Dual-regularized Feedback and Precoding for D2D Assisted MIMO Systems

Junting Chen, *Member, IEEE*, Haifan Yin,
Laura Cottatellucci, *Member, IEEE*, David Gesbert, *Fellow, IEEE*

***Abstract*—This paper considers the problem of efficient feedback design for massive multiple-input multiple-output (MIMO) downlink transmissions in frequency division duplexing (FDD) bands, where some partial channel state information (CSI) can be directly exchanged between users via device-to-device (D2D) communications. Drawing inspiration from classical point-to-point MIMO, where efficient mechanisms are obtained by feeding back directly the precoder, this paper proposes a new approach to bridge the channel feedback and the precoder feedback by the joint design of the feedback and precoding strategy following a team decision framework. Specifically, the users and the base station (BS) minimize a common mean squared error (MSE) metric based on their individual observations on the imperfect global CSI. The solutions are found to take similar forms as the regularized zero-forcing (RZF) precoder, with additional regularizations that capture any level of uncertainty in the exchanged CSI, in case the D2D links are absent or unreliable. Numerical results demonstrate superior performance of the proposed scheme for an arbitrary D2D link quality setup.**

***Index Terms*—MIMO, device-to-device, limited feedback, precoding, CSI exchange, FDD systems**

I. Introduction

MIMO systems can achieve a large throughput gain by exploiting the spatial degrees of freedom using multiple antennas. To realize such advantage, the BS needs to collect the instantaneous global CSI from the users and compute the precoder in a centralized way [1]–[9]. While there are a lot of recent works focusing on time-division duplex (TDD) systems where channel reciprocity can be exploited for downlink CSI acquisition, FDD systems still dominate commercial cellular networks where much of the licensed spectrum comes in paired chunks. In FDD massive MIMO systems, acquiring downlink CSI is challenging because the CSI, as estimated at the user side, needs to be fed back to the BS using limited feedback resources. Traditional approaches for feedback reduction include, for example, channel codebook design using quantization theories [1]–[4], robust precoding under limited CSI [5]–[7], and rate splitting encoding strategy [8], [9].

As considered to be one of the main novelties brought by 5G standards, D2D communication is a promising technology that allows users to directly communicate with each other without routing the message via the BS. In recent years, D2D assisted MIMO transmission has attracted increasing attention [10]–[20]. With D2D communications, some users can act as relays to assist the data transmission to the target users, by reusing the radio resources of the cellular network through proper power control [13], [14], or transmitting in the underlay mode with adaptive precoder design [15]–[17] or receive mode selection [18]. In addition, the works [19] and [20] studied the precoder design and analyzed the spectrum efficiency for massive MIMO systems with underlay D2D communications.

While the bulk of the literature focuses on using D2D to deliver data streams, there are some works exploiting D2D to assist the signaling in the context of MIMO cellular transmission. Specifically, the authors in [11] and [21] proposed a *precoder feedback scheme* for FDD multiuser MIMO systems, where the users first obtain the global CSI via D2D communications, and then compute and feed back the precoder to the BS through a rate-limited uplink feedback channel. It is demonstrated in [11], [21] that in the ideal case when users have perfect global CSI via infinite-rate D2D, the precoder feedback scheme can achieve signifiant throughput gain over the CSI feedback scheme. This result draws inspiration from the problem of feedback design in classical point to point MIMO under rate-limited (i.e. quantized) feedback [22]. In particular, with perfect global CSI at the user side, it is better to design a quantizer with an objective to maximize the ultimate signal-to-interference-and-noise ratio (SINR) or the sum rate, rather than to merely minimize the individual channel distortion. While computing a global precoder at the user side is generally not possible in multiuser settings (due to the locality of channel state information available at the user), it becomes possible if D2D links are available.

However, prior works [11] and [21] have several limitations. First, high quality D2D communications may not always be available in practical systems, since the D2D link is also capacity limited and subject to transmission latency. Consequently, with distortions in CSI exchange due to quantization and transmission latency, the performance of the precoder

Manuscript received December 30, 2016; revised May 14, 2017; accepted July 18, 2017.

The support of Huawei France Research Center is gratefully acknowledged. David Gesbert was also supported in part by the ERC under the European Union Horizon 2020 research and innovation program (Agreement no. 670896). The material in this paper was presented, in part, at the Asilomar Conference on Signals, Systems and Computers, Asilomar, CA, USA, Nov. 6–9, 2016.

Junting Chen was with Department of Communication Systems, EURECOM, 06410 Sophia-Antipolis, France. He is now with Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA (e-mail:juntingc@usc.edu).

Haifan Yin was with Department of Communication Systems, EURECOM, 06410 Sophia-Antipolis, France. He is now with Sequans Communications, 92700 Paris, France.

Laura Cottatellucci and David Gesbert are with Department of Communication Systems, EURECOM, 06410 Sophia-Antipolis, France (e-mail: {cottatel, gesbert}@eurecom.fr).

feedback scheme may significantly degrade. Second, as found in [21], when the D2D link quality falls too low, it is better to switch from the precoder feedback regime to the CSI feedback regime. However, a binary switch feedback strategy between the two regimes is difficult to optimize and bound to be suboptimal. In particular, even when the switch point may be found, the switch-based scheme may not perform well in the regime of medium to low D2D quality, because it simply discards the additional CSI exchanged between users and switches back to a classical CSI feedback scheme. Another desirable feature is the adaptation to spatial heterogeneities in D2D links, i.e., some pairs of users may be in close range and have good D2D capabilities while other pairs may be in long range and have poor D2D conditions. An underlying challenge is to fathom a feedback strategy which could bridge the precoder feedback and the CSI feedback into a single unified feedback strategy.

This paper addresses all the above issues. It aims at developing a new family of feedback and precoding strategies for the users and the BS in the scenarios of heterogeneous D2D link qualities between the users. Specifically, we adopt a team decision approach, whereby the users and the BS share a common objective to minimize a modified MSE criterion of the received signals at the user side. More specifically, the users compute a feedback vector to minimize the expected MSE based on their individual observations of the global CSI, and the BS computes a regularized minimum mean square error (MMSE) precoder based on the feedback from the users. We show that the solutions to such team decision problems have a *dual regularization* structure, where the feedback is given by the vector that maximizes the signal-to-interference-leakage-and-noise-ratio (SLNR) regularized by the D2D link qualities, and the precoder is given by the RZF solution that is also regularized by the D2D link qualities. The proposed dual-regularized feedback and precoding strategy bridge the gap between the conventional CSI feedback scheme [1]–[7] and the precoder feedback scheme [12], [21], and it converges to the two existing schemes in the extremes of no D2D and perfect D2D, respectively.

To summarize, this paper develops a novel D2D assisted feedback and precoding scheme that has the following advantages:

- *Throughput enhancement:* The proposed scheme achieves higher sum rate than both the CSI feedback scheme and the precoder feedback scheme over all D2D link qualities.
- *Robustness:* When users have heterogeneous D2D link qualities, all the users can still achieve higher data rate than the CSI feedback scheme.
- *Compatibility:* The scheme allows the coexistence of users with and without D2D cooperation, where the D2D cooperative users gain additional benefit through the CSI exchange, while the non-D2D users are not affected.

The rest of the paper is organized as follows. Section II introduces the channel model, the information structure of the users and the BS, the CSI exchange strategy, and the formulation for the feedback and precoding problems. The feedback strategy is studied in Section III and the precoding strategy is studied in Section IV. Section V gives the numerical results, and Section VI gives the conclusions.

Notations: The notations $\|\mathbf{a}\|$ and $\|\mathbf{A}\|$ denote the Euclidean norm of vector $\mathbf{a}$ and the matrix 2-norm of $\mathbf{A}$, respectively. In addition, $\mathbf{A}^{\mathrm{H}}$ denotes the Hermitian transpose of $\mathbf{A}$; and $\mathbf{A}^{\star}$ denotes the optimal solution to an optimization problem where $\mathbf{A}$ is the variable.

II. System Model

This section first introduces the channel model and the CSI exchange strategy via D2D communications. Then, the feedback and precoding problems based on heterogeneous CSI are formulated.

A. Channel Model

Consider a K-user downlink MIMO system, where the BS is equipped with $N_t \geq K$ antennas and the users have single antenna. Denote the downlink channel for user k as $\mathbf{h}_k^{\mathrm{H}}$, where $\mathbf{h}_k \in \mathbb{C}^{N_t}$ is a column vector and follows distribution $\mathcal{CN}(\mathbf{0}, \mathbf{R}_k)$. The channels between users are mutually independent.[1] User k knows $\mathbf{h}_k$ perfectly, and the global statistics $\{\mathbf{R}_k\}$ is known by all the users.

Let $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_K] \in \mathbb{C}^{N_t \times K}$ be the channel matrix for all the users and $\mathbf{W} \in \mathbb{C}^{N_t \times K}$ denote the precoder for the downlink transmission. The received signal $\mathbf{y} = [y_1, y_2, \ldots, y_K]^{\mathrm{T}}$ at the user side is

$$\mathbf{y} = \mathbf{H}^{\mathrm{H}}\mathbf{W}\mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^K$ is the vector of transmission symbols that satisfies $\mathbb{E}\{\mathbf{x}\mathbf{x}^{\mathrm{H}}\} = \mathbf{I}_K$, the precoder $\mathbf{W}$ satisfies the sum power constraint $\mathrm{tr}\{\mathbf{W}^{\mathrm{H}}\mathbf{W}\} \leq P$, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ is the Gaussian noise.

In general, the desired precoder $\mathbf{W}$ is a function of the CSI $\mathbf{H}$ which is originally available at the user side. To assist the precoding, each user can feedback B bits of CSI related information to the BS.

B. CSI Exchange via D2D

In parallel to cellular communications, the users exploit reliable D2D links to directly exchange the CSI with each other. For example, the D2D communication can be implemented in out-band mode with no interference to the cellular communication, using existing technologies such as WiFi Direct, Bluetooth and ZigBee.

Denote the CSI $\mathbf{h}_k$ of user k known by user j as $\hat{\mathbf{h}}_k^{(j)}$, which is modeled as follows

$$\mathbf{h}_k = \alpha_{jk}\hat{\mathbf{h}}_k^{(j)} + \sqrt{1-\alpha_{jk}^2}\boldsymbol{\xi}_k^{(j)} + \mathbf{h}_k^{(j)\perp} \tag{2}$$

where $\alpha_{jk} \in [0,1]$ is a parameter to capture the quality of the CSI obtained via D2D, $\boldsymbol{\xi}_k^{(j)}$ is a zero mean random vector with distribution $\mathcal{CN}(\mathbf{0}, \boldsymbol{\Xi}_{kj})$ to model the noise due to quantization or transmission delays, and $\mathbf{h}_k^{(j)\perp}$ is orthogonal to both $\hat{\mathbf{h}}_k^{(j)}$

[1]For dependent channels, distributed source coding can also be used to design the D2D assisted feedback strategy [23], [24]. However, this is beyond the scope of this paper.

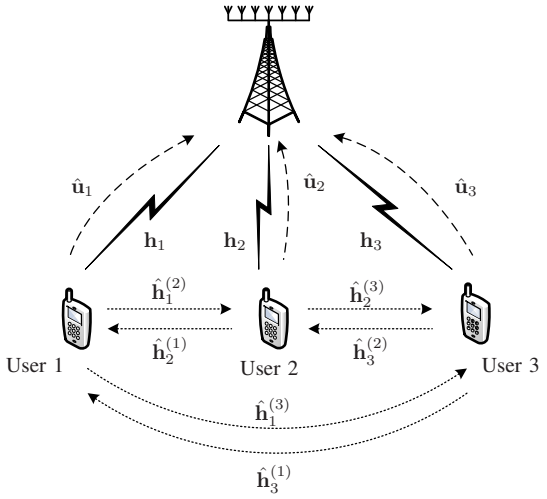


Figure 1. An example on the signaling structure for a three-user MIMO system, where the users exchange the quantized CSI via D2D.

and $\boldsymbol{\xi}_k^{(j)}$ to model the portion of CSI $\mathbf{h}_k$ that is not to be transmitted to user j. In particular, $\alpha_{jk} = 0$ means there is no D2D from user k to user j, and hence user j has no knowledge of $\mathbf{h}_k$, whereas, $\alpha_{jk} = 1$ means there is perfect D2D, and user j knows perfectly $\mathbf{h}_k - \mathbf{h}_k^{(j)\perp}$.

After the exchange of CSI, user k has the *imperfect* global CSI $\hat{\mathbf{H}}_k \in \mathbb{C}^{N_t \times K}$ given by

$$\hat{\mathbf{H}}_k = [\hat{\mathbf{h}}_1^{(k)}, \hat{\mathbf{h}}_2^{(k)}, \dots \hat{\mathbf{h}}_{k-1}^{(k)}, \mathbf{h}_k, \hat{\mathbf{h}}_{k+1}^{(k)}, \dots, \hat{\mathbf{h}}_K^{(k)}]. \quad (3)$$

Fig. 1 illustrates the signaling structure in the three-user case. Two application examples are provided below to justify the CSI exchange model (2).

1) Quantized CSI Exchange in Correlated Channels: Consider to exchange the CSI via D2D using a limited amount of bits and the transmission delays are negligible. As proposed in [21], an efficient mechanism in correlated channels is based on *signal subspace projection*. Conceptually, if the channel subspaces of user k and j are partially overlapping, then only the portion of CSI that lies in the overlapping signal subspace is needed to be exchanged. The intuition is that if the two users have non-overlapping channel subspaces, they do not need to exchange the CSI because their preferable precoding vectors would not create interference to each other.

