IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 24, 2019, accepted January 5, 2020, date of publication January 9, 2020, date of current version January 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2965280

Joint Sparse Channel Recovery With Quantized Feedback for Multi-User Massive MIMO Systems

FARZANA KULSOOM[1], ANNA VIZZIELLO[1], HASSAN NAZEER CHAUDHRY[2], AND PIETRO SAVAZZI[1]

[1]Department of Electrical, Computer and Biomedical Engineering, University of Pavia, 27100 Pavia, Italy
[2]Department of Electrical, Information and Bioengineering, Politecnico di Milano, 20133 Milan, Italy

Corresponding author: Farzana Kulsoom (farzana.kulsoom01@universitadipavia.it)

ABSTRACT Accurate channel state information (CSI) at the transmitter is an essential prerequisite for transmit beamforming in massive multiple input multiple output (MIMO) systems. However, due to a large number of antennas in massive MIMO systems, the pilot training and feedback overhead become a bottleneck. To resolve this issue, the research work presents a novel framework for frequency division duplex (FDD) based multi-user massive MIMO system. A 2-step quantization technique is employed at the user equipment (UE) and the CSI is recovered at the base station (BS) by applying the proposed compressed sensing (CS) based algorithms. The received compressed pilots are quantized by preserving 1 bit per dimension direction information as well as the partial amplitude information. Subsequently, this information is fed back to the BS, which employs the proposed quantized partially joint orthogonal matching pursuit (Q-PJOMP) or quantized partially joint iterative hard thresholding (Q-PJIHT) CS algorithms to recover the CSI from a limited and quantized feedback. Indeed, an appropriate dictionary and the hidden joint channel sparsity structure among users is exploited by the CS methods, resulting in the reduction of the feedback information required for channel estimation. Simulations are performed using singular value decomposition (SVD) and minimum mean square error (MMSE) beamforming utilizing the estimated channel. The results confirm that the proposed 2-step quantization approaches the system with channel knowledge without quantization, thus overcoming the training and feedback overhead problem. Moreover, the proposed 2-step quantization outperforms 1-bit quantization, at the cost of slightly higher complexity.

INDEX TERMS Compressed sensing, joint channel estimation, quantization, channel state information (CSI), multiple input multiple output (MIMO), sparse channel estimation, dictionary.

I. INTRODUCTION

Massive multiple input multiple output (MIMO) is one of the emerging technology in wireless communication. It employs a large number of transmit antennas at the base station (BS), which makes the system more reliable and enhances the throughput compared to traditional MIMO [1]. The high throughput is achieved by improving spectral efficiency, mitigating inter-user interference and employing more directional beams [1]. The massive MIMO system can serve various users simultaneously in the same time-frequency block, due to the large deployment of antennas at the BS. Nevertheless, serving many users concurrently is challenging due to the interference among them, which can be mitigated if each user has its aligned beam. This is obtained with appropriate precoding/beamforming techniques at BS [2]. However, the accurate beamforming requires proper channel state information (CSI) at the transmitter. Traditionally, CSI can be obtained by using time division duplex (TDD) or frequency division duplex (FDD) schemes.

In TDD, both uplink (UL) and downlink (DL) operate at the same frequency bands, but in separate time slots. A channel estimate in one band can be utilized in the other, thanks to channel reciprocity. Consequently, in TDD if the channel estimate is obtained in UL, by employing the channel reciprocity, this estimate is also applicable to DL. The primary constraint of TDD is that acquisition and utilization of CSI should be performed within the coherence time [3]. Therefore, the users have to transmit their pilots simultaneously within this coherence time. Since the pilots should be orthogonal to evade interference among them, the limited number of available orthogonal sequences may result in pilot contamination [3].

The associate editor coordinating the review of this manuscript and approving it for publication was Adnan M. Abu-Mahfouz.

Unlike TDD, in FDD the channel reciprocity is no longer applicable since UL and DL are in separate bands. Despite the advantage of channel reciprocity in TDD, FDD is important in two ways; firstly, it is considered more robust to delay-sensitive applications [4], secondly, most of the existing systems are already deployed in FDD. Therefore, it is of great significance to improve and enhance the approaches to obtain CSI in FDD systems. Since channel reciprocity can not be exploited in FDD, DL and UL channels are separately estimated. Typically, the training is done in DL by estimating CSI at the users, this CSI is sent to the BS in the uplink using a dedicated signaling link. In general, the number of pilots required for training grows linearly with the number of transmit antennas at the BS. Since massive MIMO have a large number of BS antennas, the pilot training and feedback overhead have become the bottle-neck in FDD [5].

One of the potential solutions to resolve the training overhead problem is to explore an appropriate technique for reducing the feedback overhead. The research work in [6] and the experimental studies in [7] reveal that the increase in the number of BS antennas results in limited prominent transmission directions per user. Since there are few scatterers at the BS side as compared to the number of antennas, the channel matrix will tend to be sparse [7], [8]. In this scenario, to estimate the channel, compressed sensing (CS) paradigm can be utilized to represent a high dimensional channel vector into low dimension [5], [6], [9]. The main objective of the aforementioned research works is to reduce the number of pilot training overhead by taking advantage of the channel sparsity in general, without exploiting any structure in the measurements. Besides the channel sparsity, it has been observed that the users are mostly located in the vicinity sharing common scatterers, for example, in offices, playgrounds, streets, apartments. Due to these scatterers, the channel is strongly correlated between the multi-antennas of each user, known as ''intra-user joint channel sparsity'', and among closely located distinct users, termed as ''inter-user joint channel sparsity'' [10].

The research works in [11], [12], exploit grouping among users to estimate the channel. The central idea is based on grouping users with common channel statistics, which will decrease the pilot overhead [11]. Subsequently, the inter-user joint sparsity can be applied within each of these groups. The research work [12] utilizes the beam-block sparsity model to reduce the pilots in DL, which are then fed back to UL in a quantized form and applied at BS to obtain CSI through CS. However, [11] does not consider intra-user or partial joint user sparsity. Other notable research works [13], [14] laid the theoretical foundations for modeling intra and inter-user joint sparsity, known as joint sparsity model (JSM). Utilizing the ideas of JSM, a class of algorithms under the paradigm of distributed compressed sensing (DCS) exploits Intra and the inter-channel correlation between users to unveil joint sparsity structure in massive MIMO systems [8], [15]–[17]. Although the research work in [8], [9] substantially reduced the training overhead problem, the feedback pilot bits are still considerably high. Thus, besides reducing the number of measurements, some work has also been done to lower the number of feedback bits using direct quantization [18], [19]. However, this type of direct quantization is complex and requires several bits per symbol for decoding to achieve better performance. Another extreme case of feedback reduction is 1-bit per dimension using CS algorithms as discussed in [16]. But, it hardly provides directional information and power information is lost in this manner. The power control information is very critical for CSI since it captures the effect of scatterers, multipath fading and signal strength deterioration with distance. To improve the performance, [12] presents a scheme to quantize both the amplitude and phase of the received pilots, thus preserving both directional and power information, although in a different research problem. In [20] spatial channel estimation is performed for FDD MIMO interference alignment (IA)[1] system, and a zero-feedback bit solution is proposed. However, [20] assumes some predefined channel directions and also that UL and DL spatial channel clusters bounce on similar sub-paths.

The presented work is inspired by [12] in terms of considering both amplitude and phase information. However, we have observed that it is more efficient to quantize the complex pilot phase information with 1-bit per dimension as in [16] and acquire partial amplitude information by averaging the amplitude of compressed received pilots. Therefore, the phase information, in 1-bit compressed form as [16] along with the amplitude information, is sent back to the BS using a more simplified technique than [12]. While [12] uses eigenbeam model (EBM), our work is based on virtual channel model (VCM) that uses Fourier basis instead of eigenbasis [7]. Although VCM and EBM models are comparable in terms of system capacity and performance, VCM is inherently less complex, is more suitable for uniform linear arrays (ULAs) and can be readily extendable to frequency-selective channel [7].

In more details, this research work presents quantized feedback-based algorithms for partially joint channel estimation in massive MIMO systems, preserving both power and directional information. The pilots are transmitted from BS to each user equipment (UE), which are then quantized using a two-step quantization method and fed back to BS. The BS uses the proposed CS-based recovery algorithms to recover the channel from the quantized feedback: quantized partially joint orthogonal matching pursuit (Q-PJOMP) or quantized partially joint iterative hard thresholding (Q-PJIHT). This recovered channel information is utilized by the transmit beamforming to reduce inter-user interference. Two types of transmit beamforming, namely (i) minimum mean square error (MMSE) or regularized zero-forcing (RZF) and (ii) singular value decomposition (SVD), are employed with realistic estimated imperfect channel information, as opposed

[1]The IA transforms the signal on the transmitters to be aligned in the signal subspace, for maximum multiplexing gain.

to [8], [16] that assume perfect channel knowledge to perform beamforming.

Effectively, we have focused on multiple aspects of communication and presented a complete system for CSI acquisition spanning all three main stages: (i) channel estimation, (ii) beamforming procedure, (iii) data detection.

The main contributions of this work are as follows:

- This research work presents two novel distributed CS-based algorithms, Q-PJOMP and Q-PJIHT, to estimate a partially joint channel by utilizing quantized feedback in massive MIMO systems. The feedback bits are reduced without compromising significant performance gain while preserving power and directional information, as compared to [16] that considers only directional information.
- An efficient dictionary-based sparsifying matrix has been adopted, which is more effective as compared to the square DFT based sparsifying matrix especially for the frequency selective channel.
- Besides improved channel estimation algorithms, two beamforming techniques, namely MMSE and SVD, are applied to reduced inter-user interference. Indeed we consider a scenario where the BS communicates simultaneously with multiple UEs through multiple beams.
- A realistic CSI estimate is acquired and utilized in the transmit beamforming procedure, as compared to the common perfect channel knowledge assumption as in [8], [16]. Furthermore, we have investigated a more realistic frequency selective channel, instead of a flat fading as in [8], [16].

Simulation results reflect that Q-PJOMP is the most suitable algorithm to estimate the partially joint channel from quantized feedback. Moreover, by applying MMSE based transmit beamforming, the system performance is significantly improved.

The remainder of the paper is organized as follows. Sec. II describes the massive MIMO OFDM based system model, particularly the spatial correlations of users' channels are emphasized with limited feedback per dimension. Sec. III addresses the proposed CS techniques for training and feedback schemes, i.e., Q-PJOPM and Q-PJIHT, exploiting the distributed joint channel sparsity model. Sec. IV presents the simulation results for performance evaluation, finally, conclusions are drawn in Sec. V.

II. SYSTEM MODEL

A. MASSIVE MIMO OFDM SYSTEM

Consider a single-cell multi-user massive MIMO OFDM transmission over a quasi-static frequency selective fading channel. The system is composed of a BS and K users in FDD mode. The number of antennas at the BS is equal to N_t, with N_t very large, and each user is equipped with N_r antennas. The overall system model can be represented as follows:

$$\mathbf{Y}_i = \mathbf{H}_i\mathbf{X}_i + \mathbf{N}_i \tag{1}$$

where $\mathbf{X}_i \in \mathbb{C}^{N_t \times T}$ is the transmitted signal in T time slots, $\mathbf{H}_i \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix from the BS to the i^{th} user, $\mathbf{N}_i \in \mathbb{C}^{N_r \times T}$ is the Gaussian noise, and $\mathbf{Y}_i \in \mathbb{C}^{N_r \times T}$ is the received signal on each i^{th} user. The fading channel between each transmit antenna and i^{th} user is considered as frequency-selective and has L uncorrelated channel taps. For $N_r \times N_t$ MIMO channel, the frequency response can be written as:

$$\mathbf{H}_i(f) = \sum_{l=1}^{L} \beta_l \mathbf{a}_r(\theta_{r,l})\mathbf{a}_t(\theta_{t,l})e^{-j2\pi f \tau_l} \tag{2}$$

where β_l is the complex amplitude for l^{th} path, τ_l is the l^{th} path delay, $\theta_{r,l}$ and $\theta_{t,l}$ are the transmitted and received path angles and $\mathbf{a}_r$, $\mathbf{a}_t$ represents array steering and response vectors of transmitted/received signal in the direction of θ_t/θ_r. To mitigate frequency selective fading the OFDM scheme is used [21]–[23]. An OFDM symbol duration $T_{sym} = (N_o + G) \cdot T_s$ is considered, where N_o is the number of subcarriers equally spaced in frequency at $\Delta f = \dfrac{1}{T_s}$, with the sampling time T_s and G is the guard interval in number of samples, which is set larger than the expected channel delay spread to further reduce the inter-symbol interference (ISI) [21].

