

# CSIT ESTIMATION AND FEEDBACK FOR FDD MULTI-USER MASSIVE MIMO SYSTEMS

Xiongbin Rao<sup>\*</sup>, Vincent K.N. Lau<sup>\*</sup>, and Xiangming KONG<sup>†</sup>

<sup>\*</sup>Department of ECE, the Hong Kong University of Science and Technology, Hong Kong

<sup>†</sup> Huawei Technologies CO., LTD., P.R. China

## ABSTRACT

To fully utilize the spatial multiplexing gains or array gains of massive MIMO, the channel state information must be obtained at the transmitter side (CSIT). However, conventional CSIT estimation approaches are not suitable for FDD massive MIMO systems because of the overwhelming training and feedback overhead. In this paper, we consider multi-user massive MIMO systems and deploy the compressive sensing (CS) technique to reduce the training as well as the feedback overhead in the CSIT estimation. We propose a distributed compressive CSIT estimation and feedback scheme to exploit the hidden joint sparsity structure in the user channel matrices and we obtain simple insights into how the joint channel sparsity can be exploited to improve the CSIT recovery performance.

**Index Terms**— Massive MIMO, CSIT estimation and feedback, compressive sensing (CS).

## 1. INTRODUCTION

Massive MIMO can greatly enhance the wireless communication capacity due to the increased degrees of freedom, and there is intense research interest in the applications of massive MIMO in next generation wireless systems [1]. To fully utilize the spatial multiplexing gains and the array gains of massive MIMO [2], knowledge of channel state information at the transmitter (CSIT) is essential. In TDD massive MIMO systems, the CSIT can be obtained by exploiting the channel reciprocity using uplink pilots [1]. Hence, many works have considered massive MIMO of TDD systems [1,3]. On the other hand, as FDD is generally considered to be more effective for systems with symmetric traffic and delay-sensitive applications [4] and the most cellular systems today employ FDD, it is also of great interest to consider massive MIMO of FDD systems [5]. To obtain CSIT at the base station (BS) of FDD systems, conventional works [6–8] require the BS first transmits downlink pilot symbols so that the user can estimate the downlink CSI locally using the least square (LS) [8] or minimum mean square error (MMSE) [7]. The estimated CSI are then fed back to the BS via uplink signaling channels. However, using these conventional CSI estimation techniques, the number of independent pilot symbols required at the BS has to scale linearly with the number of transmit antennas  $M$  at the BS (i.e.  $O(M)$ ). For massive MIMO, as  $M$  becomes very large, the pilot training overhead (downlink) as well as the CSI feedback overhead (uplink) would be prohibitively large. In addition, the number of independent pilot symbols available is limited by the channel coherence time [1]. Hence, a new CSIT estimation and feedback design will be needed to support FDD massive MIMO systems.

In massive MIMO systems, as has been identified by many experimental studies [9,10], the user channel matrices tend to be jointly

sparse due to the shared and limited local scatterers [11]. Hence, it is desirable to exploit the hidden joint sparsities in the CSIT estimation and feedback process. In [12], a CS-based channel estimation method is proposed to exploit the sparse multipath channels in time, frequency as well as spatial domains in MIMO systems. By exploiting the spatial sparsity using CS in massive MIMO systems, it is shown that only  $O(s \log M)$  training<sup>1</sup> overhead [12] is needed and this represents a substantial reduction of the CSIT estimation overhead compared with the conventional LS approach. However, these works [12, 13] have considered the point-to-point system only. There are several first order technical challenges associated with the extension of the existing CS-based CSIT estimation techniques to multi-user massive MIMO systems of FDD systems.

- **How to exploit the joint channel sparsity among different users distributively.** In multi-user massive MIMO systems, the user channel matrices may be jointly sparse if they share common local scattering clusters [9, 10], as illustrated in Figure 2. Therefore, it is highly desirable to exploit not only the per-link channel sparsity but also the *joint sparsity structure* to further reduce the CSIT estimation and feedback overhead.
- **Tradeoff analysis between the CSIT estimation quality and the joint channel sparsity.** Besides the algorithm development challenge above, it is also desirable to obtain design insights into how the joint channel sparsity can affect the CSIT estimation performance.

