Limited Feedback Designs for Machine-type Communications Exploiting User Cooperation

JIHO SONG¹, (Member, IEEE), BYUNGJU LEE², (Member, IEEE), SONG NOH³, (Member, IEEE) and JONG-HO LEE¹, (Member, IEEE)

¹School of Electrical Engineering, University of Ulsan, Ulsan, South Korea (e-mail: jihosong@ulsan.ac.kr)
²Samsung Research, Seoul, South Korea (e-mail: byungjulee1730@gmail.com)
³Department of Information and Telecommunication Engineering, Incheon National University, Incheon, South Korea (e-mail: songnoh@inu.ac.kr)

Corresponding author: Byungju Lee (e-mail: byungjulee1730@gmail.com).

This work was supported by the 2018 Research Fund of University of Ulsan.

ABSTRACT Multiuser multiple-input multiple-output (MIMO) systems are a prime candidate for use in massive connection density in machine-type communication (MTC) networks. One of the key challenges of MTC networks is to obtain accurate channel state information (CSI) at the access point (AP) so that the spectral efficiency can be improved by enabling enhanced MIMO techniques. However, current communication mechanisms relying upon frequency division duplexing (FDD) might not fully support an enormous number of devices due to the rate-constrained limited feedback and the time-consuming scheduling architectures. In this paper, we propose a user cooperation-based limited feedback strategy to support high connection density in massive MTC networks. In the proposed algorithm, two close-in users share the quantized version of channel information in order to improve channel feedback accuracy. The cooperation process is performed without any transmitter interventions (i.e., in a grant-free manner) to satisfy the low latency requirement that is vital for MTC services. Moreover, based on the sum-rate throughput analysis, we develop an adaptive cooperation algorithm with a view to activating/deactivating the user cooperation mode according to channel and network conditions.

INDEX TERMS Limited feedback, machine type communications, massive connectivity, multiuser diversity, user cooperation

I. INTRODUCTION

INTERNET of things (IoT), which refers to the connected future world in which every mobile device and machines are linked to the internet via wireless link, has received attention from both academia and industry in recent years [1]. IoT enables a wide range of unprecedented services such as autonomous driving, smart home/factory, and factory automation, just to name a few [2]. Massive connectivity is one of the most important requirements of a fully connected IoT society [3], [4]. In accordance with this trend, the international telecommunication union (ITU) defined massive machine-type communication (mMTC) as one representative service category.¹ In mMTC networks, data communications may occur between an MTC device and a server or directly between MTC devices [6]. It is of considerable importance to support high connection density with limited resources because the number of devices is at least two orders of magnitude higher than current human-centric communication.

From a technological standpoint, enormous number of devices in mMTC networks can be used to exploit full benefit of multiuser MIMO. It is essential to have high-resolution channel state information (CSI) at the access point (AP) to exploit multiuser diversity gain [7], [8]. In most current cellular systems relying upon frequency division duplexing (FDD), the quantized CSI is communicated to the AP via a rate-constrained feedback link [9]–[11]. One challenge of feedback-assisted multiuser systems is that low-resolution CSI overrides the multiuser diversity gain because the signal-to-interference-plus-noise ratio (SINR) is limited due to the

¹Three representative service categories include enhanced mobile broadband (eMBB), ultra-reliable and low latency communication (uRLLC), and massive machine-type communication (mMTC) [5].
channel quantization error [12], [13]. In the feedback-assisted multiuser systems, the rate-constrained feedback mechanism is the biggest obstacle to supporting a massive number of devices on an MTC network.

Antenna combining techniques, e.g., quantization-based combining (QBC) [14], has been proposed to obtain high-resolution CSI. A key feature of the QBC is that receive antennas are combined to generate an effective channel that can be quantized accurately. Employing more antennas would enhance the quantization performance. However, direct application of antenna combining techniques for mMTC is infeasible since it is difficult to employ multiple antenna elements due to the strict budget constraints on small-scale devices. In this paper, we develop a cooperative feedback strategy to obtain an additional spatial dimension for the antenna combining process without imposing an additional burden on mMTC devices.

Recently, multiuser systems incorporating user cooperation algorithms have been proposed [15], [16]. In [15], the user in the cooperative link helps other adjacent users by forwarding adjacent users’ information while achieving its own quality of service (QoS). In [16], the users in the cooperative link exchange their CSI via device-to-device (D2D) communications. The users can compute a more appropriate precoder at the receiver side because CSI exchange allows users to obtain the global CSI at the cost of increased control signalling overhead. However, it is not feasible to exchange CSI with all the other users because the number of users for the mMTC network is much larger than that of current human-centric communication. Furthermore, the downlink feedback and data exchange channels would be subject to more stringent overhead constraints as compared to current cellular systems. For these reasons, it is necessary to reevaluate multiuser MIMO techniques to handle massive connection density in mMTC networks with a minimal overhead for feedback and/or cooperative links.

The aim of this paper is to develop user cooperation strategies for an mMTC network allowing only point-to-point CSI exchange between close-in users. The rate-constrained feedback link is considered to tackle the overhead issue in FDD-based cellular systems. Only few bits are exploited to exchange CSI in order to obtain high-resolution CSI with a minimal burden on the cooperation framework. Moreover, the user cooperation process is designed to work without transmitter interventions to reduce network latency that control signal transmissions are involved in. To the best of our knowledge, a user cooperation strategy designed to reduce channel quantization error has been proposed here for the first time. The main contributions are presented as follows:

- **Cooperative limited feedback:** Adjacent users are connected to a cooperation link and these users are considered one cooperation unit (CU). Each user in CU shares the other user’s local channel information (i.e., local channel direction information (CDI) and channel quality indicator (CQI)). CSI sharing is only allowed between users in a CU. Each user generates the global channel information required for downlink transmission (i.e., global CDI and CQI) using its own channel information and the local channel information received from an adjacent user. After exchanging each other’s global CQI, the user having larger global CQI is assigned as the main user (MU) and the other user is assigned as an assistant user (AU). The AU acts as an assistant for MU by allowing MU to use its receive antennas to construct the global CDI more precisely. The MU only feeds back the global CSI so that the AP perceives the MU as the sole active user, while the AU is transparent to the AP. In the data transmission phase, both the MU and AU receive data information from the AP and then the AU forwards the received signal to the MU. The downlink channel feedback accuracy of the MU is improved due to the virtue of exploiting AU resources.

- **Automatic role assignment:** Identification of the MU and AU is an important issue since only the spectral efficiency of the MU can be increased by sacrificing the resources of the AU. In the proposed algorithm, the cooperation process between users is designed to occur without transmitter intervention (i.e., a grant-free environment) through an active decision process. An important issue behind this active decision process is the motivation for participating in cooperative communication as the AU. Under a grant-free environment, the AP regards the MU as the sole user and this identification process is transparent to the AP.

- **Adaptive cooperative feedback:** If cooperative feedback is activated, the number of active users is reduced by half because two users are combined as a single CU to obtain high-resolution CSI. Unless a massive number of users are active in the mMTC network, the cooperative feedback strategy might not be an effective solution because the multiuser diversity gain is highly limited in a small-user regime. For this reason, effective allocation of limited multiuser resources is required to obtain accurate CSI without loss of multiuser diversity gain. We analyze the sum-rate throughput of the multiuser MIMO systems relying upon the proposed cooperation algorithm. Based on the analytical studies, we develop cooperation mode-switching criteria to activate/deactivate the cooperation mode according to channel and network conditions.

In Section II, we introduce a multiuser MIMO system and review a user selection algorithm. In Section III, we present the proposed cooperative feedback algorithm. An adaptive cooperation algorithm is developed based on analytical studies on sum-rate throughput in Section IV. In Section V, we present numerical results to verify the performance of the proposed scheme. We conclude the paper in Section VI.

2The users can be any kinds of machines, devices, and mobile users.

3One possible option can be the social relationship between users [15]. If users have a close relationship in the social domain, users can readily help each other by using their own resources for cooperative feedback. Other possible scenario is the cooperation between different devices in each user.
Throughout this paper, \( \mathbb{C} \) denotes the field of complex numbers, \( \mathcal{C}(m, \sigma^2) \) denotes the complex normal distribution with mean \( m \) and variance \( \sigma^2 \), \( 0_{a,b} \) is the \( a \times b \) all zeros matrix, \( I_M \) is the \( M \times M \) identity matrix, \( \chi^2 \) is the Chi-squared random variable, \( \beta(\cdot, \cdot) \) is the Beta-distributed random variable, \( \mathbb{B}(\cdot, \cdot) \) is the Beta function, \( \Gamma(\cdot) \) is the gamma function, \( \binom{m}{\ell} \) is the binomial coefficient, \( (z)_{k} \) is the Pochmann symbol, \( \lceil \cdot \rceil \) is the ceiling function, \( \| \cdot \|_{p} \) is the \( p \)-norm, and \([a]_{\ell} \) is the \( \ell \)-th element of the column vector \( a \). Also, \( \mathbf{A}_{:, m}, \mathbf{A}^{\dagger}, \mathbf{A}^{H}, \text{Tr}(\mathbf{A}) \), and \( \mathbf{A}_{a,b} \) denote \( m \)-th column vector, pseudo-inverse, conjugate transpose, trace, and \((a, b)\)-th entry of the matrix \( \mathbf{A} \), respectively.