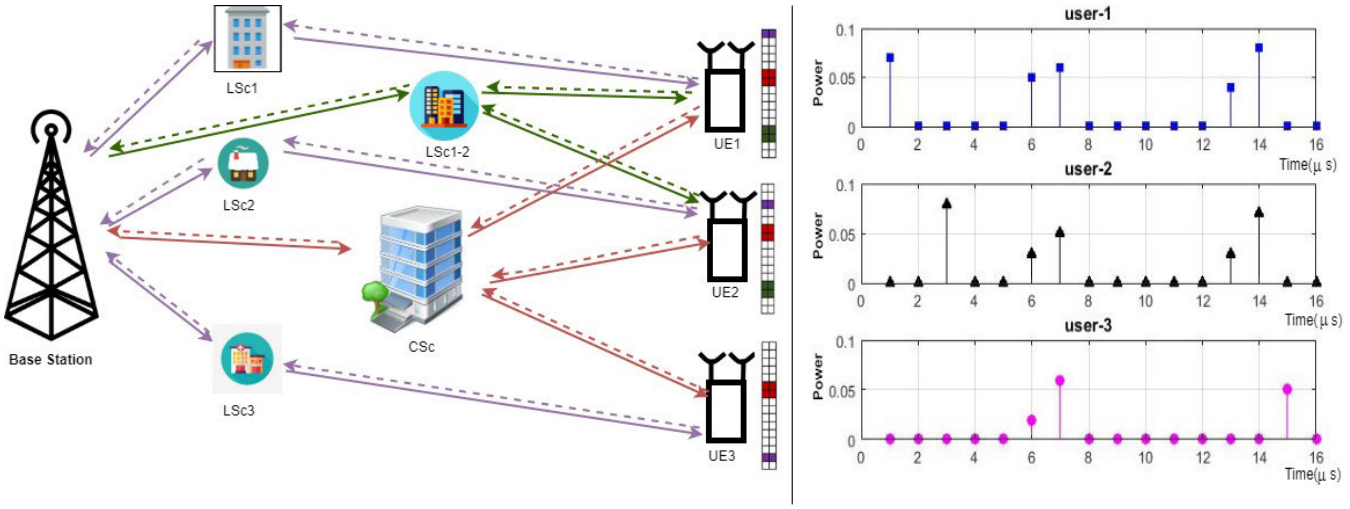
B. DISTRIBUTED JOINT CHANNEL SPARSITY MODEL

In a cell there are some dominant scattering clusters whose position is determined from the specific attributes of the cell itself, like the presence of buildings or other propagation obstacles. These scatterers can be shared by users regardless of their position [5]. Massive MIMO experiments [7] prove that channel matrices at the user side is sparse due to the limited local scatterers at the BS [15], and maybe jointly correlated due to the shared common local scattering clusters [24]. Thus, both per-link and joint channel sparsity can be exploited through CS-based solutions to reduce training and feedback overheads. The inter-user channel sparsity can be further explained based on how scatterers are shared among users:

- The scatterers shared among all users are known as common scatterers (CSc), such as, the CSc shared among all UE in Fig. 1. These common scatterers will produce *joint sparsity*.
- The scatterers for a single user or shared among the subgroup of users is known as local scatterers (LSc), and will exhibit *partial joint sparsity*. For example, in Fig. 1 LSc1, LSc2, and LSc3 are local scatterers for UE1, UE2, and UE3 respectively, while, LSc1-2 is shared among UE1 and UE2.

The presented research work considers a ULA model for both BS and user side antennas. To represent spatio correlation channel for ULA in angular domain, the virtual channel representation can be considered. The experimental studies in [7] confirm that spatial channel can be transformed into angular domain as follows:

$$\mathbf{H}_i^a = \mathbf{F}_{R_i}^H \mathbf{H}_i \mathbf{F}_{T_i}, \quad \mathbf{H}_i = \mathbf{F}_{R_i}\mathbf{H}_i^a\mathbf{F}_{T_i}^H \tag{3}$$



(a) Downlink (solid line) and feedback link (dotted line) for joint channel sparsity with local scatterers (LSc) and common scatterers (CSc)

(b) Hypothetically generated channel delay profile of three users to show the joint sparsity

FIGURE 1. Illustration of joint sparsity structure in MIMO channel.

where $\mathbf{F}_{R_i} \in \mathbb{C}^{N_r \times N_r}$ and $\mathbf{F}_{T_i} \in \mathbb{C}^{N_t \times N_t}$ represent the unitary matrices with Fourier basis. This transforms the channel at the BS and the user side from a spatial domain to the virtual angle domain. $\mathbf{H}_i^a \in \mathbb{C}^{N_r \times N_t}$ is the virtual channel representation and can be considered equivalent to the actual spatial channel $\mathbf{H}_i$ in Fourier domain [7]. The non-zero (c, d)-th entry of $\mathbf{H}_i^a$ indicates the spatial path from the c-th BS transmit direction to the d-th receive direction of the i^{th} user. Equation (3) has been used to represent the spatial channel in the angular domain with the sparsifying matrix $\mathbf{F}$ [5], [8]. Although with angular domain representation channel sparsity is guaranteed, to obtain it explicitly for each link and joint channel sparsity, the design of sparsifying matrix $\mathbf{F}$ is very critical. In [8] and [16] a sparsifying matrix $\mathbf{F}$ is considered as square DFT matrix, but in [5] and [9] it has been shown that, by introducing more redundancy in sparsifying matrix $\mathbf{F}$, an improved channel representation can be achieved, as will be detailed in Sec. III.

Fig. 1 shows the system description with local and common scatterers in a multi-user MIMO system. Fig.1(a) explains the joint sparsity structure by displaying varying support indexes (non-zeros entries) in different colors in the channel matrix. For example, the scatterers shared among all users are represented by the common support Θ_c indicated in red, the support of the scatterers shared only by some users is shown in green, and of the local scatterers is represented in blue. The individual support Θ_i represents the non-zeros entries of each user. Fig.1(a) shows the position of the support in this joint sparsity structure, while Fig.1(b) shows that the power values at these supports could be varying. Besides, Fig.1(b) also gives a snapshot of the channel delay profile for different users in this hypothetically generated propagation environment.

The sparsity of the massive MIMO channel may be explained as follows [8], [10], [16]:

1) **Intra-User Joint Sparsity (Individual joint Sparsity):**
Since there are more than one receiver antennas N_r per user, the channel will exhibit intra-user joint sparsity. This effectively means that there will be channel dependency among different antennas of each user at the same time. This dependency is produced by local scatterers. Consequently, the receiver channel matrix elements will be extremely correlated and its row vectors will usually have the same sparsity support. Specifically, the matrix $\mathbf{H}_i^a$ is a row sparse and for each user i, the j^{th} row vector $\mathbf{H}_i^a(j)$ will have same support Θ_i such that $0 < |\Theta_i| \ll N_t$, i.e.,

$$\Theta_{i1} = \Theta_{i2} = \ldots = \Theta_{iN_r} \simeq \Theta_i \tag{4}$$

2) **Inter-User Sparsity (Distributed Joint Sparsity):**
Inter-user sparsity exploits the channel dependency among different users at the same time [10]. It refers to the scenario of a massive MIMO system where different users are close enough to share some common scatterers. Therefore, the channel matrices of different users will be highly correlated and will have common sparsity support Θ_c:

$$\Theta_c = \bigcap_{i=1}^{K} \Theta_i \tag{5}$$

This common sparse support will be a subset of individual sparse support if users are sharing some common scatterers, such that $\Theta_c \subseteq \Theta_i$. However, if there are only common scatterers among users, then the individual sparsity support will be equal to the common sparsity support. There is possibility that none of the users are sharing any common scatterers, in that case, $\Theta_c = \emptyset$ [8].

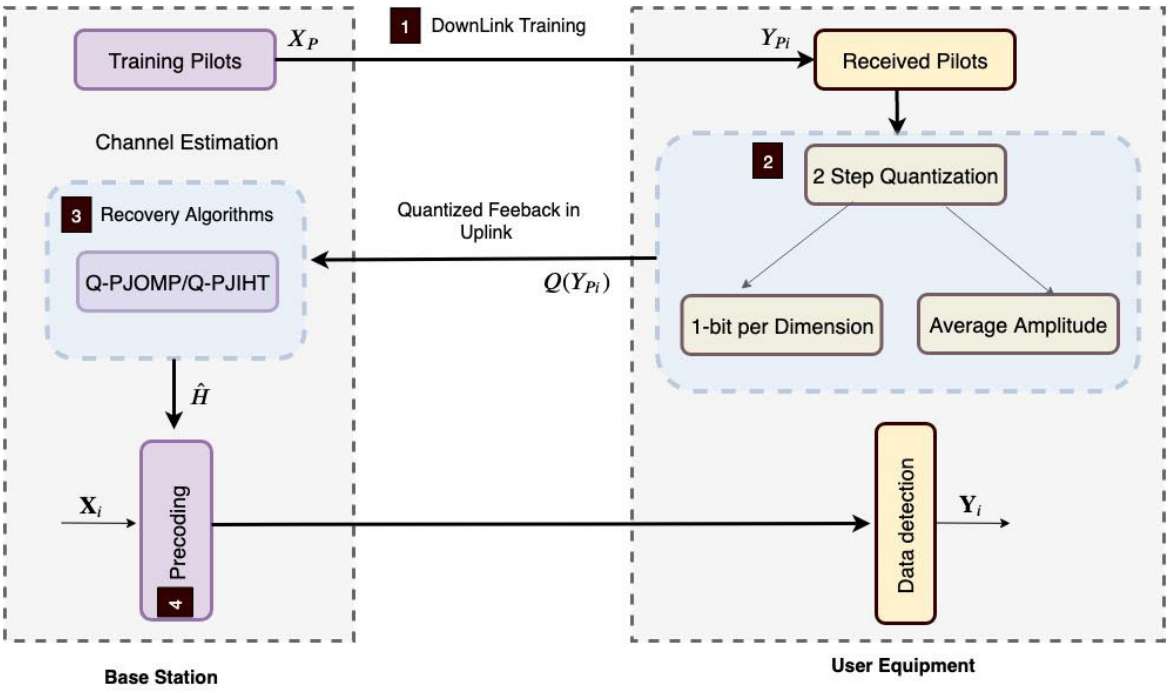


FIGURE 2. Flow chart of the overall framework of CSI acquisition and utilization.