Therefore, let $\mathbf{R}_k = \mathbf{V}_k \mathbf{\Lambda}_k \mathbf{V}_k^{\mathrm{H}}$ be the eigendecomposition, where $\mathbf{\Lambda}_k$ is an $M_k \times M_k$ diagonal matrix containing the nonzero eigenvalues of $\mathbf{R}_k$ sorted in descending order (with M_k being the rank), and $\mathbf{V}_k$ is an $N_t \times M_k$ semi-unitary matrix. The channel of user k can be written as $\mathbf{h}_k = \mathbf{h}_k^{(j)} + \mathbf{h}_k^{(j)\perp}$ where $\mathbf{h}_k^{(j)} = \mathbf{V}_j \mathbf{V}_j^{\mathrm{H}} \mathbf{h}_k$ and $\mathbf{h}_k^{(j)\perp} = (\mathbf{I} - \mathbf{V}_j \mathbf{V}_j^{\mathrm{H}}) \mathbf{h}_k$. As a result, $\mathbf{h}_k^{(j)}$ contains all the necessary information for user j and is orthogonal to $\mathbf{h}_k^{(j)\perp}$. Consider to quantize $\mathbf{h}_k^{(j)}$ into $\hat{\mathbf{h}}_k^{(j)}$ using B_{d} bits; the quantization error can be modeled by (2), where the parameters α_{kj} and $\mathbf{\Xi}$ can be computed using distortion-rate theories [21], [25]. A numerical example will be given in Section V-B.

Note that the goal of the paper is not to study quantization techniques for CSI exchange, but to exploit the general CSI exchange model (2) to develop advanced feedback and precoding strategies.

2) Delayed CSI Exchange: Consider to exchange the unquantized CSI $\mathbf{h}_k$. A practical problem is the potential delays in transmitting the CSI via D2D communications, and it induces CSI distortion due to small scale channel fading. Let $\hat{\mathbf{h}}_k^{(j)}$ be the out-dated CSI received by user j. The autoregressive time-variation model [26] for the MIMO channel yields $\mathbf{h}_k = \alpha_{jk} \hat{\mathbf{h}}_k^{(j)} + \sqrt{1 - \alpha_{jk}^2} \boldsymbol{\xi}_k^{(j)}$, where $\alpha_{jk} = J_0(2\pi f_{\mathrm{d}} \tau_{\mathrm{d}})$ is the correlation coefficient, $J_0(\cdot)$ is the zeroth order Bessel function, $f_{\mathrm{d}} = \frac{v}{c} f_{\mathrm{c}}$ is the maximum doppler frequency under carrier frequency f_{c}, propagation speed c, and user mobility speed v; τ_{d} is the D2D transmission delay, and $\boldsymbol{\xi}_k^{(j)}$ is zero mean Gaussian distributed with variance $\mathbf{\Xi} = \mathbf{R}_k$. The CSI exchange model (2) follows, with $\mathbf{h}_k^{(j)\perp} = \mathbf{0}$.

C. Feedback and Precoding under Partial CSI Exchange

Consider the transmit precoding that minimizes the modified MSE $\mathbb{E}\{\|a\mathbf{y} - \mathbf{x}\|_2^2\}$ for some positive value a. Such modified MMSE criterion was studied in [5] and observed to provide better performance than the usual MMSE criterion $\mathbb{E}\{\|\mathbf{y} - \mathbf{x}\|_2^2\}$. The intuition is that by introducing the auxiliary variable a, the sum power P can be better utilized [5].

From the signal model (1), it can be shown that the modified MSE is given by

$$\mathbb{E}\{\|a\mathbf{y} - \mathbf{x}\|_2^2\} = \left\|a\mathbf{H}^{\mathrm{H}}\mathbf{W} - \mathbf{I}\right\|_F^2 + a^2 K. \quad (4)$$

The common goal of the users and the BS is to minimize the modified MSE in (4) subject to total transmission power P.

We consider that the users and the BS make sequential team decisions to minimize the modified MSE based on their individual information. Specifically, the feedback and precoding problems are formulated as follows.

Feedback: Each user k feeds back to the BS a discretized vector $\hat{\mathbf{u}}_k = \mathcal{Q}_k(\mathbf{W}_k^\star)$ where the matrix $\mathbf{W}_k^\star$ is obtained as the solution to the following MMSE problem

$$\begin{aligned} \underset{a \geq 0,\, \mathbf{W} \in \mathbb{C}^{N_t \times K}}{\text{minimize}} \quad & \mathbb{E}\left\{ \left\|a\mathbf{H}^{\mathrm{H}}\mathbf{W} - \mathbf{I}\right\|_F^2 + a^2 K \,\middle|\, \hat{\mathbf{H}}_k \right\} \\ \text{subject to} \quad & \mathrm{tr}\{\mathbf{W}^{\mathrm{H}}\mathbf{W}\} \leq P \end{aligned} \quad (5)$$

in which the expectation is taken over the uncertainty of the global CSI $\mathbf{H}$ given the individually observed imperfect global CSI $\hat{\mathbf{H}}_k$. The quantizer $\mathcal{Q}_k(\cdot)$ is to be designed.

Precoding: Given the feedback $\hat{\mathbf{U}} = [\hat{\mathbf{u}}_1, \hat{\mathbf{u}}_2, \dots, \hat{\mathbf{u}}_K]$ from all the users, the BS computes the precoder $\mathbf{W}^\star$ as the solution to the following problem

$$\begin{aligned} \underset{a \geq 0,\, \mathbf{W} \in \mathbb{C}^{N_t \times K}}{\text{minimize}} \quad & \mathbb{E}\left\{ \left\|a\mathbf{H}^{\mathrm{H}}\mathbf{W} - \mathbf{I}\right\|_F^2 + a^2 K \,\middle|\, \hat{\mathbf{U}} \right\} \\ \text{subject to} \quad & \mathrm{tr}\{\mathbf{W}^{\mathrm{H}}\mathbf{W}\} \leq P \end{aligned} \quad (6)$$

where the expectation is taken over the uncertainty of the global CSI $\mathbf{H}$ given the feedback $\hat{\mathbf{U}}$ from the users.

III. Feedback Design

This section derives the feedback strategy of the user based on the solution of the MMSE problem (5) from the finite-rate codebook. In particular, assume that the codebooks $\mathcal{C}_k$ are generated independently using random vector quantization (RVQ) as follows

$$\mathcal{C}_k = \{\mathbf{R}_k^{\frac{1}{2}}\mathbf{f}_i/\|\mathbf{R}_k^{\frac{1}{2}}\mathbf{f}_i\| : \mathbf{f}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_t}),\, i = 1, 2, \ldots, 2^B\}. \tag{7}$$

The above RVQ codebook is also widely used in the literature for limited feedback under correlated channels [2].[2]

A. The MMSE Solution based on Imperfect Global CSI

Let $\widetilde{\mathbf{W}} = a\mathbf{W}$. Problem (5) is equivalent to minimizing $\mathbb{E}\{\|\mathbf{H}^{\mathrm{H}}\widetilde{\mathbf{W}} - \mathbf{I}\|_F^2 \,|\hat{\mathbf{H}}_k\} + a^2K$ over $\widetilde{\mathbf{W}}$, subject to $\mathrm{tr}\{\widetilde{\mathbf{W}}^{\mathrm{H}}\widetilde{\mathbf{W}}\}/P \le a^2$, which implies that given the optimal solution $\widetilde{\mathbf{W}}^\star$, the optimal solution a can be computed as $a^\star = \big(\mathrm{tr}\{\widetilde{\mathbf{W}}^{\star\mathrm{H}}\widetilde{\mathbf{W}}^\star\}/P\big)^{\frac{1}{2}}$. As a result, problem (5) is equivalent to the following unconstrained problem

$$\underset{\widetilde{\mathbf{W}}\in\mathbb{C}^{N_t\times K}}{\text{minimize}} \quad \mathbb{E}\left\{\|\mathbf{H}^{\mathrm{H}}\widetilde{\mathbf{W}} - \mathbf{I}\|_F^2 \,|\hat{\mathbf{H}}\right\} + K\frac{\mathrm{tr}\{\widetilde{\mathbf{W}}^{\mathrm{H}}\widetilde{\mathbf{W}}\}}{P} \tag{8}$$

and the optimal solution to (5) is given by $\mathbf{W}^\star = \frac{1}{a}\widetilde{\mathbf{W}}^\star = \big(\mathrm{tr}\{\widetilde{\mathbf{W}}^{\star\mathrm{H}}\widetilde{\mathbf{W}}^\star\}/P\big)^{-\frac{1}{2}}\widetilde{\mathbf{W}}^\star$, where $\widetilde{\mathbf{W}}^\star$ is the optimal solution to (8).

To compute the expectation in (8), consider to stack the column vectors in (2) for $k = 1, 2, \ldots, K$ into $N_t \times K$ matrices, and the exchanged CSI model for user k can be written as

$$\mathbf{H} = \hat{\mathbf{H}}_k\mathbf{A}_k^{\frac{1}{2}} + \mathbf{E}_k\mathbf{S}_k^{\frac{1}{2}} + \mathbf{H}_k^{\perp} \tag{9}$$

where

$$\mathbf{E}_k = [\boldsymbol{\xi}_1^{(k)}, \boldsymbol{\xi}_2^{(k)}, \ldots, \boldsymbol{\xi}_K^{(k)}], \qquad \text{and}$$

$$\mathbf{H}_k^{\perp} = [\mathbf{h}_1^{(k)\perp}, \mathbf{h}_2^{(k)\perp}, \ldots \mathbf{h}_{k-1}^{(k)\perp}, \mathbf{0}, \mathbf{h}_{k+1}^{(k)\perp}, \ldots, \mathbf{h}_K^{(k)\perp}].$$

In addition, $\mathbf{A}_k$ and $\mathbf{S}_k$ are diagonal matrices with diagonal elements given by

$$[\mathbf{A}_k]_{(j,j)} = \begin{cases} \alpha_{kj}^2 & j \ne k \\ 1 & j = k, \end{cases}$$

and $\mathbf{S}_k = \mathbf{I} - \mathbf{A}_k$.

Moreover, we define the *CSI uncertainty coefficient* as follows, which is found to be an important parameter for the solution to the feedback and precoding problem.

Definition 1 (CSI uncertainty coefficient): The *uncertainty matrix* for the global CSI $\hat{\mathbf{H}}_k$ obtained from CSI exchange by user k is defined as

$$\mathbf{Q}_k \triangleq \sum_{j\ne k}(1 - \alpha_{kj}^2)\boldsymbol{\Xi}_{jk} + \frac{K}{P}\mathbf{I}_{N_t}. \tag{10}$$

[2]Note that this paper does not aim at optimizing the codebook. The codebook design for the proposed feedback and precoding structure that exploits the CSI exchanged between users is left for future works.

In particular, under uncorrelated channels with $\boldsymbol{\Xi}_{jk} = \mathbf{I}$ in the CSI exchange model (2), the *uncertainty coefficient* for user k is defined as

$$q_k \triangleq \sum_{j\ne k}(1 - \alpha_{kj}^2) + K/P. \tag{11}$$

The optimal solution to the feedback problem (5) is given as follows.

Proposition 1 (MMSE solution under imperfect global CSI): The optimal solution $\mathbf{W}_k^\star$ to the feedback problem (5) is given by

$$\mathbf{W}_k^\star = \beta_k\Big(\hat{\mathbf{H}}_k\mathbf{A}_k\hat{\mathbf{H}}_k^{\mathrm{H}} + \mathbf{Q}_k\Big)^{-1}\hat{\mathbf{H}}_k\mathbf{A}_k^{\frac{1}{2}} \tag{12}$$

where and $\beta_k > 0$ is chosen such that $\mathrm{tr}\{\mathbf{W}_k^{\star\mathrm{H}}\mathbf{W}_k^\star\} = P$.

Proof: Please refer to Appendix A. □

As a result of Proposition 1, the kth column of $\mathbf{W}_k^\star$ is given by

$$\mathbf{w}_k^\star = \beta_k\Big(\hat{\mathbf{H}}_k\mathbf{A}_k\hat{\mathbf{H}}_k^{\mathrm{H}} + \mathbf{Q}_k\Big)^{-1}\mathbf{h}_k \tag{13}$$

where, we recall that perfect CSI $\mathbf{h}_k$ is assumed available at user k.

Note that the statistics of $\mathbf{H}_k^{\perp}$ does not play a role in the solution (12)–(10), and thus it justifies the CSI exchange strategy in Section II-B1, where the CSI $\mathbf{h}_k^{(j)\perp}$ is not needed to be exchanged.

The above solution (13) takes a similar form as the robust MMSE precoder under imperfect CSI at the BS, where the term $\mathbf{Q}_k$ performs regularization due to the uncertainty from the CSI exchange and the signal-to-noise ratio (SNR).

B. Vector Discretization

The MMSE solution (13) is computed in the continuous domain, and needs to be discretized into B bits for the feedback. In fact, we are only interested in the kth column of $\mathbf{W}_k^\star$, because in the special case of perfect D2D link qualities $\alpha_{kj} = 1$, the kth column of $\mathbf{W}_k^\star$ is the desired precoder for user k and only the kth column is needed to be fed back [12]. In the case of no D2D $\alpha_{kj} = 0$, the kth column of $\mathbf{W}_k^\star$ degenerates to the scaled channel $\mathbf{h}_k$, which is also the desired vector to be fed back.

