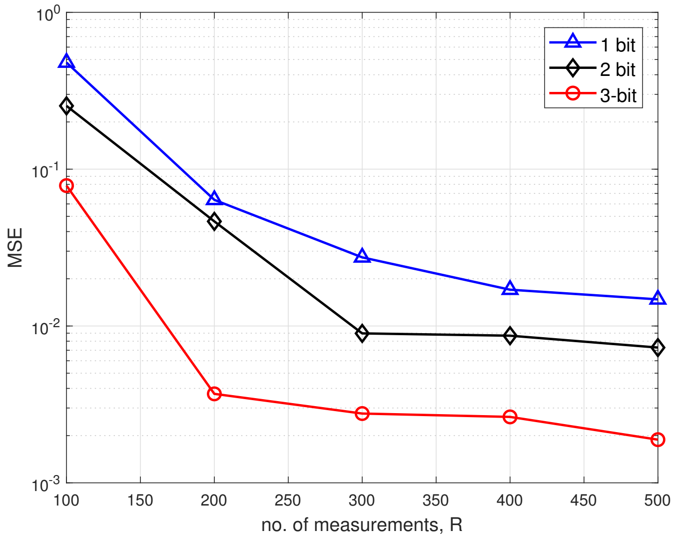
It is seen that the proposed algorithm achieves better recovery performance. Notably, the proposed method attains an NMSE $\leq 10^{-1}$ when $R = 200$ bits of feedback are used, while the most competitive baseline (i.e., CS-PG) could not attain the same NMSE using $R \leq 450$. This presents a substantial overhead saving by more than 50%.

In this case, transmitting $\boldsymbol{h}$ under float-16 format will require over $8,000$ bits. We see that our proposed approach attains reasonable performance when the number of compression measurements $R = 300$, which requires only 300 bits. As a result, over 95% feedback overhead is saved.

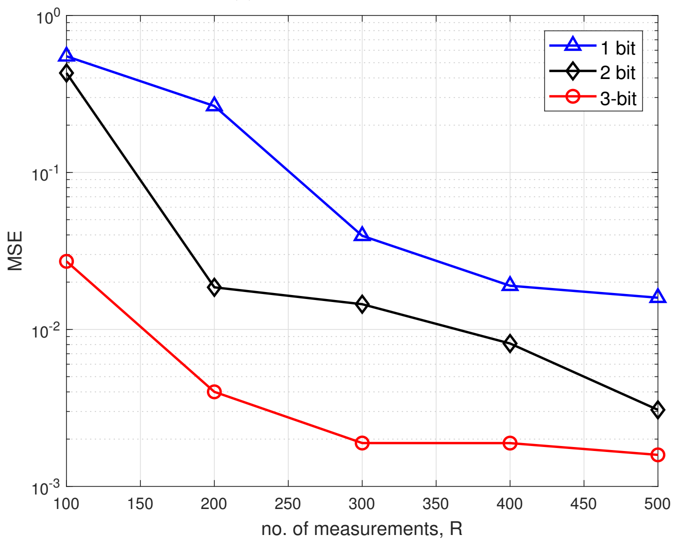
Next, we test the performance of proposed algorithm with respect to different levels of quantization and the different number of compression measurements. Fig. 3 shows the performance under two different problem sizes. First, it is seen that with increasing compression measurements R, the recovery performance gets better—as expected. Second, if each measurement of $\boldsymbol{y}$ is quantized using more bits, the performance of the proposed method is further enhanced. More specifically, our empirical results suggest that using 3-bit quantization and $R = 200$ compression measurements is more preferable than using 1-bit quantization and $R = 600$—as shown in Fig. 3. Note that both strategies feed back the same amount of bits, but the 3-bit quantization may make the recovery problem less challenging.

V. Conclusion

In this paper, we revisited the limited feedback based CSI estimation problme in FDD massive MIMO systems. We proposed a random quantization and compression strategy and an MLE-based CSI recovery algorithm at the user end and the BS side, respectively. Like prior works using CS-based formulation, our feedback strategy only requires the user to perform relatively simple matrix-vector multiplication operations. However, our recovery accuracy is substantially higher than those of the compressive sensing-based baselines.



(a) $N = 16$, $M = 16$



(b) $N = 32$, $M = 32$

Fig. 3: NMSE performance of proposed algorithm.

REFERENCES

[1] R. W. Heath, N. Gonzalez-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE Journal Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, 2016.

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

[3] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, 2012.

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

[5] D. J. Love, R. W. Heath, V. K. Lau, D. Gesbert, B. D. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341–1365, 2008.

[6] P. N. Alevizos, X. Fu, N. D. Sidiropoulos, Y. Yang, and A. Bletsas, "Limited feedback channel estimation in massive MIMO with non-uniform directional dictionaries," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5127–5141, 2018.

[7] C. Qian, X. Fu, N. D. Sidiropoulos, and Y. Yang, "Tensor-based channel estimation for dual-polarized massive MIMO systems," *IEEE Trans. on Signal Process.*, vol. 66, no. 24, pp. 6390–6403, 2018.

[8] 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, 2013.

[9] T. L. Marzetta and B. M. Hochwald, "Fast transfer of channel state information in wireless systems," *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1268–1278, 2006.

[10] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *48th Asilomar Conf. Signals, Systems and Computers*. IEEE, 2014, pp. 957–961.

[11] C. Qian, X. Fu, and N. D. Sidiropoulos, "Amplitude retrieval for channel estimation of MIMO systems with one-bit ADCs," *IEEE Signal Process. Lett.*, vol. 26, no. 11, pp. 1698–1702, 2019.

[12] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and S. Li, "Channel estimation for millimeter-wave multiuser MIMO systems via PARAFAC decomposition," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7501–7516, 2016.

[13] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, "Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 7, pp. 1524–1538, 2017.

[14] C. Qian, X. Fu, and N. D. Sidiropoulos, "Algebraic channel estimation algorithms for FDD massive MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 961–973, 2019.

[15] M. A. Davenport, Y. Plan, E. Van Den Berg, and M. Wootters, "1-bit matrix completion," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 189–223, 2014.

[16] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[17] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoustics Speech Signal Process.*, vol. 34, no. 5, pp. 1081–1089, 1986.

[18] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, 1996.

[19] C. Qian, Y. Shi, L. Huang, and H. C. So, "Robust harmonic retrieval via block successive upper-bound minimization," *IEEE Trans. Signal Process.*, vol. 66, no. 23, pp. 6310–6324, 2018.

[20] A. Zymnis, S. Boyd, and E. Candes, "Compressed sensing with quantized measurements," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 149–152, 2009.

[21] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *Proc. IEEE Int. ITG Workshop Smart Antennas (WSA)*. IEEE, 2010, pp. 143–150.

[22] M. Shao, W.-K. Ma, J. Liu, and Z. Huang, "Accelerated and deep expectation maximization for one-bit MIMO-OFDM detection," *arXiv preprint arXiv:2210.03888*, 2022.