C. LIMITED FEEDBACK BASED CHANNEL RECOVERY AND BEAMFORMING

The current subsection discusses the overall procedure for channel recovery from the quantized feedback pilots and utilization of that estimate for transmit beamforming as shown in Fig. 2. The BS sends pilots in the DL, which are quantized at each UE and are fed back to the BS via a dedicated uplink signaling channel. The BS utilizes these quantized pilots to exploit channel's hidden joint sparsity for channel estimation using distributed CS algorithms. Eventually, the recovered channel is employed for beamforming at the transmitter for more reliable data detection at the receiver. The overall system description of the quantized feedback and channel estimation is summarized in Framework and further given in Fig. 2. Each step of Fig. 2 is explained in detail as follows:

★ **Step-1 Downlink training:** As discussed in Sec. I, to obtain CSI of the DL at the transmitter in FDD, the pilots are sent in DL. For this purpose, the BS broadcasts T common training symbols to the K users over the downlink. Therefore, the received pilots for the i^{th} user can be written as:

$$\mathbf{Y}_{pi} = \mathbf{H}_i\mathbf{X}_p + \mathbf{N}_i \tag{6}$$

where $\mathbf{X}_p \in \mathbb{C}^{N_t \times T}$ is the concatenated transmitted pilots, and $\mathbf{Y}_{pi} \in \mathbb{C}^{N_r \times T}$ is the received pilots on each i^{th} user as shown in step-1 of Fig. 2.

★ **Step-2 Quantization:** The received pilots $\mathbf{Y}_{pi}$ are quantized at each UE to reduce the number of feedback bits, as shown in step-2 of Fig. 2. It is represented as $Q(\mathbf{Y}_{pi})$, however, the quantization process may also induces some error as follows:

$$\mathbf{Y}_{pi} = Q(\mathbf{Y}_{pi}) + \Gamma_i = Q(\mathbf{H}_i\mathbf{X}_p + \mathbf{N}_i) + \Gamma_i \tag{7}$$

where Q(.) is the uniform quantization function that maps the complex pilots to bits $[+1, -1]$ and Γ_i is the quantization error. In the presented work, pilots are quantized by preserving both direction and amplitude, as opposed to [16], where pilots are quantized to the extreme 1-bit case. The 1-bit scheme lacks power information, having only directional information, this results in poor channel estimate. There-

Framework Joint Channel Recovery Framework From Quantized Feedback

Step-1: A common training pilot vector $\mathbf{X}_p$ is transmitted from BS to all the users.

Step-2: Each user's received pilots are quantized using two steps:

(a) The averaged amplitude μ_i of the received pilot vector $\mathbf{Y}_{pi}$ is calculated for each user. Then these averaged amplitudes are quantized to 2 bits per user;

(b) The sign information is achieved with 1 bit per dimension from the received pilots:
$S(\mathbf{Y}_{pi}) = S(\mathbf{H}_i\mathbf{X}_p)$ such that
$S(\mathbf{Y}_{pi}) = \text{sign}(\text{Re}(\mathbf{Y}_{pi})) + j\text{sign}(\text{Im}(\mathbf{Y}_{pi}))$.

Step-3: Each user feedbacks the obtained quantized and reduced $\mathbf{Y_{pi}}$ to BS.

Step-4: The channel is jointly recovered for all users from the quantized feedback pilots $\mathbf{Y}_{pi}$ using distributed CS at BS.

fore, we have considered averaged quantized amplitudes of the received pilot for each user. The averaged amplitude $\mu_i = \frac{1}{P_t}\sum_i^{P_t} |\mathbf{Y}_{pi}|$ of the received complex pilot $\mathbf{Y}_{pi}$ is determined for each user. Subsequently, these averaged amplitudes are quantized to 2 bits per user. Furthermore, the direction information is computed using 1 bit per dimension from the received pilots, given as:

$$\mathbf{Y}_{pi} = \text{sign}(\text{Re}(\mathbf{Y}_{pi})) + j\text{sign}(\text{Im}(\mathbf{Y}_{pi})) \tag{8}$$

The amplitude information μ is fed back along with the direction information to the BS.

★ **Step-3 Channel recovery using CS:** The quantized pilots determined in step 2 are fed back to BS in UL. At the BS, CS-based algorithms are applied to recover the channel. The BS employs the 2-bit partial amplitude information per users along with 1 bit per dimension direction information of each user to recover the channel jointly for all users. It is important to mention that, since the users are sharing the channel, partial amplitude along with direction information is sufficient to jointly estimate the partially correlated channel. For this purpose, the CS-based JOMP [8] and BIHT algorithms [16] are modified. This facilitates channel recovery using quantized feedback information. We have proposed two novel distributed CS-based algorithms for better analysis in terms of complexity and performance. The Q-PJOMP given in Algorithm 1 and Q-PJIHT is given in Algorithm 2, respectively, with further details in Sec. III.

★ **Step-4 Beamforming:** During step 3 the estimate of the channel $\hat{\mathbf{H}}_i$ becomes available at the BS side, and it is employed to achieve transmit beamforming, as shown in step 4 of Fig. 2. The precoding matrix is constructed by applying this estimate, it overcomes the inter-user interference and decorrelates users in a highly correlated environment. Furthermore, the precoding matrix is beneficial for data detection to decorrelate users. In particular, we have focused on MMSE or RZF, as shown in equation (9). It attempts to maintain strong signal gain while limiting interference among users. MMSE based precoding is more reliable than ZF, since ZF amplifies the noise in the presence of highly correlated user channel. Moreover, MMSE precoder is not only robust for MU-MIMO systems where users are located in the vicinity, but it also provides stable performance at high SNRs [25]. MMSE precoder is considered as the more suitable choice for linear precoding in MIMO based wireless systems [25], [26]. For MMSE based precoding, the received signal model in equation (1) is revisited by including the beamforming matrix $\mathbf{W}$ as follows:

$$\text{MMSE}\begin{cases} \mathbf{G} = \hat{\mathbf{H}}^H(\hat{\mathbf{H}}\hat{\mathbf{H}}^H + \lambda\mathbf{I})^{-1} \\ \eta = \sqrt{\dfrac{N_t}{Tr(\mathbf{G}\mathbf{G}^H)}} \\ \mathbf{W} = \eta\mathbf{G} \\ \mathbf{Y}_i = \mathbf{X}_i\mathbf{W}_i\mathbf{H}_i + \mathbf{N}_i \end{cases} \tag{9}$$

In the above equation, $\lambda = N_t\sigma^2/K$ is the regularization factor, if $\lambda = 0$, the MMSE precoder becomes equal to a ZF. The MMSE optimal precoder $\mathbf{G}$ is obtained from the estimated channel $\hat{\mathbf{H}}$ of all users. In the next step, $\mathbf{G}$ is scaled with the power scaling factor η to minimize the mean square error (MSE) under the BS transmit power constraints [27]. Primarily, η is used for power allocation at the BS to minimize the total power consumption under some given noise and interference related constraints. While λ attempts to maintain the strong signal gain with limited interference among users. Besides the MMSE precoder in equation (9), for comparison purpose, an SVD based beamforming is also applied, as shown in equation (11). It can be noted that we do not assume perfect channel knowledge. Accordingly, the estimate of realistic channel can be decomposed using SVD, and is given as follows:

$$\hat{\mathbf{H}}_i = \hat{\mathbf{U}}_i\hat{\Lambda}_i\hat{\mathbf{V}}_i^H \tag{10}$$

where $\hat{\mathbf{U}}_i \in \mathbb{C}^{N_r\times N_r}$ and $\hat{\mathbf{V}}_i \in \mathbb{C}^{N_t\times N_t}$ represent the unitary matrices, such that $\hat{\mathbf{V}}_i^H\hat{\mathbf{V}}_i = \hat{\mathbf{U}}_i\hat{\mathbf{U}}_i^H = \mathbf{I}$ and $\hat{\Lambda}_i \in \mathbb{C}^{N_r\times N_t}$ is a diagonal matrix. Since we have considered realistic channel $\hat{\mathbf{H}}_i$, the estimated channel characteristics will not exactly match with that of the ideal channel. Consequently, the unitary property $\mathbf{V}_i^H\hat{\mathbf{V}}_i = \mathbf{I}$ does not hold any more and this may induce interference. Nevertheless, these ramifications can be combated provided the channel estimate is robust. Based on SVD decomposition, the equation (1) is revised and beamforming equations can be given as:

$$\text{SVD}\begin{cases} \hat{\mathbf{X}}_i = \hat{\mathbf{V}}_i^H\mathbf{X}_i \\ \mathbf{Y}_i = \mathbf{H}_i\mathbf{X}_i + \mathbf{N}_i = (\mathbf{U}_i\Lambda_i\mathbf{V}_i^H)\mathbf{X}_i + \mathbf{N}_i, \quad \forall_i \in K \\ \hat{\mathbf{U}}_i^H\mathbf{Y}_i = \hat{\mathbf{U}}_i^H(\mathbf{U}_i\Lambda_i\mathbf{V}_i^H)\hat{\mathbf{V}}_i\hat{\mathbf{X}}_i + \hat{\mathbf{U}}_i^H\mathbf{N}_i, \quad \forall_i \in K \\ \hat{\mathbf{Y}}_i = \Lambda_i\hat{\mathbf{X}}_i + \hat{\mathbf{N}}_i \end{cases} \tag{11}$$

where $\hat{\mathbf{X}}_i$ is the precoded signal at the transmitter, $\hat{\mathbf{N}}_i = \hat{\mathbf{U}}_i^H\mathbf{N}_i$ is the noise term, and $\hat{\mathbf{Y}}_i = \hat{\mathbf{U}}_i^H\mathbf{Y}_i$ is the decoder for the transmit beamformer at the receiver side.

III. COMPRESSIVE CHANNEL ESTIMATION WITH LIMITED FEEDBACK

The current section focuses on channel estimation at BS using CS algorithms. It is performed by employing the quantized pilots received on BS from UE, as indicated in step-3 of Fig. 2. We here discuss compressed sensing algorithms and dictionary design in more detail.

A. COMPARISON OF DICTIONARY AND DFT BASIS FOR MEASUREMENT MATRIX DESIGN

A channel that exhibits only a few dominant propagation paths may be considered sparse and is approximated as a linear combination over a known basis or dictionary, resulting in a sparse channel response [21]. Conventionally, DFT basis ϕ have been utilized for sparse channel representation [8], [12],

[16], [21], [28] as shown below.

$$\phi = \frac{\mathbf{1}}{\sqrt{\mathbf{N_t}}} \begin{bmatrix} e^{-j2\pi k_{1,1}} & \dots & e^{-j2\pi k_{1,N_t}} \\ e^{-j2\pi k_{2,1}} & \dots & e^{-j2\pi k_{2,N_t}} \\ \vdots & \ddots & \\ e^{-j2\pi k_{N_t,1}} & \dots & e^{-j2\pi k_{N_t,N_t}} \end{bmatrix}$$

The usage of the DFT basis is compliant with the theoretical results of signal estimation in CS [29], therefore, the DFT basis are employed in sparse channel representation. Considering normalized DFT basis ϕ as a sparsifying matrix, the sparse channel response can be written as, $\hat{\mathbf{H}}_i = \phi \mathbf{B}_{si}$, where $\mathbf{B}_{si} \in \mathbb{C}^{N_t \times N_r}$ is the sparse matrix. As described earlier in equation (3), such modeling is applied to represent spatial channel response into the angular domain, also known as the virtual channel model. Though, the DFT basis represents the channel only in a few directions. In a real-world scenario, the actual signal can emanate from arbitrary directions, which may differ from the one represented in the DFT matrix. This results in leakage effect and poor channel estimate. In case the channel is highly correlated or exhibits multipath, the leakage effect worsens even further [5].