An intuitive solution is to find a vector from the codebook that is "closest" to the kth *normalized* column vector of $\mathbf{W}_k^\star$. However, it is not known what a good distance measure would be for the "closeness", as $\mathbf{W}_k^\star$ has the physical meaning as a regularized precoder.[3] Alternatively, we exploit the equivalence between the solution (13) and the solution to a Rayleigh quotient maximization problem. Specifically, the result is given in the following lemma.[4]

Lemma 1 (Equivalence between MMSE and SLNR): Let $\mathbf{H}$ be an $N_t \times K$ matrix with the kth column given by vector $\mathbf{h}_k$. For a positive definite matrix $\mathbf{Q}$, the following result holds

$$(\mathbf{H}\mathbf{H}^{\mathrm{H}} + \mathbf{Q})^{-1}\mathbf{h}_k = c_k\mathbf{u}_k$$

[3]For example, a good distance measure to discretize the MIMO channel is the geodesic on the Grassmann manifold, but such measure may not be meaningful to discretizing the precoder.

[4]Similar results have also been established in [27] for the special case of $\mathbf{Q} = \alpha\mathbf{I}$.

where c_k is a complex-valued scalar and $\mathbf{u}_k$ is the solution to the following problem

$$\underset{\|\mathbf{u}\|=1}{\text{maximize}} \quad \frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}|\mathbf{h}_j^{\mathrm{H}}\mathbf{u}|^2+\mathbf{u}^{\mathrm{H}}\mathbf{Q}\mathbf{u}}. \tag{14}$$

Proof: Denote the objective function (14) as

$$g(\mathbf{u})=\frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}|\mathbf{h}_j^{\mathrm{H}}\mathbf{u}|^2+\mathbf{u}^{\mathrm{H}}\mathbf{Q}\mathbf{u}}.$$

Since $\mathbf{Q}$ is positive definite, the function $g(\mathbf{u})$ is well-defined and continuous everywhere in the compact domain $\{\mathbf{u}\in\mathbb{C}^{N_t}: \|\mathbf{u}\|=1\}$. Therefore, $g(\mathbf{u})$ attains a maximum value, denoted as T, and moreover, $g(\mathbf{u})\geq 0$.

Define $h(g)=\frac{g}{g+1}$, where $g\in[0,T]$. The function $h(g)$ is bijective, continuous, and strictly increasing, since $h^{'}=\frac{1}{(g+1)^2}>0$. Therefore, the maximum value of $h(g)$ is attained at $g=T$. As a result, the maximizer $\mathbf{u}^\star$ such that $g(\mathbf{u}^\star)=T$ is also the maximizer of the composite function

$$\begin{aligned}
&h(g(\mathbf{u}))\\
&=\frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}|\mathbf{h}_j^{\mathrm{H}}\mathbf{u}|^2+\mathbf{u}^{\mathrm{H}}\mathbf{Q}\mathbf{u}}\left(\frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}|\mathbf{h}_j^{\mathrm{H}}\mathbf{u}|^2+\mathbf{u}^{\mathrm{H}}\mathbf{Q}\mathbf{u}}+1\right)^{-1}\\
&=\frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j=1}^{K}|\mathbf{h}_j^{\mathrm{H}}\mathbf{u}|^2+\mathbf{u}^{\mathrm{H}}\mathbf{Q}\mathbf{u}}\\
&=\frac{\mathbf{u}^{\mathrm{H}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\mathbf{u}}{\mathbf{u}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}}+\mathbf{Q})\mathbf{u}}
\end{aligned}$$

and vice versa. Hence, the set of solutions to (14) is identical to the set of solutions to

$$\underset{\|\mathbf{u}\|=1}{\text{maximize}} \quad \frac{\mathbf{u}^{\mathrm{H}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\mathbf{u}}{\mathbf{u}^{\mathrm{H}}(\mathbf{H}\mathbf{H}^{\mathrm{H}}+\mathbf{Q})\mathbf{u}} \tag{15}$$

which is known as Rayleigh quotient maximization, where the solution $\mathbf{u}^\star$ is given by the dominant eigenvector of $(\mathbf{H}\mathbf{H}^{\mathrm{H}}+\mathbf{Q})^{-1}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}$. This leads to

$$c_k\mathbf{u}^\star=(\mathbf{H}\mathbf{H}^{\mathrm{H}}+\mathbf{Q})^{-1}\mathbf{h}_k$$

where c_k is some scalar such that $\|\mathbf{u}^\star\|=1$. □

As a result of Lemma 1, we can relate the pseudo-inverse solution (13) in continuous domain to a quotient maximization problem in the finite domain. The feedback strategy for user k is to choose a vector $\hat{\mathbf{u}}_k$ from $\mathcal{C}_k$ as the solution to the following problem

$$\underset{\mathbf{u}\in\mathcal{C}_k}{\text{maximize}} \quad \frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}\alpha_{kj}^2\left|(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{u}\right|^2+\mathbf{u}^{\mathrm{H}}\mathbf{Q}_k\mathbf{u}}. \tag{16}$$

Some insights can be obtained from the special case of uncorrelated channels.

Corollary 1 (Feedback strategy for uncorrelated channels): Under uncorrelated channels $\mathbf{R}_k=\mathbf{I}$, the feedback vector $\hat{\mathbf{u}}_k$ is computed as

$$\underset{\mathbf{u}\in\mathcal{C}_k}{\text{maximize}} \quad \frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}\alpha_{kj}^2\left|(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{u}\right|^2+\sum_{j\neq k}(1-\alpha_{kj}^2)+\frac{K}{P}}. \tag{17}$$

It is observed from (17) that when users have perfect global CSI, i.e., $\alpha_{kj}=1$, they feed back the discretized SLNR precoder, whereas when users have no CSI from each other, i.e., $\alpha_{kj}=0$, they feed back their own quantized CSI. The terms α_{kj}^2 and $1-\alpha_{kj}^2$ steer the feedback vector from a precoding vector to a CSI vector, according to the D2D link qualities.

IV. Precoder Design

Based on the feedback $\hat{\mathbf{U}}$ to the BS, the precoder will be obtained by solving (6), where the objective can be written as

$$\begin{aligned}
&\mathbb{E}\left\{\|a\mathbf{H}^{\mathrm{H}}\mathbf{W}-\mathbf{I}\|_F^2+a^2K\,\middle|\hat{\mathbf{U}}\right\}\\
&\quad=\mathrm{tr}\Big\{a^2\mathbf{W}^{\mathrm{H}}\mathbb{E}\big\{\mathbf{H}\mathbf{H}^{\mathrm{H}}\,\big|\hat{\mathbf{U}}\big\}\mathbf{W}\Big\}\\
&\qquad-2\mathrm{tr}\Big\{\mathrm{Re}\Big\{a\mathbb{E}\big\{\mathbf{H}^{\mathrm{H}}\,\big|\hat{\mathbf{U}}\big\}\mathbf{W}\Big\}\Big\}+a^2K+K.
\end{aligned}$$

Using the same technique in Section III, the solution can be obtained as

$$\mathbf{W}^\star=\left[\mathbb{E}\big\{\mathbf{H}\mathbf{H}^{\mathrm{H}}\,\big|\hat{\mathbf{U}}\big\}+\frac{K}{P}\mathbf{I}_{N_t}\right]^{-1}\mathbb{E}\big\{\mathbf{H}\,\big|\hat{\mathbf{U}}\big\}\mathbf{\Psi} \tag{18}$$

in which, $\mathbf{\Psi}$ is a diagonal matrix for power scaling to satisfy the power constraint.

A. Precoding for Uncorrelated Channels

The main challenge of the precoding is to evaluate the terms $\mathbb{E}\big\{\mathbf{H}\mathbf{H}^{\mathrm{H}}\,\big|\hat{\mathbf{U}}\big\}$ and $\mathbb{E}\big\{\mathbf{H}\,\big|\hat{\mathbf{U}}\big\}$. To gain intuitions, we first study the uncorrelated channel case.

Under uncorrelated channels, the channel $\mathbf{h}_k$ can be expressed in terms of the feedback vector $\hat{\mathbf{u}}_k$ in (17). The result is characterized in the following proposition.

Proposition 2 (Characterization of $\hat{\mathbf{u}}_k$): Suppose that the RVQ codebook in (7) is used to obtain the feedback vector $\hat{\mathbf{u}}_k$ from solving (17). Then,

$$\mathbf{h}_k\overset{\mathrm{d}}{=}\theta_k\mathbf{G}_k\left(\sqrt{1-\epsilon_k^2}\hat{\mathbf{u}}_k+\epsilon_k\mathbf{z}_k\right) \tag{19}$$

where $x\overset{\mathrm{d}}{=}y$ means that x and y have the same distribution, $\mathbf{z}_k\in\mathbb{C}^{N_t}$ is a unit norm vector that satisfies $\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{z}_k=0$ and $\mathbb{E}\{\mathbf{z}_k\}=\mathbf{0}$ with isotropic distribution, $\epsilon_k\in[0,1]$ is a random variable that is independent to $\hat{\mathbf{u}}_k$ and $\mathbf{z}_k$, and θ_k is a scaling factor, and satisfies $\frac{N_t-1}{N_t}2^{-\frac{B}{N_t-1}}\leq\mathbb{E}\{\epsilon_k^2\}\leq 2^{-\frac{B}{N_t-1}}$, in which, the expectation is taken over the distributions of $\mathbf{h}_k$, $\hat{\mathbf{h}}_j^{(k)}$, and $\mathcal{C}_k$. In addition,

$$\mathbf{G}_k=\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}+q_k\mathbf{I}_{N_t}. \tag{20}$$

Proof: Please refer to Appendix B. □

From (19), we have

$$\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\overset{\mathrm{d}}{=}\theta_k^2\Big[(1-\epsilon_k^2)\mathbf{G}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}+\epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\Big] \tag{21}$$

where we establish a few asymptotic results under some special cases as follows.

Lemma 2 (Properties under high SNR and high feedback resolution): The following holds

$$\begin{aligned}
&\lim_{B\to\infty}\mathbb{E}\{\theta_k^2\epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\,|\hat{\mathbf{U}}\}\\
&\qquad=\lim_{B\to\infty}\mathbb{E}\{\theta_k^2\epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\}.
\end{aligned} \tag{22}$$

In addition, for $\alpha_{kj} \in \{0,1\}$ over all $j \neq k$

$$\lim_{P\to\infty}\lim_{B\to\infty}\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{u}}_k = 0 \tag{23}$$

with probability 1. Moreover,

$$\lim_{\{\alpha_{kj}\to 0\}}\mathbb{E}\big\{\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\big|\hat{\mathbf{U}}\big\} = \varpi\mathbb{E}\Big\{(1-\epsilon_k^2)\mathbf{G}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}} + \epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\big|\hat{\mathbf{U}}\Big\} \tag{24}$$

Proof: Please refer to Appendix C. □

Inspired from Lemma 2, we make the following approximations

$$\mathbb{E}\{\theta_k^2\epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\big|\hat{\mathbf{U}}\} \approx \mathbb{E}\{\theta_k^2\epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\} \tag{25}$$

$$\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{u}}_k \approx 0 \tag{26}$$

$$\mathbb{E}\big\{\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\big|\hat{\mathbf{U}}\big\} \approx \varpi\mathbb{E}\Big\{(1-\epsilon_k^2)\mathbf{G}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}} + \epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\big|\hat{\mathbf{U}}\Big\} \tag{27}$$

and

$$\mathbb{E}\{\epsilon_k^2\} \approx 2^{-\frac{B}{N_t-1}} \tag{28}$$

which are asymptotically accurate at high SNR P, high feedback resolution B, small α_{kj} and large N_t.

To characterize the asymptotic accuracy of the approximations, define the relation $\overset{\Theta}{\approx}$ for variables $\mathbf{X}$ and $\mathbf{Y}$ that depend on P, B, $\{\alpha_{kj}\}$, and N_t, where $\mathbf{X} \overset{\Theta}{\approx} \mathbf{Y}$, if $\mathbf{X}$ and $\mathbf{Y}$ satisfy $\|\mathbf{X}-\mathbf{Y}\| \to 0$ as P, B, α_{kj}^{-1}, and N_t all go to infinity. In addition, let

$$\mathbf{\Omega} = \mathrm{diag}(q_1^2, q_2^2, \ldots, q_K^2) \tag{29}$$

and

$$\phi = \sum_{k=1}^{K}\Big(\sum_{j\neq k}\alpha_{kj}^4 + \frac{q_k^2}{N_t}\Big). \tag{30}$$

We have the following approximation results for $\mathbb{E}\big\{\mathbf{H}\mathbf{H}^{\mathrm{H}}\big|\hat{\mathbf{U}}\big\}$ and $\mathbb{E}\big\{\mathbf{H}\big|\hat{\mathbf{U}}\big\}$.