II. SYSTEM MODEL AND BACKGROUND

We briefly review the FDD-based multiuser MIMO systems. We first present the system model and then discuss antenna combining-based limited feedback. An overview of conventional multiuser MIMO systems is depicted in Fig. 1(a).

A. SYSTEM MODEL

We consider multiuser MIMO systems employing \( M \) transmit antennas at the AP and \( N \) receive antennas at each of \( K \) users. Assuming a block-fading channel, an input-output expression for the \( k \)-th user is defined as

\[
y_k = z_k^H (\sqrt{\rho} \mathbf{h}_k x + n_k),
\]

where \( y_k \in \mathbb{C} \) is the received baseband signal, \( \rho \) is the signal-to-noise ratio (SNR), \( z_k \in \mathbb{C}^N \) is the unit-norm combiner,

\[
\mathbf{H}_k = \begin{bmatrix} h_{k,1}^1, \ldots, h_{k,N}^1 \end{bmatrix}^H \in \mathbb{C}^{N \times M}
\]

is the MIMO channel matrix, \( h_{k,1}^n \in \mathbb{C}^M \) is the channel vector between the AP and the \( n \)-th receive antenna consisting of independent and identically distributed (i.i.d.) entries following \( \mathcal{C}(0, 1) \), \( x \in \mathbb{C}^M \) is the transmit signal vector (with the power constraint \( \mathbb{E}[\|x\|^2] \leq 1 \)), and \( n_k \in \mathbb{C}^N \) is the additive noise vector with entries following \( \mathcal{C}(0, 1) \). We note that the user index is subscripted and the set of user indices in the network is written by \( \mathcal{K} = \{1, \ldots, K\} \).

We consider a single layer data transmission for each user. The transmit signal vector is rewritten as \( x = \mathbf{F}s \), where

\[
\mathbf{F} = \frac{1}{\sqrt{M}} \begin{bmatrix} f_1, \ldots, f_M \end{bmatrix} \in \mathbb{C}^{M \times M}
\]

is the precoder and \( s = [s_1, \ldots, s_M]^T \in \mathbb{C}^M \) is the transmit symbol vector. Note that \( f_m \in \mathbb{C}^M \) and \( s_m \in \mathbb{C} \) denote the transmit beamformer and the data stream for the \( m \)-th scheduled user with the power constraints \( \|f_m\|^2 = 1 \) and \( \mathbb{E}[|s_m|^2] \leq 1 \).

In FDD-based cellular systems, an AP obtains channel information through receiver feedback from each user. In feedback-assisted MIMO architectures, channel vectors are quantized using the predefined codebook

\[
\mathcal{C} = \{e^1, \ldots, e^Q\}.
\]

where \( Q \) is the number of codewords in the global codebook. To facilitate the multiuser signaling framework, the quantized channel information is fed back to the AP via a rate-constrained feedback link.

We employ the opportunistic random beamforming approach that utilizes a set of unitary matrices to construct the global codebook [7], [17]. Similar to the LTE-Advanced codebooks in [18], [19], exploiting more sets of unitary matrices enables the AP to obtain high-resolution CSI. Furthermore, the new radio (NR) codebooks in [20] significantly increases the channel quantization performance with measurably large (oversampled) discrete Fourier transform (DFT) codebooks. However, an ultra low latency requirement for MTC services restricts the use of the large channel quantization codebooks. Assuming an intensely rate-constrained feedback link, we consider \( M \) codewords for channel feedback and beamforming, meaning that the number of codewords equals to the number of transmit antennas. If only \( M \) codewords are allowed for CSI quantization, it would be optimal to consider a set...
of orthonormal vectors for random beamforming [21]. We use a single unitary matrix \( C \) for defining codewords \( c^m \). We denote the combined noise. The SINR of the \( k \)-th user is given by

\[
\gamma_k = \frac{\| \tilde{h}^\text{eff}_k c^m \|^2}{\rho + \sum_{\ell=1, \ell \neq k}^M \| \tilde{h}^\text{eff}_k c^\ell \|^2}
\]

Finally, the unit-norm combiner is computed according to

\[
\bar{u}_{k|m} = \frac{\tilde{u}_{k|m}}{\| \tilde{u}_{k|m} \|_2} \in \mathbb{C}^N.
\]

B. QBC-BASED LIMITED FEEDBACK FOR MULTIUSER MIMO

One major issue of the feedback architecture using \( M \) codewords is that the channel quantization performance is expected to be poor in a single receive antenna scenario. When multiple receive antennas are available, antenna combining techniques can be applied to enhance the channel quantization performance [14], [22]. We consider a QBC-based antenna combining algorithm [14]. The objective of the QBC is to compute an effective channel vector that can be quantized accurately using a small-sized codebook. For a given target codeword \( c^m \), the receive combiner \( \bar{z}_{k|m} \) is computed such that a cross-correlation between the effective channel

\[
\bar{h}^\text{eff}_k c^m = H^H_k z_{k|m}
\]

and the target codeword is maximized. As discussed in [14], the effective channel that maximizes the cross-correlation is obtained by projecting \( c^m \) onto the channel subspace such as

\[
c_{\text{proj}, m} = \left[ q^1, \ldots, q^N \right] \left[ q^1, \ldots, q^N \right]^H c^m \in \mathbb{C}^M,
\]

where \( q^\ell \) is the orthonormal basis that spans \( H_k \). Using the QBC algorithm, the receive combiner \( \bar{u}_{k|m} \), which satisfies \( H_k^H \bar{u}_{k|m} = c_{\text{proj}, m} \), can be computed by multiplying the pseudo-inverse of the channel matrix such as

\[
\begin{align*}
(H_k H_k^H)^{-1} H_k H_k^H \bar{u}_{k|m} &= (H_k H_k^H)^{-1} H_k c_{\text{proj}, m}. \\
\text{Finally, the unit-norm combiner is computed according to}
\end{align*}
\]

\[
\bar{z}_{k|m} = \frac{\bar{u}_{k|m}}{\| \bar{u}_{k|m} \|_2} \in \mathbb{C}^N.
\]

Assuming the user \( k \) is scheduled to use the \( m \)-th beamformer (codeword), the received signal is written by

\[
y_{k|m} = \sqrt{\rho/\bar{m}} H^H_k \left[ c^1, \ldots, c^M \right] s + \bar{z}_{k|m} n_k
\]

\[
= \sqrt{\frac{\rho}{M}} \left( \bar{h}^\text{eff}_k c^m s^m + \sum_{\ell=1, \ell \neq m}^M \bar{h}^\text{eff}_k c^\ell s^\ell \right) + \bar{n}_{k|m},
\]

because the \( m \)-th beamformer is identical to the \( m \)-th codeword such that \( f^m = c^m \) in our random beamforming architecture. Notice that \( \bar{n}_{k|m} = \bar{z}_{k|m} n_k \sim \mathcal{CN}(0,1) \) denotes the combined noise. The SINR of the \( k \)-th user is then defined by

\[
\gamma_k = \frac{\| \bar{h}^\text{eff}_k c^m \|^2}{\rho + \sum_{\ell=1, \ell \neq m}^M \| \bar{h}^\text{eff}_k c^\ell \|^2}.
\]

III. COOPERATIVE LIMITED FEEDBACK ARCHITECTURE

One of the key challenges in developing feedback-assisted multiuser systems for mMTC is to obtain accurate CSI at the AP while overcoming the following restrictions:

- Ultra low latency requirement for mMTC restricts the use of large codebook and this makes it difficult to achieve robust channel quantization performance.

The index of the selected codeword is dropped for the rest of the sections.
Considering the strict budget constraints of small-scale devices, it is not practical to employ a multitude of antenna elements for the antenna combining. In QBC, the resolution of an effective channel increases as the spatial dimensions of a channel matrix for an antenna combining increases [14]. In this paper, we propose user cooperation strategies to obtain high-resolution CSI while limiting the number of antenna elements at the receiver. The main feature of the proposed algorithm is that users in a CU employ an additional spatial dimension for the antenna combining by allowing a limited amount of CSI exchange. The key difference between the conventional system in Fig. 1(a) and the proposed system in Fig. 1(b) is the existence of Wi-Fi peer-to-peer sidelink, as presented in [26]. Furthermore, an aim of this step is to compute the quantized version of variables in the proposed cooperative feedback architecture: combining by allowing a limited amount of CSI exchange.

