IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 13, 2020, accepted June 22, 2020, date of publication June 30, 2020, date of current version July 20, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3006087*

# Simultaneous Channel Estimation and Data Detection Based on Superimposed Training for Many Access MIMO System in Uplink

**KUN ZHANG, WENDI WANG, (Graduate Student Member, IEEE), AND HUARUI YIN, (Member, IEEE)**

CAS Key Laboratory of Wireless-Optical Communications, University of Science and Technology of China, Hefei 230026, China

Corresponding author: Huarui Yin (yhr@ustc.edu.cn)

This work was supported in part by the National Science and Technology Major Project of China Ministry of Industry and Information Technology (MIIT) under Grant 2017ZX03001007, and in part by the National Science Foundation of China under Grant 61571412.

**ABSTRACT** Massive Multiple Input Multiple Output (MIMO) has great potential to improve spectrum efficiency in the fifth generation (5G) wireless communication systems. However, the efficiency was disastrously reduced by the heavy burden of overhead for device detection and channel estimation of the large amount of small data packets in the uplink channel. In the paper, we proposed a novel transmission scheme by superposing the training symbols for active device detection and channel estimation on the data symbols in the uplink transmission to improve the efficiency. More specifically, in order to mitigate the cross interference among the superposed signals, we proposed to split the transmission into the training phase and the traffic phase. Then we superposed the training phase in the next transmission with the traffic phase in the current transmission. Furthermore, we give the optimization of power allocation ratio between training phase and traffic phase to obtain the optimal overall performance. The analytic and simulation results show that, with the help of spatial isolation among devices, our proposed transmission scheme can significantly improve the transmission efficiency compared with the existing schemes.

**INDEX TERMS** Many access, channel estimation, active device detection, superimposed training.

## I. INTRODUCTION

The fifth generation (5G) system is expected to provide high quality communication service. Compared with current technologies (4G), 10 to 100 times increment of connected devices is one of the 5G technical requirements [1]. In 5G, Multiple Input Multiple Output (MIMO) technology is a key technology for achieving ultra-high system capacity [2]. To make the full use of the spatial multiplexing gains or array gains in MIMO, obtaining accurate channel state information (CSI) [2]–[5] is important under massive connections. Besides, in order to increase the capacity of 5G systems, a variety of non-orthogonal multiple access (NOMA) schemes have been proposed [6]–[9]. The resource allocation schemes of some NOMA methods (including the power domain NOMA [6], code domain sparse code multiple access (SCMA) [7]–[9]) also depends on accurate CSI.

The associate editor coordinating the review of this manuscript and approving it for publication was A. Taufiq Asyhari.

The training-based channel estimation is widely applied in wireless communications. Massive connectivity for Internet-of-Things (IoT) and machine-type communications (MTC) is one of the most important characteristics of 5G communication systems. As the number of devices grows, the system efficiency will be deteriorated because of the increasing training overhead [3], especially in the uplink channel. Though the transmission request of small packets of each individual device is fairly sparse, the massive connections of the system will still result in very frequent transmission requests [10]. Since the training cost increases linearly with the number of all potential online devices, the time-frequency resource set aside for traffic data decreases greatly.

How to reduce the training cost has attracted significant attention in recent years. Considering that only a small fraction of potential devices are active in a massive device connectivity scenario at any given time slot, joint active device detection (ADD) and channel estimation (CE) based on compressed sensing (CS) technology has been investigated by many researchers [11]–[14]. Using CS technology,

the length of the training sequence could be less than the number of total online devices. Furthermore, the block sparse structure of massive MIMO is especially well suited for active device detection, thus the length of training symbols can be further reduced. This block structure advantage of MIMO has been studied in [15], [16]. Zvika Ben-Haim has shown that block sparse methods are particularly successful when the atoms within each block are nearly orthogonal in paper [16]. However, this approach still requires dedicated resources for channel estimation. To ensure the performance of active device detection and channel estimation, the length of training symbols should be larger than the number of active devices. The number of active devices is large and growing for many access system, thus the time frequency resources reserved for training is unavoidably high and lead to low overall efficiency of data transmission.

As an alternative method, superimposed training (ST) scheme is proposed, where training symbols are superimposed with data symbols to reduce training costs [17]–[25]. No dedicated resource is necessarily reserved for training, thus the transmission efficiency can be improved. In recent years, a lot of works focus on study the application of ST scheme in MIMO [22], [23], [26], [27]. The ST scheme can outperform the traditional training-based scheme with massive connectivity. This point has been discussed and simulated completely in [23]–[25]. A few articles discussed multi-user issues [18], [19], but they are limited to dozens of devices. In [23], more devices can be supported, but the number of supported devices is still less than the number of antennas.

However, for many access scenario, the total number of users $N$ will far exceed the number of antennas $M$, even for massive MIMO. In [10], the case of the number of active devices being greater than $M$ is studied, but the channel information is assumed to be completely known. Massive devices can cause tremendous mutual interference between training symbols and data symbols. The continuous increase of active devices may lead to the failure of traditional ST scheme: the channel estimation error is still too huge, which makes it difficult to eliminate the interference.

In this paper, to overcome the limitation of the traditional ST scheme, we proposed a cross transmission superimposed training (CTST) scheme combining active device detection and superimposed training to accommodate more devices creatively. In the scheme, we split the whole transmission into 2 pipelined phases: the first phase is training phase, and the second is traffic phase. Unlike the traditional ST scheme, we superposed the training phase (training symbols) of the next transmission with the data phase (data symbols) of the current transmission to mitigate the cross interference between the training symbols and data symbols of current transmission. When recovering the current active device signal, we can detect the active devices for the next transmission and estimate their CSI simultaneously.

The CSI of the current active devices has been estimated accurately at the previous transmission. Once the data of active device is correctly recovered, their impact on channel estimation can be eliminated, which makes it possible to improve the performance of active device detection and channel estimation. This in turn improves the performance of data recovery.

The main contributions of our work are listed as follows:

1) **Proposing the cross transmission superimposed training scheme for mitigation of the cross interference between data and training symbols in uplink massive connection scenario.**
2) **Improving the performance and transmission efficiency by power allocation optimization between data phase and training phase.**

Compared with the traditional ST scheme [22]–[27], we first perform active device detection combined with superimposed training. The new CTST scheme can eliminate interference more effectively to support massive connections where $N \gg M$. Compared with the joint channel estimation and active device detection scheme [11]–[14], we proposed a scheme combined with superimposed training, so that data recovery and channel estimation can be conducted simultaneously.

The rest of our paper is organized as follows: in Section II, we describe the system model of superimposed training scheme for uplink many access MIMO. In Section III, the specific steps of Block Orthogonal Matching Pursuit (BOMP) based active device detection and channel estimation are given. Miss detection probability (MDP) of active devices, mean square error (MSE) are analyzed. In Section IV, we introduce the specific steps of iterative algorithm for data recovery and channel estimation. The performance of data recovery is analyzed. The signal to Interference plus Noise Ratio (SINR) expression is analyzed. The optimal power allocation factor to improve overall performance is also given. In Section V, we present simulation results, which demonstrate the effectiveness of our scheme, and verify the theoretical analysis results. Finally, the conclusion of our paper is given in Section VI.

*Notations:* In the paper, we use boldface letters to denote matrices (upper case) and vectors (lower case). And $(.)^T$, $\mathbb{C}^{M\times N}$ denote transpose, the set of complex-valued matrix with dimension $M \times N$, respectively. The operators $(.)^T$, $(.)^H$ and $\otimes$ are the transpose, conjugate transpose and the Kronecker product of matrix respectively. $||\mathbf{A}||_F$ and $||\mathbf{a}||_2$ denote the Frobenius norm of matrix $\mathbf{A}$ and Euclidean norm of vector $\mathbf{a}$ respectively.

## II. SYSTEM MODEL

We consider uplink transmission of a single-cell massive connectivity scenario in MIMO. A base station (BS), equipped with $M$ antennas, serves $N$ single-antenna intelligent devices simultaneously. The antennas at the BS, as well as the antennas among devices, are sufficiently apart to yield spatially independent channels.

Active devices are a small fraction of total number of potential devices at any given time. The number of active

devices in the current transmission $N_c$ and in the next transmission $N_n$ is far less than $N$ as shown in Fig. 1 ($N_c \ll N, N_n \ll N$). However, considering the rapid growth of the number of active devices in future 5G system, it is reasonable assumed that $N_c > M, N_n > M$ [10].

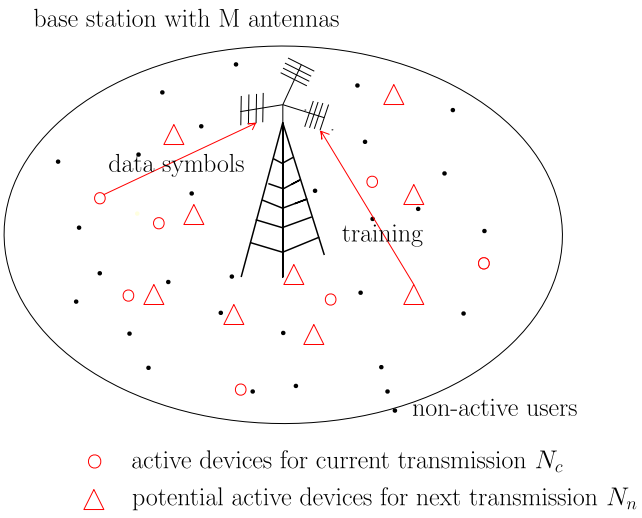


**FIGURE 1. System model of superimposed training for many access MIMO system.**

For any device $n \in [1, N]$, we define its active indicator of the current transmission as below

$$a_n = \begin{cases} 1, & \text{if device } n \text{ sends data at the current transmission} \\ 0, & \text{otherwise} \end{cases}$$

This device sparse model is widely used in the joint scheme of active device detection and channel estimation [11]–[14].

Define the indices set of active devices within the current transmission as

$$I = \{n : a_n = 1\}, \tag{1}$$

and $|I| = N_c$. $\mathbf{h}_n \in \mathbb{C}^{M\times 1}$ represents the channel coefficients from the device $n$ to the BS, and the elements of $\mathbf{h}_n$ follow independent and identically distributed (i.i.d.) complex Gaussian distribution with zero mean and unit variance.