Proposition 3 (Approximation results): Suppose (25)–(28) hold under relation $\overset{\Theta}{\approx}$. Then,

$$\mathbb{E}\big\{\mathbf{H}\mathbf{H}^{\mathrm{H}}\big|\hat{\mathbf{U}}\big\} \overset{\Theta}{\approx} \varpi\Big[\big(1-\sigma_{\mathrm{fb}}^2\big)\hat{\mathbf{U}}\mathbf{\Omega}\hat{\mathbf{U}}^{\mathrm{H}} + \sigma_{\mathrm{fb}}^2\phi\mathbf{I}\Big] \tag{31}$$

$$\mathbb{E}\big\{\mathbf{H}\big|\hat{\mathbf{U}}\big\} \overset{\Theta}{\approx} \hat{\mathbf{U}}\mathbf{\Upsilon} \tag{32}$$

where $\sigma_{\mathrm{fb}}^2 = 2^{-\frac{B}{N_t-1}}$, ϖ is some positive scalar, and $\mathbf{\Upsilon}$ is some diagonal matrix.

Proof: Please refer to Appendix D. □

Using the approximations (31) and (32), the modified MMSE precoder $\mathbf{W}^\star$ based on the feedback $\hat{\mathbf{U}}$ can be computed as

$$\Big[\big(1-\sigma_{\mathrm{fb}}^2\big)\hat{\mathbf{U}}\mathbf{\Omega}\hat{\mathbf{U}}^{\mathrm{H}} + \Big(\sigma_{\mathrm{fb}}^2\phi + \frac{K}{P}\Big)\mathbf{I}\Big]^{-1}\hat{\mathbf{U}}\mathbf{\Psi} \tag{33}$$

where $\mathbf{\Psi}$ is a diagonal matrix such that all the columns of $\mathbf{W}^\star$ satisfy $\|\mathbf{w}_k^\star\|_2^2 = P/K$. The intuition that equal power allocation is used in (33) is because $\mathbf{\Upsilon}$ in (32) is unknown.

Despite that the approximations (25)–(28) were derived under extreme parameters, the precoder (33) shows good performance under moderate SNR P, small feedback resolution B, moderate number of antennas N_t, and all range of CSI exchange quality $0 \leq \alpha_{kj} \leq 1$, as will be seen from the numerical result section.

Furthermore, it is observed that the feedback and precoding solutions (16) and (33) have a dual-regularized structure, where the regularization terms capture the quality of CSI exchange and the SNR.

B. Case Study

Here we discuss how the existing feedback and precoding strategies are the special cases of the proposed dual-regularized feedback and precoding strategy (17) and (33).

1) No CSI Exchange via D2D: This corresponds to $a_{kj} = 0$ in the CSI model (2). As a result, the feedback strategy (17) degenerates to

$$\underset{\mathbf{u}\in\mathcal{C}_k}{\text{maximize}} \quad |\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2 \tag{34}$$

which is the conventional CSI feedback scheme as studied in [1]–[7]. In this case, the feedback matrix from all the K users is given by $\hat{\mathbf{U}} = \hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \ldots, \hat{\mathbf{h}}_K]$.

From (33), the precoding strategy at the BS yields,

$$\mathbf{W}^\star = \Big[\big(1-2^{-\frac{B}{N_t-1}}\big)\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}} + \Big(\frac{K}{N_t}2^{-\frac{B}{N_t-1}} + \frac{K}{P}\Big)\mathbf{I}\Big]^{-1}\hat{\mathbf{H}}\mathbf{\Psi}. \tag{35}$$

As a comparison from the literature, the robust MMSE precoding scheme in [5] gives

$$\mathbf{W}^{\mathrm{RB}} = \psi\Big[\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}} + \Big(K2^{-\frac{B}{N_t}} + \frac{K}{P}\Big)\mathbf{I}\Big]^{-1}\hat{\mathbf{H}} \tag{36}$$

where ψ is a scalar for the sum power constraint. Here, $\mathbf{W}^{\mathrm{RB}}$ takes a similar form as the dual-regularized precoding (35), and the difference is mainly due to the different quantization technique used for the feedback in [5].[5]

2) Perfect CSI Exchange via D2D: This corresponds to $\alpha_{kj} = 1$ in the CSI model (2). The feedback strategy (17) takes the form

$$\underset{\mathbf{u}\in\mathcal{C}_k}{\text{maximize}} \quad \frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}\big|(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{u}\big|^2 + \frac{K}{P}} \tag{37}$$

and the precoding from (33) yields

$$\Big[\big(1-\sigma_{\mathrm{fb}}^2\big)\big(\frac{K}{P}\big)^2\hat{\mathbf{U}}\hat{\mathbf{U}}^{\mathrm{H}} + \big(\sigma_{\mathrm{fb}}^2\phi + \frac{K}{P}\big)\mathbf{I}\Big]^{-1}\hat{\mathbf{U}}\mathbf{\Psi} \approx \sqrt{\frac{P}{K}}\hat{\mathbf{U}} \tag{38}$$

for high SNR $\frac{P}{K} \gg 1$, since $\phi = \big(K(K-1) + \frac{K^3}{PN_t}\big) \gg \big(\frac{K}{P}\big)^2$. The precoder (38) is the same as that in the cooperative precoder feedback scheme studied in [12] and [21].

[5]For (36), the generalized Lloyd algorithm is applied to design the codebook $\mathcal{C}_k$ rather than using RVQ as in Section II. Moreover, (36) assumes the quantization of $\mathbf{h}_k$ including its magnitude, whereas the proposed feedback scheme only considers the channel direction $\mathbf{h}_k/\|\mathbf{h}_k\|$. As a result, the precoder in (36) has the term $K2^{-B/N_t}$ rather than $\frac{K}{N_t}2^{-B/(N_t-1)}$ in the proposed scheme (35).

C. Extension to Correlated Channels

One of the challenges in correlated channels is that Proposition 2 does not hold. Here, we develop a heuristic precoding strategy using the insights obtained from the precoding solution (33). A rigorous derivation and the associated analysis are left for future works.

In the precoding strategy (33) for uncorrelated channels, there are two regularization terms for the projection of the feedback matrix $\hat{\mathbf{U}}$. The first term $\big(1-2^{-\frac{B}{N_t-1}}\big)\hat{\mathbf{U}}\mathbf{\Omega}\hat{\mathbf{U}}^{\mathrm{H}}$ determines how much the matrix $\hat{\mathbf{U}}$ should be inverted. The columns $\hat{\mathbf{u}}_k$ of $\hat{\mathbf{U}}$ are weighted by q_k^2 that capture the uncertainty from CSI exchange, and $\big(1-2^{-\frac{B}{N_t-1}}\big)$ that captures the uncertainty due to finite-rate feedback to the BS. To look into the two extreme cases, (i) for no CSI exchange, q_k^2 are large, which means that the inversion of $\hat{\mathbf{U}}$ is desired, because in this case the feedback $\hat{\mathbf{U}} \approx \hat{\mathbf{H}}$. (ii) For perfect CSI exchange, q_k^2 are small, which means that the inversion of $\hat{\mathbf{U}}$ is not desired.

The second term $(2^{-\frac{B}{N_t-1}}\phi+\frac{K}{P})\mathbf{I}$ determines how much the inversion of $\hat{\mathbf{U}}$ should be regularized, and plays the opposite role to the first term.

Correspondingly, in correlated channels, one may want to weight the feedback vectors $\hat{\mathbf{u}}_k$ by $\mathbf{Q}_k = \sum_{j\neq k}(1-\alpha_{kj}^2)\mathbf{\Xi}_{jk}+\frac{K}{P}\mathbf{I}$. This is because the uncertainty from CSI exchange is only in the subspace determined by $\mathbf{\Xi}_{jk}$. In addition, the uncertainty due to finite-rate feedback to the BS can be approximated by $2^{-\frac{B}{M_k-1}}$. This is because from the codebook construction (7), the codewords only lies in the M_k-dimensional subspace, where M_k is the rank of the channel covariance matrix $\mathbf{R}_k$.

Let $\widetilde{\mathbf{u}}_k = \mathbf{Q}_k\hat{\mathbf{u}}_k$ and $\widetilde{\mathbf{U}} = [\widetilde{\mathbf{u}}_1, \widetilde{\mathbf{u}}_2, \ldots, \widetilde{\mathbf{u}}_K]$. From the above insights, a heuristic precoding strategy can be developed as follows

$$\mathbf{W}^\star = \left[\widetilde{\mathbf{U}}\mathbf{B}\widetilde{\mathbf{U}}^{\mathrm{H}} + \Big(\sum_{k=1}^{K} 2^{-\frac{B}{M_k-1}}\phi + \frac{K}{P}\Big)\mathbf{I}\right]^{-1}\hat{\mathbf{U}}\mathbf{\Psi} \quad (39)$$

where $\mathbf{B} = \mathrm{diag}\big((1-2^{-\frac{B}{M_1-1}}), (1-2^{-\frac{B}{M_2-1}}), \ldots, (1-2^{-\frac{B}{M_K-1}})\big)$ and $\mathbf{\Psi}$ is a diagonal matrix for power scaling.

We verify the rationale of (39) by checking the two extreme cases. For no CSI exchange, $\alpha_{kj} = 0$, we have the CSI feedback $\hat{\mathbf{U}} = \hat{\mathbf{H}}$ as from (16). On the other hand, the precoder from (39) becomes $\mathbf{W}^\star = \big[\sum_{k=1}^{K}(1-2^{-\frac{B}{M_k-1}})\mathbf{Q}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k\mathbf{Q}_k^{\mathrm{H}} + \big(\sum_{k=1}^{K}2^{-\frac{B}{M_k-1}}\phi+\frac{K}{P}\big)\mathbf{I}\big]^{-1}\hat{\mathbf{U}}\mathbf{\Psi}$. It can be easily verified that when the users have identical signal subspace, $\mathbf{Q}_k\hat{\mathbf{u}}_k = \hat{\mathbf{h}}_k$ and $\mathbf{W}^\star$ becomes a typical robust MMSE precoder based on the CSI feedback. When users have non-overlapping signal subspaces, $\widetilde{\mathbf{u}}_k = \mathbf{Q}_k\hat{\mathbf{u}}_k \approx \mathbf{0}$, and $\mathbf{W}^\star \approx \hat{\mathbf{H}}$, which is reasonable because maximum ratio combining (MRC) is optimal when users have non-overlapping signal subspaces.

For perfect CSI exchange, $\alpha_{kj} = 1$, we have the precoder feedback as from (16). In this case, $\mathbf{Q}_k\hat{\mathbf{u}}_k = \frac{K}{P}\hat{\mathbf{u}}_k$ and $\widetilde{\mathbf{u}}_k\widetilde{\mathbf{u}}_k^{\mathrm{H}} = (\frac{K}{P})^2\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}$, which is far less significant than the second regularization term $\frac{K}{P}\mathbf{I}$, and hence $\mathbf{W}^\star \approx \sqrt{\frac{P}{K}}\hat{\mathbf{U}}$, which converges to the precoder feedback scheme.

V. Numerical Results

In this section, we evaluate the performance of the proposed dual-regularized feedback and precoding scheme, and compare it with the existing CSI feedback and precoder feedback schemes. We focus on the service in a local area where the users that are close to each other. Note that the results can be easily extended to system level implementations using user grouping and two-layer precoding techniques [28].

Consider the system topology in Fig. 1, where the BS has $N_t = 20$ antennas and transmits downlink signals to $K = 3$ users using spatial multiplexing. Each user has $B = 10$ bits for the feedback to the BS. For the proposed scheme, consider that each user has B_{d} bits to exchange a CSI vector with another user via D2D. The total downlink transmission power is $P = 20$ dB.

We compare the proposed scheme with the following baselines:

- **Baseline 1 (Robust MMSE based on CSI Feedback [5]):** Each user quantizes the channel according to (34) and conveys the CSI feedback to the BS. The BS computes the robust MMSE precoder according to (36).
- **Baseline 2 (Cooperative Precoder Feedback [21]):** Each user computes the precoder according to (37) and conveys the precoder feedback to the BS. The BS directly applies the feedback vectors as the precoder as in (38).

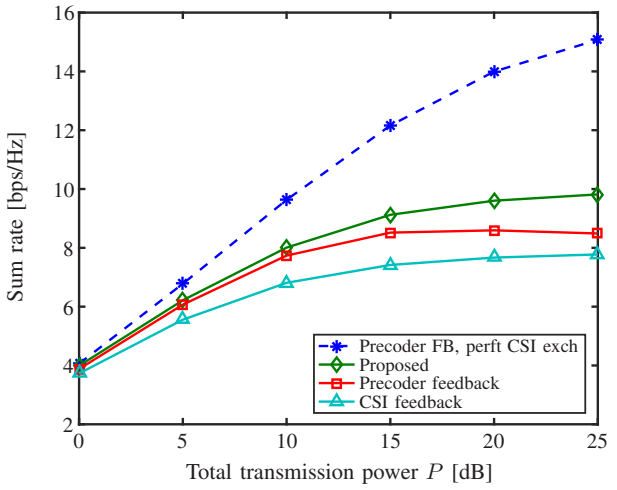
A. Uncorrelated Channels

Assume that the MIMO channel is uncorrelated and $\mathbf{R}_k = \mathbf{I}$ for all k. The model of the exchanged CSI in (2) degenerates to $\mathbf{h}_k = \alpha_{jk}\hat{\mathbf{h}}_k^{(j)} + \sqrt{1-\alpha_{jk}^2}\boldsymbol{\xi}_k^{(j)}$, where $\boldsymbol{\xi}_k^{(j)} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and the CSI quality is modeled using the distortion-rate theory as $\alpha_{jk}^2 = \alpha^2 = 1-2^{-B_{\mathrm{d}}/N_t}$ [25].