The term local is used to designate the variables and signal processing operations within CU, while the term global is used to designate the variables and signal processing operations within AU.

The bar symbol \( \bar{\cdot} \) will be used to highlight variables corresponding to the global signal processing operation.

The tilde symbol \( \tilde{\cdot} \) will be used to highlight variables corresponding to the downlink transmission.

We present the details of the cooperative feedback algorithm based on the assumption that two close-in users have already been combined as a single CU according to \( \{a, b\} \). An overview of the cooperative feedback is depicted in Fig. 2.

One viable approach to achieve reduction of both local CSI quantization error and cooperation link overhead is to share only a single effective channel that is combined based on the QBC algorithm [14]. The combined channel is quantized using the random vector quantization (RVQ) codebook

\[
D = \{d^1, \ldots, d^{Q_{cl}}\},
\]

which consists of \( Q_{cl} = 2^{2\beta_{cl}} \) codewords. The distance of the cooperation link is much shorter than that of the feedback link. The cooperation link would be subject to less stringent overhead constraints compared to that for the feedback link such that \( B_{cl} \gg B \).

First, user \( k \in \{a, b\} \) in the CU computes a virtual vector

\[
h_{k,q}^{\text{virt}} = H_k^H z_{k,q}
\]

that can be quantized more accurately using a target codeword \( \hat{d} \in D \), where the local combiner \( z_{k,q} \in \mathbb{C}^N \) and the projected codeword \( d^{\text{proj},q} \in \mathbb{C}^M \), needed for the local channel combining, are computed using similar method in (5) and (6).

Second, each user selects a local codeword that maximizes the cross-correlation of the normalized virtual channel vector and the codeword, i.e.,

\[
\cos^2 \phi_{k|q} = \left| (d^\hat{q})^H h_{k,q}^{\text{virt}} / \| h_{k,q}^{\text{virt}} \|_2 \right|^2,
\]

where \( \phi_{k|q} \) is the difference in angle. We call the selected codeword local CDI and its corresponding cross-correlation coefficient local CQI. The local CDI and CQI are given by

\[
(v_k, \tau_k) = (d^\hat{q}, \| h_{k,q}^{\text{virt}} \|_2 \cos \phi_{k|q}),
\]

where the index of the selected codeword is

\[
\hat{q} = \arg \max_{q \in \{1, \ldots, Q_{cl}\}} \cos^2 \phi_{k|q}.
\]

Under the assumption the index of the selected codeword is dropped, the selected local combiner, the virtual channel vector, and the difference in angle can be rewritten as

\[
(z_k, h_{k}^{\text{virt}}, \phi_{k}) = (z_{k|\hat{q}}, h_{k|\hat{q}}^{\text{virt}}, \phi_{k|\hat{q}}).
\]

The quantized virtual channel vector is then defined with the local CDI and CQI according to

\[
\hat{h}_{k}^{\text{virt}} = \tau_k v_k = \| h_{k}^{\text{virt}} \|_2 \cos \phi_k v_k.
\]
Finally, users in the CU exchange the local CDI and CQI, i.e., quantized virtual channel vector $\mathbf{h}_{k}^{\text{virt}}$ in (15), with its cooperation partner via a $B^3$-bits cooperation link. The local CDI and CQI will be included in a global channel matrix of its cooperation partner. Assuming the user $k$ is the AU $b$, the local combining and channel quantization, and channel sharing processes are depicted in the left side of Fig. 2.

2) Global CSI acquisition & Role assignment

An aim of this step is to compute global channel information and to assign the role of MU and AU. The global CSI will be fed back to the AP.

First, each user constructs a global channel matrix

$$
\mathbf{H}_k^{\text{qu}} = \left[ \begin{array}{c} \mathbf{H}_k \\ \mathbf{h}_{k}^{\text{virt}} \end{array} \right]^H \in \mathbb{C}^{(N+1)\times M},
$$

(16)

which includes one’s own channel matrix $\mathbf{H}_k$ and the quantized virtual channel vector $\mathbf{h}_{k}^{\text{virt}}$ from a cooperation partner, where $k \in \{a, b\} \setminus \{k\}$. Each user regards the virtual channel vector as an additional channel vector between the AP and the virtual antenna element at the receiver. Assuming oneself is selected as MU, each user computes the global effective channel vector

$$
\mathbf{h}_{k|m}^{\text{eff,qu}} = (\mathbf{H}_k^{\text{qu}})^H \mathbf{z}_{k|m}
$$

(17)

that can be quantized accurately with a target codeword $c^m \in \mathcal{C}$. The global combiner $\mathbf{z}_{k|m} \in \mathbb{C}^{N+1}$ is computed using the combining method in (5) and (6). The difference with the combining process in (11) is that the rank of the channel and the dimension of the combiner are increased to $N+1$.

Second, each user selects a single global codeword that maximizes the SINR

$$
\gamma_{k|m} = \frac{|(\mathbf{h}_{k|m}^{\text{eff,qu}})^H \mathbf{c}^m|^2}{\rho + \sum_{l=1, l \neq m}^{M} |(\mathbf{h}_{k|m}^{\text{eff,qu}})^H \mathbf{c}^l|^2}.
$$

We call the selected codeword global CDI and its corresponding SINR global CQI. The global CDI and CQI are

$$\left( \bar{v}_k, \bar{\gamma}_k \right) = \left( \mathbf{c}^m, \gamma_{k|m} \right), \quad (18)$$

where the index of the selected codeword is $m = \arg\max_m \gamma_{k|m}$. Assuming the index of the selected codeword is dropped, the selected global combiner and the effective channel vector are rewritten according to

$$\left( \bar{z}_k, \mathbf{h}_k^{\text{eff,qu}} \right) = \left( \mathbf{z}_{k|m}, \mathbf{h}_{k|m}^{\text{eff,qu}} \right). \quad (19)$$

Third, users in the CU exchange their global CQIs with cooperating users via a cooperation link to determine who’s best for maximizing the data-rate throughput. The user having a larger CQI is assigned to MU and the unselected user is assigned to AU. The role assignment are made at the beginning of the transmission frame and will continue for the duration of the channel coherence time. We assume that the $a$-th (odd number indexed) user is assigned to MU and the $b$-th (even number indexed $b = a + 1$) user is assigned to AU. The set of indices of MUs is then written by $\bar{K} = \{1, 3, \cdots, K - 1\}$.

Algorithm 1 Cooperative feedback algorithm

1) Local CSI acquisition

1. Compute local combiner and virtual channel vector

$$
\mathbf{z}_{k|q} = \frac{(\mathbf{H}_k + \mathbf{H}_k^{\text{qu}}\mathbf{d}^\text{drop})\mathbf{a}^q_{\text{delay}}}{\|\mathbf{H}_k + \mathbf{H}_k^{\text{qu}}\mathbf{d}^\text{drop}\|^2/2}, \quad \mathbf{h}_{k|q}^{\text{virt}} = \mathbf{H}_k^H \mathbf{z}_{k|q}
$$

2. Select local CDI and CQI ($\bar{v}_k, \bar{\gamma}_k$)

3. Exchange local CDI and CQI $\mathbf{h}_{k|q}^{\text{virt}} = \bar{\tau}_k \mathbf{v}_k$

2) Global CSI acquisition & Role assignment

4. Construct global channel matrix $\mathbf{H}_k^{\text{qu}} = [\mathbf{H}_k^H, \mathbf{h}_{k}^{\text{virt}}]^H$

5. Compute global combiner and effective channel vector

$$
\mathbf{z}_{k|m} = \frac{(\mathbf{H}_k^H + \mathbf{H}_k^{\text{qu}}\mathbf{d}^\text{drop})\mathbf{a}^\text{m} \mathbf{H}_\ell^H}{\|\mathbf{H}_k^H + \mathbf{H}_k^{\text{qu}}\mathbf{d}^\text{drop}\|^2/2}, \quad \mathbf{h}_{k|m}^{\text{eff,qu}} = \left( \mathbf{H}_k^H \right)^H \mathbf{z}_{k|m}
$$

6. Select global CDI and CQI ($\bar{v}_k, \bar{\gamma}_k$)

7. Exchange global CQI

8. Assign MU having a larger global CQI

9. MU reports global CDI and CQI to the AP

3) User scheduling

10. Schedule $M$ MUs $U = \{ \bar{u}_1, \cdots, \bar{u}_M \}$

4) Post-signal processing

11. Save received signals $y_k$

12. AU combines received signals $y_b = \mathbf{z}_b^H y_b$

13. AU reports $y_b$ to MU

14. MU obtains virtual received signals $\mathbf{y}_a = [\mathbf{y}_a^T, \mathbf{y}_b^T]^T$

15. MU combines received signals $\mathbf{\hat{y}}_a = \overline{\mathbf{z}_a^H} \mathbf{y}_a$

Finally, MU transmits the global CSI to the AP via a feedback link. The number of active users in the network is reduced by half $K/2$. Assuming the user $k$ is the MU $a$, the global combining, channel quantization, and channel feedback processes are depicted in the right side of Fig. 2.