*Notations:* Uppercase and lowercase boldface letters denote matrices and vectors respectively. The operators  $(\cdot)^H$ ,  $(\cdot)^\dagger$ ,  $|\cdot|$ ,  $I_{\{\cdot\}}$ ,  $\|\cdot\|_F$ ,  $O(\cdot)$ ,  $o(\cdot)$  and  $\Pr(\cdot)$  are the conjugate transpose, Moore-Penrose pseudoinverse, cardinality, indicator function, Frobenius norm, big-O notation, little-o notation, and probability operator respectively;  $\text{supp}(\mathbf{h})$  is the index set of the non-zero entries of vector  $\mathbf{h}$ ;  $\mathbf{A}_\Omega$  and  $\mathbf{A}^\Omega$  denote the sub-matrices formed by collecting the columns and rows, respectively, of  $\mathbf{A}$  whose indexes are in set  $\Omega$ .

## 2. SYSTEM MODEL

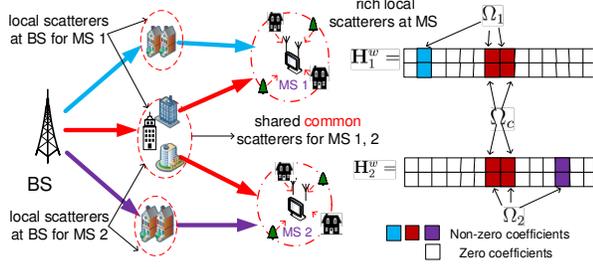
### 2.1. Multi-user Massive MIMO System

Consider a flat block-fading multi-user massive MIMO system operating in FDD mode. There is one BS and  $K$  users in the network as illustrated in Figure 1, where the BS has  $M$  antennas ( $M$  is large) and each user has  $N$  antennas. To estimate the downlink CSI, the BS broadcasts a sequence of  $T$  training pilot symbols on its  $M$  antennas, as illustrated in Figure 1. Denote the transmitted pilot signal<sup>2</sup> from the BS in the  $j$ -th time slot as  $\mathbf{x}_j \in \mathbb{C}^{M \times 1}$ ,  $j = 1, \dots, T$ . The received signal vector at the  $i$ -th user in the  $j$ -th time slot  $\mathbf{y}_{ij} \in \mathbb{C}^{N \times 1}$

<sup>1</sup> $s$  denotes the sparsity level, i.e., the number of non-zero spatial channel paths.

<sup>2</sup>We discuss in Section 3.2 on how to choose the pilot symbols  $\{\mathbf{x}_j\}$ .

This work is funded by RGC614913 and Huawei Technologies CO., LTD., P.R. China.



**Fig. 1.** Illustration of joint channel sparsity structure due to the limited and shared local scattering effect at the BS side.  $\Omega_c$  is the support of the *common* scatterers shared by all users, while  $\Omega_i$  is the support of the individual scatterers for the  $i$ -th user.

can be expressed as

$$\mathbf{y}_{ij} = \mathbf{H}_i \mathbf{x}_j + \mathbf{n}_{ij}, j = 1, \dots, T, \quad (1)$$

where  $\mathbf{H}_i \in \mathbb{C}^{N \times M}$  is the quasi-static channel matrix from the BS to the  $i$ -th user and  $\mathbf{n}_{ij} \in \mathbb{C}^{N \times 1}$  is the complex Gaussian noise with zero mean and unit variance. Let  $\mathbf{X} = [\mathbf{x}_1 \ \dots \ \mathbf{x}_T] \in \mathbb{C}^{M \times T}$ ,  $\mathbf{Y}_i = [\mathbf{y}_1 \ \dots \ \mathbf{y}_T] \in \mathbb{C}^{N \times T}$  and  $\mathbf{N}_i = [\mathbf{n}_{i1} \ \dots \ \mathbf{n}_{iT}] \in \mathbb{C}^{N \times T}$  be the concatenated transmitted pilots, received signal, and noise vectors respectively; the signal model (1) can be equivalently written as

$$\mathbf{Y}_i = \mathbf{H}_i \mathbf{X} + \mathbf{N}_i, \quad (2)$$

where  $\text{tr}(\mathbf{X}\mathbf{X}^H) = PT$  and  $P$  is the average transmit power at the BS per time slot.