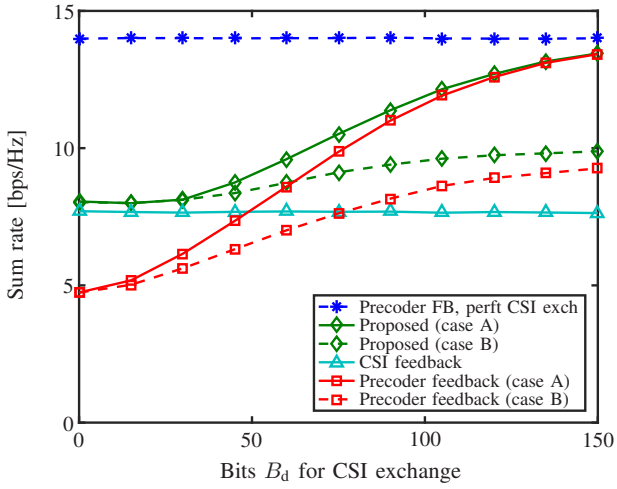
Three cases are consider for the evaluation.

- **Case A:** All the users have the same D2D link quality for CSI exchange, i.e., $\alpha_{kj} = \alpha$, $k \neq j$.
- **Case B:** There is no D2D link between user 2 and 3, while all the other links have identical D2D link quality for CSI exchange, i.e., $\alpha_{23} = \alpha_{32} = 0$, and $\alpha_{kj} = \alpha$ for $(k, j) \notin \{(2,3),(3,2)\}$.
- **Case C:** User 3 has no D2D links with user 1 and 2, while user 1 and 2 have identical D2D link quality for CSI exchange, i.e., $\alpha_{12} = \alpha_{21} = \alpha$, and $\alpha_{kj} = 0$ otherwise.

1) Sum Rate: Fig. 2 (a) shows the sum rate versus the total downlink transmission power P under $B_{\mathrm{d}} = 60$ bits for each CSI vector exchanged via D2D. The proposed scheme outperforms both the precoder feedback and CSI feedback scheme from low to high SNR. In the high SNR regime, the performance of the precoder feedback scheme suffers from inter-user interference due to the quantization noise from the rate-limited CSI exchange. By contrast, the proposed scheme demonstrates performance improvement as being aware of the noise from CSI exchange and regularizes the feedback and precoding according to the noise statistics. Note that although the proposed scheme shows conservative gains over the precoder feedback scheme under moderate D2D capacity for CSI exchange, as will be shown later, it demonstrates robustness when the D2D for CSI exchange is poor, where the performance of the precoder feedback scheme significantly deteriorates.



(a)



(b)

Figure 2. Sum rate performance: (a) sum rate versus the total downlink transmission power P under $B_\mathrm{d} = 60$ bits for CSI exchange, and (b) sum rate versus B_d. The performance is benchmarked with the ideal case where the users have perfect CSI exchange with each other (star dashed curves).

Fig. 2 (b) shows the sum rate versus the number of bits B_d for CSI exchange. First, the proposed scheme outperforms both the precoder feedback scheme and the CSI feedback scheme over all D2D capacity for CSI exchange. In the regime of no CSI exchange, $B_\mathrm{d} = 0$, it approaches the CSI feedback scheme with robust MMSE precoder. In the regime of high quality CSI exchange, $B_\mathrm{d} = 150$ bits, the proposed scheme converges to the precoder feedback scheme, and performs significantly better than the CSI feedback scheme. Second, by removing one D2D link in case B, the performance of the proposed scheme and the precoder feedback scheme both degrade, but the precoder feedback scheme degrades more at the high SNR regime. Note that at $B_\mathrm{d} = 0$, the proposed scheme performs slightly better than the CSI feedback scheme with robust MMSE precoding. This is because the precoder developed in [5] assume the channel gain $\|\mathbf{h}_k\|$ is also (implicitly) conveyed to the BS, whereas, in our implementation (for fair comparison), only the channel direction $\mathbf{h}_k/\|\mathbf{h}_k\|$ is quantized and reported to the BS.

2) Robustness: The performance under heterogeneous D2D link qualities for CSI exchange is demonstrated in Fig. 3, which shows the achievable downlink data rate for each user versus B_d under case B, where there is no D2D link between user 2 and user 3, while all the other links have identical quality, i.e., $\alpha_{23} = \alpha_{32} = 0$, and $\alpha_{kj}^2 = 1 - 2^{-B_\mathrm{d}/N_t}$ for $(k,j) \notin \{(2,3),(3,2)\}$. The precoder feedback scheme boosts the performance for user 1, but significantly sacrifices user 2 and 3, who achieve much lower data rate than the CSI feedback scheme. By contrast, the proposed scheme outperforms the CSI feedback scheme over all D2D capacity for all users. Specifically, user 2 and 3 achieve similar data rate, and user 1 performs slightly better. This shows that the proposed scheme is able to maintain fairness and robustness when users have heterogeneous D2D link qualities.

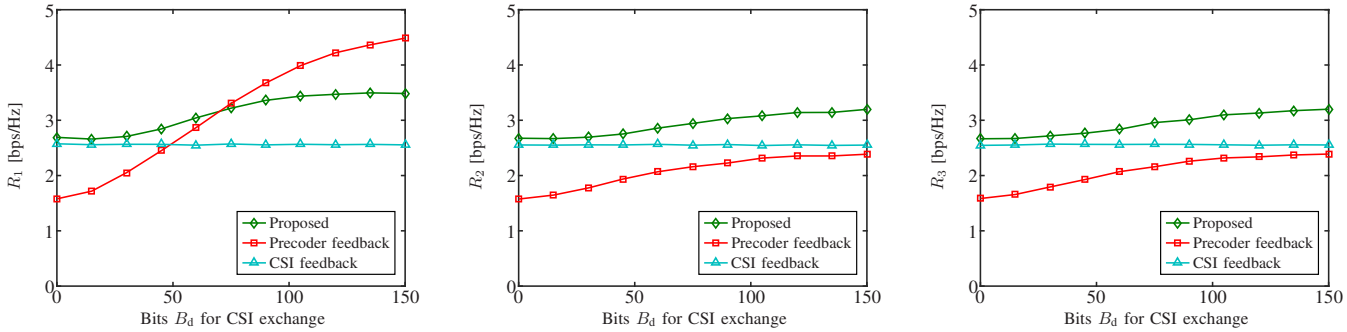
3) Backward Compatibility: Considering that user 3 has no D2D at all, it has to feedback the CSI. Fig. 4 shows the achievable downlink data rate for all the users, where the precoder feedback scheme yields performance worse than the CSI feedback scheme for all the users. By contrast, for the proposed scheme, user 1 and 2 still benefit from D2D and perform better than the CSI feedback scheme. Meanwhile, the performance of user 3 is not harmed, although user 1 and 2 do not have any CSI from user 3. This shows that the proposed scheme is compatible with the conventional CSI feedback scheme, where the BS can serve users with D2D and without D2D simultaneously without performance degradation for any user.

B. Correlated Channels

Consider $K = 2$ user case, where user 1 and user 2 are both away from the BS by 60 meters. The inter-user distance between user 1 and user 2 ranges from 1 to 25 meters. Consider the urban macro scenario specified in the 3GPP standard [29], where the power azimuth spectrum is modeled to have a Laplacian distribution and the angular spread is modeled as a log-normal random variable $\sigma_\mathrm{AS} = 10^{0.34x+0.81}$ with x being standard Gaussian distributed. Uniform linear antenna array (ULA) is assumed when calculating the antenna correlations. In addition, the two users independently experience in log-normal shadowing with 8 dB standard deviation.

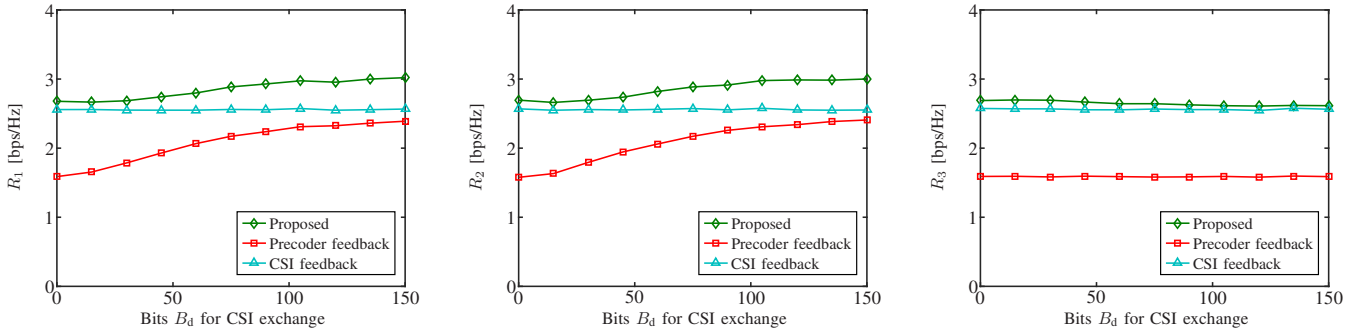
Assume that the SNR for D2D communication is P_d when the inter-user distance is $d = 1$ meter, and the SNR varies as $P_\mathrm{d} d^{-2}$ for $d \geq 1$. Suppose that for each time slot, 20 sec $\cdot$ Hz radio resource is allocated for CSI exchange for each user, for example, to transmit in a 1 ms time frame using 20 kHz bandwidth. Therefore, the number of bits for CSI exchange at each time slot is modeled as $B_\mathrm{d} = 20\log_2(1 + P_\mathrm{d} d^{-2})$. For example, $B_\mathrm{d} = 133$ bits for $d = 1$ meter and $P_\mathrm{d} = 20$ dB. The users exchange the CSI using *entropy-coded scalar quantization* based on subspace projections as described in Section II-B. It can be shown using distortion-rate theories that the CSI distortion via D2D is given by [21]

$$1 - \alpha_{jk}^2 = \frac{M_{kj}}{\mathrm{tr}\{\mathbf{R}_{kj}\}} \Big(\prod_{i=1}^{M_{kj}} \lambda_{kj}^{(i)} \Big)^{1/M_{kj}} 2^{-B_\mathrm{d}/M_{kj}} \qquad (40)$$



(a) User 1 (b) User 2 (c) User 3

Figure 3. Robustness: achievable downlink data rate for each user versus the number of bits B_d for CSI exchange under case B, where there is no D2D link between user 2 and user 3, while all the other links have identical link quality, i.e., $\alpha_{23} = \alpha_{32} = 0$, and $\alpha_{kj}^2 = 1 - 2^{-B_\mathrm{d}/N_t}$ for $(k, j) \notin \{(2, 3), (3, 2)\}$.



(a) User 1 (b) User 2 (c) User 3

Figure 4. Backward compatibility: achievable downlink data rate for each user versus the number of bits B_d for CSI exchange under case C, where user 3 has no D2D at all, but user 1 and user 2 have identical D2D link quality, i.e., $\alpha_{12}^2 = \alpha_{21}^2 = 1 - 2^{-B_\mathrm{d}/N_t}$, and $\alpha_{kj} = 0$ otherwise.

and the quantization error variance is $\boldsymbol{\Xi}_{kj} = \frac{\mathrm{tr}\{\mathbf{R}_{kj}\}}{M_{kj}} \mathbf{V}_{kj} \mathbf{V}_{kj}^\mathrm{H}$, where $\lambda_{kj}^{(i)}$ is the ith largest eigenvalue of $\mathbf{R}_{kj}$.

1) High SNR for D2D: Fig. 5 shows the sum rate for the MIMO downlink transmission versus the inter-user distance d between the two users. Note that as d increases, the quality for CSI exchange deteriorates, but the signal subspaces of the two users are less overlapped due to the one-ring model. For $P_\mathrm{d} = 30$ dB, the proposed scheme achieves similar performance to the precoder feedback scheme. This is because the proposed scheme operates at the regime of high CSI exchange quality. First, when the inter-user distance d is small, the SNR for D2D communication is high, and hence there are sufficient number of bits for high quality CSI exchange. Second, when d is large, the SNR for D2D communication drops, but the rank M_{kj} for the overlapped signal subspace characterized by $\mathbf{R}_{kj}$ also drops. As a result, the quality for CSI exchange α_{kj} in (40) remains close to 1. Third, it is observed that both the proposed scheme and the precoder feedback scheme significantly outperform the CSI feedback scheme.

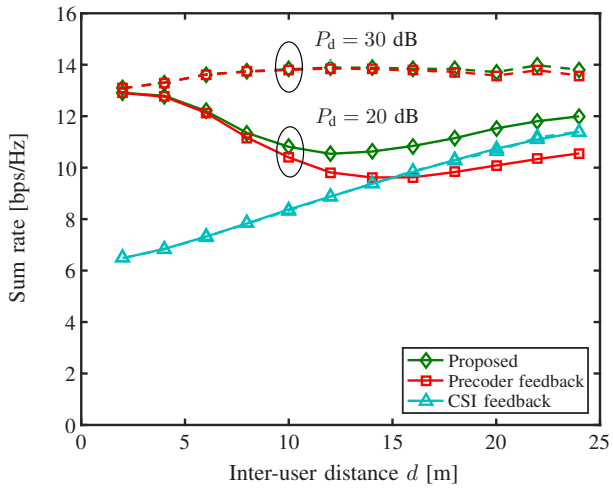


Figure 5. Sum rate of MIMO downlink transmission versus the inter-user distance d between the two users.