To estimate the channel accurately, a robust and effective basis is an essential prerequisite. To construct an improved and substantial basis or dictionary, we extended the squared DFT matrix by proposing added redundancy and considering the channel delay profile as in [9]. The objective is to build a more flexible and enhanced representation of the channel. In the dictionary formulation, the channel delay profile τ_{ch} can be represented using OFDM sample time T_s and guard interval G, for each dictionary point as follows:

$$\tau_{ch} = [0, \alpha, 2\alpha, \cdots\cdots (M-1)\alpha] \tag{12}$$

where M is the length of dictionary column, α is step size in the range of 0 to M and is given as $\alpha = [G \cdot T_s - (G \cdot T_s/M)]$. To avoid inter-symbol interference (ISI), the step size α is constrained with $\tau_{\max}$ not exceeding G, i.e., $(\tau_{\max} < G)$ [21]. To formulate the dictionary basis $\mathbf{D}$ for the channel response, the OFDM symbol duration T_{sym} and the delay profile can be utilized as follows:

$$\mathbf{D} = \begin{bmatrix} e^{-j2\pi k_1 \tau_1} & e^{-j2\pi k_1 \tau_2} & \dots & e^{-j2\pi k_1 \tau_M} \\ e^{-j2\pi k_2 \tau_1} & e^{-j2\pi k_2 \tau_2} & \dots & e^{-j2\pi k_2 \tau_M} \\ \vdots & \dots & & \\ e^{-j2\pi k_{N_t} \tau_1} & e^{-j2\pi k_{N_t} \tau_2} & \dots & e^{-j2\pi k_{N_t} \tau_M} \end{bmatrix}$$

where $\tau(m) = \tau_{\text{ch}}(m)/T_{\text{sym}}$ is the channel delay normalized by OFDM symbol duration T_{sym}, having $m = \{1, 2, \cdots M\}$. To effectively estimate a channel with multiple paths, the guard interval and channel delay profiles must be taken into account. Comparable to DFT basis, the channel response can be sparsely represented in dictionary $\mathbf{D}$ based sparsifying matrix, $\hat{\mathbf{H}}_i = \mathbf{D}\mathbf{B}_{si}$. The $\mathbf{B}_{si} \in \mathbb{C}^{M \times N_r}$ is the sparse matrix such that $(\|\mathbf{B}_{si}\|_0 \ll N_t)$. Equation (1) can be rewritten as:

$$\mathbf{Y}_{pi} = \mathbf{A}\hat{\mathbf{H}}_i + \mathbf{N}_i = \mathbf{A}\mathbf{D}\mathbf{B}_{si} + \mathbf{N}_i \tag{13}$$

where $\mathbf{A}$ is the measurement matrix, it can be constructed by utilizing the transmitted pilots $\mathbf{X}_p$ and sub DFT basis or sub-dictionary. In further detail, a sub-dictionary $\mathcal{D} \in \mathbb{C}^{P_t \times N_t}$ can be derived from dictionary $\mathbf{D} \in \mathbb{C}^{N_t \times M}$, subsequently, the measurement matrix $\mathbf{A}$ can be expressed as:

$$\mathbf{A}_{\mathcal{D}} = \mathbf{X}_p^H \otimes \mathcal{D} \tag{14}$$

The basis $\mathcal{D}$ comprises of specially selected rows of $\mathbf{D}$ associated with the pilot locations [21], as shown in equation (15). The transmitted pilots $\mathbf{X}_p$ are built according to Rademacher distribution (RD) consisting of equally likely symbols taken from $[+1, -1]$. Subsequently, they are multiplied by a phase rotation and can be formulated as $\mathbf{X}_{P_n} = [P_1 e^{-j\pi l_1} P_2 e^{-j\pi l_2} \dots P_t e^{-j\pi l_{P_t}}]$, where P_t represents the total number of transmitted pilots for the n^{th} transmit antenna.

$$\mathbf{A}_{\mathcal{D}} = \mathbf{X}_p^H \otimes \begin{bmatrix} e^{-j2\pi k_1 \tau_1} \dots e^{-j2\pi \cdot k_1 \tau_{N_t}} \\ e^{-j2\pi k_2 \tau_1} \dots e^{-j2\pi \cdot k_2 \tau_{N_t}} \\ \vdots \\ e^{-j2\pi k_{P_t} \tau_1} \dots e^{-j2\pi \cdot k_{P_t} \tau_{N_t}} \end{bmatrix} \tag{15}$$

where $\mathbf{A}_{\mathcal{D}} \in \mathbb{C}^{P_t \times N_t}$ represents the measurement matrix constructed from known pilots and dictionary. For comparison, the measurement matrix $\mathbf{A}$ is computed using two equivalent approaches; firstly, starting from $\mathbf{D}$ as already discussed in equation (15), and secondly, through DFT basis ϕ, that is shown in equation (16).

$$\mathbf{A}_{\phi} = \mathbf{X}_p^H \otimes \begin{bmatrix} e^{-j2\pi k_{1,1}} & \dots & e^{-j2\pi k_{1,N_t}} \\ e^{-j2\pi k_{2,1}} & \dots & e^{-j2\pi k_{2,N_t}} \\ \vdots & \ddots & \\ e^{-j2\pi k_{P_t,1}} & \dots & e^{-j2\pi k_{P_t,N_t}} \end{bmatrix} \tag{16}$$

$\mathbf{A}_{\phi} \in \mathbb{C}^{P_t \times N_t}$ is calculated from DFT sub matrix. It is noteworthy that the construction of the measurement matrix $\mathbf{A}_{\mathcal{D}}$ is different from $\mathbf{A}_{\phi}$ used by [8], [16].

In the case $\mathbf{D}$, $\mathbf{A}$ and $\mathbf{Y}_{pi}$ are known in equation (13), the solution of $\|\mathbf{B}_{si}\|_0$ will yield the channel estimate. Solving $\|\mathbf{B}_{si}\|_0$ will provide an exact solution, however, it is NP-hard problem. The severity of predicament can be reduced if we relax l_0 norm to some higher norm such that l_2 norm i.e., min $\|\mathbf{B}_{si}\|_{1,2}$. Where, the $\|\mathbf{B}_{si}\|_{1,2}$ is the summation of l_2 norm of each row in $\mathbf{B}_{si}$ [5]. The aforementioned optimization problem can be formulated for joint channel estimation at BS [30], as shown in equation (17)

$$\min_{\{\mathbf{B}_{si} \forall_i\}} \sum_{i=1}^{K} \|\mathbf{Y}_{pi} - \mathbf{A}\mathbf{D}\mathbf{B}_{si}\|_2 \leq \epsilon \tag{17}$$

The above mentioned minimization problem is very challenging, due to both individual and joint sparsity among different users. By solving the above optimization problem, instead of obtaining a channel estimate with training overhead proportional to N_t, it is possible to obtain a good channel

estimate which is proportional to the sparsity level s of the sparse channel with $s \ll N_t$.

B. CONSTRAINTS FOR STABLE RECOVERY

Several constraints and properties have been discussed in the literature for the stable sparse signal reconstruction of the measurement matrix $\mathbf{A}$. For example, the accuracy of recovery parameters can be determined in case $\mathbf{A}$ satisfies the null space property (NSP), restricted isometry property (RIP) or mutual coherence, however, NSP or RIP is difficult to measure. For guaranteed sparse signal recovery, it is adequate to only satisfy the mutual coherence property for the measurement matrix $\mathbf{A}$ [21], [31].

Theorem 1: In mutual coherence, the maximum absolute and normalized inner product between the columns of measurement matrix $\mathbf{A}$ is analyzed. It determines the correlation among the columns of the measurement matrix and can be formulated as in [21]:

$$\mu\{\mathbf{A}\} = \max_{1\le i\le j\le N_t, i\neq j} \frac{|\mathbf{A}_i, \mathbf{A}_j|}{\|\mathbf{A}_i\|_2 \|\mathbf{A}_j\|_2} = \max_{1\le i\le j\le N_t, i\neq j} \frac{|\mathbf{A}_i^H \mathbf{A}_j|}{\|\mathbf{A}_i\|_2 \|\mathbf{A}_j\|_2} \tag{18}$$

The mutual coherence of $\mathbf{A}$ at the pilot locations can be written as:

$$\mu\{A\} = \max_{1\le i\le j\le N_t, i\neq j} \times \frac{\sum_{l=1}^{P_t} (\mathbf{X}_p(k_l) e^{-j2\pi k_l \tau_i/N_t})(\mathbf{X}_p(k_l) e^{j2\pi k_l \tau_j/N_t})}{\sum_{l=1}^{P_t} |\mathbf{X}_{pl}|^2} \tag{19}$$

$$= \max_{1\le i\le j\le N_t, i\neq j} \frac{\sum_{l=1}^{P_t} |\mathbf{X}_p(k_l)|^2 e^{j2\pi k_l \tau (j-i)/N_t})}{\sum_{l=1}^{P_t} |\mathbf{X}_p(k_l)|^2} \tag{20}$$

Since our complex pilots are equi-power, we can conveniently assume that: $(|(\mathbf{X}_p(k_1)| = |(\mathbf{X}_p(k_2)| ..... = |(\mathbf{X}_p(k_{P_t})| = 1)$. Consequently, the normalized equation (20) can be re-written as:

$$\mu\{\mathbf{A}\} = \max_{1.\le i\le j\le N_t, i\neq j} \sum_{l=1}^{P_t} \frac{1}{P_t} 1 e^{j2\pi k_l \tau (j-i)/N_t}) \tag{21}$$

From equation (21) it is evident that the value of $\mu\{\mathbf{A}\}$ will not rely on the pilot symbols, instead it only depends on the selected columns of dictionary $\mathbf{D}$. The dictionary $\mathbf{D}$ is a sparsifying matrix and is employed to realize the channel, it is mostly designed before channel estimation. Consequently, if $\mathbf{D}$ is constructed tactfully, during the channel estimation $\mu\{\mathbf{A}\}$ will remain small and fixed. In short, from theorem 1 it is obvious that, for efficient channel estimation, a minimum coherence value is desired, therefore, the objective is to minimize $\mu\{\mathbf{A}\}$. The value of μ is generally in the range of [0-1], and the smaller value is desired to achieve better recovery [31].

The sparse channel reconstruction is based on the assumption that the channel response is sparse. In other words, the channel energy is uniformly distributed among the few dominant taps. However, the exact positions of these dominant taps are not known a priori and must be estimated for the effective channel measurement [32]. To estimate the exact positions, a redundant basis $\mathbf{D}$ is required. In conclusion, from the reduced and quantized pilots $\mathbf{Y}_{pi}$ and the dictionary $\mathbf{D}$, the channel estimate $\hat{\mathbf{H}}_i$ is obtained at the BS by utilizing the CS algorithms (i. e., Q-PJOMP, and Q-PJIHT), discussed in detail in the following subsection.

C. ALGORITHMS FOR SPARSE CHANNEL ESTIMATION

In some cases, it has been observed that the sparse data exhibit special structure in its measurement, for example, in a certain measurement, the sparse coefficients may form a group or block of zero and non zero entries. Such sparsity structure is known as group/block sparsity [12], [33]. This sparsity structure can be formulated by employing multiple measurements, known as joint sparsity [8], [16], [34]. In joint sparsity, each row of N vectors tends to be zero or non-zero simultaneously. The research work [35] demonstrates that the performance gain can be achieved if such structures are exploited in sparse measurements of the observed data. Consequently, the hidden joint sparsity structure of the correlated multi-users MIMO channel is exploited in the proposed DCS algorithms Q-PJOMP and Q-PJIHT. In both methods, it has been assumed that the sparsity information is known at the BS, which can be determined using slow-timescale stochastic learning [36], [37]. The proposed Q-PJOMP and Q-PJIHT are presented in Algorithm 1 and 2, respectively, where 2-step quantization methods are labeled as A, 1-bit binary algorithm is indicated with B, and the common lines to both algorithms are marked with AB. If the algorithm 1 and 2 are viewed without label A, they transform into 1-bit Q-PJOMP/Q-PJIHT. Similarly, inspecting the algorithms without label B will convert them into the 2-step Q-PJOMP/Q-PJIHT. Both of the algorithms are presented subsequently in the following subsections.