3) User scheduling

After collecting global CDI and CQI from $K/2$ MUs, the AP schedules $M$ MUs such that

$$
\bar{U} = \{ \bar{u}_1, \cdots, \bar{u}_M \},
$$

where $\bar{u}_m$ denotes the scheduled MU exploiting the $m$-th codeword $c^m$. The $m$-th scheduled MU is given by

$$
\bar{u}_m = \arg\max_{k \in U^m} \bar{\gamma}_k,
$$

where $U^m$ denotes the set of indices of MUs who choose $c^m$ as their global CDI.

4) Post-signal processing

An aim of this step is to decode the received signals

$$
\mathbf{y}_k = \sqrt{\rho} \mathbf{H}_k \mathbf{x} + \mathbf{n}_k \in \mathbb{C}^{N}, \quad k \in \{a, b\}.
$$

First, AU combines the received signal with the local combiner

$$
\mathbf{y}_b = \sqrt{\rho} \mathbf{z}_b^H \mathbf{H}_b \mathbf{x} + \mathbf{n}_b = \sqrt{\rho} (\mathbf{h}_{b|q}^{\text{virt}})^H \mathbf{x} + \mathbf{n}_b \in \mathbb{C},
$$

$^6$A memory in the receiver allows retention of received signals. The post-signal processing has no effect on a variation of the sum-rates because this process is conducted using the received signals stored in memory.
The proposed cooperative feedback algorithm exploits some multiuser resources to improve channel quantization performance. High-resolution CSI can be obtained because more antenna elements (spatial dimensions) are used for global antenna combining. However, the proposed approach would restrict options that can be used to improve a network throughput due to the following reasons: First, the squared norm of the global effective channel vector decreases as the number of antennas used for a combining process increases [14, Lemma 3]. Second, the sum-rate grows like $M \log_2(\log K/2)$ because user candidates are reduced by half.

In this section, we develop an analytical framework weighing the pros and cons of the proposed cooperative feedback algorithm. Based on the analytical framework, an adaptive cooperation algorithm is proposed in order to activate/deactivate the proposed cooperation strategy according to channel and network conditions.

## IV. ADAPTIVE COOPERATION ALGORITHM

The proposed cooperative feedback algorithm exploits some of the pros of the proposed cooperation strategy according to channel and network conditions.

### A. LOSS IN LOCAL CHANNEL QUANTIZATION

Before investigating the received signal of MU, we pause to analyze the channel quantization error induced in the process of local combining. As discussed in Section II-B, the channel quantization error between the normalized virtual channel vector $h^\text{virt}_{b/q}$ and the target codeword $d^q$ is quantified by

$$S = \sin^2 \phi_{b/q} = 1 - \left| (d^q)^H h^\text{virt}_{b/q} / \|h^\text{virt}_{b/q}\|_2 \right|^2,$$

(25)

where $\phi_{b/q}$ is the difference in angle defined in (12). It is verified in [14] that each error follows $\beta(M-N,N)$ random variable and its cumulative distribution function (cdf) can be approximated for small $s$, with $\delta = (M-N-1)\frac{s}{M-N}$, according to

$$F_S(s) \approx \left(\frac{M-1}{N-1}\right) s^{M-N} \mathbb{I}_{[0,\delta]}(s) + \mathbb{I}_{(\delta,1]}(s).$$

(26)

The local CDI is obtained by selecting the codeword corresponding to the smallest quantization error from among $Q^\text{el}$ error terms $\sin^2 \phi_{b/q}$, $q \in \{1, \cdots, Q^\text{el}\}$. The channel quantization error corresponding to the local CDI can be studied by deriving the distribution of the smallest quantization error

$$\sin^2 \hat{\phi}_b = \sin^2 \phi_{b/q},$$

where the index of the codeword in (13) is rewritten by

$$\hat{q} = \arg\min_{q \in \{1, \cdots, Q^\text{el}\}} \sin^2 \phi_{b/q},$$

(27)

Based on the largest order statistics, we derive the expectation of the smallest quantization error in the following proposition.

**Proposition 1**: The expectation of the channel quantization error corresponding to the local CDI is approximated by

$$E[\sin^2 \phi_b] \approx Q^\text{el} \left(\frac{M-1}{N-1}\right)^{\frac{1}{1+\frac{1}{M-N}}} B\left(Q^\text{el}, \frac{M-N+1}{M-N}\right).$$

Proof: Minimizing the quantization error is the same as maximizing the normalized beamforming gain

$$C = \cos^2 \phi_{b/q} = \left| (d^q)^H h^\text{virt}_{b/q} / \|h^\text{virt}_{b/q}\|_2 \right|^2.$$ 

The cdf of the normalized beamforming gain is given by

$$F_C(c) = 1 - F_{S}(1-c) \simeq 1 - \left(\frac{M-1}{N-1}\right) (1-c)^{M-N} \mathbb{I}_{[0,\delta]}(1-c) + \mathbb{I}_{(\delta,1]}(1-c)$$

\[\text{(a)}\]

$$= 1 - \left(\frac{M-1}{N-1}\right) (1-c)^{M-N} \mathbb{I}_{[1-\delta,1]}(1-c) + \mathbb{I}_{[0,1-\delta]}(c)$$

\[\text{(b)}\]

$$= \left(\frac{M-1}{N-1}\right) (1-c)^{M-N} \mathbb{I}_{[1-\delta,1]}(c),$$

where (a) is derived because $\mathbb{I}_{[0,\delta]}(1-c) = \mathbb{I}_{[1-\delta,1]}(c)$,
Numerical results

Quantization error, $E[\sin^2 b]$ (where $b$ is derived based on [29, 6.6.8]).

The probability that the largest beamforming gain is $g \approx E[G]$ is derived such that

$$ Pr\{\cos^2 \phi_b < g\} = \prod_{q=1}^{Q^cl} Pr\{\cos^2 \phi_{b|q} < g\}. $$

Therefore, the cdf of $G$ is defined with the cdf of $C$ such as

$$ F_G(g) = (F_C(g) Q^cl) \cdot \mathbb{I}_{[1,\delta,1]}(g). $$

Because $G$ is a non-negative random variable, the expectation of $G$ can be derived such that

$$ E[G] \approx 1 - \int_0^1 F_G(g) \, dg $$

$$ = 1 - \delta \sum_{q=0}^{Q^cl} \frac{(-1)^q}{q+1} \frac{1}{M-N} q! \, (1-g)^q \frac{M-N}{N-1} \, Q^cl $$

$$ = 1 - \delta \sum_{q=0}^{Q^cl} \frac{(-1)^q}{q+1} \frac{1}{M-N} (1-g)^q \frac{M-N}{N-1} $$

$$ = \mathbb{I}_{[0,1]}(1-c) = \mathbb{I}_{[0,1-\delta]}(c), \text{ and } (b) \text{ is derived because } F_C(c) = 0 \text{ when } 0 \leq c < 1 - \delta. $$

The normalized beamforming gain corresponding to the selected codeword (local CDI) is written as

$$ G \equiv \cos^2 \phi_b = \max_{q \in \{1, \ldots, Q^cl\}} \cos^2 \phi_{b|q}. $$

The probability that the largest beamforming gain is smaller than an arbitrary number $g$ is defined according to

$$ Pr\{\cos^2 \phi_b < g\} = \prod_{q=1}^{Q^cl} Pr\{\cos^2 \phi_{b|q} < g\}. $$

Therefore, the cdf of $G$ is defined with the cdf of $C$ such as

$$ F_G(g) = (F_C(g) Q^cl) \cdot \mathbb{I}_{[1,\delta,1]}(g). $$

Because $G$ is a non-negative random variable, the expectation of $G$ can be derived such that

$$ E[G] \approx 1 - \int_0^1 F_G(g) \, dg $$

$$ = 1 - \delta \sum_{q=0}^{Q^cl} \frac{(-1)^q}{q+1} \frac{1}{M-N} (1-g)^q \frac{M-N}{N-1} $$

$$ = 1 - \delta \sum_{q=0}^{Q^cl} \frac{(-1)^q}{q+1} \frac{1}{M-N} (1-g)^q \frac{M-N}{N-1} $$

The expectation of the quantization error corresponding to the local CDI is approximated such that

$$ E[\sin^2 \phi_b] \approx Q^cl C \left(\begin{array}{c} M-N+1 \\ M-N \end{array}\right) B \left(\begin{array}{c} M-N+1 \\ M-N \end{array}\right) $$

because $\sin^2 \phi_b = 1 - \cos^2 \phi_b$. This completes the proof.