The channels are assumed to be block-fading: they remain constant for a certain duration and then change independently. We assume the channel response matrix $\mathbf{H}_I := \left[\mathbf{h}_{i_1}, \mathbf{h}_{i_2}, \cdots, \mathbf{h}_{i_{N_c}}\right]$ of active devices has been obtained in the previous transmission period. $\mathbf{H}_I = \tilde{\mathbf{H}}_I + \Delta\mathbf{H}_I$. $\tilde{\mathbf{H}}_I$ is the estimation of the channel. $\Delta\mathbf{H}_I$ is the channel estimation error. Using the least square method, $\tilde{\mathbf{H}}_I$ and $\Delta\mathbf{H}_I$ can be considered uncorrelated.

Similarly, for any given device $n \in [1, N]$ at the next transmission, we define the active indicator $b_n$ as below:

$$b_n = \begin{cases} 1, & \text{if device } n \text{ will send training sequence} \\ & \text{at the next transmission} \\ 0, & \text{otherwise} \end{cases}$$

Define the set of active devices within the next transmission as

$$I^{\text{ne}} = \{n : b_n = 1\}, \tag{2}$$

and $|I^{\text{ne}}| = N_n$, where $N_n$ is the number of active devices at the next transmission. The channel response matrix $\mathbf{H}_{I^{\text{ne}}} := \left[\mathbf{h}_{j_1}, \mathbf{h}_{j_2}, \cdots, \mathbf{h}_{j_{N_n}}\right]$ are all unknown to the BS.

We assume that each frame of transmission consists of $T$ symbols. Let $\mathbf{x}_{tn} \in \mathbb{C}^{1\times T}$ denote the training symbols of device $n$. Define $\mathbf{X}_t := \left[\mathbf{x}_{t1}^T, \mathbf{x}_{t2}^T, \cdots, \mathbf{x}_{tN}^T\right]^T$, which is all known by BS. Specially designed sequence will be assigned to each online device for channel estimation and active device detection. The signal received by the BS only contains the training sequences of the active devices. Define $\mathbf{X}_{tI^{\text{ne}}} := \left[x_{tj}, \forall j \in I^{\text{ne}}\right]$ as the training matrix of active devices at the next transmission, which is used for least square method to get the CSI of active devices.

Let $\mathbf{s}_{di} \in \mathbb{C}^{d\times 1}$ denote the data symbols from device $i$. We view $d$ as the maximum length of the messages for all devices within a frame. For the users whose message lengths are less than $d$, we assume their messages have been zero-padded to $d$ before precoding. $\mathbf{x}_{di} = \mathbf{s}_{di}^T\mathbf{P}_i^T \in \mathbb{C}^{1\times T}$ stands for the spreading data sequence. $\mathbf{P}_i \in \mathbb{C}^{T\times d}$ is the complex precoding matrix, which should have full column rank for data recovery.

Additionally, we assume each column of the precoding matrices is normalized to unit energy. $\|\mathbf{P}_{im}\|_2^2 = 1$, where $m$ is the column index of $\mathbf{P}_i$. The precoding method is also adopted in [10]. Define $\mathbf{X}_{dI} := \left[\mathbf{x}_{di_1}^T, \mathbf{x}_{di_2}^T, \cdots, \mathbf{x}_{di_{N_c}}^T\right]^T$ as the data matrix for all current active devices.

For ease of analysis, we assume

$$E\left[\|\mathbf{x}_{tn}\|_2^2\right] = T, \quad E\left[\|\mathbf{s}_{di}\|_2^2\right] = d. \tag{3}$$

The training sequence and data are QPSK modulation symbols, and the energy of each symbol is 1.

Traditional training based schemes are composed of two phases: training phase and data phase. Unlike traditional ST scheme, in a certain transmission, the training symbols of active devices in the next transmission are transmitted together with the data symbols of active devices in the current transmission. Then no extra dedicated resource is required for activity detection and channel estimation. The comparison of different transmission superposition schemes is shown in Fig.2.

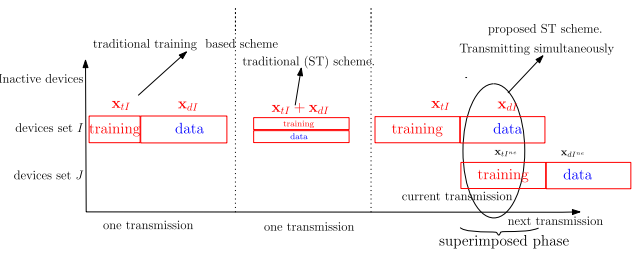


**FIGURE 2. Comparison between the proposed CTST scheme and the traditional transmission schemes.**

As shown in Fig.2, active devices in the next transmission will send their training sequence in advance, no longer waiting for the end of the data phase of the current transmission. If the traditional ST scheme is adopted, the system model and receiving equation are similar to those in [23], from which our proposed CTST scheme evolved. The received signal $\mathbf{Y} \in \mathbb{C}^{M\times T}$ at the receiver within one frame can be written as

$$\begin{aligned}\mathbf{Y} &= \sqrt{\rho_1}\sum_{i=1}^{N}\mathbf{h}_i a_i \mathbf{x}_{di} + \sqrt{\rho_2}\sum_{j=1}^{N}\mathbf{h}_j b_j \mathbf{x}_{tj} + \mathbf{Z} \\ &= \sqrt{\rho_1}\sum_{i\in I}^{N}\mathbf{h}_i \mathbf{x}_{di} + \sqrt{\rho_2}\sum_{j\in I^{\mathrm{ne}}}^{N}\mathbf{h}_j \mathbf{x}_{tj} + \mathbf{Z} \\ &= \sqrt{\rho_1}\mathbf{H}_I \mathbf{X}_{dI} + \sqrt{\rho_2}\mathbf{H}_{I^{\mathrm{ne}}}\mathbf{X}_{tI^{\mathrm{ne}}} + \mathbf{Z} \\ &= \sqrt{\rho_1}\tilde{\mathbf{H}}_I \mathbf{X}_{dI} + \sqrt{\rho_2}\mathbf{H}_{I^{\mathrm{ne}}}\mathbf{X}_{tI^{\mathrm{ne}}} + \underbrace{\sqrt{\rho_1}\Delta\mathbf{H}_I \mathbf{X}_{dI} + \mathbf{Z}}_{\mathbf{Z}'}, \end{aligned} \quad (4)$$

where $\rho_1$ is the energy of data symbols, $\rho_2$ is the energy of training symbols. $\mathbf{Z} \in \mathbb{C}^{M\times T}$ represents the additive noise, with i.i.d. circularly symmetric complex Gaussian distributed random entries of zero mean and unit variance. $\mathbf{Z}'$ is the sum of the additive noise and residual error due to imperfect channel estimation.

In ST scheme, training sequence can have the same length as the extended data sequence. In reality, $T$ is usually much larger than the number of active devices, i.e. $T \gg N_c, N_n$. This ensures the performance of active device detection and channel estimation. Also, the precoding scheme contributes to solving data recovery with $N_c > M$ [10]. [1]

As a comparison, we give a model of traditional ST scheme with active device detection, where the training symbols are superimposed to its own data symbols for each current active device. The received signal of traditional superimposed transmission scheme $\mathbf{Y}_{\mathrm{tra}}$ can be written as

$$\begin{aligned}\mathbf{Y}_{\mathrm{tra}} &= \sum_{i=1}^{N}\mathbf{h}_i a_i(\sqrt{\rho_1}\mathbf{x}_{di} + \sqrt{\rho_2}\mathbf{x}_{ti}) + \mathbf{Z} \\ &= \sum_{i\in I}^{N}\mathbf{h}_i(\sqrt{\rho_1}\mathbf{x}_{di} + \sqrt{\rho_2}\mathbf{x}_{ti}) + \mathbf{Z} \\ &= \sqrt{\rho_1}\mathbf{H}_I \mathbf{X}_{dI} + \sqrt{\rho_2}\mathbf{H}_I \mathbf{X}_{tI} + \mathbf{Z} \\ &= \sqrt{\rho_1}\tilde{\mathbf{H}}_I \mathbf{X}_{dI} + \sqrt{\rho_2}\mathbf{H}_I \mathbf{X}_{tI} + \underbrace{\sqrt{\rho_1}\Delta\mathbf{H}_I \mathbf{X}_{dI} + \mathbf{Z}}_{\mathbf{Z}'}. \end{aligned} \quad (5)$$

The index set of active devices $I$ and the channel information of the active devices $\mathbf{H}_I$ are both unknown to the BS. This is the combination of ST scheme given in [23] and active device detection given in [11], [14].

The mutual interference between training sequence and data sequence sent by active devices is the key factor affecting system performance in the proposed CTST scheme. We design an efficient iterative algorithm to continuously mitigate the interaction contamination between signal and training sequence. The results of data recovery and channel estimation are updated by the iterative algorithm. After each iteration, the channel estimation is more accurate, and more device data are recovered correctly.

[1] By the way, this proposed CTST scheme is easy to be extended to other NOMA methods like SCMA or PDMA.

## III. BOMP BASED ACTIVE DEVICE DETECTION AND CHANNEL ESTIMATION

Device activity detection and channel estimation are the prerequisites of data recovery. In this section, Block Orthogonal Matching Pursuit (BOMP) [15] is adopted to perform device activity detection, Least Square (LS) method to obtain the channel state information.[2]

Both the analytic results of device activity detection and channel estimation are given in the section.

In a particular iteration, it is assumed that the data of some active devices in the current transmission has already been correctly recovered and subtracted. Define $I_r$ as the collection of devices whose data have been correctly recovered. Then the equivalent residual signal $\mathbf{r}_t$ when doing active device detection and channel estimation can be written as

$$\begin{aligned}\mathbf{r}_t &= \mathbf{Y} - \sqrt{\rho_1}\sum_{i\in I_r}^{N}\tilde{\mathbf{h}}_i \mathbf{x}_{di} \\ &= \sqrt{\rho_2}\sum_{n=1}^{N}\mathbf{h}_n b_n \mathbf{x}_{tn} + \underbrace{\sqrt{\rho_1}\sum_{i\in I_w}^{N}\tilde{\mathbf{h}}_i \mathbf{x}_{di} + \mathbf{Z}'}_{\mathbf{Z}'_t}, \end{aligned} \quad (6)$$

where $I_w$ is the collection of devices whose data have not been correctly recovered ($I = [I_r, I_w]$), and $\mathbf{Z}'_t = \sqrt{\rho_1}\Delta\mathbf{H}_I \mathbf{X}_{dI} + \sqrt{\rho_1}\sum_{i\in I_w}^{N}\tilde{\mathbf{h}}_i \mathbf{x}_{di}^T + \mathbf{Z}$ is the equivalent noise for active device detection and channel estimation. Define $\bar{\sigma}_{h_i}^2$ as the mean square error of channel estimation in the current transmission and $\bar{\sigma}_{h_j}^2$ as the mean square error of channel estimation in the next transmission. The mean value of effective noise energy can be expressed as

$$\begin{aligned}E\left[\left\|\mathbf{Z}'_t\right\|_2^2\right] &= MT + \rho_1 MN_c\bar{\sigma}_{h_i}^2 d + (N_c - k)\rho_1 Md \\ &\approx MT + (N_c - k)\rho_1 Md, \end{aligned} \quad (7)$$

where $k$ is the number of active devices whose data have been correctly recovered, $|I_w| = N_c - k$. The proof of (7) is shown in Appendix A in details.