1) CHANNEL RECOVERY WITH Q-PJOMP (ALGORITHM 1)

The current subsection describes Q-PJOMP for partially joint channel estimation. Although several research works present the solution with joint sparsity through OMP based approaches [38], [39], the system with individual and partially joint sparsity still requires additional investigation. In [39] simultaneous joint OMP (SOMP) for multiple measurements sharing joint support is investigated. In the current work, the system with individual and partially joint support is considered. To handle this kind of structured sparsity, in the literature some works are already present. For example, in [40] a generalized joint sparsity model (GJSM) is developed.

Another research work [8] applies GJSM to exploit the hidden joint sparsity in massive MIMO channels for efficient resource utilization. Their main objective is to reduce the training overhead, which turns out in lowering the feedback overhead, although the number of feedback bits is still very

Algorithm 1 Quantized Partially Joint-OMP

1 (AB): **Input:** $\mathbf{Y}_{pi} : i \in K, \mathbf{D}$
2 (AB): Measurement matrix $\leftarrow \mathbf{A}$, support $= s_c, s_i$
3 (A) : Averaged quantized amplitude $\leftarrow Q(\mu)$
4 (AB): $\{Thresholds : \eta_1 < 1, \eta_2 > 1\}$
5 (AB): **Output:** Channel estimate $\hat{\mathbf{H}}_i$

Procedure:
Step 1 - Initialization:
6 (AB): support :$\{\Theta_i = 0, \Theta_c = 0\}$
7 (A) : $\mu' \leftarrow Q'(\mu)$
8 (AB): Compute $\mathbf{Y}_{pi}$ from (8) and $\mathbf{A}$ from (15) or (16)
9 (A) : The angle of $\mathbf{Y}_{pi}$ and the de-quantized amplitude are used to rebuilt $\mathbf{Y}_{pi}$
$\mathbf{Y}_{pi} \leftarrow \mu'\left(cos\theta_{\mathbf{Y}_{pi}} + jsin\theta_{\mathbf{Y}_{pi}}\right)$ and
10 (AB): Initialize residual $\mathbf{Res}_i = \mathbf{Y}_{pi}$
11 (AB): **Step 2 - Identify the common support:**
while $t \leq s_c$ **or** $\|\mathbf{A}^H \mathbf{Res}_i\|_F^2 \geq \eta_1$ **then**
identify the common support Θ_c that solves optimization problem
$\Theta_t \leftarrow Argmax \sum_i^K |\langle \mathbf{A}_s^H \mathbf{Res}_i \rangle|$
$\Theta_c \leftarrow \Theta_c \overset{s}{\cup} \Theta_t$
12 (AB): Determine the orthogonal projector $\mathbf{P}_o$ onto the span of the atoms indexed in Θ_c
$\mathbf{P}_o \leftarrow (\mathbf{A}_s)(\mathbf{A}_s)^\dagger$
$\mathbf{y}_o \leftarrow \mathbf{P}_o \mathbf{Y}_{pi}$
Residual Update:
13 (B) : $\mathbf{Res}_i \leftarrow \mathbf{Y}_{pi} - \text{sign}(\text{Re}(\mathbf{y}_o)) + j\text{sign}(\text{Im}(\mathbf{y}_o))$
14 (A) : $\mathbf{Res}_i \leftarrow \mathbf{Y}_{pi} - \mu'\left(\cos\theta_{y_o} + j\sin\theta_{y_o}\right)$
$t \leftarrow t + 1$

end
15 (AB): **Step 3 - Identify the individual support: initialize** $\Theta_i \leftarrow \Theta_c$
while $t_i \leq |s_i - s_c|$ **or** $\|\mathbf{Res}_i\|_F^2 \leq \eta_2$ **then**
$\Theta_{ti} \leftarrow \text{Arg max}|\langle \mathbf{A}_s^H \mathbf{Res}_i \rangle|$
$\Theta_i \leftarrow \Theta_i \overset{s}{\cup} \Theta_{ti}$
Residual Update:
16 (B) : $\mathbf{Res}_i \leftarrow \mathbf{Y}_{pi} - (\text{sign}(\text{Re}(\mathbf{y}_o)) + j\text{sign}(\text{Im}(\mathbf{y}_o))$
17 (A) : $\mathbf{Res}_i \leftarrow \mathbf{Y}_{pi} - \mu'\left(\cos\theta_{y_o} + j\sin\theta_{y_o}\right)$
$t_i \leftarrow t_i + 1$
end

Step 4 - Channel estimate The channel estimate is obtained
18 (AB): $\mathbf{B}_{si} = (\mathbf{A}_{\Theta_i})^\dagger \mathbf{Y}_{pi}$
$\hat{\mathbf{H}}_i = \mathbf{D}\mathbf{B}_{si}$

high. Therefore, our focus is not only to reduce the number of measurements by taking advantage of joint sparsity structure among users channel but also to limit the number of feedback bits through a 2-step quantization process, while ensuring efficient reconstruction.

To reduce the training overhead, our approach is inspired by [8] and [9]. In [8] the training overhead is lowered by considering the partially joint support structure among the user's channel. While in our previous work [9] it is achieved by utilizing a flexible and robust dictionary-based measurement matrix. Besides reducing the training overhead, this work also lowers the number of feedback bits by introducing a 2-step quantization method. Q-PJOMP is based on a greedy algorithm select-discard OMP (SD-OMP) [40]. It can perform individual as well as partially joint sparse channel approximation by choosing the ''best match'' projection of multi-user quantized received pilots onto the span of the measurement matrix.

Initially, the received pilots are recomputed applying the de-quantized mean vector μ and the angle of 1-bit quantized vector $\mathbf{Y}_{pi}$ (line 9 in algorithm 1). Subsequently, the algorithm detects the common support for all the users, which consists of jointly selecting one column from the measurement matrix $\mathbf{A}$ in each iteration (line 11 step-2 in algorithm 1). The main intuition behind this selection is to determine an atom (column index) that contributes to the maximum amount of residual energy across all signals [41]. The indices that appear in common support Ω_c are likely to be estimated by most of the users. In each iteration, new support is appended to the existing support set unless a stopping criterion is satisfied. Another important step is to orthogonalize the residual $\mathbf{Res}_i$ with the selected atom (column) of the measurement matrix (line 12 of Algorithm 1) so that each atom is chosen only once as in [39], [42]. In the next step, the individual support of each user is identified as in standard OMP (lines 15 step-3 in algorithm 1) [42]. Indeed, besides the common support, also a few individual supports may be present in each signal. Note that the indices of the measurement matrix which were already been taken by common support, are discarded by the individual support. The rest of the procedure is the same as the common support identification. Afterward, the complex pilot at each support index is recomputed and updated (lines 13 and 14 for common support and lines 16 and 17 for individual support in algorithm 1) in a design similar to [43], although the overall procedure and problem are different from the proposed. Finally, the channel is estimated using the LS approach (lines 18 step-4 in algorithm 1). To achieve stable recovery, there are different stopping criteria. For example, the algorithm will run for a particular number of iteration or will be executed until the norm of residual remains positive $\|\mathbf{Res}_i\| \geq \eta_2$. In our case, the system will halt if the correlation between the atom of a matrix $\mathbf{A}$ and the residual drops below a threshold η_1;, which we set ($\eta_1 = 0.04$) in our simulation.

2) CHANNEL RECOVERY WITH Q-PJIHT (ALGORITHM 2)

We here describe the Q-PJIHT algorithm to estimate partially joint sparse channel, whose performance is then compared with the previously detailed Q-PJOMP.

Algorithm 2 Quantized Partially Joint-IHT

1 (AB): **Input:**$\{\mathbf{Y}_{pi} : i \in K\}$, $\mathbf{D}$
2 (AB): Measurement matrix $\leftarrow \mathbf{A}\}$
3 (AB): Sparsity level S $\leftarrow s_c, s_i$,
4 (A) : Averaged quantized amplitude $\leftarrow Q(\mu)$
5 (AB): Step size: $\eta \leftarrow \frac{1-\sqrt{\frac{S}{M}}}{M}$
6 (AB): **Output:** Channel estimate $\hat{\mathbf{H}}_i$

Procedure:
Step 1 - Initialization:
7 (A) : $\mu' \leftarrow Q'(\mu)$
8 (AB): Compute $\mathbf{Y}_{pi}$ from (8) and $\mathbf{A}$ from (15) or (16)
9 (AB): The angle of $\mathbf{Y}_{pi}$ and the de-quantized amplitude are used to rebuilt $\mathbf{Y}_{pi}$
10 (A) : $\mathbf{Y}_{pi} \leftarrow \mu'\left(\cos\theta_{\mathbf{Y}_{pi}} + j\sin\theta_{\mathbf{Y}_{pi}}\right)$
11 (AB): Initialize residual $\mathbf{Res}_i^o = \mathbf{Y}_{pi}$
$\mathbf{B}_{si}^o \leftarrow \mathbf{A}^H\mathbf{Res}_i^o$
$\Theta_i \leftarrow 0$
$\Theta_c \leftarrow 0$

Step 2 **while(maximum iteration) do**
12 (AB): $\mathbf{B}_{si}^{n+1} = \mathbf{B}_{si}^n + \eta\mathbf{A}^H\mathbf{Res}_i^n$
13 (AB): Identify the common support Θ_c that solves the optimization problem
$\Theta_t \leftarrow \text{Arg max} \sum_i^K |\langle \mathbf{A}_s^H\mathbf{Res}_i\rangle|$
$\Theta_c \leftarrow \Theta_c \overset{s}{\cup} \Theta_t$
Repeat the process s_c times
14 (AB): Identify the individual support:
Initialize $\Theta_i \leftarrow \Theta_c$
$\Theta_{ti} \leftarrow \text{Arg max}|\langle \mathbf{A}_s^H\mathbf{Res}_i\rangle|$
$\Theta_i \leftarrow \Theta_i \overset{s}{\cup} \Theta_{ti}$
Repeat the process $|s_i - s_c|$ times
15 (AB): Hard threshold all the entries except those indexed in Θ_i
$\mathbf{B}_{si}^{n+1} \leftarrow H_T(\mathbf{B}_{si}^{n+1}, \Theta_i)$
Residual Update :
16 (AB): $\mathbf{y}_o \leftarrow \mathbf{A}_\Theta\ \mathbf{B}_{si}^{n+1}$
17 (B) : $\mathbf{Res}^{n+1} \leftarrow \mathbf{Y}_{pi} - \text{sign}(\text{Re}(\mathbf{y}_o) + j\text{sign}(\text{Im}(\mathbf{y}_o)$
18 (A) : $\mathbf{Res}^{n+1} \leftarrow \mathbf{Y}_{pi} - \mu'\left(\cos\theta_{y_o} + j\sin\theta_{y_o}\right)$
end

19 (AB): **Step 3 - Channel estimate**
The spare channel is estimated by repeating the process until $\mathbf{Y}_{pi}$ is consistent with $\mathbf{B}_{si}$ or the stopping criteria is met. $\hat{\mathbf{H}}_i \leftarrow \mathbf{DB}_{si}$

The research work [16] presents the 1-bit solution with joint sparsity through IHT based approaches. However, the system with joint sparsity and quantized feedback still require improvements. In [44] 1-bit quantization is proposed, later to reduce hardware cost and system processing burden, the research work [16] utilizes this 1-bit per dimension to estimate the partially joint channel in a massive MIMO system. The proposed Q-PJIHT algorithm explores both individual and joint sparsity support as in [16], although we apply a different quantization procedure that includes the amplitude as well as the sign information. The proposed work also selects the optimal step size (η) depending on the sparsity level and dictionary length M. Indeed, since IHT is a gradient descent based algorithm, its step size must be chosen optimally for better convergence [45].