Finally, the accuracy of the quantization error in Proposition 1 and the formulation in [14], i.e.,

$$ \left(\begin{array}{c} M-1 \\ N-1 \end{array}\right) \frac{M-1}{M-N}, $$

are evaluated by numerical results in Fig. 3. It is shown that our formulation in Proposition 1 shows better accuracy than the formulation in [14]. Further, we point out that the quantization error is better fitted to the numerical results as the feedback bits $B^cl$ increase because the cdf in (26) is valid for small $s$.

**B. RECEIVED SIGNAL OF MU**

We take a closer look at the received signal of MU in (24). Assuming MU $a$ is scheduled to use the $m$-th beamformer $c^n_m$ (meaning that $\tilde{a}^m = a$), the received signal is written by

$$ y_{n|m} = \sqrt{\rho_{a|m}} \cdot \bar{H}_a \cdot \left(\begin{array}{c} c^1_m \\ \cdot \cdot \cdot \\ c^M_m \end{array}\right) + z^H_{a|m} \cdot \bar{n}_a $$

$$ = \sqrt{\frac{\rho}{M}} \left(\begin{array}{c} \tilde{H}^H_{a|m} \cdot c^m_s \cdot s^m \\ + \sum_{\ell=1,\ell \neq m}^{M} \left(\begin{array}{c} \tilde{H}^H_{a|m} \cdot c^\ell \cdot s^\ell \end{array}\right) \right) + \bar{n}_{a|m}, $$

where $\tilde{H}^H_{a|m} = \tilde{H}^H_{a|m} \cdot z_{a|m}$ is the global effective channel vector, and $\bar{n}_{a|m} = z_{a|m} \cdot \sqrt{C N(0,1)}$ is the combined noise. In order to examine beamforming gain and interuser interference, we must investigate the cross-correlations between the global effective channel vector $\tilde{H}^H_{a|m}$ and codewords in $C$.

The global effective channel vector is computed through two antenna combining processes (locally in AU and globally in MU) and each antenna combining process causes an individual quantization error. Before investigating both quantization errors jointly, we discuss each quantization error separately.

First, we discuss the channel quantization error caused in the process of virtual user quantization using the local codewbook $D$. Note that the local quantization error is presented in Section III-1. As illustrated in Fig. 4(a), the virtual channel vector $\tilde{h}_{b|v}^T$ in (21) is divided into the quantized virtual channel vector $\tilde{h}_{b|v}^T$ in (15) and the local error vector $e_b$ according to

$$ \tilde{h}_{b|v}^T = (||\tilde{h}_{b|v}^T||_2 \cos \phi_b) v_b + (||\tilde{h}_{b|v}^T||_2 \sin \phi_b) e_b $$

$$ = \tilde{h}_{b|v}^T + (||\tilde{h}_{b|v}^T||_2 \sin \phi_b) e_b, $$

where $\sin^2 \phi_b = 1 - ||\tilde{h}_{b|v}^T||_2^2$ quantifies the local quantization error. Second, we discuss the channel quantization error caused in the process of effective vector quantization using the global codewbook $C$. Note that the global quantization error is presented in Section III-2. As depicted in Fig. 4(b), the effective channel vector $\tilde{h}_{a|m}^T$ in (17) is divided into the target codeword $c^n$ and the global error vector $e_{a|m}$ such that

$$ \tilde{h}_{a|m}^T = (||\tilde{h}_{a|m}^T||_2 \cos \theta_{a|m}) c^n $$

$$ + (||\tilde{h}_{a|m}^T||_2 \sin \theta_{a|m}) e_{a|m}, $$

where $\cos^2 \phi_{b|q} = 1 - \cos^2 \phi_{b|q}$. This completes the proof.
where $\sin^2 \bar{\theta}_{a|m} = 1 - \| (c_m) H \bar{h}_{a|m}^{\text{eff,qu}} / \| \bar{h}_{a|m}^{\text{eff,qu}} \|^2$ quantifies the global quantization error.

We now consider both quantization errors jointly. When MU conducts post-signal processing, the global combiner $\mathbf{z}_{a|m}$ is used to combine spatial dimensions of the global channel matrix $\mathbf{H}_g$ in (23) for downlink transmissions. We call $\mathbf{H}_a$ the downlink channel matrix. We noted that the downlink channel matrix $\mathbf{H}_b$ includes the unquantized virtual vector $\bar{h}_b^{\text{virt}}$ in (14). However, the global combiner is computed using another global channel matrix $\mathbf{H}_g$ in (16) including the quantized virtual vector $\bar{h}_b^{\text{qu}}$ in (15). The correspondence between the downlink channel matrix and the global channel matrix must be investigated because the combined quantization error occurs due to the difference between $\mathbf{H}_a$ and $\mathbf{H}_g$.

First, we rewrite the downlink channel matrix in (23) by plugging the virtual channel vector in (29) according to

$$\mathbf{H}_a = \begin{bmatrix} \mathbf{H}_a^H H \bar{h}_b^{\text{virt}} \\ \bar{h}_b^{\text{virt}} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{H}_a^H H \bar{h}_b^{\text{virt}} \\ \bar{h}_b^{\text{virt}} \end{bmatrix} = \mathbf{H}_a^{\text{qu}} + \begin{bmatrix} 0_{N,M} \\ \| \bar{h}_b^{\text{virt}} \|^2 \sin \phi_b \end{bmatrix} e_b^{\text{qu}} H.$$  (31)

As depicted in Fig. 4(c), the effective channel vector in (28), corresponding to downlink transmissions, is written by

$$\mathbf{h}_{a|m}^{\text{eff,qu}} = \mathbf{H}_a^{\text{qu}} \mathbf{z}_{a|m}$$

$$= (\mathbf{H}_a^{\text{qu}}) H \mathbf{z}_{a|m} + \begin{bmatrix} 0_{N,M} \\ \| \bar{h}_b^{\text{virt}} \|^2 \sin \phi_b \end{bmatrix} \bar{e}_b^{\text{qu}} \mathbf{z}_{a|m}$$

$$= \mathbf{h}_{a|m}^{\text{eff,qu}} + \mathbf{e}_{b|m}^{\text{qu}}.$$  (32)

where the effective channel vector $\mathbf{h}_{a|m}^{\text{eff,qu}} = (\mathbf{H}_a^{\text{qu}}) H \mathbf{z}_{a|m}$ is defined in (17) and the error vector is given by

$$\mathbf{e}_{b|m}^{\text{qu}} = \begin{bmatrix} \| \mathbf{z}_{a|m} N + 1 \| \bar{h}_b^{\text{virt}} \|^2 \sin \phi_b \end{bmatrix} e_b^{\text{qu}} \in \mathbb{C}^M.$$  (33)

We call $\mathbf{h}_{a|m}^{\text{eff,qu}}$ the downlink channel vector.

Second, we rewrite the downlink channel vector $\mathbf{h}_{a|m}^{\text{eff,qu}}$ by plugging the effective channel vector $\mathbf{h}_{a|m}^{\text{eff,qu}}$ in (30) such that

$$\bar{h}_{a|m}^{\text{eff}} = (\| \mathbf{h}_{a|m}^{\text{eff,qu}} \|^2 \sin \bar{\theta}_{a|m} \mathbf{e}_c^{a|m})^m$$

$$+ \begin{bmatrix} \| \bar{h}_{a|m}^{\text{eff,qu}} \|^2 \sin \bar{\theta}_{a|m} \mathbf{e}_c^{a|m} \\ \mathbf{e}_{b|m}^{\text{qu}} + \mathbf{e}_{b|m}^{\text{qu}} \end{bmatrix}$$  (34)

where $\| \bar{h}_{a|m}^{\text{eff,qu}} \|^2 \sin^2 \bar{\theta}_{a|m}$ in (a) denotes the beamforming gain for data transmissions, (b) denotes the global quantization error, and (c) denotes the local quantization error. The relationship between the codeword $c_m$ and the global effective vector $\mathbf{h}_{a|m}^{\text{eff}}$ is summarized in Fig. 4(c).