## 2.2. Joint Sparse Massive MIMO Channel

Assume a uniform linear array model for the antennas at the BS and the users, and use the angular domain representation [14], the channel matrix  $\mathbf{H}_i$  can be expressed as

$$\mathbf{H}_i = \mathbf{A}_R \mathbf{H}_i^w \mathbf{A}_T^H,$$

where  $\mathbf{A}_R \in \mathbb{C}^{N \times N}$  and  $\mathbf{A}_T \in \mathbb{C}^{M \times M}$  denote the unitary matrices for the angular domain transformation at the user side and BS respectively,  $\mathbf{H}_i^w \in \mathbb{C}^{N \times M}$  is the angular domain channel matrix (see physical explanation of this model in [14]). In multi-user massive MIMO systems, due to the limited local scattering effects at the BS side, the angular domain channel matrices  $\{\mathbf{H}_i^w\}$  are in general sparse. Furthermore, from extensive experimental studies of the multi-user massive MIMO systems [9, 10], we have the following two important observations on the angular domain channel matrices  $\{\mathbf{H}_i^w\}$ :

- **Observation I (Sparsity Support within Individual Channel Matrix):** The row vectors within an  $\mathbf{H}_i^w$  usually have the same sparsity support, i.e.,  $\text{supp}(\mathbf{h}_{i1}) = \text{supp}(\mathbf{h}_{i2}) = \dots = \text{supp}(\mathbf{h}_{iN}) \triangleq \Omega_i$ , where  $\mathbf{h}_{ij}$  is the  $j$ -th row of  $\mathbf{H}_i^w$ . This is due to the limited scattering at the BS side, and relatively rich scattering at the users (as illustrated in Figure 1).
- **Observation II (Partially Shared Support between Different Channel Matrices):** When the users are physically close to each other, they share some local scattering clusters at the BS [15], as illustrated in Figure 1. Hence, the channel matrices  $\{\mathbf{H}_i^w\}$  may have a common support. Specifically, there exists

an index set  $\Omega_c$  of the common support, such that  $\Omega_c \subseteq \Omega_i$ , for all  $i$ .

Based on the above Observations, we formally give the following model for the channel matrices in multi-user massive MIMO systems.

**Definition 1 (Joint Sparse Massive MIMO Channel).** *The channel matrices  $\{\mathbf{H}_i^w : \forall i\}$  have the following properties, with parameter  $\mathbb{S} = \{s_c, \{s_i : \forall i\}\}$ .*

(a) *Individual joint sparsity due to local scattering at the BS:* Denote  $\mathbf{h}_{ij}$  as the  $j$ -th row vector of  $\mathbf{H}_i^w$ ; then there exists an index set  $\Omega_i$ , with  $|\Omega_i| \leq s_i \ll M$ ,  $\forall i$ , such that

$$\text{supp}(\mathbf{h}_{i1}) = \text{supp}(\mathbf{h}_{i2}) = \dots = \text{supp}(\mathbf{h}_{iN}) \triangleq \Omega_i. \quad (3)$$

(b) *Distributed joint sparsity due to common scattering at the BS:* There exists an index set  $\Omega_c$ , with  $|\Omega_c| \geq s_c$ , such that

$$\bigcap_{i=1}^K \Omega_i = \Omega_c. \quad (4)$$

Furthermore, the entries of  $(\mathbf{H}_i^w)_{\Omega_i}$  are i.i.d. complex Gaussian distributed with zero mean and unit variance.

Note that  $\mathbb{S} = \{s_c, \{s_i : \forall i\}\}$  in Definition 1, where  $|\Omega_i| \leq s_i$ ,  $\forall i$ ,  $|\Omega_c| \geq s_c$ , is the statistical knowledge of the channel sparsity levels and we assume  $\mathbb{S}$  is available<sup>3</sup> at the BS. Note that the channel support profile  $\mathcal{P} = \{\Omega_c, \{\Omega_i : \forall i\}\}$  is *unknown* to both the BS and the  $K$  MSs. Based on the above channel model, we shall elaborate our distributed CSIT estimation and feedback framework in the next section.

## 2.3. CSIT Estimation and Feedback

In this section, we propose a novel CSIT estimation and feedback framework and the details are given in Algorithm 1. This novel scheme allows us to exploit the distributed joint sparsity in the user channel matrices to enhance the CSIT estimation quality. Obviously, the pilot training and feedback overhead in Algorithm 1 are characterized by  $T$ . Our goal is to exploit the hidden joint channel sparsity in the CSIT recovery in Step 3 of Algorithm 1 to reduce the required training and feedback overhead  $T$  in multi-user massive MIMO systems.