2) Medium or Low SNR for D2D: For $P_\mathrm{d} = 20$ dB, the proposed scheme performs similarly to the precoder feedback scheme for small d, but it performs significantly better than the precoder feedback scheme when d increases. This is because, the SNR for D2D communication drops faster than the shrinkage of the overlapped signal subspace. For large d, the proposed scheme enters into the regime of low CSI exchange quality, and it demonstrates better performance by using the dual-regularization technique.

VI. Conclusion

This paper proposes a dual-regularized feedback and precoding strategy for multiuser MIMO systems with limited feedback from the users to the BS, where the users can exploit short range D2D communications to exchange a portion of CSI with each other. The proposed strategy exploits such imperfect global CSI at the user side to feedback a regularized vector to the BS. Based on the feedback, the BS computes an MMSE type precoder regularized by the CSI uncertainty at the user side due to the noisy CSI exchange via D2D. The feedback and precoding strategy was derived by formulating an MMSE problem for the whole network and solving it using a team decision concept, where each user tries to solve the MMSE problem given the individual observation of the global CSI, and the BS solves the MMSE problem given the feedback from the users. It was shown that the solutions to such sequential MMSE problems take similar forms to the maximum SLNR precoding and RZF precoding, with additional regularizations to capture the uncertainty of the exchanged CSI between users. Numerical results show that in terms of sum rate performance, the proposed feedback and precoding strategy performs uniformly better than both the CSI feedback scheme and precoder feedback scheme from low to high SNR and from low to high D2D communication qualities. In addition, the proposed scheme demonstrates robustness and backward compatibility over heterogenous D2D link qualities, where all the users can gain benefits even though some users may have poor or no D2D.

Appendix A
Proof of Proposition 1

From (9), we need to evaluate

$$\mathbb{E}\Big\{\Big\|(\hat{\mathbf{H}}_k\mathbf{A}_k^{\frac{1}{2}}+\mathbf{E}_k\mathbf{S}_k^{\frac{1}{2}}+\mathbf{H}_k^{\perp})^{\mathrm{H}}\widetilde{\mathbf{W}}-\mathbf{I}\Big\|_F^2\,\Big|\hat{\mathbf{H}}_k\Big\}.$$

First, due to the CSI exchange strategy in Section II-B, we have $\mathbb{E}\{\mathbf{H}_k^{\perp}\,|\hat{\mathbf{H}}_k\}=\mathbb{E}\{\mathbf{H}_k^{\perp}\}=\mathbf{0}$, and

$$\mathbb{E}\{\mathbf{H}_k^{\perp}(\mathbf{H}_k^{\perp})^{\mathrm{H}}\}=\sum_{j\neq k}\mathbb{E}\{\mathbf{h}_j^{(k)\perp}(\mathbf{h}_j^{(k)\perp})^{\mathrm{H}}\}=\sum_{j\neq k}\mathbf{R}_{jk}^{\perp}$$

where $\mathbf{R}_{jk}^{\perp}\triangleq(\mathbf{I}-\mathbf{V}_k\mathbf{V}_k^{\mathrm{H}})\mathbf{R}_j(\mathbf{I}-\mathbf{V}_k\mathbf{V}_k^{\mathrm{H}})$ is the covariance matrix of $\mathbf{h}_j$ projected onto the null space of $\mathbf{R}_k$. In addition,

$$\begin{aligned}
&\mathrm{tr}\Big\{\mathbb{E}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\mathbf{E}_k\mathbf{S}_k\mathbf{E}_k^{\mathrm{H}}\widetilde{\mathbf{W}}\,\Big|\hat{\mathbf{H}}_k\Big\}\Big\}\\
&\quad=\mathrm{tr}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\mathbb{E}\Big\{\mathbf{E}_k\mathbf{S}_k\mathbf{E}_k^{\mathrm{H}}\Big\}\widetilde{\mathbf{W}}\Big\}\\
&\quad=\mathrm{tr}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\sum_{j\neq k}(1-\alpha_{kj}^2)\mathbb{E}\Big\{\boldsymbol{\xi}_j^{(k)}(\boldsymbol{\xi}_j^{(k)})^{\mathrm{H}}\Big\}\widetilde{\mathbf{W}}\Big\}\\
&\quad=\mathrm{tr}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\sum_{j\neq k}(1-\alpha_{kj}^2)\boldsymbol{\Xi}_{jk}\widetilde{\mathbf{W}}\Big\}. \qquad (41)
\end{aligned}$$

Using the property of the Frobenius norm, we have

$$\begin{aligned}
&\mathbb{E}\Big\{\Big\|(\hat{\mathbf{H}}_k\mathbf{A}_k^{\frac{1}{2}}+\mathbf{E}_k\mathbf{S}_k^{\frac{1}{2}}+\mathbf{H}_k^{\perp})^{\mathrm{H}}\widetilde{\mathbf{W}}-\mathbf{I}\Big\|_F^2\,\Big|\hat{\mathbf{H}}_k\Big\}\\
&\quad=\mathrm{tr}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\hat{\mathbf{H}}_k\mathbf{A}_k\hat{\mathbf{H}}_k^{\mathrm{H}}\widetilde{\mathbf{W}}\Big\}\\
&\qquad+\mathrm{tr}\Big\{\mathbb{E}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\mathbf{E}_k\mathbf{S}_k\mathbf{E}_k^{\mathrm{H}}\widetilde{\mathbf{W}}\,\Big|\hat{\mathbf{H}}_k\Big\}\Big\}\\
&\qquad+\mathrm{tr}\Big\{\mathbb{E}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\mathbf{H}_k^{\perp}(\mathbf{H}_k^{\perp})^{\mathrm{H}}\widetilde{\mathbf{W}}\Big\}\Big\}\\
&\qquad-2\mathrm{tr}\Big\{\mathrm{Re}\big\{\mathbf{A}_k^{\frac{1}{2}}\hat{\mathbf{H}}_k^{\mathrm{H}}\widetilde{\mathbf{W}}\big\}\Big\}+K\\
&\quad=\mathrm{tr}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\hat{\mathbf{H}}_k\mathbf{A}_k\hat{\mathbf{H}}_k^{\mathrm{H}}\widetilde{\mathbf{W}}\Big\}\\
&\qquad+\mathrm{tr}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\sum_{j\neq k}\Big[(1-\alpha_{kj}^2)\boldsymbol{\Xi}_{jk}+\mathbf{R}_{jk}^{\perp}\Big]\widetilde{\mathbf{W}}\Big\}\\
&\qquad-2\mathrm{tr}\Big\{\mathrm{Re}\big\{\mathbf{A}_k^{\frac{1}{2}}\hat{\mathbf{H}}_k^{\mathrm{H}}\widetilde{\mathbf{W}}\big\}\Big\}+K
\end{aligned}$$

where the last equality exploits the property (41).

As a result, the objective function (8) becomes

$$\begin{aligned}
&\mathbb{E}\Big\{\Big\|(\hat{\mathbf{H}}_k\mathbf{A}_k^{\frac{1}{2}}+\mathbf{E}_k\mathbf{S}_k^{\frac{1}{2}}+\mathbf{H}_k^{\perp})^{\mathrm{H}}\widetilde{\mathbf{W}}-\mathbf{I}\Big\|_F^2\,\Big|\hat{\mathbf{H}}_k\Big\}+K\frac{\mathrm{tr}\{\widetilde{\mathbf{W}}^{\mathrm{H}}\widetilde{\mathbf{W}}\}}{P}\\
&=\mathrm{tr}\Big\{\widetilde{\mathbf{W}}^{\mathrm{H}}\Big[\hat{\mathbf{H}}_k\mathbf{A}_k\hat{\mathbf{H}}_k^{\mathrm{H}}+\sum_{j\neq k}\mathbf{R}_{jk}^{\perp}\\
&\qquad+\sum_{j\neq k}(1-\alpha_{kj}^2)\boldsymbol{\Xi}_{jk}+\frac{K}{P}\mathbf{I}_{N_t}\Big]\widetilde{\mathbf{W}}\Big\}\\
&\qquad-2\mathrm{tr}\Big\{\mathrm{Re}\big\{\mathbf{A}_k^{\frac{1}{2}}\hat{\mathbf{H}}_k^{\mathrm{H}}\widetilde{\mathbf{W}}\big\}\Big\}+K
\end{aligned}\qquad(42)$$

which is a quadratic function in $\widetilde{\mathbf{W}}$. It is see to see from (42) that the Hessian matrix (second order derivative) is positive definite; therefore, the global minimizer is obtained at setting the derivative of (42) to $\mathbf{0}$. In the special case of $\mathbf{R}_{jk}^{\perp}=\mathbf{0}$ for all (k,j), the optimal solution is thus given by

$$\widetilde{\mathbf{W}}^{\star}=\Big(\hat{\mathbf{H}}_k\mathbf{A}_k\hat{\mathbf{H}}_k^{\mathrm{H}}+\mathbf{Q}_k\Big)^{-1}\hat{\mathbf{H}}_k\mathbf{A}_k^{\frac{1}{2}}\qquad(43)$$

where $\mathbf{Q}_k=\sum_{j\neq k}(1-\alpha_{kj}^2)\boldsymbol{\Xi}_{jk}+\frac{K}{P}\mathbf{I}_{N_t}$.

In the general case of $\mathbf{R}_{jk}^{\perp}\neq\mathbf{0}$, we apply the following lemma.

Lemma 3: Let $\mathbf{U}\in\mathbb{C}^{N\times m_1}$ and $\mathbf{V}\in\mathbb{C}^{N\times m_2}$ be two semi-unitary matrices that satisfy $\mathbf{U}^{\mathrm{H}}\mathbf{V}=\mathbf{0}$. The following holds

$$(\mathbf{I}+\mathbf{U}\mathbf{D}_1\mathbf{U}^{\mathrm{H}})^{-1}\mathbf{U}\mathbf{b}=(\mathbf{I}+\mathbf{U}\mathbf{D}_1\mathbf{U}^{\mathrm{H}}+\mathbf{V}\mathbf{D}_2\mathbf{V}^{\mathrm{H}})^{-1}\mathbf{U}\mathbf{b}\qquad(44)$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are two positive semidefinite diagonal matrices, and $\mathbf{b}$ is an m_1-dimensional vector.

Proof: The matrix inversion lemma yields

$$(\mathbf{I}+\mathbf{U}\mathbf{D}_1\mathbf{U}^{\mathrm{H}})^{-1}=\mathbf{I}-\mathbf{U}(\mathbf{D}_1^{-1}+\mathbf{U}^{\mathrm{H}}\mathbf{U})\mathbf{U}^{\mathrm{H}}$$

and hence

$$\begin{aligned}
(\mathbf{I}+\mathbf{U}\mathbf{D}_1\mathbf{U}^{\mathrm{H}})^{-1}\mathbf{U}\mathbf{b}&=\mathbf{U}\mathbf{b}-\mathbf{U}(\mathbf{D}_1^{-1}+\mathbf{I}_{m_1})\mathbf{b}\\
&=-\mathbf{U}\mathbf{D}_1^{-1}\mathbf{b}.
\end{aligned}$$

Thus, $\mathbf{x}=-\mathbf{U}\mathbf{D}_1^{-1}\mathbf{b}$ is the unique solution to the equation $(\mathbf{I}+\mathbf{U}\mathbf{D}_1\mathbf{U}^{\mathrm{H}})\mathbf{x}=\mathbf{U}\mathbf{b}$.

From the orthogonality, we have $\mathbf{V}\mathbf{D}_2\mathbf{V}^{\mathrm{H}}\mathbf{x} = -\mathbf{V}\mathbf{D}_2\mathbf{V}^{\mathrm{H}}\mathbf{U}\mathbf{D}_1^{-1}\mathbf{b} = \mathbf{0}$. Therefore, $\mathbf{x}$ is also the unique solution to the equation $(\mathbf{I}+\mathbf{U}\mathbf{D}_1\mathbf{U}^{\mathrm{H}}+\mathbf{V}\mathbf{D}_2\mathbf{V}^{\mathrm{H}})\mathbf{x} = \mathbf{U}\mathbf{b}$, where the solution can also be written as $\mathbf{x} = (\mathbf{I}+\mathbf{U}\mathbf{D}_1\mathbf{U}^{\mathrm{H}}+\mathbf{V}\mathbf{D}_2\mathbf{V}^{\mathrm{H}})^{-1}\mathbf{U}\mathbf{b}$. This confirms identity (44). □

Note that $\mathbf{R}_{jk}^{\perp}$ and $\hat{\mathbf{H}}_k$ are in orthogonal subspaces for each $j \neq k$. To see this, one may observe that $\mathbf{h}_m^{(k)\perp}$ and $\hat{\mathbf{h}}_n^{(k)}$ are orthogonal for all m, n. This is because by the CSI exchange strategy in Section II-B, $\hat{\mathbf{h}}_n^{(k)}$ lies in the subspace of $\mathbf{R}_k$, whereas, $\mathbf{h}_m^{(k)\perp}$ lies in the subspace that is orthogonal to the subspace of $\mathbf{R}_{mk}$, and thus must be orthogonal to the subspace of $\mathbf{R}_k$. As a result, Lemma 3 can be applied, and (43) is also the solution for the case $\mathbf{R}_{jk}^{\perp} \neq \mathbf{0}$. □

APPENDIX B
PROOF OF PROPOSITION 2

In the ideal case of infinite-rate feedback, the feedback vector $\mathbf{u}_k$ is identical to the kth column of $\mathbf{W}_k^{\star}$ and using Lemma 1, the vector can be computed as

$$\begin{aligned}\mathbf{u}_k &= \arg\max_{\|\mathbf{u}\|=1} \frac{|\mathbf{h}_k^{\mathrm{H}}\mathbf{u}|^2}{\sum_{j\neq k}\alpha_{kj}^2\left|(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{u}\right|^2 + q_k} \\ &= \arg\max_{\|\mathbf{u}\|=1} \frac{\mathbf{u}^{\mathrm{H}}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}\mathbf{u}}{\mathbf{u}^{\mathrm{H}}\Big(\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}} + q_k\mathbf{I}_{N_t}\Big)\mathbf{u}} \quad (45)\\ &= \theta_k^{-1}\left[\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}} + q_k\mathbf{I}_{N_t}\right]^{-1}\mathbf{h}_k \quad (46)\\ &= \theta_k^{-1}\mathbf{G}_k^{-1}\mathbf{h}_k \quad (47)\end{aligned}$$

where the complex scalar[6]

$$\theta_k = e^{j\vartheta_k}\left\|\left[\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}} + q_k\mathbf{I}_{N_t}\right]^{-1}\mathbf{h}_k\right\|_2 \quad (48)$$

is to make sure $\|\mathbf{u}_k\| = 1$.