Finally, we rewrite the received signal of MU by plugging the downlink vector $\mathbf{h}_{a|m}^{\text{eff}}$ in (34) into the input-output expression in (28). As presented in (35), the desired signal (1), global (interuser) interferences (II), local (interuser) interferences (III), and noise (IV) are clearly distinguished.

### C. ANALYSIS OF SINR

We analyze the SINR for all MUs in the network. Similar to [13], the lower bound of SINR is defined in (36) because $\sum_{\ell=1}^{M} | e_{b|m}^{H} c^\ell |^2 \geq \sum_{\ell=1, \ell \neq m}^{M} | e_{b|m}^{H} c^\ell |^2$. Note that the equality holds when the local quantization error $\mathbf{e}_{b|m}$ is orthogonal to the target codeword $c_m$. We noted that the error vector $\mathbf{e}_{b|m}$ is replaced by (33) in order to show all the local interference terms clearly. To verify the second equality, we should recall that the unit-norm error vector $\mathbf{e}_{a|m}$ is orthogonal to the target codeword $c_m$. Since we consider a single unitary matrix for constructing the global codebook, the error vector is on

If global codebook consists of more than $M$ codewords, $| e_{b|m}^{H} c^\ell |^2$ can be modeled by $\beta (1, M - 2)$ because $e_{b|m}$ and $c^\ell$ independent and isotropically distributed on the $M - 1$ dimensional hyperplane orthogonal to $c_m$ [12].
SINR\_{a|m} \geq \frac{\rho}{M} \sum_{\ell=1}^{M} \left| \mathbf{h}^{\text{eff,qu}}_{a|m} \right|^2 \frac{1}{2} \cos^2 \hat{\theta}_{a|m} \left[ \mathbb{E} \left[ |\hat{r}_{a|m}|^2 \right] + \frac{\rho}{M} \left| \mathbf{h}^{\text{eff,qu}}_{a|m} \right|^2 \right. \\
+ \frac{\rho}{M} \left| \mathbf{z}_{a|m} \right|^2 \sum_{\ell=1, \ell \neq \ell'}^{M} \left( |\mathbf{e}^H_{\ell}| \mathbf{e}_{\ell'} \right)^2 + \frac{\rho}{M} \left| \mathbf{z}_{a|m} \right|^2 \sum_{\ell=1}^{M} \left| \mathbf{e}^H_{\ell} \mathbf{c}^\ell \right|^2 \\
= \frac{\rho}{M} \left| \mathbf{h}^{\text{eff,qu}}_{a|m} \right|^2 \frac{1}{2} \cos^2 \hat{\theta}_{a|m} \left[ 1 + \frac{\rho}{M} \left| \mathbf{h}^{\text{eff,qu}}_{a|m} \right|^2 \right. \\
+ \frac{\rho}{M} \left| \mathbf{z}_{a|m} \right|^2 \sum_{\ell=1, \ell \neq \ell'}^{M} \left( |\mathbf{e}^H_{\ell}| \mathbf{e}_{\ell'} \right)^2 \\
\left. + \frac{\rho}{M} \left| \mathbf{z}_{a|m} \right|^2 \sum_{\ell=1}^{M} \left| \mathbf{e}^H_{\ell} \mathbf{c}^\ell \right|^2 \right].

(36)

\begin{align*}
\text{(a) Cumulative distributions of local and global quan. errors} & \\
\text{(b) Cumulative distributions of local and global interferences} & \\
\end{align*}

FIGURE 5. CDF of local and global variables (M = 4, N = 2, B^4 = 8).

The hyperplane, which is made up of M − 1 orthonormal codewords that are orthogonal to \( \mathbf{e}^m \). The error vector is thus written as a linear combination of the orthonormal codewords as \( \mathbf{e}_{a|m} = \sum_{\ell=1, \ell \neq \ell'}^{M} (\mathbf{e}^H_{\ell}) \mathbf{e}_{\ell'} \). The sum of terms in (i) is computed according to \( \left| \mathbf{e}_{a|m} \right|^2 = \sum_{\ell=1, \ell \neq \ell'}^{M} |\mathbf{e}^H_{\ell} \mathbf{c}^\ell|^2 = 1 \), because \( \mathbf{c}^\ell \) and \( \mathbf{e}_{a|m} \) are unit-norm vectors. The third equality follows because the sum of terms in (ii) is computed as

\[ \sum_{\ell=1}^{M} |\mathbf{e}^H_{\ell} \mathbf{c}^\ell|^2 = \mathbf{e}^H (\mathbf{C}^H \mathbf{C}) \mathbf{e}_b = \left| \mathbf{e}_b \right|^2 = 1, \]

because \( \mathbf{C} = [\mathbf{c}^1, \ldots, \mathbf{c}^M] \in \mathbb{C}^{M \times M} \) is the unitary matrix.

Unfortunately, it is not easy to derive the joint distribution between the local and global error terms in (36). For purposes of analysis, it is necessary to obtain a reasonable criterion for approximating the expression of (36). First, we should point out that the size of the local codebook is much bigger than that of the global codebook such as \( B^{\text{cl}} \gg B \). Moreover, \( \sin^2 \phi_b \) is the minimum of \( Q^2 \) local error terms, while \( \cos^2 \hat{\theta}_{a|m} \) is an general global error term. For these reasons, the variance and magnitude of the local quantization error are smaller than those of the global quantization error, as shown in Fig. 5(a). It is also verified in Fig. 5(b) that the variance and magnitude of the local interference are small compared to those of the global interference. From these observations, we infer that the local interferences would have minimal effect on the SINR.

For tractability of the analysis, we consider an approximated SINR by evaluating an average squared magnitude of local interference. Note that the accuracy of this approximation is evaluated in the following paragraph. The expected value of local interference is given by

\[ \mathbb{E} \left[ \left| \mathbf{z}_{a|m} \right|^2 \right] \mathbb{E} \left[ \left| \mathbf{h}^\text{virt}_{b} \right|^2 \sin^2 \phi_b \right] \]

(\text{a}) \quad \frac{\rho}{M} \left| \mathbf{h}^\text{virt}_{b} \right|^2 \sin^2 \phi_b

\[ \frac{(M - N + 1) \mathbb{E} \left[ \sin^2 \phi_b \right]}{N + 1} \]

(\text{b}) \quad \frac{(M - N + 1) \omega}{N + 1} \]

(\text{c}) \quad \frac{(M - N + 1) \omega}{N + 1} \]

because \( \left| \mathbf{z}_{a|m} \right|^2, \left| \mathbf{h}^\text{virt}_{b} \right|^2 \), and \( \sin^2 \phi_b \) are independent. Note that (a) is derived because the absolute square of each entry of the global combiner \( \mathbf{z}_{a|m} \in \mathbb{C}^{N+1} \) is expected to be one over the number of entries, (b) is derived because the quantity \( \left| \mathbf{h}^\text{virt}_{b} \right|^2 \) of effective channel follows
Chi-squared distribution $\chi^2_{(M-N+1)}$ and its expectation is $M - N + 1$ [14, Lemma 3], and $(c)$ is derived in Proposition 1, where the expectation of the quantization error is $\omega = Q^c(\frac{M-1}{N-1}) - 1 \frac{M}{N} B(Q^c(\frac{M-1}{N-1}), M-N+1)$.

The SINR is approximated in (37), where $\alpha = 1 + \rho \nu / M$ denotes the noise-plus-local interference for a given SNR $\rho$.

In Fig. 6, we evaluate the accuracy of our approximation by comparing the cdf of the approximated SINR in (37) with that of the SINR expression in (36). As the comparison with numerical results without replacing the local interference by its expectation value, the SINR approximation does not significantly affect our results. For easy of analysis, we use the approximated SINR in (37) for the rest of sections.

Further, the distribution of the quantity $\| \mathbf{h}_{ua|m}^{\text{eff,qu}} \|^2$ in (37) is derived in the following proposition.