### Algorithm 1 CSIT Estimation and Feedback

- **Step 1 (Pilot Training):** The BS sends the compressive training symbols  $\mathbf{X} \in \mathbb{C}^{M \times T}$ , with  $T \ll M$ .
- **Step 2 (Compressive Measurement and Feedback):** The  $i$ -th mobile user observes the compressed measurements  $\mathbf{Y}_i$  from the pilot symbols given in (2) and feeds back to the BS side.
- **Step 3 (Joint CSIT Recovery at BS):** The BS recovers the CSIT  $\{\mathbf{H}_1^e, \dots, \mathbf{H}_K^e\}$  jointly based on the compressed feedback  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_K\}$ .

<sup>3</sup>Note that the sparsity level depends on the scattering environment and changes slowly (changes over a very long timescale). Hence, knowledge of  $\mathbb{S}$  can be obtained easily based on the prior knowledge of the propagation environment (e.g., can be acquired from offline channel propagation measurement at the BS [16] or long term stochastic learning and estimation [17]).

### 3. JOINT CSIT RECOVERY ALGORITHM DESIGN

In this section, we shall propose a joint orthogonal matching pursuit (J-OMP) algorithm to conduct the CSIT recovery at the BS (Step 3 in Algorithm 1) by exploiting the hidden sparsity structures of the channel matrices (Definition 1). To achieve this, we first rewrite (2) into the standard CS model. Denote the following new variables:

$$\bar{\mathbf{Y}}_i = \sqrt{\frac{M}{PT}} \mathbf{Y}_i^H \mathbf{A}_R \in \mathbb{C}^{T \times N}, \quad \bar{\mathbf{X}} = \sqrt{\frac{M}{PT}} \mathbf{X}^H \mathbf{A}_T \in \mathbb{C}^{T \times M}, \quad (5)$$

$$\bar{\mathbf{H}}_i = (\mathbf{H}_i^w)^H \in \mathbb{C}^{M \times N}, \quad \bar{\mathbf{N}}_i = \sqrt{\frac{M}{PT}} \mathbf{N}_i^H \mathbf{A}_R \in \mathbb{C}^{T \times N}. \quad (6)$$

Substituting these variables into (2), we obtain

$$\bar{\mathbf{Y}}_i = \bar{\mathbf{X}} \bar{\mathbf{H}}_i + \bar{\mathbf{N}}_i, \quad \forall i. \quad (7)$$

Then (7) matches the standard CS measurement model, where  $\bar{\mathbf{X}}$  is the measurement matrix with  $\text{tr}(\bar{\mathbf{X}}^H \bar{\mathbf{X}}) = M$  and  $\{\bar{\mathbf{H}}_i\}$  are the joint sparse matrices.

#### 3.1. Proposed J-OMP Algorithm

The proposed J-OMP is designed by extending conventional OMP [18] to the specific sparsity structures of massive MIMO channels and the details are given in Algorithm 2. Note that  $\eta$  ( $\eta > 1$ ) in the input of Algorithm 2 is a threshold parameter (as in Step 3). In Algorithm 2, Step 2 and 3 aim to identify the common support  $\Omega_c$  and the individual support  $\Omega_i$  respectively. Based on the estimated individual support  $\Omega_i^e$ , Step 4 recovers the channel matrices using the LS approach. Let  $s_i = s, \forall i$ ; then the overall complexity of Algorithm 2 is  $O(KsMNT)$ .

#### 3.2. Discussion of the Pilot Training Matrix $\mathbf{X}$

Under the proposed CSIT estimation scheme, one issue that remains to be discussed is how to design the entries of the  $M \times T$  pilot training matrix  $\mathbf{X}$ . In the CS literature, it is shown that efficient and robust CS recovery can be achieved when the measurement matrix satisfies a proper restricted isometry property (RIP) [19]. On the other hand, matrices randomly generated from the sub-Gaussian distribution [19] can satisfy the RIP property with overwhelming probability [19]. Hence, the pilot training matrix  $\mathbf{X} \in \mathbb{C}^{M \times T}$  can be designed as  $\mathbf{X} = \mathbf{A}_T \mathbf{X}_a$ , where  $\mathbf{X}_a \in \mathbb{C}^{M \times T}$  is i.i.d. drawn from  $\left\{ -\sqrt{\frac{P}{M}}, \sqrt{\frac{P}{M}} \right\}$ , with equal probability.