However, the feedback vector $\hat{\mathbf{u}}_k$ is computed from a finite set $\mathcal{C}_k$. To model the relationship between $\mathbf{u}_k$ and $\hat{\mathbf{u}}_k$, observe that $\mathbf{u}_k$ is isotropically distributed. This is because $\mathbf{h}_k$ and $\hat{\mathbf{h}}_j^{(k)}$ are isotropically distributed and mutually independent, and hence from (45), $\mathbf{u}_k$ is isotropic. Since the vectors in $\mathcal{C}_k$ are also isotropic (see (7)), we can write

$$\mathbf{u}_k = \sqrt{1-\epsilon_k^2}\hat{\mathbf{u}}_k + \epsilon_k\mathbf{z}_k \quad (49)$$

where there is a unique pair of variables $0 \leq \epsilon_k \leq 1$ and $\mathbf{z}_k \in \mathbb{C}^{N_t}$, such that $\|\mathbf{z}_k\| = 1$ and $\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{z}_k = 0$. Note that the marginal distribution of $\mathbf{z}_k$ is isotropic, because both $\mathbf{u}_k$ and $\hat{\mathbf{u}}_k$ are isotropic.

The discretization error between an isotropic vector and an isotropic codebook was studied in [3], and the result is given in the following lemma.

[6]The phase ϑ_k is to capture the difference between the solution to maximizing the Rayleigh quotient (45) and the solution from pseudo-inverse (46).

Lemma 4 (Discretization error [3]): The following result holds

$$\frac{N_t-1}{N_t}2^{-\frac{B}{N_t-1}} \leq \mathbb{E}\{\epsilon_k^2\} \leq 2^{-\frac{B}{N_t-1}}$$

where the expectation is taken over the distributions of $\mathbf{h}_k$, $\hat{\mathbf{h}}_j^{(k)}$, and $\mathcal{C}_k$.

Note that $\mathbf{G}_k$ is invertible because $q_k > 0$. As a result, we have $\mathbf{h}_k = \theta_k\mathbf{G}_k\big(\sqrt{1-\epsilon_k^2}\hat{\mathbf{u}}_k + \epsilon_k\mathbf{z}_k\big)$, which leads to (19). □

APPENDIX C
PROOF OF LEMMA 2

To show (22), first, we note that from (49) in Appendix B, ϵ_k is modeled as the magnitude of the discretization error from $\mathbf{u}_k$, and hence, is statistically independent to $\hat{\mathbf{u}}_k$ and $\mathbf{z}_k$. Second, from Lemma 4, $\mathbb{E}\{\epsilon_k^2\}$ is upper bounded by $2^{-\frac{B}{N_t-1}}$, which implies that $\mathbb{E}\{\epsilon_k^2\} \to 0$ as $B \to \infty$. Third, $\theta_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}$ is always bounded. Therefore, the limits in both sides of (22) exit and equal to 0 as $B \to \infty$.

To show (23), consider $\alpha_{kj} = 0$; the result trivially holds. Consider $\alpha_{kj} = 1$, which yields $1-\alpha_{kj}^2 = 0$ in (17). We then observe that

$$\lim_{B\to\infty}\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{u}}_k = \sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{u}_k$$

where $\mathbf{u}_k$ is given by (46), which converges to the zero-forcing (ZF) vector that is orthogonal to $\hat{\mathbf{h}}_j^{(k)}$ for $j \neq k$ with probability 1 (since the matrix $\hat{\mathbf{H}}_k$ has full column rank with probability 1), as $P \to \infty$. This proves that the limit converges to 0.

To show (24), we observe from (48) that the term $\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}$ becomes negligible when parameters α_{kj} approach to 0. This suggests that θ_k mainly depends only on $\mathbf{h}_k$ in its asymptotic distribution. As a result, as $\alpha_{kj} \to 0$, θ_k is independent to $\hat{\mathbf{U}}$ and has identical distribution over all k's. Thus, the equation holds. □

APPENDIX D
PROOF OF PROPOSITION 3

It can be shown that for finite x and a, if $x \overset{\Theta}{\approx} y$ and $a \overset{\Theta}{\approx} b$, then $ax \overset{\Theta}{\approx} by$. To see this, denote $x = y+\Theta_1$ and $a = b+\Theta_2$, where Θ_1 and Θ_2 goes to 0 as P, B, $\{\alpha_{kj}^{-1}\}$, and N_t all go to infinity. Then $ax = by + b\Theta_1 + y\Theta_2 + \Theta_1\Theta_2$, where $b\Theta_1+y\Theta_2+\Theta_1\Theta_2$ goes to 0 as the aforementioned parameters go to infinity.

A. First Term of (21)

$$\begin{aligned}&\mathbf{G}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}} \\ &\quad= \sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\sum_{l\neq k}\alpha_{kl}^2\hat{\mathbf{h}}_l^{(k)}(\hat{\mathbf{h}}_l^{(k)})^{\mathrm{H}} \\ &\qquad+q_k^2\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}} + q_k\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}} \\ &\qquad+q_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\sum_{l\neq k}\alpha_{kl}^2\hat{\mathbf{h}}_l^{(k)}(\hat{\mathbf{h}}_l^{(k)})^{\mathrm{H}} \\ &\overset{\Theta}{\approx} q_k^2\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\end{aligned}$$

where approximation (26) is used. As a result, $\mathbb{E}\{\mathbf{G}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\,|\hat{\mathbf{U}}\} \overset{\Theta}{\approx} q_k^2\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}$.

From the noisy feedback model in (49), since the unit norm vector $\hat{\mathbf{u}}_k$ characterizes the direction and ϵ_k characterizes the discretization error in terms of magnitude, they are statistically independent. Therefore,

$$\begin{aligned}\mathbb{E}\Big\{(1-\epsilon_k^2)\mathbf{G}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\,\Big|\hat{\mathbf{U}}\Big\}&=\mathbb{E}\{(1-\epsilon_k^2)\}\mathbb{E}\{\mathbf{G}_k\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\,|\hat{\mathbf{U}}\}\\&\overset{\Theta}{\approx}(1-2^{-\frac{B}{N_t-1}})q_k^2\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}}\end{aligned}\tag{50}$$

where the approximation (28) is used.

B. Second Term of (21)

$$\begin{aligned}&\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\\&\quad=\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\sum_{l\neq k}\alpha_{kl}^2\hat{\mathbf{h}}_l^{(k)}(\hat{\mathbf{h}}_l^{(k)})^{\mathrm{H}}\\&\qquad+q_k^2\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}+q_k\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\\&\qquad+q_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\sum_{l\neq k}\alpha_{kl}^2\hat{\mathbf{h}}_l^{(k)}(\hat{\mathbf{h}}_l^{(k)})^{\mathrm{H}}\end{aligned}\tag{51}$$

where

$$\mathbb{E}\{\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\}=\frac{1}{N_t}\mathbf{I}\tag{52}$$

since $\mathbf{z}_k$ is the discretization noise that follows an isotropic distribution from (49).

The expectation of (51) can be calculated term-by-term as follows.

For the first term of (51), since $\mathbf{h}_j$ is independent to $\mathbf{h}_l$, we have that $\hat{\mathbf{h}}_j^{(k)}$ is independent to $\hat{\mathbf{h}}_l^{(k)}$, and both of them are independent to $\mathbf{z}_k$. Thus, for $j\neq l$, we have

$$\begin{aligned}&\mathbb{E}\Big\{\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\alpha_{kl}^2\hat{\mathbf{h}}_l^{(k)}(\hat{\mathbf{h}}_l^{(k)})^{\mathrm{H}}\Big\}\\&\quad=\alpha_{kj}^2\alpha_{kl}^2\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\Big\}\mathbb{E}\Big\{\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\Big\}\mathbb{E}\Big\{\hat{\mathbf{h}}_l^{(k)}(\hat{\mathbf{h}}_l^{(k)})^{\mathrm{H}}\Big\}\\&\quad=\frac{\alpha_{kj}^2\alpha_{kl}^2}{N_t}\mathbf{I}\end{aligned}$$

For $j=l$, we have

$$\begin{aligned}&\mathbb{E}\Big\{\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\Big\}\\&\quad=\alpha_{kj}^4\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbb{E}\{\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\}\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\Big\}\\&\quad=\frac{\alpha_{kj}^4}{N_t}\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\Big\}\end{aligned}\tag{53}$$

To evaluate the expectation in (53), we apply the law of total expectation $\mathbb{E}\{X\}=\mathbb{E}\{\mathbb{E}\{X|Y\}\}$. Specifically, denote $\varrho_j\triangleq(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{h}}_j^{(k)}$. Note that according to the model for CSI exchange in (2), the (marginal) distribution of $\hat{\mathbf{h}}_j^{(k)}$ is complex Gaussian with zero mean and unit variance. Hence $2\varrho_j$ is a Chi-square random variable with $2N_t$ degrees of freedom, and the second order moment of ϱ_j is $\mathbb{E}\{\varrho_j^2\}=\frac{1}{4}\mathbb{E}\{(2\varrho_j)^2\}=N_t(N_t+1)$.[7] Then,

$$\begin{aligned}&\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\Big\}\\&\quad=\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}\varrho_j(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\Big\}\\&\quad=\mathbb{E}\Big\{\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}\varrho_j(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\,\Big|\varrho_j\Big\}\Big\}\\&\quad=\mathbb{E}\Big\{\varrho_j\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\,\Big|\varrho_j\Big\}\Big\}\\&\quad=\mathbb{E}\{\varrho_j^2\}\frac{1}{N_t}\mathbf{I}\\&\quad=(N_t+1)\mathbf{I}\end{aligned}\tag{54}$$

where equality (54) uses the fact that $\hat{\mathbf{h}}_j^{(k)}$ is isotropically distributed given the magnitude ϱ_j.

As a result, the first term of (51) becomes

$$\begin{aligned}&\mathbb{E}\Big\{\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\sum_{l\neq k}\alpha_{kl}^2\hat{\mathbf{h}}_l^{(k)}(\hat{\mathbf{h}}_l^{(k)})^{\mathrm{H}}\Big\}\\&\quad=\sum_{j\neq k}\frac{\alpha_{kj}^2}{N_t}\Big(\alpha_{kj}^2(N_t+1)+\sum_{l\neq k,j}\alpha_{kl}^2\Big)\mathbf{I}\\&\quad=\sum_{j\neq k}\frac{\alpha_{kj}^2}{N_t}\Big(\alpha_{kj}^2N_t+\sum_{l\neq k}\alpha_{kl}^2\Big)\mathbf{I}.\end{aligned}$$

The third and fourth terms of (51) can be evaluated as

$$\begin{aligned}&\mathbb{E}\Big\{q_k\sum_{j\neq k}\alpha_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\Big\}\\&\quad=\sum_{j\neq k}q_k\alpha_{kj}^2\mathbb{E}\Big\{\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\Big\}\mathbb{E}\Big\{\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\Big\}\\&\quad=\sum_{j\neq k}q_k\frac{\alpha_{kj}^2}{N_t}\mathbf{I}.\end{aligned}$$

Moreover, the discretization error ϵ_k in magnitude is independent of the error $\mathbf{z}_k$ in direction, and therefore, we have

$$\begin{aligned}&\mathbb{E}\Big\{\epsilon_k^2\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\Big\}\\&\quad=\mathbb{E}\{\epsilon_k^2\}\mathbb{E}\Big\{\mathbf{G}_k\mathbf{z}_k\mathbf{z}_k^{\mathrm{H}}\mathbf{G}_k^{\mathrm{H}}\Big\}\\&\quad\overset{\Theta}{\approx}2^{-\frac{B}{N_t-1}}\Big[\sum_{j\neq k}\frac{\alpha_{kj}^2}{N_t}\Big(\alpha_{kj}^2N_t+\sum_{l\neq k}\alpha_{kl}^2+2q_k\Big)+\frac{q_k^2}{N_t}\Big]\mathbf{I}\end{aligned}\tag{55}$$

where approximation (28) is used.

[7] The kth moment of the central Chi-square random variable with $2m$ degrees of freedom is given by $2^k\frac{(m+k-1)!}{(m-1)!}$.