**Proposition 2:** The squared norm of global effective vector

$$U \triangleq \| \mathbf{h}_{ua|m}^{\text{eff,qu}} \|^2 = (\mathbf{H}_{qu}^H \mathbf{H}_{ua|m})^2$$

follows Chi-squared distribution $\chi^2_{(M-N)}$ and its pdf is $f_U(u) \approx \frac{u^{M-N-1} e^{-u/\omega}}{\omega^{M-N} B(M-N+1)}$ with the variance $\omega^2 \approx \frac{1}{N+1} \left[ N + \frac{M-N}{\bar{M}-N+1} \right]$ and $\omega \triangleq \frac{\rho}{M} \| \mathbf{h}_{ua|m}^{\text{eff,qu}} \|^2 = \tilde{\gamma}_{am}$.

\[
\tilde{\gamma}_{am} \approx \rho^2 \frac{M}{\alpha} \left( (M-\frac{N-1}{N}) \frac{Q^{c}(\frac{M}{N-1})}{\rho/M(M-N)} - (M-N) \log \left( \left( \frac{M-1}{N-1} \right) Q^{c}(\frac{M}{N-1}) \left( \frac{\rho/M}{M-N} \right) + \left( \frac{\rho}{M} \right)^{-1} \right) \right).
\]

Finally, the cdf of approximated SINR $X \triangleq \tilde{\gamma}_{am}$ in (37) is derived based on the pdf of $\| \mathbf{h}_{ua|m}^{\text{eff,qu}} \|^2$ according to

\[
F_X(x) \approx 1 - \left( (M-1) e^{-\frac{\rho}{M} x} \right) x + 1 \frac{1}{(M-N) M-N+1}.
\]

**D. COOPERATION MODE SWITCHING ALGORITHM**

We develop a cooperation mode-switching algorithm based on the expectation of a sum-rate. To estimate sum-rate throughput, we derive the cdf of the SINR for the scheduled MUs in $\mathcal{U}$. By using a similar logic from [23, Theorem 1], the SINR for the $m$-th selected MU is estimated as in (39) with the cdf of SINR in (38). We now take a closer look at $Q^c(m, M-N)$ meaning the number of CQI candidates for the $m$-th user selection process. Since each user generates $M$ CQIs, the total number of CQI candidates is given by $KM$. Because the user having the largest CQI is selected from the remaining CQI candidates, the scheduled user and the codeword is excluded for the following user selection process. The number of CQI candidates in the $m$-th user selection process is defined by

\[
Q^c(m) = (K - 2(m - 1)) (M - (m - 1)).
\]

Finally, the sum-rate of the multiuser MIMO system relying upon the proposed cooperative feedback is estimated as

\[
R_{\text{prop}} \approx \sum_{m=1}^{M} \log_2 (1 + \tilde{\gamma}_{am}) \tag{40}
\]

with the estimated SINR for the scheduled MUs in (39).

**Remark 1:** When the cooperative feedback is not activated, the sum-rate of the multiuser MIMO system is estimated in [23] according to $R_{\text{conv}} \approx \sum_{m=1}^{M} \log_2 (1 + \gamma_{am})$. The SINR for the scheduled user is defined in (41) and the number of CQI candidates is given by $Q^c(m) = (K - 2(m - 1)) (M - (m - 1))$. In the proposed cooperation mode-switching algorithm, multiuser MIMO systems activate the cooperative feedback mode when the differential sum-rate is positive such that

\[
\Delta R \equiv R_{\text{prop}} - R_{\text{conv}} > 0. \tag{42}
\]
V. NUMERICAL RESULTS

In this section, we evaluate the performance of the cooperative feedback algorithm based on the sum-rate defined as

$$R_{\text{num}} = \sum_{m=1}^{M} \log_2(1 + \bar{\gamma}_{|m|}),$$

where the SINR for the $m$-th scheduled user is computed numerically according to

$$\bar{\gamma}_{|m|} = \frac{|(\mathbf{h}_{n|m}^{\text{ef}})^H \mathbf{c}^m|^2}{\bar{\rho} + \sum_{\ell=1, \ell\neq m}^{M} |(\mathbf{h}_{n|m}^{\text{ef}})^H \mathbf{c}^\ell|^2}.$$  

Note that $\mathbf{h}_{n|m}^{\text{ef}}$ denotes the scheduled MU exploiting the $m$-th codeword $\mathbf{c}^m$. The sum-rate performance is evaluated numerically from Monte-carlo simulations with 10,000 independent channels (solid blue lines in Figs. 7). Moreover, the sum-rate performance is verified analytically with the formulation derived in Section IV-D (dotted black lines in Figs. 7).

We first investigate the accuracy of the sum-rate formulation $R_{\text{prop}}$ derived in (40). In Figs. 7(a) and 7(b), the numerical results and the sum-rate formulation are compared for a variety of user numbers against SNR $\rho$. The accuracy of the sum-rate formulation is verified by assuming the proposed cooperation mode is activated. Note that the sum-rate formulation is derived in Section IV-D based on the largest order statistics [30]. In Figs. 7(c) and 7(d), we also compare the numerical results and the sum-rate formulation against the number of users $K$. It should be noted that the numerical results of the previous multiuser systems relying upon the conventional feedback algorithm are depicted by using the solid red lines. According to the extreme value theory, the differences between the numerical results and the sum-rate in (40) decrease as the
number of users $K$ increases. In Figs. 7, it is shown that the differences between the numerical results and the sum-rate formulation are negligible, especially when there are a large number of users on the MTC network. Moreover, the error term $\nu$ in (37) is matched to the numerical results when the overhead for the cooperation link $B^{c3}$ is large. It is expected that the sum-rate formulation is better fitted to the numerical results as the size of the local codebook $B^{c3}$ increases.

In Figs. 8(a) and 8(b), we evaluate the sum-rate performances of the proposed mode-switching algorithm. We first note that the numerical results of conventional feedback algorithm, proposed cooperative feedback algorithm, and proposed mode-switching algorithm are depicted by using solid red lines, solid blue lines, and dotted black lines, respectively. The proposed algorithm shows better sum-rate estimation performance when the number of users is sufficiently large because the cooperation mode-switching algorithm is developed based on the extreme value theory [30]. Be aware that the mode-switching algorithm in Fig. 8(a) always triggers the proposed cooperation mode. For this reason, the cooperative feedback algorithm and mode-switching algorithm produce the same numerical results and the numerical results overlap. In Fig. 8(c), we take a closer look at the cross point between the cooperative feedback activation mode and the deactivation mode in Fig. 8(b). The estimated mode-switching point in the mode-switching algorithm (circle) and the mode-switching point in the numerical results (square) are within the range of a 0.08-dB window. This result means that the proposed adaptive cooperation algorithm finds mode-switching points well based on the given information of SNR $\rho$ and system parameters, i.e., $K$, $B^{c3}$, $M$, and $N$. Numerical simulations verify that the mode-switching algorithms outperform conventional multiuser MIMO systems that do not trigger the proposed cooperative feedback mode.

VI. CONCLUSION

In this paper, we developed limited feedback frameworks suitable for multiuser systems in MTC networks. First, we proposed the user cooperation-based limited feedback strategy to obtain high-resolution CSI with minimal additional burden on the current FDD-based communication architecture. We focused on reducing channel quantization errors by allowing a limited amount of CSI exchange between close-in users. In the proposed algorithm, some multiuser resources are used to enhance channel quantization performance while minimizing multiuser diversity gain degradation that this approach entails. Second, we carried out sum-rate throughput analysis to solve the trade-off problem between channel quantization performance and multiuser diversity gain. Based on the analytical studies, we developed the cooperation mode-switching algorithm in order to activate/deactivate cooperation mode according to channel and network conditions without transmitter interventions. Numerical results verified that the proposed algorithm improves sum-rate throughput because the multiuser resources are used to obtain high-resolution CSI.
We study statistical properties of the global channel matrix $\tilde{H}_a^{qu} = [\tilde{H}_a^H, \tilde{h}_b^{virt}]^H$ in (16) to analyze the quantity
\begin{equation}
U \doteq \| \tilde{h}_{a[m]}^{eff, qu} \|^2 = \| (\tilde{H}_a^{qu})^H \tilde{z}_{a[m]} \|^2_2
\end{equation}
of the global effective channel vector. The global channel matrix is composed of $\tilde{H}_a$ including $N$ channel vectors of MU and the (quantized) virtual channel vector $\tilde{h}_b^{virt}$ transferred from AU. It is already known that the entries in $\tilde{H}_a$ follow $CN(0, 1)$. We focus on analyzing the virtual channel in (15),
\begin{equation}
\tilde{h}_b^{virt} = \cos \phi_b \tilde{h}_b^{virt} \parallel 2v_b
\end{equation}
where we define $v_b^{virt} \doteq \| \tilde{h}_b^{virt} \|_2 v_b$. Notice that $\cos \phi_b$, $\| \tilde{h}_b^{virt} \|_2$, and $v_b$ are independent. Since $v_b$ is an unit-norm vector with isotropically distributed entries, the quantity $A \doteq \| v_b^{virt} \|^2 = \| \tilde{h}_b^{virt} \|^2$ follows Chi-squared distribution $\chi^2_M(M - N + 1)$, where its pdf is given by
\begin{equation}
f_A(a) = \frac{\alpha^M - N e^{-\alpha/2} \Gamma(M - N + 1)}{(2\alpha^2)^{(M - N + 1)/2} \Gamma(M - N + 1)}
\end{equation}
with the variance $\sigma^2 = \frac{1}{2}$ [31], [32].