Initially, in Algorithm 2, the received pilots are recomputed by applying the de-quantized mean vector and the angle of 1-bit quantized vector $\mathbf{Y}_{pi}$ (step-1 line 10 in Algorithm 2). Subsequently, after initializing the residual and the support, the algorithm detects the common support (step-2 line 13 in Algorithm 2) for all the users. In the next step, the individual support is calculated (step-2 line 14 in Algorithm 2) in the same way performed in Algorithm 1. The algorithm is based on gradient descent technique that works iteratively to realize a robust estimate. Although standard IHT based algorithms are computationally simple and take less memory than standard OMP based methods, they have several challenges while dealing with the practical problem. As an example, the stable signal recovery highly depends on the selection of step size η, sparsity parameter [45] and number of iterations.

IV. PERFORMANCE EVALUATION

The performance of the proposed scheme is evaluated using various CS techniques and different metrics. The proposed 2-step Q-PJIHT is compared with 1-bit Q-PJIHT whose algorithmic construction is inspired by 1-bit BIHT [16]. The 2-step Q-PJIHT differs from 1-bit BIHT in terms of considering channel direction and partial amplitude information. Similarly, the Q-PJOMP is proposed and its 1-bit and 2-step versions are taken into account for analysis. Q-PJOMP has considered the equivalent joint channel sparsity model as in JOMP [8]. Moreover, both Q-PJIHT and Q-PJOMP differ from [8], [16] in terms of employing more robust sparsifying basis $\mathbf{D}$. The primary goal of this presented framework is to reduce the number of measurements and feedback bits while preserving high performance.

A. SIMULATION SETUP

We consider a MIMO-OFDM system, with $N_t = 128$ transmit antennas, $N_r = 2$ receive antenna and $K = 10$ number of users. The BS uses 256 OFDM subcarriers for downlink transmission while some sub-carriers are employed for pilot sequences and others for data. Specifically, during the first phase of training, only training pilots are sent without data, while in the second phase both data and pilots are transmitted in combo fashion. The dictionary length M is set to 300, whose value can be increased to reduce the error rate at the cost of higher processing. After varying M up to 512 we selected 300 as a good compromise between performance and computation. We assume a system without frequency offset among the local oscillators at the transmitter and the

receivers. Moreover, the multipath fading channel with iid additive white Gaussian noise with zero mean and unit variance is considered. The multipath channel length L_{taps} is set equal to 6, whose consequent ISI is directly solved by the OFDM guard interval G, moreover, the length of G is set to 16. The joint sparsity s_c and the individual sparsity s_i parameters are chosen to be 6 and 10 respectively, where s_c and s_i are defined in Sec. II-B. Table 1 summarizes the system parameters. In the following subsections, the performance analysis is presented evaluating different metrics, i. e., signal to noise ratio (SNR) degradation, normalized mean squared error (NMSE) as in [8] for the channel state information at the transmitter (CSIT), and bit error rate (BER) while detecting the data at each user terminal.

TABLE 1. Simulation parameters.

Parameter	Symbol	Value
Transmit antennas	N_t	128
Receive antennas	N_r	2
Users	K	10
OFDM subcarriers	N_o	256
OFDM guard interval	G	16
QAM Modulations	QAM	4
Multipath channel length	L	6
Dictionary length	M	300

B. SNR DEGRADATION

Although we developed the full system including channel estimation, beamforming procedure and data detection, our main aim is the channel estimation. Thus, in this subsection we specifically evaluate the effect of both the received pilots quantization and the number of employed pilots on the channel estimate through SNR degradation, as shown in Fig. 3. In particular, equation (22) gives the averaged SNR loss obtained using the estimated channel with different numbers of pilot symbols as in [16]. For channel $\mathbf{H}_i$ and its estimate $\hat{\mathbf{H}}_i$, the precoder $\mathbf{w}_i$ is the maximizer of $\|\mathbf{w}_i^H \mathbf{H}_i \mathbf{H}_i^H \mathbf{w}_i\|_2^2$ and $\hat{\mathbf{w}}_i$ is the maximizer of $\|\hat{\mathbf{w}}_i^H \hat{\mathbf{H}}_i \hat{\mathbf{H}}_i^H \hat{\mathbf{w}}_i\|_2^2$, respectively. The overall SNR loss in dB can be calculated as follows

$$\mathrm{SNR}_{deg} = 10\log_{10} \frac{\|\mathbf{w}_i^H \mathbf{H}_i \mathbf{H}_i^H \mathbf{w}_i\|_2^2}{\|\hat{\mathbf{w}}_i^H \hat{\mathbf{H}}_i \hat{\mathbf{H}}_i^H \hat{\mathbf{w}}_i\|_2^2} \tag{22}$$

Around 200 simulations have been performed to obtain the channel estimate $\hat{\mathbf{H}}_i$ and the precoder $\hat{\mathbf{w}}_i$, which are used in equation (22) to calculate the averaged SNR degradation. Fig. 3 analyzes the SNR degradation employing Q-PJIHT or Q-PJOMP having 1-bit or 2-step quantization using either DFT or dictionary-based sparsifying matrices. It can be observed that when the feedback overhead approaches 48 pilots, the SNR degradation exceeds 2 dB in 1-bit method [16], which is more than the degradation achieved by the proposed 2 bit Q-PJOMP.

Overall, the proposed 2-step Q-PJOMP outperforms the other schemes, which is due to several reasons: firstly, Q-PJOMP is more efficient than the IHT based algorithm (Q-PJIHT). Secondly, since the choice of η is critical in IHT [45], a flexible gradient step size η is employed in the Q-PJIHT algorithm. The η is used as per sparsity level requirements in each iteration, contrary to a fixed-size of 0.01 used in [16]. Thirdly, the 2-step quantization preserves both amplitude and phase information rather than having only direction information as in the compared 1-bit feedback method [16]. Finally, our system employs a dictionary-based sparsifying matrix, which, as shown in [9], is more robust to estimate the channel as compared to earlier research works [8], [16]. From Fig. 3 it can be easily noticed that with 64 pilots in the magnified view, the proposed quantized feedback Q-PJOMP with a dictionary-based approach surpasses the performance of almost all feedback based algorithms. Moreover, by increasing the number of pilots, the SNR loss drops down to almost zero for all the solutions. Finally, focusing on the proposed Q-PJOMP method in Fig. 3, the 2-step quantization reflects less degradation as compared to the 1-bit quantization.

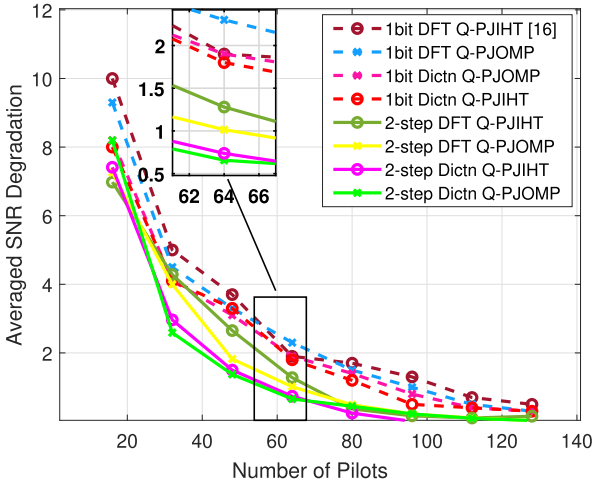


FIGURE 3. Averaged SNR degradation for 1-bit and 2-step quantization using DFT and Dictionary basis.

C. CSIT NMSE ANALYSIS VERSUS COMMON SUPPORT

Fig. 4 illustrates the NMSE of CSIT while increasing the common support s_c, under the same simulation parameters [8] for the comparison purpose. The simulation parameters are enlisted as followed: number of transmit pilot P = 45, $N_t = 160$, $N_r = 2$, K = 40, $s_i = 17$ and $SNR = 28$ dB. NMSE has been calculated as follows:

$$\mathrm{NMSE} = \mathrm{E}\left(\frac{\|\mathbf{H}_i - \hat{\mathbf{H}}_i\|_F}{\|\mathbf{H}_i\|_F}\right) \tag{23}$$

The individual and joint sparsity levels are generated using the spatial channel model (SCM)[2] as in [8], [16]. Fig. 4 reveals as the common support s_c increases among users, the system performance improves, i. e., when the channel

[2]3rd generation partnership project(3GPP) and the international telecommunication union(ITU) has developed a spatial channel model(SCM) to model various urban and rural propagation scenarios [46]

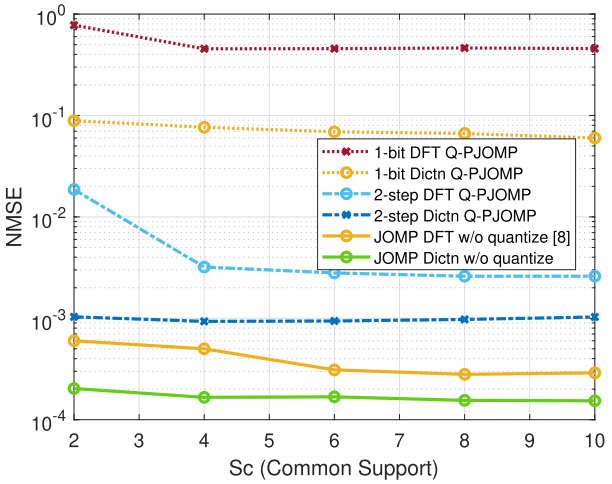


FIGURE 4. NMSE of CSIT vs common support for 1-bit, 2-step quantization using Q-PJOMP and without quantization, under P = 45, $N_t = 160$, $N_r = 2$, K = 40, $s_i = 17$ and transmit SNR = 28 dB.

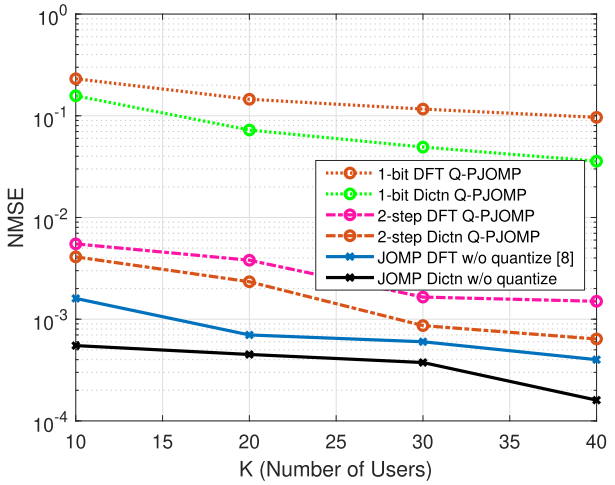


FIGURE 5. NMSE of CSIT vs number of users for 1-bit, 2-step quantization using Q-PJOMP and without quantization, under P = 45, $N_t = 160$, $N_r = 2$, $s_c = 10$, $s_i = 17$ and transmit SNR = 28 dB.

between users is more correlated. The results confirm that the 2-step quantization provides a better estimation than the 1-bit scheme and approaches the techniques without quantization with a much lesser number of bits. The 1-bit quantization method (1-bit DFT Q-PJOMP) in Fig. 4 has the worst performance because it uses only direction information and also DFT basis, instead of the more robust dictionary as in 1-bit Dictn Q-PJOMP. Comparing only the methods without quantization, it can also be observed that the proposed JOMP algorithm with a dictionary basis gives better performance than DFT based JOMP [8]. This point validates one of our contributions that joint channel estimation is more reliable with dictionary basis.