C. Derivation for $\mathbb{E}\{\mathbf{H}\mathbf{H}^H|\hat{\mathbf{U}}\}$

Using the fact that $\mathbf{H}\mathbf{H}^{\mathrm{H}} = \sum_{k=1}^{K}\mathbf{h}_k\mathbf{h}_k^{\mathrm{H}}$ and $\sum_{k=1}^{K} q_k^2\hat{\mathbf{u}}_k\hat{\mathbf{u}}_k^{\mathrm{H}} = \hat{\mathbf{U}}\boldsymbol{\Omega}\hat{\mathbf{U}}^{\mathrm{H}}$, we have

$$\begin{aligned}
&\mathbb{E}\Big\{\mathbf{H}\mathbf{H}^{\mathrm{H}}\big|\hat{\mathbf{U}}\Big\}\\
&\overset{\Theta}{\approx} \varpi\Bigg[\big(1-2^{-\frac{B}{N_t-1}}\big)\hat{\mathbf{U}}\boldsymbol{\Omega}\hat{\mathbf{U}}^{\mathrm{H}}\\
&\quad + 2^{-\frac{B}{N_t-1}}\sum_{k=1}^{K}\Bigg[\sum_{j\neq k}\frac{\alpha_{kj}^2}{N_t}\Big(\alpha_{kj}^2N_t+\sum_{l\neq k}\alpha_{kl}^2+2q_k\Big)+\frac{q_k^2}{N_t}\Bigg]\mathbf{I}\\
&\overset{\Theta}{\approx} \varpi\Bigg[\big(1-2^{-\frac{B}{N_t-1}}\big)\hat{\mathbf{U}}\boldsymbol{\Omega}\hat{\mathbf{U}}^{\mathrm{H}}+2^{-\frac{B}{N_t-1}}\sum_{k=1}^{K}\Big(\sum_{j\neq k}\alpha_{kj}^4+\frac{q_k^2}{N_t}\Big)\mathbf{I}\Bigg]
\end{aligned}$$

where $\boldsymbol{\Omega}$ is given in (29), and the last approximation is to use the fact that

$$\begin{aligned}
&\sum_{j\neq k}\frac{\alpha_{kj}^2}{N_t}\Big(\alpha_{kj}^2N_t+\sum_{l\neq k}\alpha_{kl}^2+2q_k\Big)+\frac{q_k^2}{N_t}\\
&\quad=\sum_{j\neq k}\alpha_{kj}^4+\frac{q_k^2}{N_t}+\sum_{j\neq k}\frac{\alpha_{kj}^2}{N_t}\Big(\sum_{l\neq k}\alpha_{kl}^2+2q_k\Big)\\
&\quad\overset{\Theta}{\approx}\sum_{j\neq k}\alpha_{kj}^4+\frac{q_k^2}{N_t}.
\end{aligned}$$

D. Derivation for $\mathbb{E}\{\mathbf{H}|\hat{\mathbf{U}}\}$

To compute $\mathbb{E}\{\mathbf{h}_k\big|\hat{\mathbf{U}}\}$, we have

$$\begin{aligned}
&\mathbb{E}\Big\{\theta_k\mathbf{G}_k\big(\sqrt{1-\epsilon_k^2}\hat{\mathbf{u}}_k+\epsilon_k\mathbf{z}_k\big)\,\big|\hat{\mathbf{U}}\Big\}\\
&\quad=\mathbb{E}\Big\{\sqrt{1-\epsilon_k^2}\Big\}\mathbb{E}\Big\{\theta_k\sum_{j\neq k}a_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\hat{\mathbf{u}}_k+\theta_kq_k\hat{\mathbf{u}}_k\,\big|\hat{\mathbf{U}}\Big\}\\
&\qquad+\mathbb{E}\{\epsilon_k\}\mathbb{E}\Big\{\theta_k\sum_{j\neq k}a_{kj}^2\hat{\mathbf{h}}_j^{(k)}(\hat{\mathbf{h}}_j^{(k)})^{\mathrm{H}}\mathbf{z}_k+\theta_kq_k\mathbf{z}_k\,\big|\hat{\mathbf{U}}\Big\}\\
&\quad\overset{\Theta}{\approx}\mathbb{E}\big\{\sqrt{1-\epsilon_k^2}\big\}\mathbb{E}\big\{\theta_kq_k\,\big|\hat{\mathbf{U}}\big\}\hat{\mathbf{u}}_k \qquad (56)\\
&\quad=\upsilon_k\hat{\mathbf{u}}_k
\end{aligned}$$

where $\upsilon_k \triangleq \mathbb{E}\big\{\sqrt{1-\epsilon_k^2}\big\}\mathbb{E}\big\{\theta_kq_k\,\big|\hat{\mathbf{U}}\big\}$. Approximation (56) is from (26)–(27) and the fact that $\mathbf{z}_k$ is zero mean and independent to $\hat{\mathbf{h}}_j^{(k)}$.

To stack the column vectors $\mathbf{h}_k$ into $\mathbf{H}$ and let $\boldsymbol{\Upsilon} = \mathrm{diag}(\upsilon_1,\upsilon_2,\ldots,\upsilon_K)$, the approximation (31) follows. □

REFERENCES

[1] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 7, pp. 1478–1491, 2007.

[2] D. J. Love and R. W. Heath, "Limited feedback diversity techniques for correlated channels," *IEEE Trans. Veh. Technol.*, vol. 55, no. 2, pp. 718–722, 2006.

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

[4] 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, 2007.

[5] A. D. Dabbagh and D. J. Love, "Multiple antenna MMSE based downlink precoding with quantized feedback or channel mismatch," *IEEE Trans. Commun.*, vol. 56, no. 11, pp. 1859–1868, 2008.

[6] M. Ding and S. D. Blostein, "MIMO minimum total MSE transceiver design with imperfect CSI at both ends," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 1141–1150, 2009.

[7] J. Chen and V. K. N. Lau, "Two-tier precoding for FDD multi-cell massive MIMO time-varying interference networks," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1230–1238, Jun. 2014.

[8] C. Hao, Y. Wu, and B. Clerckx, "Rate analysis of two-receiver MISO broadcast channel with finite rate feedback: A rate-splitting approach," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3232–3246, 2015.

[9] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.*, no. 7, pp. 4611–4624, 2016.

[10] L. Cottatellucci, "D2D CSI feedback for D2D aided massive MIMO communications," in *Proc. IEEE Sensor Array and Multichannel Signal Process. Workshop*, Jul. 2016, pp. 1–5.

[11] H. Yin, L. Cottatellucci, and D. Gesbert, "Enabling massive MIMO systems in the FDD mode thanks to D2D communications," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2014, pp. 656–660.

[12] J. Chen, H. Yin, L. Cottatellucci, and D. Gesbert, "Precoder feedback versus channel feedback in massive MIMO under user cooperation," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2015, pp. 1449–1453.

[13] J. C. F. Li, M. Lei, and F. Gao, "Device-to-device (D2D) communication in MU-MIMO cellular networks," in *IEEE Global Commun. Conf.*, Dec. 2012, pp. 3583–3587.

[14] G. Fodor and N. Reider, "A distributed power control scheme for cellular network assisted D2D communications," in *Proc. IEEE Global Commun. Conf.*, Dec. 2011, pp. 1–6.

[15] H. Tang, C. Zhu, and Z. Ding, "Cooperative MIMO precoding for D2D underlay in cellular networks," in *IEEE Int. Conf. Commun.*, Jun. 2013, pp. 5517–5521.

[16] K. Jayasinghe, P. Jayasinghe, N. Rajatheva, and M. Latva-Aho, "Linear precoder-decoder design of MIMO device-to-device communication underlaying cellular communication," *IEEE Trans. Commun.*, vol. 62, no. 12, pp. 4304–4319, Dec. 2014.

[17] W. Zhong, Y. Fang, S. Jin, K. K. Wong, S. Zhong, and Z. Qian, "Joint resource allocation for device-to-device communications underlaying uplink MIMO cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 1, pp. 41–54, Jan. 2015.

[18] H. Min, W. Seo, J. Lee, S. Park, and D. Hong, "Reliability improvement using receive mode selection in the device-to-device uplink period underlaying cellular networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 2, pp. 413–418, Feb. 2011.

[19] X. Lin, R. W. Heath, and J. G. Andrews, "The interplay between massive MIMO and underlaid D2D networking," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3337–3351, Jun. 2015.

[20] S. Shalmashi, E. Björnson, M. Kountouris, K. W. Sung, and M. Debbah, "Energy efficiency and sum rate when massive MIMO meets device-to-device communication," in *IEEE Int. Conf. Commun. Workshop*, Jun. 2015, pp. 627–632.

[21] J. Chen, H. Yin, L. Cottatellucci, and D. Gesbert, "Feedback mechanisms for FDD massive MIMO under D2D-based limited CSI sharing," *IEEE Trans. Wireless Commun.*, 2017, to appear, preprint http://arxiv.org/abs/1611.07224.

[22] 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.

[23] Q. Li, D. Gesbert, and N. Gresset, "A cooperative channel estimation approach for coordinated multipoint transmission networks," in *Proc. IEEE Int. Conf. Commun. Workshop*, 2015, pp. 94–99.

[24] D. Rebollo-Monedero, S. Rane, and B. Girod, "Wyner-Ziv quantization and transform coding of noisy sources at high rates," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, vol. 2, 2004, pp. 2084–2088.

[25] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Springer Science & Business Media, 1992, vol. 159.

[26] K. Baddour and N. Beaulieu, "Autoregressive models for fading channel simulation," in *Proc. IEEE Global Telecomm. Conf.*, vol. 2, 2001, pp. 1187 –1192 vol.2.

[27] P. Patcharamaneepakorn, S. Armour, and A. Doufexi, "On the equivalence between SLNR and MMSE precoding schemes with single-antenna receivers," *IEEE Commun. Lett.*, vol. 16, no. 7, pp. 1034–1037, 2012.

[28] J. Nam, A. Adhikary, J.-Y. Ahn, 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.

[29] "Spatial channel model for Multiple Input Multiple Output (MIMO) simulations," 3GPP, Tech. Rep. TR 25.996 (Release 14), Mar. 2017.

Junting Chen (S'11–M'16) received the Ph.D. degree in electronic and computer engineering from Hong Kong University of Science and Technology (HKUST), Hong Kong SAR China, in 2015, and the B.Sc. degree in electronic engineering from Nanjing University, Nanjing, China, in 2009.

He is a Postdoctoral Research Associate with Ming Hsieh Department of Electrical Engineering, University of Southern California (USC), Los Angeles, CA, USA. Prior to that, he was a postdoctoral researcher with Department of Communication Systems at EURECOM, France, in 2015–2016. From 2014–2015, he was a visiting student with the Wireless Information and Network Sciences Laboratory at MIT, Cambridge, MA, USA. He was an recipient of the HKTIIT Post-Graduate Excellence Scholarships in 2012 from HKUST. His research interests include signal processing, optimizations, nonlinear control, and statistical learning, with applications to wireless communications and localization.

Haifan Yin obtained the Ph.D. degree from Télécom ParisTech, France in 2015. He received the B.Sc. degree and the M.Sc. degree from Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2012 respectively. He is currently a research engineer in Sequans Communications, France. His research interests include massive MIMO, signal processing, and cooperative networks. H. Yin was a recipient of the 2015 Chinese Government Award for Outstanding Self-financed Students Abroad.

Laura Cottatellucci (S'01–M'07) obtained the Ph.D. from Technical University of Vienna, Austria (2006) and the Master degree from La Sapienza University, Rome, Italy (1995). Specialized in networking at Guglielmo Reiss Romoli School (1996, Italy), she worked in Telecom Italia (1995–2000) as responsible of industrial projects. From April 2000 to September 2005 she worked as senior research in ftw Austria on CDMA and MIMO systems. She was research fellow in INRIA (Sophia Antipolis, France) from October to December 2005 and at the University of South Australia, Australia in 2006. Since December 2006 she is assistant professor in EURECOM. Cottatellucci is currently associate editor for IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE TRANSACTIONS ON SIGNAL PROCESSING and served as guest editor for EURASIP JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING (special issue on cooperative communications). Her research interests lie in the areas of large system analysis and algorithm design for multiuser wireless communications and complex networks, random matrix theory, and game theory.

David Gesbert (F'11) is Professor and Head of the Communication Systems Department, EURECOM. He obtained the Ph.D degree from Ecole Nationale Superieure des Telecommunications, France, in 1997. From 1997 to 1999 he has been with the Information Systems Laboratory, Stanford University. He was then a founding engineer of Iospan Wireless Inc, a Stanford spin off pioneering MIMO-OFDM (now Intel). Before joining EURECOM in 2004, he has been with the Department of Informatics, University of Oslo as an adjunct professor. D. Gesbert has published about 270 papers and 25 patents, some winning the 2015 IEEE Best Tutorial Paper Award (Communications Society), 2012 SPS Signal Processing Magazine Best Paper Award, 2004 IEEE Best Tutorial Paper Award (Communications Society), 2005 Young Author Best Paper Award for Signal Proc. Society journals, and paper awards at conferences 2011 IEEE SPAWC, 2004 ACM MSWiM. He is a Technical Co-chair for ICC2017 in Paris and WSA2017 in Berlin. He was named in the 2014 Thomson-Reuters List of Highly Cited Researchers in Computer Science. Since 2015, he holds the ERC Advanced grant PERFUME on the topic of smart device Communications in future wireless networks.