All other entries in the global channel matrix $\tilde{H}_a^{qu}$ follow Gaussian distribution so that it would be good in analyzing the quantity $\| \tilde{h}_{a[m]}^{eff, qu} \|^2_2$ if the entries in $v_b^{virt}$ also follow Gaussian distribution. However, the quantity $\| v_b^{virt} \|^2$ has the distribution of the $(M - N + 1)$-dimensional vector, while the vector in the $M$-dimensional complex space. For this reason, $v_b^{virt}$ cannot be modeled by Gaussian random variables although its phase follows uniform distribution in the interval $[0, 2\pi)$ and the quantity $\| v_b^{virt} \|^2$ follows Chi-squared distribution. For ease of analysis, we made the approximations as follows:

- First, we assume that the quantity $\| \tilde{v}_b^{virt} \|^2$ can be approximated by $B \doteq \| \tilde{v}_b^{virt} \|^2_2$ that follows Chi-squared distribution $\chi^2_M$ with $M$ degree of freedom (DoF), where its pdf is given by $f_B(b) = \frac{b^{M-1}e^{-b/(2\alpha^2)}}{(2\alpha^2)^M \Gamma(M)}$ with the variance $\sigma^2 = \frac{M - N}{2M}$. The accuracy of this approximation increases as the ratio of DoFs between $A$ and $B$, i.e., $\frac{M - N}{2M}$, approaches one. The approximation is the result of changing DoF while adjusting its variance to make $E[B]$ equal to $E[A]$ such that
\begin{equation}
E[B] = \int_0^\infty b f_B(b) \, db = \frac{\Gamma(M - 1)\Gamma(M)}{(2\alpha^2)^M \Gamma(M)} = \frac{M - N}{2M}.
\end{equation}

- Second, we assume that the random variable $\cos^2 \phi_b$ can be replaced by its expectation value $E[\cos^2 \phi_b] = 1 - \omega$ by using similar method in Section IV-C. It should be pointed out that the variance of the minimum local error is very small compared to that of $\| \tilde{v}_b^{virt} \|^2_2$. We thus conclude that the operation of replacing the local error term by its expectation will lead a minimal effect. The expectation of the quantization error is derived in Proposition 1 such that $\omega = Q_{cl}(L - 1 - 1) - 1 - \pi B(Q_{cl}, \frac{M - N + 1}{M - N - 1})$.

As a result of the above approximations, the virtual vector can be approximated by $\tilde{h}_b^{virt} \doteq \hat{h}_b^{virt}$, where its entries follow $CN(0, 1 - \omega)$. After all the discussions of the entries in $\tilde{H}_a^{qu}$, the global channel matrix is modeled by
\begin{equation}
\tilde{H}_a^{qu} \doteq R \tilde{h}_a^{virt} \doteq \hat{H}_a^{qu}
\end{equation}
where $\hat{H}_a^{qu}$ is the global channel matrix having entries that follow $CN(0, 1)$, and the covariance matrix is given by
\begin{equation}
R = \begin{bmatrix} I_N & 0_{N, 1} \\ 0_{1, N} & 0_{N, 1} + (1 - \omega)(M - N + 1) \end{bmatrix} \in \mathbb{C}^{(N + 1) \times (N + 1)}.
\end{equation}

The squared norm of the global effective channel vector is now approximated by
\begin{equation}
\| \tilde{h}_{a[m]}^{eff, qu} \|^2_2 \approx \frac{\| (\hat{H}_a^{qu})^H \hat{u}_{a[m]} \|^2_2}{\| \tilde{u}_{a[m]} \|^2_2} = \| \hat{u}_{a[m]} \|^2_2 \frac{1}{D} \doteq \hat{D},
\end{equation}
because $(\hat{H}_a^{qu})^H \hat{u}_{a[m]} = e^{\phi_{v,M}}$, as shown in (5) and (6). We derive the distribution of $\| \hat{u}_{a[m]} \|^2_2$.

**Proposition 3:** The distribution of $\| \hat{u}_{a[m]} \|^2_2$ is the same as the distribution of $\frac{w^H (\hat{H}_a^{qu})^H}{w^H (\hat{H}_a^{qu})^H} w$, where $w \in \mathcal{W}$ is the column vector that is subject to
\begin{equation}
\mathcal{W} \doteq \left\{ w \in \mathbb{C}^{N + 1} : \| w \|_2 = e^{\psi w}, \psi_n \sim U(0, 2\pi) \right\}.
\end{equation}

**Proof:** For a given channel matrix $\hat{H}_a^{qu} \in \mathbb{C}^{(N + 1) \times M}$, the covariance matrix of $\hat{u}_{a[m]}$ is computed according to
\begin{equation}
Z = (\hat{H}_a^{qu})^H \hat{E}[e^{\phi_{v,M}} (e^{\phi_{v,M}})^H] (\hat{H}_a^{qu})^H
\end{equation}
where (a) is derived with $E[e^{\phi_{v,M}} (e^{\phi_{v,M}})^H] = I_M$ because $e^{\phi_{v,M}}$ is an independent (unit-norm) vector that is isotropically distributed in $\mathbb{C}^M$. To study statistical distributions of the quantity $\| \hat{h}_{a[m]}^{eff, qu} \|^2_2$, we model the receive combiner as
\begin{equation}
u = Z^\dagger \tilde{w},
\end{equation}
where $w \in \mathcal{W}$ is the isotropically distributed column vector. By plugging the receive combiner in (47) into (46), the approximated norm square of the global effective channel vector is rewritten by
\begin{equation}
D \doteq \frac{\| (\hat{H}_a^{qu})^H \hat{u}_{a[m]} \|^2_2}{\| \hat{u}_{a[m]} \|^2_2} = \frac{w^H (Z^\dagger Z^\dagger)^H Z^\dagger w}{w^H (Z^\dagger Z^\dagger) w}.
\end{equation}
Based on the numerical results in Fig. 9, we conclude that the global effective channel vector is well approximated by (46). In this paper, we use the approximated channel vector with the pdf defined in (49).

REFERENCES


JIHO SONG received the B.S. and M.S. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2009 and 2011, respectively, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2017. He is currently an Assistant Professor with the School of Electrical Engineering, University of Ulsan, Ulsan, South Korea. Before joining the University of Ulsan, he was a Senior Researcher with Motorola Mobility, Chicago, IL, USA. His research interests include the design and analysis of millimeter-wave communication, multiuser MIMO communication, and limited feedback strategies for massive MIMO systems. He was the recipient of the Bronze Prize in Samsung Electronics’ 23rd Humantech Paper Contest in 2017.

BYUNGJU LEE received the B.S. and Ph.D. degrees from the School of Information and Communication, Korea University, Seoul, South Korea, in 2008 and 2014, respectively. From 2014 to 2015, he was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, Seoul National University. He was a Post-Doctoral Scholar with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, from 2015 to 2017. He joined Samsung Electronics in 2017, has been involved in research and standardization for 3GPP 5G NR. His research interests include physical layer system design for 5G wireless communications.

SONG NOH received the B.S. degree in electrical engineering from Soongsil University, Seoul, South Korea, in 2008, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2010, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 2015. From 2015 until 2018, he was a System Engineer in the Next Generation and Standards Group, Intel Corporation, Hillsboro, OR, USA, and participated in design of PHY/MAC layer protocol for millimeter-wave mobile broadband systems. Since 2018, he has been on the faculty of the Department of Information Telecommunication Engineering in Incheon National University, Incheon, South Korea. His research interests include the design and analysis of millimeter-wave communication systems and adaptive signal processing for wireless communications. Along with co-authors, he was a co-recipient of the Communication Theory Symposium Best Paper Award at the 2014 IEEE Global Communications Conference (Globecom).

JONG-HO LEE received his BS degree in Electrical Engineering, and his MS degree and PhD in Electrical Engineering and Computer Science from Seoul National University, Seoul, Rep. of Korea in 1999, 2001, and 2006, respectively. From 2006 to 2008, he was a Senior Engineer at Samsung Electronics, Suwon, Rep. of Korea. From 2008 to 2009, he was a Postdoctoral Researcher at the Georgia Institute of Technology, Atlanta, GA, USA. From 2009 to 2012, he was an Assistant Professor at the Division of Electrical, Electronic, and Control Engineering, Kongju National University, Cheonan, Rep. of Korea. From 2012 to 2018, he was an Associate Professor at the Department of Electronic Engineering, Gachon University, Seongnam, Rep. of Korea. Since 2018, he has been a faculty member at the School of Electronic Engineering, Soongsil University, Seoul, Rep. of Korea. His research interests include wireless communication systems and signal processing for communication, with an emphasis on multiple-antenna techniques, multi-hop relay networks, physical layer security, and full-duplex wireless communications.