[2] Here we use the most commonly used algorithms for channel estimation and active device detection. When the number of antennas and devices are large, the matrix dimension of the system is too large, so we try to use the algorithm with low complexity.

For channel estimation, the linear minimum mean-square error (LMMSE) [28] can also be used to obtain better performance. After the data of most active devices is recovered correctly, data-aided (DA) method can be used to improve the performance [23], but it also increases the complexity. We will continue to analyze in the subsequent work.

For active device detection, CoSaMP algorithm can also be improved to make use of block sparse structure, which may have better performance [29], but algorithm improvement is not the focus of this paper. BP [30] and convex programming [31] method may have better detection performance, but for the proposed scheme, the complexity may be too high.

Active device set can be obtained from $\mathbf{r}_t$ with compressive sensing method, then we can estimate their channel state information with the least square algorithm.

Among various compressive sensing (CS) recovery algorithms [30]–[34](including basis pursuit (BP) [30], Convex programming method [31], Orthogonal Matching Pursuit (OMP) [32], Compressive Sampling Matching Pursuit (CoSaMP) [33] and Subspace Pursuit (SP) [34]), BOMP, extended from OMP, is chosen for its competitive good performance and low complexity compared with other algorithms. Detailed steps of BOMP algorithm is presented in the algorithm flow chart as shown in Fig.3. Define $\epsilon$ as the residual energy threshold of algorithm termination. This threshold can be set as the energy of the equivalent noise: $\epsilon = E\left[\left\|\mathbf{Z}_t'\right\|_2^2\right] \approx MT + (N_c - k)\rho_1 Md$.

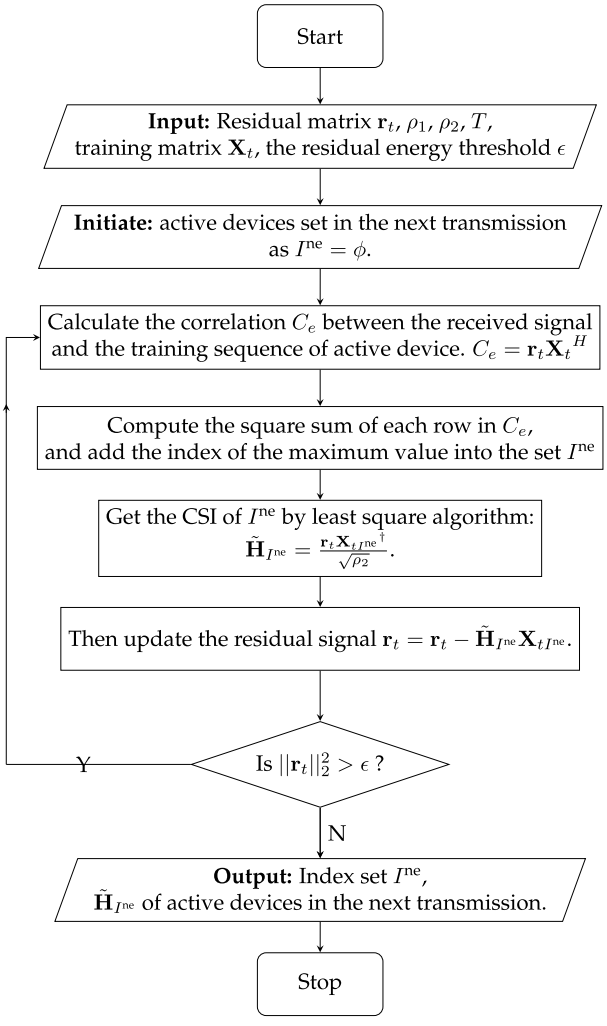


**FIGURE 3.** Algorithm 1 BOMP algorithm for active devices detection.

### A. THE PERFORMANCE OF ACTIVE DEVICE DETECTION

According to the above BOMP algorithm, the correlation between the received signal and the training sequence of devices is a key factor to decide whether the device is active.

Denote the correlation coefficient between the received signal and the training sequence of active and inactive devices as $C_{e0}$ and $C_{e1}$, respectively. $C_{e0}$ and $C_{e1}$ follow a chi-square distribution with $2M$ degrees of freedom.

Define $e_t$ is the average energy of each element of $\mathbf{Z}_t'$, $e_t = \frac{E\left[\left\|\mathbf{Z}_t'\right\|_2^2\right]}{MT} = 1 + \rho_1 N_c \bar{\sigma}_{\hat{h}_i}^2 \frac{d}{T} + (N_c - k)\rho_1 \frac{d}{T}$.

Define $\mu_0$ and $\mu_1$ as the expectation of $C_{e0}$ and $C_{e1}$. They can be expressed as

$$\begin{aligned} \mu_0 &= \left[T^2 + (N_n - 1)T + e_t T\right] M, \\ \mu_1 &= \left[N_n T + e_t T\right] M. \end{aligned} \tag{8}$$

Define $\sigma_0^2$ and $\sigma_1^2$ as the variances of $C_{e0}$ and $C_{e1}$, respectively.

$$\begin{aligned} \sigma_0^2 &= 4M \left\{\frac{1}{2}\left[T^2 + (N_n - 1)T + e_t T\right]\right\}^2, \\ \sigma_1^2 &= 4M \left\{\frac{1}{2}\left[N_n T + e_t T\right]\right\}^2. \end{aligned} \tag{9}$$

According to our assumptions $T \gg N_n$, it can be confirmed that $\mu_0 \gg \mu_1$ and $\sigma_0^2 \gg \sigma_1^2$.

$$P_1(C_{e0} > C_{e1}) = \Phi\left(\frac{\mu_0 - \mu_1}{\sqrt{\sigma_0^2 + \sigma_1^2}}\right) \approx \Phi(\sqrt{M}). \tag{10}$$

The proof of (8), (9), (10) is shown in Appendix A in details.

When an active device is detected, its correlation must be greater than that of all inactive devices. So the MDP can be approximately expressed as

$$\text{MDP} \approx 1 - P_1^{(N - N_n)}. \tag{11}$$

If one inactive device has larger correlation coefficient than any one active device, it may be mistaken as active. So the false alarm probability (FAP) can be expressed as

$$\text{FAP} \approx 1 - P_1^{N_n}. \tag{12}$$

The exact values of MDP and FAP are related to the specific algorithm. However, according to the approximate results given in the equation above, we can draw a conclusion that the performance of activity detection will be improved greatly with the increment of the antenna numbers $M$ at the base station. With the block sparse structure advantage of massive MIMO systems, the MDP of active device detection will tend to zero, and its impact on the system performance is fairly small and negligible according to [12], [13] and [35].

### B. MSE: THE INTERFERENCE OF DATA ON CHANNEL ESTIMATION

With the prior known training sequence and active devices set $I^{\text{ne}}$, the channel can be estimated with (13)

$$\tilde{\mathbf{H}}_{I^{\text{ne}}} = \frac{\mathbf{r}_t \mathbf{X}_{tI^{\text{ne}}}^{\dagger}}{\sqrt{\rho_2}}. \tag{13}$$

where $\mathbf{X}_{tI^{\text{ne}}}^{\dagger} = \mathbf{X}_{tI^{\text{ne}}}{}^H (\mathbf{X}_{tI^{\text{ne}}} \mathbf{X}_{tI^{\text{ne}}}{}^H)^{-1}$ is the pseudo inverse of $\mathbf{X}_{tI^{\text{ne}}}$.

When all active devices are identified, the equivalent channel matrix error can be written as (14)

$$\Delta \mathbf{H}_{I^{\mathrm{ne}}} = \mathbf{H}_{I^{\mathrm{ne}}} - \tilde{\mathbf{H}}_{I^{\mathrm{ne}}} = \frac{\mathbf{Z}'_t \mathbf{X}^{\dagger}_{tI^{\mathrm{ne}}}}{\sqrt{\rho_2}}. \tag{14}$$

The mean square error (MSE) can be written as (15), the trace of matrix expressed by sign tr

$$\mathrm{MSE} = \mathrm{tr}(\Delta \mathbf{H}_{I^{\mathrm{ne}}}{}^{H} \Delta \mathbf{H}_{I^{\mathrm{ne}}})/N_n/M. \tag{15}$$

For LS method:

$$\mathrm{MSE} = \frac{1 + \rho_1 N_c \bar{\sigma}^2_{h_i} \frac{d}{T} + \frac{(N_c-k)\rho_1 d}{T}}{\rho_2 (T - N_n + 1)}. \tag{16}$$

The proof is shown in Appendix B in details.

For any two adjacent transmissions, the number of active devices is approximately the same. Using the same CTST scheme, it can be assumed that the MSE of the channel estimation of the previous stage is nearly the same as the current stage

$$\bar{\sigma}^2_{h_i} \approx \bar{\sigma}^2_{h_j}. \tag{17}$$

The MSE can be obtained by using (16) and (17)

$$\mathrm{MSE} \approx \frac{1 + \frac{(N_c-k)\rho_1 d}{T}}{\rho_2 (T - N_n + 1) - \rho_1 N_c \frac{d}{T}}. \tag{18}$$

When the data recovery frame error rate (FER) is low enough (FER $\ll \frac{T}{N_c \rho_1 d}$)

$$\mathrm{MSE} \approx \frac{1}{\rho_2 (T - N_n + 1) - \rho_1 N_c \frac{d}{T}}. \tag{19}$$

When MSE approaches the result given by equation (19), the additional interference energy ($\|\sqrt{\rho_2}\Delta \mathbf{H}_{I^{\mathrm{ne}}} \mathbf{X}_{tI^{\mathrm{ne}}}\|_2^2 = \mathrm{MSE}\rho_2\ N_c MT$) of training will be far less than the noise energy ($\|\mathbf{Z}\|_2^2 = MT$). In this way, the channel estimation error has little effect on the system performance. In order to analyze the conditions under which MSE performance can reach (19), we will give the performance of data recovery below.