### 4. CSIT ESTIMATION QUALITY

In this section, we analyze the CSIT estimation performance in terms of the normalized mean squared error [20] (NMSE) of the channel matrices. First of all, let  $s_i = s, \forall i$ , to obtain simple expression. Denote  $K_o \triangleq \max_{j \notin \Omega_c} \sum_{i=1}^K I_{\{j \in \Omega_i\}}$ ; then  $K_o < K$  from (4). Denote  $\gamma \triangleq \frac{K_o}{K} < 1$ . Denote  $\Theta_c, \{\Theta_i\}$  as the following the events<sup>4</sup>:

$$\Theta_c : \text{In Step 2 of Alg. 2, support } \Omega_c^e \text{ is correctly identified, i.e., } \Omega_c^e \subseteq \Omega_c. \quad (9)$$

<sup>4</sup>Note that  $|\Omega_c| \geq s_c$  (as in Definition 1), and hence, we can at most identify a *subset* of the common support  $\Omega_c$  in Step 2 of Algorithm 2, i.e.,  $\Omega_c^e \subseteq \Omega_c$ , as in (9).

---

#### Algorithm 2 J-OMP for CSIT Recovery

---

**Input:**  $\{\mathbf{Y}_i : \forall i\}, \mathbf{X}, \mathbb{S} = \{s_c, \{s_i : \forall i\}\}, \eta$  ( $\eta > 1$ ).

**Output:** Estimated  $\{\mathbf{H}_i^e\}$  for  $\{\mathbf{H}_i : \forall i\}$ .

- **Step 1 (Initialization):** Compute  $\bar{\mathbf{Y}}_i, \forall i, \bar{\mathbf{X}}$  from  $\{\mathbf{Y}_i : \forall i\}$  and  $\mathbf{X}$ , as in (5).
- **Step 2 (Common Support Identification):** Set  $\mathbf{R}_i = \bar{\mathbf{Y}}_i, \forall i, \Omega_c^e = \emptyset$  and repeat the following procedures  $s_c$  times.
  - **A (Support Estimate):** Estimate the remaining index set by  $\Omega_i = \arg \max_{|\Omega|=s_i-|\Omega_c^e|} \|(\bar{\mathbf{X}}_\Omega)^H \mathbf{R}_i\|_F, \forall i$ .
  - **B (Support Update):** Update the estimated *common* support as  $\Omega_c^e = \Omega_c^e \cup \left\{ \arg \max_j \sum_{i=1}^K I_{\{j \in \Omega_i^e\}} \right\}$ .
  - **C (Residual Update):**  $\mathbf{R}_i = (\mathbf{I} - \mathbf{P}_{\Omega_c^e}) \bar{\mathbf{Y}}_i$ , where  $\mathbf{P}_{\Omega_c^e}$  is a projection matrix and is given by

$$\mathbf{P}_{\Omega_c^e} = (\bar{\mathbf{X}}_{\Omega_c^e})(\bar{\mathbf{X}}_{\Omega_c^e})^\dagger. \quad (8)$$

- **Step 3 (Individual Support Identification):** For user  $i$ , set  $\Omega_i^e = \Omega_c^e$ , then stop if  $\|\mathbf{R}_i\|_F^2 \leq \frac{\eta NM}{P}$  or the following procedures have been repeated  $(s_i - s_c)$  times,  $\forall i$ .
    - **A (Support Update):** Update the estimated *individual* support as  $\Omega_i^e = \Omega_i^e \cup \left\{ \arg \max_j \|\bar{\mathbf{X}}(j)^H \mathbf{R}_i\|_F \right\}$ .
    - **B (Residual Update):**  $\mathbf{R}_i = (\mathbf{I} - \mathbf{P}_{\Omega_i^e}) \bar{\mathbf{Y}}_i$ .
  - **Step 4 (Channel Estimation by LS):** The estimated channel for user  $i$  is  $\mathbf{H}_i^e = \mathbf{A}_R (\bar{\mathbf{H}}_i^e)^H \mathbf{A}_T^H$ , where  $\bar{\mathbf{H}}_i^e$  is given by  $(\bar{\mathbf{H}}_i^e)^{\Omega_i^e} = (\bar{\mathbf{X}}_{\Omega_i^e})^\dagger \bar{\mathbf{Y}}_i, (\bar{\mathbf{H}}_i^e)^{[M] \setminus \Omega_i^e} = \mathbf{0}, \forall i$ .
- 

$\Theta_i$  : Conditional on  $\Theta_c$ , in Step 3 of Alg. 2, support  $\Omega_i^e$  is correctly identified, i.e.,  $\Omega_i^e = \Omega_i$ . (10)

we obtain the following theorem on the NMSE of the CSI in terms of the probabilities of event  $\Theta_c$  and  $\{\Theta_i : \forall i\}$ .