D. CSIT NMSE ANALYSIS VERSUS NUMBER OF USERS

Fig. 5 illustrates the NMSE of CSIT while increasing the numbers of users. The NMSE is calculated by using equation (23). The simulation parameters are enlisted as followed: number of transmit pilot P = 45, $N_t = 160$, $N_r = 2$, $s_c = 10$, $s_i = 17$ and $SNR = 28$ dB and number of users are increased from 10 to 40. From Fig. 5 it can be noticed that the performance improves when the number of users increase. The performance gain is due to the fact that the proposed algorithm utilized the distributed joint sparsity between user's channel matrices to jointly detect the common support. With more number of users, the system will have adequate amplitude knowledge and improved common support resulting in a reliable estimate. It can be observed that the performance is improved if both amplitude and directional information is considered, as in 2-step quantization. All the algorithms utilizing the dictionary basis for the estimate of CSIT have superior performance.

E. CSIT NMSE ANALYSIS VERSUS NUMBER OF RECEIVE ANTENNAS

Fig. 6 illustrates the NMSE of CSIT while increasing the number of antennas at each user by using equation (23). The

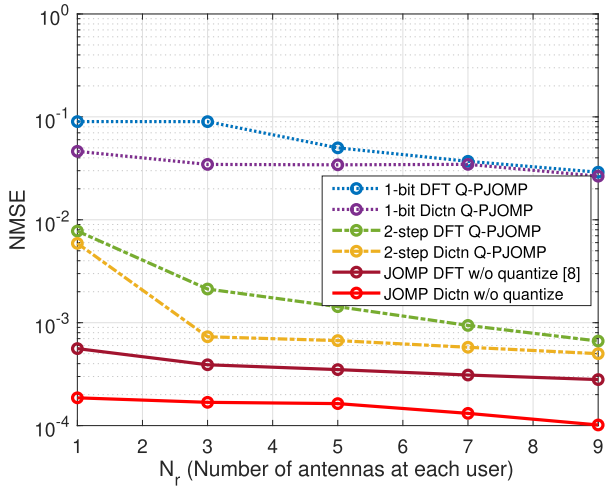


FIGURE 6. NMSE of CSIT vs number of antennas at each user for 1-bit, 2-step quantization using Q-PJOMP and without quantization, under P = 45, $N_t = 160$, user = 40, $s_c = 10$, $s_i = 17$ and transmit SNR = 28 dB.

simulation parameters are enlisted as followed: number of transmit pilot P = 45, $N_t = 160$, users = 40, $s_c = 10$, $s_i = 17$ and SNR = 28 dB and number of antennas at each user are increased from 1 to 9. From Fig. 6 it can be seen that with the increase in number of receive antenna at each user the performance improves, due to the individual joint sparsity present in the channel matrix. Hence, the CSIT quality will be enhanced by increasing N_r. As expected, similarly to Fig. 5 and 4, the performance of 2-step quantization is better than 1-bit quantization. Moreover, with the dictionary basis, the CSIT estimate is improved as compared to the DFT basis in all cases. Hence, by increasing the number of antennas at the mobile station more reliable individual joint support is identified, resulting in a robust channel estimate.

F. BER ANALYSIS

Fig. 7 and 8 present the comparison of BER versus SNR for data detection techniques utilizing SVD and MMSE

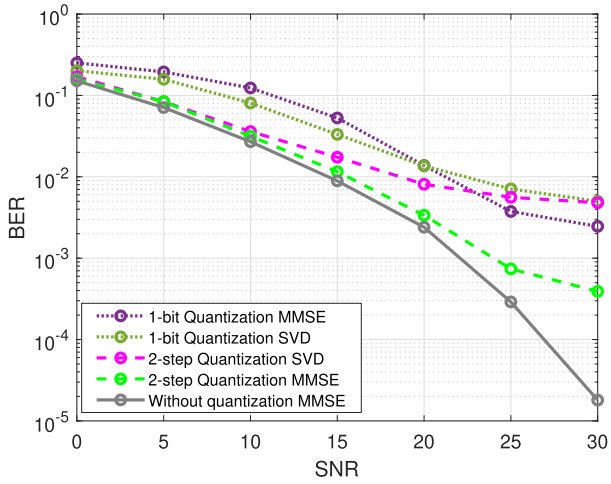


FIGURE 7. Averaged Multi-user detection with quantized CSIT based MMSE beamforming (1-bit, 2-step Q-PJOMP) and without quantization.

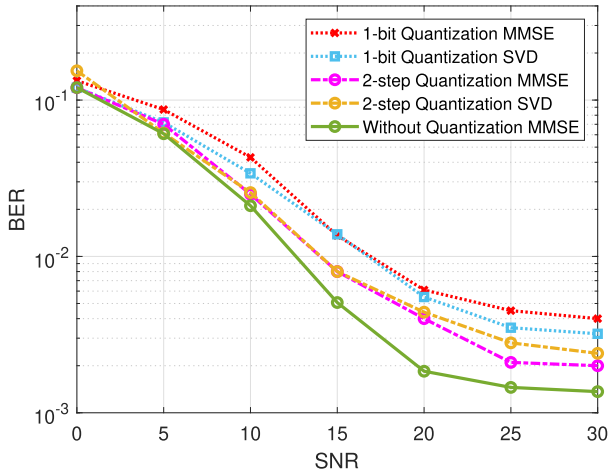


FIGURE 8. Averaged Multi-user data detection with quantized CSIT based beamforming (1-bit, 2-step Q-PJIHT) and without quantization.

based beamforming. The channel estimate required for beamforming is obtained using quantized feedback and applying CS techniques based on JOMP and JIHT. Moreover, the beamforming obtained without quantized feedback is presented for comparison purposes. It is worth noting in Fig. 7 that beamforming based on channel estimate acquired using 1-bit feedback has a wider gap from the algorithms without quantization. Moreover, Q-PJOMP with MMSE based precoding gives better results than its SVD counterpart. As shown in Fig. 7 and 8, the proposed 2-step quantization methods show better performance when applying MMSE beamforming than SVD.

Specifically, at high SNR MMSE outperforms SVD since it is more power-efficient. On the contrary, Fig. 7 and 8 illustrate that both 1-bit versions of the proposed algorithms, i. e., 1-bit Q-PJIHT and Q-PJOMP, show lower performance when employing MMSE than SVD in all the SNR range. The reason may be due to the fact that 1-bit quantization lacks power

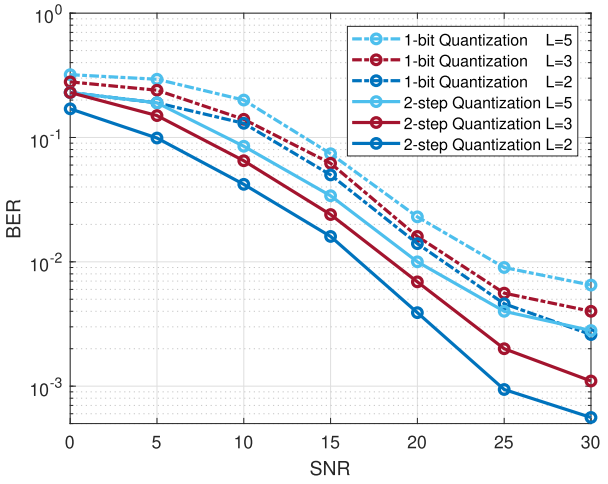


FIGURE 9. Averaged Multi-user detection with quantized CSIT based MMSE beamforming (1-bit, 2-step Q-PJOMP) with varying channel taps *L*.

information while MMSE is a power efficient beamformer, therefore without power information it is not performing well. By comparing Fig. 7 and 8, it can be observed that Q-PJOMP exhibits better performance as compared to Q-PJIHT for almost all the cases with and without quantization. Fig. 9 shows the effect of multipath fading over data detection by employing MMSE beamforming utilizing quantized 1-bit and 2-step feedback using Q-PJOMP based channel estimates. It can be observed that, as expected, by increasing the number of channel taps from 2 to 5 the BER performance in all cases is reduced, though the proposed technique yields the improved results.

G. COMPUTATIONAL COMPLEXITY ANALYSIS

In this subsection, the computational complexity comparison between the proposed IHT based algorithm Q-PJIHT and OMP based algorithm Q-PJOMP is given. The overall computational complexity for Q-PJOMP method is $\mathcal{O}(KsMN_rT)$ and for Q-PJIHT is $\mathcal{O}(IKMN_rT)$, where I is the total number of iterations. The complexity in terms of Big-O notation will be the same for the 1-bit and the 2-step quantization versions of the proposed algorithms, since Big-O notation shows how an algorithm's complexity scale when the number of parameters are increased. However, while processing the algorithms with real hardware, the lower computational complexity of 1-bit method compared to the 2-bit algorithm will be more noticeable in terms of processing time. Table 2 presents the comparison of computational time per user for 2-step and 1-bit Q-PJOMP/Q-PJIHT. Three different transmit antennas setting has been considered: $N_t = [50, 100, 150]$. As expected, the computational time increases with the number of transmit antennas. Furthermore, the 2-step algorithms (Q-PJOMP/Q-PJIHT) are more expensive than 1-bit methods. This is due to the processing of the channel amplitude information and rebuilding complex pilots from both amplitude and direction information, nevertheless, the analysis

TABLE 2. Complexity comparison for mu-massive MIMO under parametric setting P = 45, N = 2, K = 40, s_c = 9, s_i = 17, P = 28 dB AND N_t = 50,100, 150.

Time (s)	N_t=50	N_t=100	N_t=150
2-step Q-PJOMP	1.20×10^{-3}	1.40×10^{-3}	1.60×10^{-3}
1-bit Q-PJOMP	9.24×10^{-4}	1.20×10^{-3}	1.30×10^{-3}
2-step Q-PJIHT	6.86×10^{-2}	7.19×10^{-2}	7.38×10^{-2}
1-bit Q-PJIHT [16]	5.82×10^{-2}	6.94×10^{-2}	7.02×10^{-2}

reveals that the difference is not very significant. Finally, comparing Q-PJOMP with Q-PJIHT, it has been observed that iterative hard thresholding based techniques (1-bit and 2-step quantization) are generally much slower than the orthogonal matching pursuit based techniques since they require a certain number of iteration to reach an optimal point.

V. CONCLUSION

We have presented distributed compressed sensing based channel estimation techniques for the partially joint channel in a massive MIMO system. A novel Q-PJOMP and Q-PJIHT based channel estimation algorithms are proposed that utilize limited quantized feedback. Our system reduces training and feedback overhead for channel estimation by employing a fewer number of pilots along with a limited number of feedback bits. The channel is jointly recovered for all users by applying DCS at BS from 2-step quantized feedback. Results revealed that SNR degradation with 2-step quantized feedback is less than 1 dB for 64 pilots. Furthermore, as the number of pilots grows, the SNR degradation approaches closer to zero. Additionally, the presented dictionary-based system aids in reducing the training and feedback overhead. This is achieved by exploiting an improved CS algorithm and pilot design. Finally, when the channel among users is highly correlated and exhibits added common support, the jointly estimated channel at BS will reduce the computational resources and time. Future work comprises of extending the proposed massive MIMO scheme to localization issues for improving performance [47], [48], which could be achieved by exploiting a large number of serving antennas.

REFERENCES

[1] L. Daza and S. Misra, "Fundamentals of massive MIMO (Marzetta, T., et al; 2016) [book reviews]," *IEEE Wireless Commun.*, vol. 25, no. 1, p. 9, Feb. 2018.

[2] A. Vizziello, P. Savazzi, and K. R. Chowdhury, "A Kalman based hybrid precoding for multi-user millimeter wave MIMO systems," *IEEE Access*, vol. 6, pp. 55712–55722, 2018.

[3] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019.

[4] P. C. Chan, E. Lo, R. Wang, E. S. Au, V. N. Lau, R. Cheng, W. Mow, R. Murch, and K. Letaief, "The evolution path of 4G networks: FDD or TDD?" *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 42–50, Dec. 2006.

[5] Y. Ding and B. D. Rao, "Dictionary learning-based sparse channel representation and estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5437–5451, Aug. 2018.