For traditional ST scheme, the residual signal for active device detection and channel estimation are very similar to (6)

$$\mathbf{r}_{t,\mathrm{tra}} = \sqrt{\rho_2} \sum_{n=1}^{N} \mathbf{h}_n a_n \mathbf{x}_{\mathbf{t}n} + \underbrace{\sqrt{\rho_1} \sum_{i \in I_w}^{N} \tilde{\mathbf{h}}_i \mathbf{x}_{\mathbf{d}i} + \mathbf{Z}'}_{\mathbf{Z}'_t}. \tag{20}$$

Without the prior knowledge of $\tilde{\mathbf{h}}_i$, even if the data of some devices are recovered correctly, its influence on channel estimation cannot be eliminated completely. Under the same conditions, the corresponding MSE $\bar{\sigma}^2_{h_i}$ of this method will be larger. The energy of equivalent noise (7) of the traditional ST scheme is slightly larger.

Fortunately, the two schemes can be solved by similar algorithms. For the traditional ST scheme, the set $I^{\mathrm{ne}}$ in the algorithm should be replaced by the set $I$.

## IV. DATA RECOVERY ALGORITHM

In this section, we propose a data recovery algorithm and analyze its performance. The specific steps of iterative algorithm for data recovery and channel estimation are given. The expression for Signal-to-Interference-Plus-Noise Ratio (SINR) is given to analyze data recovery performance. We optimize the power allocation factor to improve the overall performance.

The residual signal after the processing of active device detection and channel estimation is given by

$$\begin{aligned} \mathbf{r}_d &= \mathbf{Y} - \sqrt{\rho_2} \tilde{\mathbf{H}}_{I^{\mathrm{ne}}} \mathbf{X}_{I^{\mathrm{ne}}t} \\ &= \sqrt{\rho_1} \mathbf{H}_{I_w} \mathbf{X}_{dI_w} + \underbrace{\sqrt{\rho_2} \Delta \mathbf{H}_{I^{\mathrm{ne}}} \mathbf{X}_{I^{\mathrm{ne}}t} + \mathbf{Z}'}_{\mathbf{Z}'_d}. \end{aligned} \tag{21}$$

Define $\mathbf{Z}'_{\mathbf{d}} = \sqrt{\rho_2} \Delta \mathbf{H}_{I^{\mathrm{ne}}} \mathbf{X}_{tI^{\mathrm{ne}}} + \mathbf{Z}'$ as the effective noise vector for data recovery.

In the data decoding process, with the formula $\mathrm{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\,\mathrm{vec}(\mathbf{B})$, we rewrite (21) as:

$$\mathrm{vec}(\mathbf{r}_d) = \sqrt{\rho_1} \sum_{i \in I_w} (\mathbf{P}_i \otimes \mathbf{h}_i)\mathbf{s}_i + \mathrm{vec}(\mathbf{Z}'_d). \tag{22}$$

We define $\mathbf{y} := \mathrm{vec}(\mathbf{r}_d)$, $\mathbf{B}_n := (\mathbf{P}_n \otimes \mathbf{h}_n)$, $\mathbf{z} = \mathrm{vec}(\mathbf{Z}'_d)$and $\mathbf{B} := [\mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_N]$, $\mathbf{S} := [\mathbf{s}_1^T, \mathbf{s}_2^T, \cdots, \mathbf{s}_N^T]^T$, then we have

$$\mathbf{y} = \sqrt{\rho_1} \mathbf{B}\mathbf{S} + \mathbf{z}. \tag{23}$$

We extend the first-order statistics-based approach for data recovery. Assume that the channels are sorted as $||\mathbf{h}_1||_2^2 \geqslant ||\mathbf{h}_2||_2^2 \geqslant \cdots \geqslant ||\mathbf{h}_n||_2^2$. In the successive interference cancellation (SIC) scheme, the device with the larger channel response should be recovered earlier.

We use similar method to mitigate the interference between training symbols and data symbols, and relax the pressure of error propagation as given in paper [10]. Error detection codes such as cyclic redundancy check (CRC) code can be utilized to indicate whether the decoded packets are error-free. Once the data of an active device is recovered and confirmed error free, its impact on other devices can be subtracted from the total receiving matrix (6). Then the active device set $I^{\mathrm{ne}}$ for the next transmission and their channel state information can be updated. Detailed steps of this iterative algorithm is presented in the algorithm flow chart as shown in Fig.4.

Using SIC scheme, the pseudo-inverse of the Kronecker product has a simplified algorithm. In order to alleviate the recovery complexity, we recover the data of just one device with equation below

$$\mathbf{B}_n^{\dagger} = (\mathbf{P}_n \otimes \mathbf{h}_n)^{\dagger} = \mathbf{P}_n^{\dagger} \otimes \mathbf{h}_n^{\dagger}. \tag{24}$$

Since the dimension of the pseudo-inverse operation is greatly reduced, the complexity of each iteration is greatly reduced with a slight performance degradation. To trade off the cost and system performance, the joint scheme can be used when SIC scheme fails to recover data. In this way, data recovery can finally achieve the performance of joint scheme.

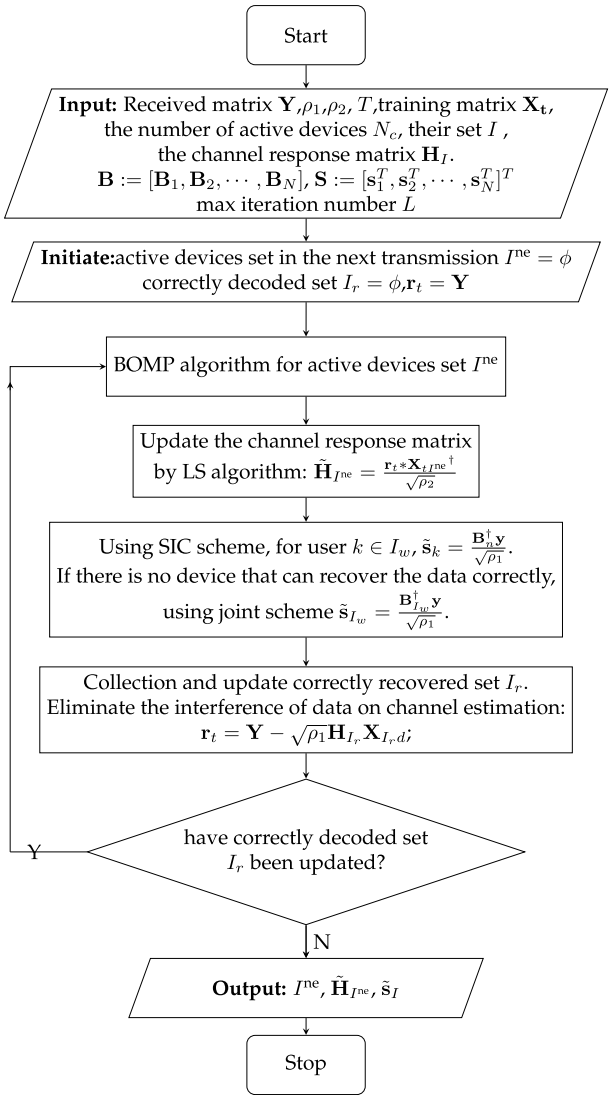


**FIGURE 4. Algorithm 2 iterative algorithm for data recovery and channel estimation.**

In order to maximize the overall performance of our scheme, we will study the SINR of data recovery for each device. Then, the optimal power allocation factor can be given according to the expression of SINR.

When the data of current active devices from 1 to $k$−1, have been recovered, the data that has been correctly recovered can be eliminated from the receiving matrix, and the equivalent receiving matrix can be expressed as

$$\mathbf{r}_k = \sqrt{\rho_1}\mathbf{h}_k\mathbf{x}_{dk} + \sqrt{\rho_1}\sum_{i=k+1}^{N}\mathbf{h}_i\mathbf{x}_{di} + \underbrace{\sqrt{\rho_2}\sum_{j\in I^{\text{ne}}}^{N}\Delta\mathbf{h}_j\mathbf{x}_{tj} + \mathbf{Z}'}_{\mathbf{Z}'_d}. \tag{25}$$

The expression of the effective noise matrix for data recovery is

$$\begin{aligned}\mathbf{Z}'_d &= \sqrt{\rho_1}\Delta\mathbf{H}_{I^{\text{ne}}}\mathbf{X}_{I^{\text{ne}}t} + \mathbf{Z}' \\ &= \sqrt{\rho_1}\Delta\mathbf{H}_I\mathbf{X}_{Id} + \sqrt{\rho_2}\Delta\mathbf{H}_{I^{\text{ne}}}\mathbf{X}_{I^{\text{ne}}t} + \mathbf{Z}. \end{aligned}\tag{26}$$

Using joint scheme, $\tilde{\mathbf{s}}_{I_w} = \frac{\mathbf{B}_{I_w}^{\dagger}\mathbf{y}}{\sqrt{\rho_1}}$. Then the equivalent SINR can be expressed as (27)

$$\begin{aligned}\text{SINR}_k &= \frac{\rho_1(MT-(N_c-k)d)\mathbf{h}_k^H\mathbf{h}_k}{(1+\frac{\rho_2 N_n(1+\frac{(N_c-k)\rho_1 d}{T})+\rho_1 N_c\frac{d}{T}}{\rho_2(T-N_n+1)-\rho_1 N_c\frac{d}{T}})MT} \\ &= \rho_{\text{eff}}\frac{(MT-(N_c-k)d)\mathbf{h}_k^H\mathbf{h}_k}{MT}, \end{aligned}\tag{27}$$

where $\rho_{\text{eff}} = \rho_1\frac{\rho_2(T-N_n+1)-\rho_1 N_c\frac{d}{T}}{\rho_2(T+1)+\rho_2(\frac{(N_c-k)\rho_1 d}{T})}$. The specific derivation process of SINR is shown in Appendix C.

The sum achievable transmission rate of the system can be expressed as

$$R = \sum_{k=1}^{N_c}\log_2(1+\text{SINR}_k). \tag{28}$$

System performance depends on SINR. To observe the influence of power allocation scheme, assume that the total power of each transmission including training phase for training symbols and data phase for data symbols is fixed. $\rho_0 d = \rho_1 d + \rho_2 T$. Define $\lambda \in [0, 1]$ as the data power allocation factor. We optimize the parameter $\lambda = \frac{\rho_1 d}{\rho_1 d+\rho_2 T}$ to maximize the SINR when performing signal recovery.