**Theorem 1** (CSIT Estimation Quality). *The NMSE of  $\mathbf{H}_i$  is bounded by*

$$\mathbb{E} \left( \frac{\|\mathbf{H}_i - \mathbf{H}_i^e\|_F^2}{\|\mathbf{H}_i\|_F^2} \right) \leq \frac{MNs}{PT(1 - \delta_s)(\tilde{s}_i N - 1)} + E_i, \quad (11)$$

where  $\tilde{s}_i = |\Omega_i|$ ,  $E_i = (2 - \text{Pr}(\Theta_c) - \text{Pr}(\Theta_i)) \left( \frac{1 - \delta_s + \delta_{2s}^2}{1 - \delta_s} \right)$ ,  $\text{Pr}(\Theta_c)$  and  $\text{Pr}(\Theta_i)$  are the probability of the events  $\Theta_c$  and  $\Theta_i$  respectively, and  $\delta_s$  and  $\delta_{2s}$  are the  $s$ -th and  $2s$ -th restricted isometry constants (RIC) [19] of  $\bar{\mathbf{X}}$  respectively.

*Proof.* Please refer to [21]. □

Recall that in Section II, we observe (i) that the channel matrix  $\mathbf{H}_i^w$  is simultaneously zero or non-zero on its columns, with dimension  $N \times 1$ , and (ii) that the  $K$  users share a partial common channel support  $\Omega_c$ . Therefore, it is interesting to see how the CSIT estimation quality is affected by these joint sparsity parameters ( $N, \Omega_c$ ). Based on Theorem 1 and by analyzing the probabilities of  $\text{Pr}(\Theta_c)$  and  $\text{Pr}(\Theta_i)$  with respect to (w.r.t.) the randomness of the channel matrices in Algorithm 2, we obtain the following results (Corollary 1-2).

**Corollary 1** (CSIT Quality w.r.t.  $N$ ). Suppose  $|\Omega_i| = s, \forall i$ . If

$$\theta \triangleq \frac{(1 - 2\delta_s)}{\left(\delta_{s+1} + 2(1 - \delta_s)\sqrt{\frac{(1+\delta_1)\eta M}{P}}\right)} > 1, \quad (12)$$

where  $\delta_1, \delta_s$  and  $\delta_{s+1}$  are the 1-th,  $s$ -th and  $2s$ -th RICs [19] of  $\bar{\mathbf{X}}$  respectively, then

$$\lim_{N \rightarrow \infty} -\frac{1}{N} \ln(E_i) \geq \beta > 0, \quad (13)$$

where  $\beta$  is a positive parameter that depends on  $\gamma, \theta, K, P, M$  and  $\bar{\mathbf{X}}$ .

*Proof.* Please refer to [21].  $\square$

Equation (13) can be re-written as

$$E_i \leq \exp(-N \cdot \beta + o(N)). \quad (14)$$

From (14), we conclude that  $E_i$  in (11) decay at least exponentially w.r.t.  $N$  and hence, a larger  $N$  turns out to have a smaller CSIT estimation error from Theorem 1. This result indicates that individual joint sparsity in the user channel matrices (Observation I) is indeed captured by the proposed recovery algorithm to improve the CSIT estimation quality.

**Corollary 2** (CSIT Quality w.r.t.  $\Omega_c$ ). Suppose  $|\Omega_i| = s, \forall i$ . Scale the threshold parameter  $\eta$  in Algorithm 2 as  $\eta = \sqrt{P}$ . Let the transmit power  $P \rightarrow \infty$  and the number of users  $K \rightarrow \infty$ . If (12) holds and  $p \triangleq s \cdot \exp(-N(\ln \theta - 1 + \frac{1}{\theta})) + M \exp(-N(\theta - 1 - \ln \theta)) < \frac{1}{2}(1 - \gamma)$ , we have

$$\mathbb{E} \left( \frac{\|\mathbf{H}_i - \mathbf{H}_i^e\|_F^2}{\|\mathbf{H}_i\|_F^2} \right) \leq \left( \sum_{t=s_c}^s \binom{s}{t} - 1 \right) \rho, \quad (15)$$

where  $\rho$  is a positive parameter that depends on  $\theta, M, N$  and  $\bar{\mathbf{X}}$ .