[6] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

[7] Y. Zhou, M. Herdin, A. M. Sayeed, and E. Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep., 2007, vol. 5, pp. 10–15. [Online]. Available: https://scholar.googleusercontent.com/scholar.bib?q=info:yHsB0lQdIqUJ:scholar.google.com/&output=citation&scisdr=CgXTn0WFEK638vLvxJQ:AAGBfm0AAAAAXhjq3JRBDwpRjJ3VLbYPMyut4l4B-4g7&scisig=AAGBfm0AAAAAXhjq3L46go4XDRifaO-pYZGFmLR5oZge&scisf=4&ct=citation&cd=-1&hl=it

[8] X. Rao and V. K. N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3261–3271, Jun. 2014.

[9] F. Kulsoom, A. Vizziello, H. N. Chaudhry, and P. Savazzi, "Pilot reduction techniques for sparse channel estimation in massive MIMO systems," in *Proc. 14th Annu. Conf. Wireless On-Demand Netw. Syst. Services (WONS)*, Feb. 2018, pp. 111–116.

[10] G. Coluccia, C. Ravazzi, and E. Magli, *Compressed Sensing for Distributed Systems*. Singapore: Springer, 2015.

[11] Y. Xu, G. Yue, and S. Mao, "User grouping for massive MIMO in FDD systems: New design methods and analysis," *IEEE Access*, vol. 2, pp. 947–959, 2014.

[12] W. Huang, Y. Huang, W. Xu, and L. Yang, "Beam-blocked channel estimation for FDD massive MIMO with compressed feedback," *IEEE Access*, vol. 5, pp. 11791–11804, 2017.

[13] M. Duarte, S. Sarvotham, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Joint sparsity models for distributed compressed sensing," in *Proc. Workshop Signal Process. Adaptative Sparse Struct. Represent. (SPARS)*, Dec. 2005.

[14] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," 2009, *arXiv:0901.3403*. [Online]. Available: https://arxiv.org/abs/0901.3403

[15] Z. Gao, L. Dai, Z. Wang, and S. Chen, "Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO," *IEEE Trans. Signal Process.*, vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[16] Z. Zhou, X. Chen, D. Guo, and M. L. Honig, "Sparse channel estimation for massive MIMO with 1-Bit feedback per dimension," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[17] K. S. Hassan, M. Kurras, and L. Thiele, "Performance of distributed compressive sensing channel feedback in multi-user massive MIMO," in *Proc. IEEE 11th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2015, pp. 430–436.

[18] J. Choi, D. J. Love, and U. Madhow, "Limited feedback in massive MIMO systems: Exploiting channel correlations via noncoherent trellis-coded quantization," in *Proc. 47th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2013, pp. 1–6.

[19] T. Inoue and R. W. Heath, "Geodesic prediction for limited feedback multiuser MIMO systems in temporally correlated channels," in *Proc. IEEE Radio Wireless Symp.*, Jan. 2009, pp. 167–170.

[20] A. A. Esswie, M. El-Absi, O. A. Dobre, S. Ikki, and T. Kaiser, "Spatial channel estimation-based FDD-MIMO interference alignment systems," *IEEE Wireless Commun. Lett.*, vol. 6, no. 2, pp. 254–257, Apr. 2017.

[21] A. N. Uwaechia and N. M. Mahyuddin, "A review on sparse channel estimation in OFDM system using compressed sensing," *IETE Tech. Rev.*, vol. 34, no. 5, pp. 514–531, Sep. 2017.

[22] A. Vizziello, I. F. Akyildiz, R. Agusti, L. Favalli, and P. Savazzi, "Cognitive radio resource management exploiting heterogeneous primary users," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–5.

[23] A. Vizziello, I. F. Akyildiz, R. Agustí, L. Favalli, and P. Savazzi, "Characterization and exploitation of heterogeneous OFDM primary users in cognitive radio networks," *Wireless Netw.*, vol. 19, no. 6, pp. 1073–1085, Aug. 2013, doi: 10.1007/s11276-012-0519-z.

[24] A. Mohydeen, P. Chargé, Y. Wang, O. Bazzi, and Y. Ding, "Spatially correlated sparse MIMO channel path delay estimation in scattering environments based on signal subspace tracking," *Sensors*, vol. 18, no. 5, p. 1451, May 2018.

[25] N. Fatema, G. Hua, Y. Xiang, D. Peng, and I. Natgunanathan, "Massive MIMO linear precoding: A survey," *IEEE Syst. J.*, vol. 12, no. 4, pp. 3920–3931, Dec. 2018.

[26] A. Kammoun, A. Muller, E. Bjornson, and M. Debbah, "Linear precoding based on polynomial expansion: Large-scale multi-cell MIMO systems," *IEEE J. Sel. Top. Signal Process.*, vol. 8, no. 5, pp. 861–875, Oct. 2014.

[27] R. Zhang, C. Choy Chai, and Y.-C. Liang, "Joint beamforming and power control for multiantenna relay broadcast channel with QoS constraints," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 726–737, Feb. 2009.
[28] J. Dai, A. Liu, and V. K. N. Lau, "Joint channel estimation and user grouping for massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 67, no. 3, pp. 622–637, Feb. 2019.
[29] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
[30] R. Alesii, P. D. Marco, F. Santucci, P. Savazzi, R. Valentini, and A. Vizziello, "Multi-reader multi-tag architecture for UWB/UHF radio frequency identification systems," in *Proc. Int. EURASIP Workshop RFID Technol. (EURFID)*, Oct. 2015, pp. 28–35.
[31] S. Foucart and H. Rauhut, "A mathematical introduction to compressive sensing," *Bull. Amer. Math.*, vol. 54, pp. 151–165, May 2017.
[32] Y. Ding and B. D. Rao, "Compressed downlink channel estimation based on dictionary learning in FDD massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.
[33] S. Banou, M. Swaminathan, G. Reus Muns, D. Duong, F. Kulsoom, P. Savazzi, A. Vizziello, and K. R. Chowdhury, "Beamforming galvanic coupling signals for IoMT implant-to-relay communication," *IEEE Sensors J.*, vol. 19, no. 19, pp. 8487–8501, Oct. 2019.
[34] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse MIMO channels with common support," *IEEE Trans. Commun.*, vol. 60, no. 12, pp. 3705–3716, Dec. 2012.
[35] J. Huang, T. Zhang, and D. Metaxas, "Learning with Structured Sparsity," Mar. 2009, *arXiv:0903.3002*. [Online]. Available: https://arxiv.org/abs/0903.3002
[36] L. Bottou and N. Murata, "Stochastic approximations and efficient learning," in *The Handbook of Brain Theory and Neural Networks*, 2nd ed. Cambridge, MA, USA: MIT Press, 2002.
[37] A. F. Molisch, A. Kuchar, J. Laurila, K. Hugl, and R. Schmalenberger, "Geometry-based directional model for mobile radio channels—Principles and implementation," *Eur. Trans. Telecommun.*, vol. 14, no. 4, pp. 351–359, 2003.
[38] W. Shen, L. Dai, Y. Shi, B. Shim, and Z. Wang, "Joint channel training and feedback for FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8762–8767, Oct. 2016.
[39] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168405002227
[40] J. Liang, Y. Liu, W. Zhang, Y. Xu, X. Gan, and X. Wang, "Joint compressive sensing in wideband cognitive networks," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2010, pp. 1–5.
[41] W. Xu, Z. Li, Y. Tian, Y. Wang, and J. Lin, "Perturbation analysis of simultaneous orthogonal matching pursuit," *Signal Process.*, vol. 116, pp. 91–100, Nov. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165168415001449
[42] J. Wen, Z. Zhou, Z. Liu, M.-J. Lai, and X. Tang, "Sharp sufficient conditions for stable recovery of block sparse signals by block orthogonal matching pursuit," *Appl. Comput. Harmon. Anal.*, vol. 47, no. 3, pp. 948–974, Nov. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1063520318300253
[43] S. Sparrer and R. F. Fischer, "Soft-feedback OMP for the recovery of discrete-valued sparse signals," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 1461–1465.
[44] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2082–2102, Apr. 2013.
[45] Y. Zhang, Y. Huang, H. Li, P. Li, and X. Fan, "Conjugate gradient hard thresholding pursuit algorithm for sparse signal recovery," *Algorithms*, vol. 12, no. 2, p. 36, Feb. 2019.
[46] *Universal Mobile Telecommunications System (UMTS); Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations (3GPP tr 25.996 version 11.0.0 Release 11) Intellectual Property Rights*, document G. T. 25.996, 2017. [Online]. Available: http://www.3gpp.org/DynaReport/25996.htm
[47] A. Vizziello, S. Kianoush, L. Favalli, and P. Gamba, "Location based routing protocol exploiting heterogeneous primary users in cognitive radio networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2013, pp. 2890–2894.
[48] S. Kianoush, A. Vizziello, and P. Gamba, "Energy-efficient and mobile-aided cooperative localization in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3450–3461, May 2016.

FARZANA KULSOOM received the B.Sc. degree in computer engineering and the master's degree in telecommunication engineering from the University of Engineering and Technology, Taxila, Pakistan, in 2006 and 2012, respectively. She was a Research Associate with the Center of Excellence for ASIC Design and DSP and participated in several industrial projects and research activities. She is currently pursuing the Ph.D. degree with the University of Pavia, Italy, in the areas of signal processing and wireless communication. Her current research interests are MIMO-based communication, next-generation communication systems, and machine learning applied to signal processing.

ANNA VIZZIELLO received the Ph.D. degree in electronics and computer science from the University of Pavia, Italy, in 2011. She is currently a Research Assistant with the Telecommunication and Remote Sensing Laboratory, University of Pavia, Italy. From 2007 to 2009, she also collaborated with the European Centre for Training and Research in Earthquake Engineering (EUCENTRE), working on the transmission of biomedical data. From 2009 to 2010, she was a Visiting Researcher with the Broadband Wireless Networking Laboratory, Georgia Institute of Technology, Atlanta, GA, USA, from summer 2009 and summer 2010, with the Universitat Politècnicade Catalunya, Barcelona, Spain, and with Northeastern University, Boston, MA, USA, from winter 2011 to summer 2016. Her research interests are intrabody networks, cognitive radio networks, and 5G systems. Recently, she has been included in the 2018 list of "N2Women: Rising Stars in Computer Networking and Communications" supported by the IEEE Communication Society.

HASSAN NAZEER CHAUDHRY received the B.Sc. degree in computer engineering from the University of Engineering and Technology at Taxila, Pakistan, in 2006, and the M.Sc. degree in electrical and ICE from the University of Engineering and Technology at Taxila, Pakistan, and Technische Universität Darmstadt, Germany, in 2009 and 2015. He is currently pursuing the Ph.D. degree with the Department of Electronics, Information, and Bioengineering (DEIB), Politecnico di Milano, Italy. He worked at several international companies, including the Center of Excellence for ASIC design and DSP, UET Taxila (Pakistan), Siemens (Germany), MIMO On GmbH (Germany), Fiat (Italy), Basic Net (Italy), and Algo Systems S.a.S. (Italy) and has over ten years of experience in industry. His current topic of research includes large scale time-evolving data sets coming from different domains, including telecommunications, data analytic over big data using machine learning techniques and pattern detection in temporarily changing graphs.

PIETRO SAVAZZI received the Laurea degree in electronics engineering and the Ph.D. degree in electronics and computer science from the University of Pavia, Italy, in 1995 and 1999, respectively. In 1999, he joined Ericsson Lab, Milan, Italy, as a System Designer, working on broadband microwave systems. In 2001, he moved to Marconi Mobile, Genoa, Italy, as a System Designer in the field of 3G wireless systems. Since 2003, he has been working at the University of Pavia, where he is currently an Assistant Professor, teaching two courses on signal processing and wireless sensor networks. His main research interests are in wireless communication and sensor systems.