SINR is determined by noise, the interference of data and training sequence sent by other devices, the interference from the training symbols depends on the channel estimation error. Among them, the energy of data symbols $\rho_1$ and channel estimation error $\sigma_{h_i}$, $\sigma_{h_j}$ are affected by power allocation scheme. Seen from (27), to maximize SINR is to maximize $\rho_{\text{eff}}$.

Only when the interference between data and training sequence is small, the scheme is feasible. When SNR is high enough, it is reasonable to assume that the impact from residual error of data recovery on the channel estimation is negligible as we mentioned in the former section. At this point MSE is given approximately by (19). According to (27), $\rho_{\text{eff}}$ can be represented as a function of $\lambda$

$$\rho_{\text{eff}} = \frac{\lambda[(1-\lambda)(T-N_n+1)-\lambda N_c]\rho_0}{(1-\lambda)(T+1)}. \tag{29}$$

Define: $\eta = \frac{N_c}{T-N_n+1}$. The derivative of $\rho_{\text{eff}}$ about $\lambda$ can be expressed as

$$\frac{\mathrm{d}\rho_{\text{eff}}}{\mathrm{d}\lambda} = \frac{[1-2(1+\eta)\lambda+(1+\eta)\lambda^2](T-N_n+1)\rho_0}{(1-\lambda)^2(T+1)}. \tag{30}$$

When the derivative is zero, the optimal data power allocation factor $\lambda$ can be calculated

$$\lambda_{opt} = 1-\sqrt{1-\frac{1}{1+\eta}} = 1-\sqrt{\frac{N_c}{T-N_n+1+N_c}}. \tag{31}$$

According to (31), because $T \gg N_c$, the power that needs to be allocated to the training part is only a small percentage. However, for low SNR case, FER is inevitably high, MSE is larger than (19). In order to reduce the influence of channel estimation error on data recovery, more energy should

be allocated to training symbols to obtain accurate channel estimation.

Compared with the data recovery performance of ideal situation (perfect CSI, without superimposed training, all the energy is allocated to the data), the performance loss caused by superimposed training can be given. We first give the received signal in the ideal situation.

$$\mathbf{Y} = \sqrt{\rho_0} \sum_{i \in I}^{N} h_i \mathbf{x}_{\mathbf{d}i} + \mathbf{Z}. \tag{32}$$

In this case, the active devices set $I$, and their channel response vector $h_i$ are known to base station. For comparison, we give the SINR expression under perfect CSI as (33)

$$\mathrm{SINR}_k^p = \frac{(MT - (N_c - k)d)\mathbf{h}_k^H \mathbf{h}_k \rho_0}{MT}. \tag{33}$$

The proof of this equation is almost the same as (27).

Through the comparison of (27) and (33), we can analyze the SNR loss caused by superimposed training sequence. The MSE of channel response will be regarded as interference for data recovery. The SINR gap caused by superimposed training is the biggest for the first device.

In the first iteration, all devices' signal have not been recovered correctly. The minimum value of $\rho_{\text{eff}}$ can be expressed as (34)

$$\rho_{\text{eff, min}} = \frac{\rho_1}{1 + \frac{\rho_2 N_n (1 + \frac{N_c \rho_1 d}{T}) + \rho_1 N_c \frac{d}{T}}{\rho_2 (T - N_n + 1) - \rho_1 N_c \frac{d}{T}}}. \tag{34}$$

Define $\delta$ as the gap with or without the influence of channel estimation error in dB and its upper bound can be derived as

$$\begin{aligned} \delta &\leqslant 10 \log_{10} \rho_0 - 10 \log_{10} \rho_{\text{eff, min}} \\ &= 10 \log_{10} \left[ \left( 1 + \frac{\rho_2 N_n (1 + \frac{N_c \rho_1 d}{T}) + \rho_1 N_c \frac{d}{T}}{\rho_2 (T - N_n + 1) - \rho_1 N_c \frac{d}{T}} \right) \frac{1}{\lambda} \right]. \end{aligned} \tag{35}$$

On the contrary, if the data of all devices have been recovered correctly, then the value of $\rho_{\text{eff}}$ will reach the maximum.

$$\rho_{\text{eff, max}} = \frac{\rho_1}{1 + \frac{(\rho_2 N_n + \rho_1 N_c \frac{d}{T})}{\rho_2 (T - N_n + 1) - \rho_1 N_c \frac{d}{T}}}. \tag{36}$$

Then $\delta$ will reach its lower bound.

$$\begin{aligned} \delta &\geqslant 10 \log_{10} \rho_0 - 10 \log_{10} \rho_{\text{eff, max}} \\ &= 10 \log_{10} \left[ \left( 1 + \frac{(\rho_2 N_n + \rho_1 N_c \frac{d}{T})}{\rho_2 (T - N_n + 1) - \rho_1 N_c \frac{d}{T}} \right) \frac{1}{\lambda} \right]. \end{aligned} \tag{37}$$

Compared with the perfect CSI, our proposed CTST scheme has an SNR loss range in data recovery given by the (35) and (37).

In the case of high SNR, $\delta$ will be very close to the lower bound. Under actual circumstances, $T \gg N_n$ is satisfied, the interference of channel estimation error on data recovery is very small. Then the performance of active device detection, channel estimation and data recovery will all achieve their desired boundaries. This is exactly what this program is pursuing, the interference between signals and training sequence can be basically eliminated.

By analyzing SINR, we can give the conditions under which the scheme can be implemented efficiently. To ensure the performance of active devices detection and channel estimation, the SINR should satisfy the equation (38) when each device is recovering its data. For all $k$, the SINR must meet the condition (38)

$$\mathrm{SINR}_k > \gamma. \tag{38}$$

where $\gamma$ is the SINR threshold to ensure the system performance. The channel response parameters of each device on each antenna satisfy the complex Gaussian random distribution, the sum of the energy responses of all antennas ($\mathbf{h}_k^H \mathbf{h}_k$) satisfies the chi-square distribution with a dimension of $2M$.

It can be approximated that the energy of the channel response matrix of the active devices is evenly distributed in the equal probability interval. The probability of having $i$ active devices in a certain equal probability interval is $P(i) = \binom{N_c}{i} (\frac{1}{N_c})^i (\frac{N_c - 1}{N_c})^{N_c - i}$. Then the expectation can be expressed as

$$\sum i \binom{N_c}{i} (\frac{1}{N_c})^i (\frac{N_c - 1}{N_c})^{N_c - i} = 1. \tag{39}$$

That is to say, there is one device in each section on average. Further more, even in a single simulation, the probability of each equal probability interval exceeding two users is relatively small when the number of active devices is large enough.

$$P(i > 2) \approx 1 - \frac{2}{e} - \frac{2}{e^2}, \tag{40}$$

where $e$ is the natural constant.

Then the average value of $k$-th largest response energy can be approximately expressed as (41).

$$E(\mathbf{h}_k^H \mathbf{h}_k) \approx \frac{1}{2} \mathrm{chi2inv} \left( \frac{2(N_c - k) + 1}{2N_c}, 2M \right). \tag{41}$$

Average SINRs of all active devices can be evaluated by substituting (41) into the expression of SINR.

In order to give a simple approximate expression, we take the channel response energy as the average given by (41). For all $k \in [1, N_c]$, equation (38) can be rewritten as (42)

$$\frac{(MT - (N_c - k)d) E(\mathbf{h}_k^H \mathbf{h}_k) \rho_1}{MT \left( 1 + \frac{\rho_2 N_n (1 + \frac{N_c \rho_1 d}{T}) + \rho_1 N_c \frac{d}{T}}{\rho_2 (T - N_n + 1) - \rho_1 N_c \frac{d}{T}} \right)} > \gamma. \tag{42}$$

To get SINR threshold $\gamma$, we can draw a simulation image of the relationship between SINR and FER for one single device.

As shown in Fig.5, the SINR threshold that can ensure the data recovery performance should be set at around $\gamma = 2.5$.

In addition, the number of equations in (4) must be at least equal to the number of variables.

$$MT > N_c d + N_n M. \tag{43}$$

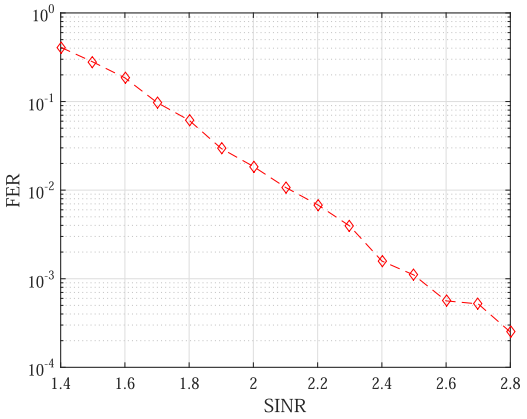


**FIGURE 5.** Relationship between SINR and FER for one single device.

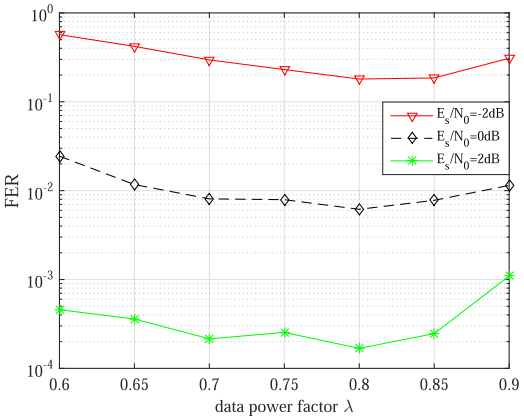


**FIGURE 6.** The influence of optimal data power ratio on data recovery performance.

From (42) and (43), the required training length $T$ or the number of active devices $N_c, N_n$ that the system can support can also be calculated.

## V. NUMERICAL RESULTS

This section presents the simulation studies to evaluate the presented theoretical analysis of the proposed scheme. In all simulations, Quadrature Phase Shift Keying (QPSK) is used for training and data modulation. The elements in channel matrices are independent and identically distributed.

The BS is equipped with $M$ antennas. The active device is randomly selected from all $N$ online devices. The number of current active devices $N_c$ and their set $I$ are known, and the number of active devices $N_n$ and their set $I^{\text{ne}}$ are unknown at the BS. The data frame length T is also the training length, $T = 5d$ in all of our simulations.

The parameters are set as $(T, d, N, N_c, N_n, M) = (1000, 200, 1280, 32, 32, 8)$. From the simulations, we can see that the system performance with eight antennas is good enough. From (27), we can see that the signal recovery performance will increase with the number of antennas. Thanks to the block sparse structure of the MIMO system and the long enough training sequence, the probability of missed detection of active devices in the next transmission is nearly zero in our simulations. This is also in line with our theoretical analysis.