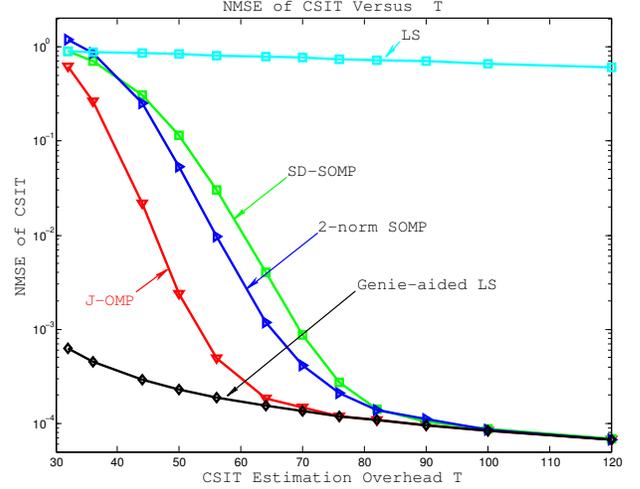
*Proof.* Please refer to [21].  $\square$

From (15),  $\mathbb{E} \left( \frac{\|\mathbf{H}_i - \mathbf{H}_i^e\|_F^2}{\|\mathbf{H}_i\|_F^2} \right) \rightarrow 0$  as  $s_c \rightarrow s \triangleq |\Omega_i|$  and a larger size (larger  $s_c, s_c \leq s$ ) of the common support  $\Omega_c$  shared by the users tends to have a smaller CSIT estimation error. This result indicates that the distributed joint sparsity in the user channel matrices (Observation II) is indeed captured by the proposed recovery algorithm to improve the CSIT estimation quality.

## 5. NUMERICAL RESULTS

In this section, we verify the performance advantages of our proposed CSIT estimation scheme via simulation. Four baselines are considered, namely baseline 1 of conventional LS [8], baseline 2 of 2-norm SOMP [22], baseline 3 of SD-SOMP [23], and baseline 4 of *Genie-aided LS* which directly recovers the CSI using LS with  $\Omega_i$  replacing  $\Omega_i^e$  in Step 4 of Algorithm 2 (baseline 4 serves as an performance upper bound). Consider a multi-user massive MIMO system with one BS and  $K = 40$  users, where the BS has  $M = 300$  antennas and each user has  $N = 2$  antennas. The average transmit SNR at the BS is  $P = 35$  dB and the channel sparsity level statistic is  $\mathbb{S} = \{14, \{s_i = 22 : \forall i\}\}$ .

In Figure 2, we compare the NMSE of the estimated CSI versus the training and feedback overhead  $T$ . From this figure, we



**Fig. 2.** NMSE of CSIT versus the CSIT training and feedback overhead  $T$  in a multi-user massive MIMO network with 1 BS and  $K = 40$  users, where the BS and the user has  $M = 300, N = 2$  antennas respectively, and the average transmit SNR at the BS is  $P = 35$  dB.

observe that the CSIT estimation quality increases as  $T$  increases. On the other hand, the proposed J-OMP algorithm achieves a substantial performance gain over the baselines. This is because the proposed J-OMP exploits the hidden joint sparsity among the user channel matrices to better recover the CSI. Furthermore, we observe that the proposed J-OMP, 2-norm SOMP and SD-SOMP all approach the genie-aided LS scheme as  $T$  increases. This is because the channel support recovery probabilities of these schemes all go to 1 as  $T$  increases. This fact also highlights the importance of having a higher probability of support recovery in the CSIT reconstruction.

## 6. CONCLUSION

In this paper, we consider multi-user massive MIMO of FDD systems and we deploy the compressive sensing (CS) technique to reduce the training as well as the feedback overhead in the CSIT estimation. We propose a distributed compressive CSIT estimation scheme so that the compressed measurements are observed at the users locally, while the CSIT recovery is performed at the base station jointly. We develop joint OMP algorithm to conduct the CSIT reconstruction which exploits the joint sparsity in the user channel matrices. We also analyze the estimated CSIT equality in terms of the normalized mean squared error. From the results, we show that the joint channel sparsity can be exploited to enhancing the CSIT estimation quality in multi-user massive MIMO systems of FDD systems.