### A. THE INFLUENCE OF DATA POWER ALLOCATION FACTOR ($\lambda$) ON DATA RECOVERY PERFORMANCE

We first simulate the power allocation scheme to obtain the optimal allocation factor. The performance of data recovery for different values of $\lambda$ is shown in Fig. 6: the optimal data power ratio $\lambda$ is about 0.80, this is consistent with the conclusion given by the theoretical formula (31). When the parameters $(T, d, N, N_c, N_n, M) = (1000, 200, 1280, 32, 32, 8)$ are introduced into (31), the theoretical result $\lambda_{opt} = 0.82$ can be obtained. The energy of the actual superimposed training is only a small part of the total energy, then its impact on data recovery will be small. It can be seen that after the power allocation factor is optimized, the performance of data recovery will be significantly improved.

At low SNR, because of the influence of the data that cannot be recovered correctly on the channel estimation, more energy should be allocated to the training part. Optimal value of $\lambda$ is a little bit smaller than 0.80. This is also consistent with our analysis.

In order to get the best performance, we set the power allocation factor to $\lambda = \frac{\rho_1 d}{\rho_1 d + \rho_2 T} = 0.8$ in the later simulation. That is $\rho_1 = 0.8\rho_0$, $\rho_2 = 0.2d\rho_0/T$. The SNR in all our simulations is defined as $E_s/N_0$, where $E_s$ is the overall symbol energy ($\rho_0$) and $N_0$ is the noise spectral density.

### B. CHANNEL ESTIMATION AND DATA RECOVERY PERFORMANCE CHANGES WITH THE INCREASE OF ITERATIONS

Then we simulate the performance of the proposed algorithm, and we can see that the performance of channel estimation and data recovery will be improved after each iteration. Both simulation and theoretical analysis show that the performance gap between the proposed superimposed training scheme and perfect CSI is small enough.

As shown in Fig. 7, after each iteration, the data of more active devices can be recovered and verified correctly, its impact on channel estimation is smaller, and the performance

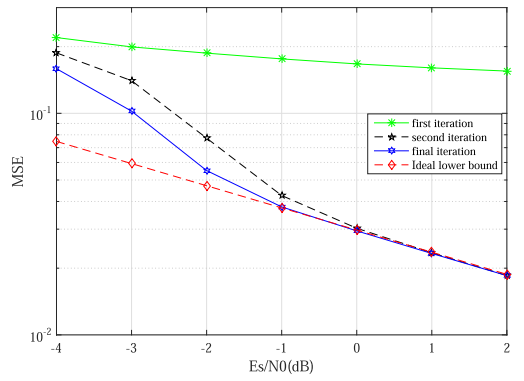


**FIGURE 7.** MSE performance: the interference of data on channel estimation.

of the channel estimation is better. At high SNR, the MSE performance of the channel estimate is very close to the ideal theoretical value (19). In this case, the data symbols has negligible effect on the channel estimation. When the SNR reaches 0 dB, the influence of data recovery error is small enough. As shown in Fig.8, FER is around 0.01 at 0 dB which satisfies the condition FER $\ll \frac{T}{N_c \rho_1 d}$. On the contrary, when the SNR is low, the influence of data symbols on channel estimation is too large.

As shown in Fig. 8, since the MSE of the channel estimation is getting smaller and smaller, each iteration can further reduce the impact of the training sequence on data recovery, then the performance of data recovery will get better and better. Compared to the ideal situation (perfect CSI, without superimposed training), the SNR loss of the decoding performance is only a gap of about 1.8 dB. This just falls within the range given by the theoretical formula (35) and (37) which is [1.7, 4.8], and it is very close to its theoretical lower bound.

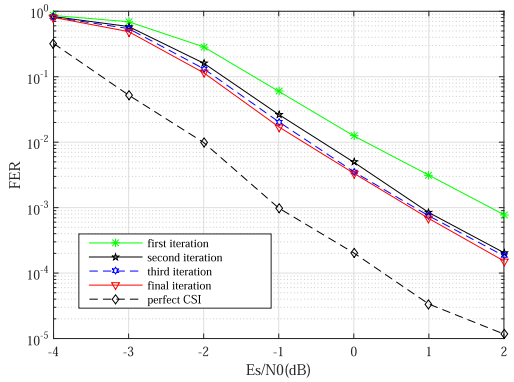


**FIGURE 8.** FER performance: the influence of channel estimation error on data recovery.

After a few iterations, the performance of the channel estimation (MSE) and data recovery performance (FER) will not change. After the third circulation, the MSE and FER performance is very close to the final performance.

### C. FER AND MSE PERFORMANCE: COMPARISON OF A SIMILAR SUPERIMPOSED TRAINING SCHEME

And then, we compared the gap between the performance of the proposed CTST scheme and traditional ST scheme. In order to compare our proposed scheme with traditional ST scheme, we set the same total energy $\rho_0$, the same power distribution factor $\lambda = 0.8$, the same training and data symbols length $T$, the same number of active devices $N_c, N_n$ and antennas $M$ for both scenarios. As shown in Fig. 9, our proposed scheme has better performance than the traditional superimposed training scheme, whether at high SNR or low SNR. Compared to the ideal situation (perfect CSI, without superimposed training), the SNR loss of the proposed CTST scheme is smaller.

Accurate data recovery is also a guarantee of active device detection and channel estimation performance. The channel estimation error of the proposed CTST scheme is also lower

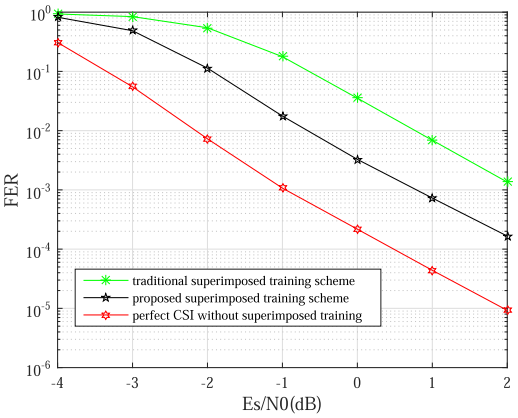


**FIGURE 9.** FER the comparison of a similar superimposed training scheme.

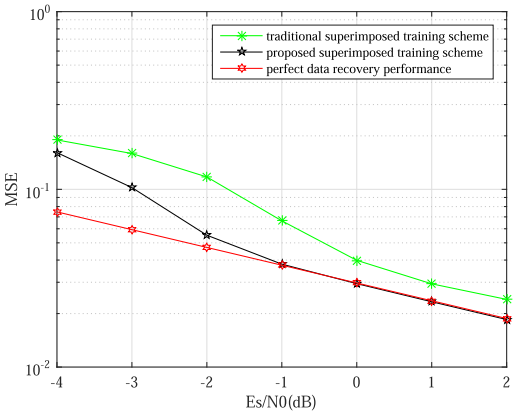


**FIGURE 10.** MSE: comparison of different superimposed training scheme.

than traditional ST scheme as shown in Fig. 10. In turn, the performance of data recovery has also improved.

### D. THE NUMBER OF ACTIVE DEVICES THAT THE SYSTEM CAN SUPPORT

Finally, we simulate the performance of the system with the number of active devices. The number of active devices that the system can accommodate is given. We assume that the number of active devices in two adjacent transmissions is approximately the same $N_c = N_n$.

As shown in Fig. 11, when the number of active devices is less than 32, the result of MSE is close to its ideal value. At this time, active device detection and channel estimation are very little disturbed by data recovery error. This satisfied the condition that guarantees system performance.

As shown in Fig. 12, the performance of data recovery will greatly decrease with the increase of the number of active devices when $N_c, N_n > 32$. According to Fig.5, the SINR threshold should be around $\gamma = 2.5$. Under the given simulation parameters, the number of active devices $N_c, N_n$ that the system can support can be calculated: $N_c, N_n < 33$ according to (42) and $N_c, N_n < 38$ according to (43). The simulation result of the number of active devices $N_c, N_n$ that the system

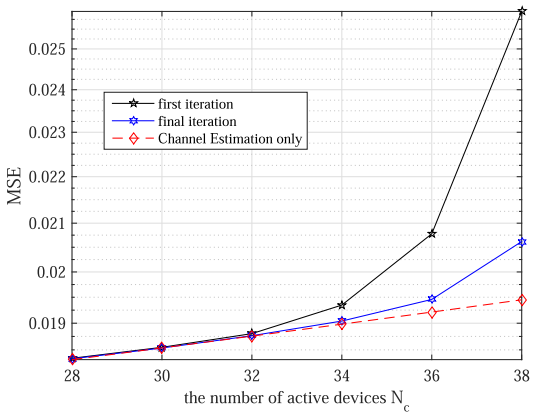


**FIGURE 11.** MSE with different number of active devices $N_c$. The parameters as $(T, d, N, M, E_s/N_0) = (1000, 200, 1280, 8, 2)$.

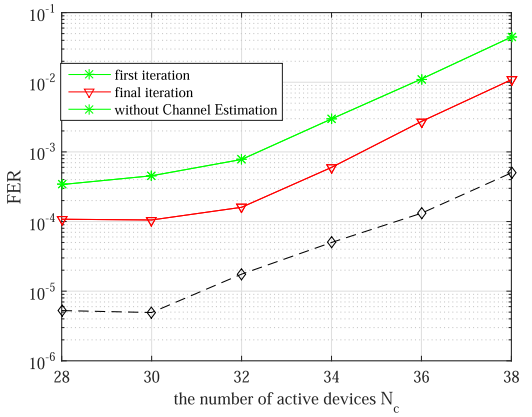


**FIGURE 12.** FER with different number of active devices $N_c$. The parameters as $(T, d, N, M, E_s/N_0) = (1000, 200, 1280, 8, 2)$.

can support is also in agreement with the theoretical results. By increasing the number of antennas $M$ or the extended length of training sequence $T$, the system can accommodate more active devices.

## VI. CONCLUSION

In this paper, the new superimposed training scheme can efficiently improve the utilization of system resources, where active device detection, channel estimation and data recovery can be performed simultaneously, and no extra dedicated resource need to be allocated for channel estimation.

Moreover, the performance of active device detection performance and channel estimation mean square error is also better due to the use of longer training sequences. Furthermore, because of the good performance of active device detection, low complexity algorithm can be used, and the performance of data recovery can also be improved. Simulation results show that the proposed CTST scheme can guarantee the performance of data transmission, channel estimation and active device detection.