## 7. REFERENCES

- [1] Erik G Larsson, Fredrik Tufvesson, Ove Edfors, and Thomas L Marzetta, "Massive MIMO for next generation wireless systems," *arXiv preprint arXiv:1304.6690*, 2013.
- [2] Quentin H Spencer, A Lee Swindlehurst, and Martin Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, 2004.

- [3] Sinh Le Hong Nguyen and Ali Ghayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, IEEE, 2013, pp. 2890–2895.
- [4] Peter WC Chan, Ernest S Lo, Ray R Wang, Edward KS Au, Vincent KN Lau, Roger S Cheng, Wai Ho Mow, Ross D Murch, and Khaled Ben Letaief, "The evolution path of 4g networks: FDD or TDD?," *IEEE Commun. Mag.*, vol. 44, no. 12, pp. 42–50, 2006.
- [5] Thomas L Marzetta, Giuseppe Caire, Merouane Debbah, I Chih-Lin, and Saif K Mohammed, "Special issue on massive MIMO," *Journal of Commun. and Networks*, vol. 15, no. 4, pp. 333–337, 2013.
- [6] Qinfang Sun, Donald C Cox, Howard C Huang, and Angel Lozano, "Estimation of continuous flat fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 549–553, 2002.
- [7] Anna Scaglione and Azadeh Vosoughi, "Turbo estimation of channel and symbols in precoded MIMO systems," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 2004, vol. 4, pp. iv–413–iv–416 vol.4.
- [8] Mehrzad Biguesh and Alex B Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–893, 2006.
- [9] Yan Zhou, Markus Herdin, Akbar M Sayeed, and Ernst Bonek, "Experimental study of MIMO channel statistics and capacity via the virtual channel representation," *IEEE Trans. Wireless Commun.*, 2006.
- [10] Jakob Hoydis, Cornelis Hoek, Thorsten Wild, and Stephan ten Brink, "Channel measurements for large antenna arrays," in *Proc. IEEE Int. Symp. Wireless Commun. Systems (ISWCS)*, 2012, pp. 811–815.
- [11] Christian R Berger, Zhaohui Wang, Jianzhong Huang, and Shengli Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, 2010.
- [12] Waheed U Bajwa, Jarvis Haupt, Akbar M Sayeed, and Robert Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.
- [13] C.R. Berger, Shengli Zhou, J.C. Preisig, and P. Willett, "Sparse channel estimation for multicarrier underwater acoustic communication: From subspace methods to compressed sensing," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1708–1721, 2010.
- [14] D. Tse and P. Viswanath, *Fundamentals of wireless communication*, Cambridge Univ Pr, 2005.
- [15] J. Poutanen, K. Haneda, J. Salmi, V. Kolmonen, F. Tufvesson, T. Hult, and P. Vainikainen, "Significance of common scatterers in multi-link indoor radio wave propagation," in *Proc. IEEE European Conf. Antennas and Propagation (EuCAP)*, 2010, pp. 1–5.
- [16] J-J Fuchs, "Detection and estimation of superimposed signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, 1998, vol. 3, pp. 1649–1652.
- [17] Léon Bottou and Noboru Murata, "Stochastic approximations and efficient learning," *The Handbook of Brain Theory and Neural Networks, Second edition.*, The MIT Press, Cambridge, MA, 2002.
- [18] T Tony Cai and Lie Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," vol. 57, no. 7, pp. 4680–4688, 2011.
- [19] Emmanuel J Candès and Michael B Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.
- [20] David Wisell and Magnus Isaksson, "Derivation of a behavioral RF power amplifier model with low normalized mean-square error," in *Proc. 37th. Eur. Microw. Conf.*, Munich, Germany, Oct. 2007, pp. 485–488.
- [21] X. Rao and V.K.N. Lau, "Distributed compressive CSIT estimation and feedback for FDD multi-user massive MIMO systems," *submitted to IEEE Trans. Signal Process.*, 2013.
- [22] Joel A Tropp, Anna C Gilbert, and Martin J Strauss, "Algorithms for simultaneous sparse approximation. part I: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.
- [23] Junhua Liang, Yang Liu, Wenjun Zhang, Youyun Xu, Xiaoying Gan, and Xinbing Wang, "Joint compressive sensing in wide-band cognitive networks," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC)*, 2010, pp. 1–5.