In conclusion, the transmission scheme can achieve massive connectivity, which is applicable to future wireless communication system.

## APPENDIX A
## BRIEF DERIVATIONS OF (7), (8), (9) AND (10)

$\mathbf{Z}_t$, $\Delta\mathbf{H}_I$ and $\mathbf{H}_I$ are independent of each other, the energy of each term of equivalent noise can be expressed as

$$E\left[\|\mathbf{Z}_t\|_2^2\right] = MT, \tag{44}$$

$$E\left[\|\mathbf{h}_n\mathbf{x_d}\|_2^2\right] = Md, \tag{45}$$

$$E\left[\|\Delta\mathbf{H}_I\mathbf{X}_{dI}\|_2^2\right] = N_c M\tilde{\sigma}_{h_i}^2 d. \tag{46}$$

Channel estimation error is small enough, we can get $E\left[\|\Delta\mathbf{H}_I\mathbf{X}_{dI}\|_2^2\right] \ll E\left[\|\mathbf{Z}_t\|_2^2\right]$. With results above, we can have result in (7).

Define $\mathbf{Z}_t' = \sqrt{e_t}\mathbf{Z}_0$, then each element of $\mathbf{Z}_0$ is i.i.d. circularly symmetric complex Gaussian distributed random of zero mean and unit variance. For all $i \in I^{ne}, j \notin I^{ne}$

$$\begin{aligned} C_{e0} &= ||\mathbf{r}_t\mathbf{x}_{ti}^H||_2^2 = ||\sqrt{\rho_2}\sum_{n\in I^{ne}}\mathbf{h}_n\mathbf{x}_{tn}\mathbf{x}_{ti}^H + \sqrt{e_t}\mathbf{Z}_0\mathbf{x}_{ti}^H||_2^2, \\ C_{e1} &= ||\mathbf{r}_t\mathbf{x}_{tj}^H||_2^2 = ||\sqrt{\rho_2}\sum_{n\in I^{ne}}\mathbf{h}_n\mathbf{x}_{tn}\mathbf{x}_{tj}^H + \sqrt{e_t}\mathbf{Z}_0\mathbf{x}_{tj}^H||_2^2. \end{aligned} \tag{47}$$

Each device's channel response vectors are independent, as well as between them and noise. We can analyze the energy of each item separately.

$$\begin{aligned} &||\mathbf{h}_i\mathbf{x}_{ti}\mathbf{x}_{ti}^H||_2^2 = T^2||\mathbf{h}_i||_2^2, n = i, \\ &||\mathbf{h}_n\mathbf{x}_{tn}\mathbf{x}_{ti}^H||_2^2 \approx ||\mathbf{h}_n||_2^2||\mathbf{x}_{tn}\mathbf{x}_{ti}^H||_2^2 \approx T||\mathbf{h}_n||_2^2, n \neq i \\ &||\sqrt{e_t}\mathbf{Z}_0\mathbf{x}_{tj}^H||_2^2 = e_t T||\frac{\mathbf{Z}_0\mathbf{x}_{tj}^H}{T}||_2^2. \end{aligned} \tag{48}$$

$\mathbf{h}_i \in \mathbb{C}^{M\times 1}, \mathbf{h}_n \in \mathbb{C}^{M\times 1}, \frac{\mathbf{Z}_0\mathbf{x}_{tj}^H}{T} \in \mathbb{C}^{M\times 1}$ in the above formula, are all with i.i.d. circularly symmetric complex Gaussian distributed random entries of zero mean and unit variance.

$$\begin{aligned} C_{e0} &= \frac{1}{2}\left[T^2 + (N_n - 1)T + e_t T\right]||x_0||_2^2, \\ C_{e1} &= \frac{1}{2}\left[N_n T + e_t T\right]||x_1||_2^2, \end{aligned} \tag{49}$$

where $x_0 \in \mathbb{C}^{M\times 1}$ and $x_1 \in \mathbb{C}^{M\times 1}$ are all with i.i.d. circularly symmetric complex Gaussian distributed random entries of zero mean and unit variance. Then, $||x_0||_2^2$, $||x_1||_2^2$ all follow a chi-squared distribution with $2M$ degrees of freedom. Their expectation is $2M$, and their variance is $4M$. With results above, we can have result in (8) and (9).

Because $\mathbf{x}_{ti}$ and $\mathbf{x}_{tj}$ are mutually independent, after a linear transformation, they are uncorrelated, Then $C_{e0}$ and $C_{e1}$ can be considered uncorrelated. It can be explained by the following equation:

$$\begin{aligned} &Cov(C_{e0}, C_{e1}) \\ &= Cov(\mathbf{x}_{ti}\mathbf{r}_t{}^H\mathbf{r}_t\mathbf{x}_{ti}^H, \mathbf{x}_{tj}\mathbf{r}_t{}^H\mathbf{r}_t\mathbf{x}_{tj}^H) \\ &= Cov(\mathbf{x}_{ti}\mathbf{R}\mathbf{x}_{ti}^H, \mathbf{x}_{tj}\mathbf{R}\mathbf{x}_{tj}^H) \\ &= Cov(\sum_{m=1}^{T}\sum_{n=1}^{T}R_{mn}x_{tim}x_{tin}, \sum_{p=1}^{T}\sum_{q=1}^{T}R_{pq}x_{tjp}x_{tjq}) \end{aligned}$$

$$= \sum_{m=1}^{T}\sum_{n=1}^{T}\sum_{p=1}^{T}\sum_{q=1}^{T} R_{mn}R_{pq}Cov(x_{tim}x_{tin}, x_{tjp}x_{tjq})$$
$$= 0, \tag{50}$$

where $\mathbf{R} = \mathbf{r}_t{}^H\mathbf{r}_t \in \mathbb{C}^{T\times T}$, $R_{mn}$ and $R_{pq}$ are elements of matrix $\mathbf{R}$.

Chi-square distribution of high dimension can be approximately treated as Gaussian distribution according to central limit theorem. We use the following parameters to show the distinction between active and inactive devices

$$P(C_{e0} > C_{e1}) = P(C_{e0} - C_{e1} > 0) = \Phi\left(\frac{\mu_0 - \mu_1}{\sqrt{\sigma_0^2 + \sigma_1^2}}\right).$$

From the equation above, we can easily conclude that with larger data length, the cross correlation coefficient can be reduced, so superposition is a good solution to extend the training symbols to whole time-frequency resource block.

According to our assumptions $T \gg N_n$, it can be confirmed that $\mu_0 \gg \mu_1$ and $\sigma_0^2 \gg \sigma_1^2$. With results above, we can have result in (10).

## APPENDIX B
## BRIEF DERIVATIONS OF MSE THEORETICAL FORMULA (16)

Due to space limitation, we only present the important steps of the derivations. the equivalent channel matrix error can be written as (51)

$$\Delta\mathbf{H}_{I^{\mathrm{ne}}} = \mathbf{H}_{I^{\mathrm{ne}}} - \tilde{\mathbf{H}}_{I^{\mathrm{ne}}} = \frac{\mathbf{Z}_t'\mathbf{X}_{tI^{\mathrm{ne}}}^{\dagger}}{\sqrt{\rho_2}}. \tag{51}$$

Equivalent noise $\mathbf{Z}_t'$ is independent of the training sequence, $\left\|\mathbf{Z}_t'\right\|_2^2 = ||\mathbf{Z}' + \sqrt{\rho_1}\sum_{n\in I_w}^{N} h_n\mathbf{x}_{\mathbf{d}n}^T||_2^2 = MT + \rho_1 MN_c\bar{\sigma}_{h_i}^2 d + (N_c - k)\rho_1 Md$.

$$\begin{aligned}\mathrm{MSE} &= \mathrm{tr}(\Delta\mathbf{H}_{I^{\mathrm{ne}}}^H\Delta\mathbf{H}_{I^{\mathrm{ne}}})/N_n/M \\ &= \frac{1}{\rho_1 N_n M}\mathrm{tr}((\mathbf{Z}_t'\mathbf{X}_{I^{\mathrm{ne}}}{}^{\dagger})^H(\mathbf{Z}_t'\mathbf{X}_{I^{\mathrm{ne}}}{}^{\dagger})) \\ &= \frac{1}{\rho_1 N_n M}\mathrm{tr}((\mathbf{X}_{I^{\mathrm{ne}}}{}^{\dagger})^H(\mathbf{Z}_t')^H(\mathbf{Z}_t'\mathbf{X}_{I^{\mathrm{ne}}}{}^{\dagger})).\end{aligned} \tag{52}$$

$(\mathbf{Z}_t')^H(\mathbf{Z}_t')$ can be approximated as a diagonal matrix

$$E[(\mathbf{Z}_t')^H(\mathbf{Z}_t')] \approx Me_t\mathbf{I}, \tag{53}$$

where $e_t = 1 + \rho_1 N_c\bar{\sigma}_{h_i}^2\frac{d}{T} + \frac{(N_c-k)\rho_1 d}{T}$ is the average energy of each element of $\mathbf{Z}_t'$, substituting (53) into equation (52),

$$\begin{aligned}\mathrm{MSE} &= \frac{e_t}{\rho_2 N_n}\mathrm{tr}((\mathbf{X}_{tI^{\mathrm{ne}}}{}^{\dagger})^H(\mathbf{X}_{tI^{\mathrm{ne}}}{}^{\dagger})) \\ &= \frac{e_t}{\rho_2 N_n}\mathrm{tr}((\mathbf{X}_{tI^{\mathrm{ne}}}{}^H\mathbf{X}_{tI^{\mathrm{ne}}})^{-1}) \\ &= \frac{e_t}{\rho_2}\bar{\lambda}.\end{aligned} \tag{54}$$

$\bar{\lambda}$ is the average of the eigenvalues of matrix $(\mathbf{X}_{tI^{\mathrm{ne}}}{}^H\mathbf{X}_{tI^{\mathrm{ne}}})^{-1}$, according to the theory of random matrices.

$$\bar{\lambda} = \frac{1}{T - N_n + 1}. \tag{55}$$

With results from (54) to (55), we can have result in (16) and (19).

## APPENDIX C
## BRIEF DERIVATIONS OF EXPRESSION OF SINR USING JOINT DATA RECOVERY SCHEME

The data symbols will be got by LS method, using equation (23)

$$\tilde{\mathbf{S}} = \frac{\mathbf{B}^{\dagger}\mathbf{r}_d}{\sqrt{\rho_1}}(\mathbf{B}^{\dagger} = \mathbf{B}^H(\mathbf{B}\mathbf{B}^H)^{-1}). \tag{56}$$

Furthermore, we have the following two results $\overline{\mathbf{S}} = \mathbf{B}^{\dagger}\mathbf{y} = \mathbf{S} + \mathbf{B}^{\dagger}\mathbf{z}$.

The corresponding estimation error can be written as

$$\Delta\overline{\mathbf{S}} = \mathbf{B}^{\dagger}\mathbf{y} - \mathbf{S} = \mathbf{B}^{\dagger}\mathbf{z}, \tag{57}$$

where $\mathbf{z} = \mathrm{vec}(\mathbf{Z}_d')$. After normalization, each element of matrix B satisfies the Gaussian random distribution. Through the asymptotic result of RMT theorem 2.35 in [36], this is also a good approximation of the rational small-scale matrix dimension, and the pseudo-inverse of B. The eigenvalues of the matrix have an empirical distribution. Define $\lambda_i$ as the eigenvalue of the matrix $(\mathbf{B}^{\mathrm{H}}\mathbf{B})^{-1}$. $\mathrm{E}[\lambda_i] = \frac{MT}{MT-(N_c-k)d}$.

$$\begin{aligned}\mathrm{E}\left[\left\{\|\Delta S\|_2^2\right\}\right] &= \mathrm{E}\left[\mathrm{tr}\left\{\mathbf{B}^{\dagger}\mathbf{z}\mathbf{z}^{\mathrm{H}}\left(\mathbf{B}^{\dagger}\right)^{\mathrm{H}}\right\}\right] \\ &\approx e_d\mathrm{E}\left[\mathrm{tr}\left\{\mathbf{B}^{\dagger}\left(\mathbf{B}^{\dagger}\right)^{\mathrm{H}}\right\}\right] \\ &= e_d\mathrm{E}\left[\mathrm{tr}\left\{\left(\left(\mathbf{B}^{\mathrm{H}}\mathbf{B}\right)^{-1}\right)^{\mathrm{H}}\right\}\right] \\ &= e_d\mathrm{E}\left[\mathrm{tr}\left\{\left(\mathbf{B}^{\mathrm{H}}\mathbf{B}\right)^{-1}\right\}\right] \\ &= (N_c - k)d\mathrm{E}[\lambda_i] \\ &= \frac{MT(N_c-k)de_d}{MT - (N_c-k)d},\end{aligned} \tag{58}$$

where $\mathbf{z}\mathbf{z}^{\mathrm{H}} = e_d\mathrm{I}_{MT} + e_d\mathrm{G}$. Among them, the diagonal element of G is a random small Gaussian distribution. And $e_d$ is the energy of each element of $\mathbf{z} = \mathrm{vec}(\mathbf{Z}_d')$.

$$\mathbf{Z}_d' = \sqrt{\rho_1}\Delta\mathbf{H}_I\mathbf{X}_{Id} + \sqrt{\rho_2}\Delta\mathbf{H}_{I^{\mathrm{ne}}}\mathbf{X}_{I^{\mathrm{ne}}t} + \mathbf{Z}. \tag{59}$$

$\mathbf{Z}$, $\Delta\mathbf{H}_I\mathbf{X}_{Id}$ and $\Delta\mathbf{H}_I^{\mathrm{ne}}\mathbf{X}_{I^{\mathrm{ne}}t}$ are independent of each other.

$$\begin{aligned}E\left[\left\|\mathbf{Z}_d'\right\|_2^2\right] &= E\left[\left\|\sqrt{\rho_1}\Delta\mathbf{H}_I\mathbf{X}_{Id} + \sqrt{\rho_2}\Delta\mathbf{H}_{I^{\mathrm{ne}}}\mathbf{X}_{I^{\mathrm{ne}}t} + \mathbf{Z}\right\|_2^2\right] \\ &= \rho_1 N_c\mathrm{MSE}Md + \rho_2 N_n\mathrm{MSE}MT + MT \\ &= \left[1 + \frac{\rho_2 N_n(1 + \frac{(N_c-k)\rho_1 d}{T}) + \rho_1 N_c\frac{d}{T}}{\rho_2(T - N_n + 1) - \rho_1 N_c\frac{d}{T}}\right]MT \\ &= e_d MT.\end{aligned}$$

We can get $e_d = 1 + \frac{\rho_2 N_n(1+\frac{(N_c-k)\rho_1 d}{T})+\rho_1 N_c \frac{d}{T}}{\rho_2(T-N_n+1)-\rho_1 N_c \frac{d}{T}}$.

After normalization, the energy of each element should be approximately equal, so the channel response for the $k$-th device is the average of each device. The estimated error of the corresponding $k$-th device is

$$\Delta\bar{\mathbf{s}}_k = \Delta\bar{\mathbf{S}}[1:d]. \tag{60}$$

Its energy can be written as

$$\begin{aligned} \mathrm{E}\left[\left\{\|\Delta\mathbf{s}_k\|_2^2\right\}\right] &= \mathrm{E}\left[\left\{\|\Delta S\|_2^2\right\}\right]/(N_c-k) \\ &= \frac{\mathrm{MTd}}{\mathrm{MT}-(N_c-k)d}. \end{aligned} \tag{61}$$

The corresponding SINR of the $k$-th device is

$$\mathrm{SINR}_k = \frac{\left\|\mathbf{s}_k'\right\|_2^2}{\|\Delta\mathbf{s}_k\|_2^2} = \frac{(\mathrm{MT}-(N_c-k)d)\left(\mathbf{h}_k^{\mathrm{H}}\mathbf{h}_k\right)\rho_{\mathrm{eff}}}{MT}, \tag{62}$$

where $\rho_{\mathrm{eff}} = \frac{\rho_1}{e_d}$.

## REFERENCES


[1] P. Pirinen, "A brief overview of 5G research activities," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity*, Nov. 2014, pp. 17–22.

[2] M. Biguesh and A. 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, Mar. 2006.

[3] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[4] A. Grant, "Joint decoding and channel estimation for linear MIMO channels," in *Proc. IEEE Wireless Commun. Netw. Conference. Conf. Rec.*, Sep. 2000, pp. 1009–1012.

[5] C. Budianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2515–2528, Oct. 2002.

[6] L. Dai, B. Wang, Y. Yuan, S. Han, C.-L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: Solutions, challenges, opportunities, and future research trends," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 74–81, Sep. 2015.

[7] H. Nikopour and H. Baligh, "Sparse code multiple access," in *Proc. IEEE Int. Symp. Pers. Indoor Mobile Radio Commun.*, Sep. 2013, pp. 332–336.

[8] M. Taherzadeh, H. Nikopour, A. Bayesteh, and H. Baligh, "SCMA codebook design," in *Proc. IEEE 80th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2014, pp. 1–5.

[9] K. Au, L. Zhang, H. Nikopour, E. Yi, A. Bayesteh, U. Vilaipornsawai, J. Ma, and P. Zhu, "Uplink contention based SCMA for 5G radio access," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2014, pp. 900–905.

[10] R. Xie, H. Yin, X. Chen, and Z. Wang, "Many access for small packets based on precoding and sparsity-aware recovery," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4680–4694, Nov. 2016.

[11] X. Xu, X. Rao, and V. K. N. Lau, "Active user detection and channel estimation in uplink CRAN systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2727–2732.

[12] G. Hannak, M. Mayer, A. Jung, G. Matz, and N. Goertz, "Joint channel estimation and activity detection for multiuser communication systems," in *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, Jun. 2015, pp. 2086–2091.

[13] X. Chen and D. Guo, "Many-access channels: The Gaussian case with random user activities," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun. 2014, pp. 3127–3131.

[14] L. Liu and W. Yu, "Massive connectivity with massive MIMO—Part I: Device activity detection and channel estimation," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2933–2946, Jun. 2018.

[15] H. Kwon and B. D. Rao, "On the benefits of the block-sparsity structure in sparse signal recovery," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 3685–3688.

[16] Z. Ben-Haim and Y. C. Eldar, "Near-oracle performance of greedy block-sparse estimation techniques from noisy measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 5, pp. 1032–1047, Sep. 2011.

[17] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami, "Channel estimation and symbol detection for block transmission using data-dependent superimposed training," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 226–229, Mar. 2005.

[18] H. Zhang, S. Gao, D. Li, H. Chen, and L. Yang, "On superimposed pilot for channel estimation in multicell multiuser MIMO uplink: Large system analysis," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1492–1505, Mar. 2016.

[19] N. Garg, A. Jain, and G. Sharma, "Partially loaded superimposed training scheme for large MIMO uplink systems," *Wireless Pers. Commun.*, vol. 100, pp. 1–26, Mar. 2018.

[20] Z. Gao, Z. Wang, and L. Dai, "Structured compressive sensing based superimposed pilot design in downlink large-scale MIMO systems," *Electron. Lett.*, vol. 50, no. 12, pp. 896–898, Jun. 2014.

[21] J. Cai, X. He, and R. Song, "Pilot optimization for structured compressive sensing based channel estimation in large-scale MIMO systems with superimposed pilot pattern," *Wireless Pers. Commun.*, vol. 100, no. 3, pp. 977–993, Jun. 2018.

[22] H. Xing, D. Wei, and F. Yin, "Channel estimation using orthogonal superimposed pilot for MIMO systems," in *Proc. 2nd Int. Conf. Signal Process. Syst.*, vol. 2, Jul. 2010, pp. 625–628.

[23] J. Ma, C. Liang, C. Xu, and L. Ping, "On orthogonal and superimposed pilot schemes in massive MIMO NOMA systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2696–2707, Dec. 2017.

[24] W. Xin, Y. Wang, and J. Lu, "Superimposed or time-division multiplexed pilots in time-varying channels: An informative consideration," in *Proc. Int. Conf. Wireless Commun. Mobile Comput. Connecting World Wirelessly (IWCMC)*. New York, NY, USA: Association Computing Machinery, 2009, pp. 770–774, doi: 10.1145/1582379.1582546.

[25] A. T. Asyhari and S. ten Brink, "Orthogonal or superimposed pilots? A rate-efficient channel estimation strategy for stationary MIMO fading channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2776–2789, May 2017.

[26] X. Jing, M. Li, H. Liu, S. Li, and G. Pan, "Superimposed pilot optimization design and channel estimation for multiuser massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11818–11832, Dec. 2018.

[27] K. Upadhya, S. A. Vorobyov, and M. Vehkapera, "Superimposed pilots: An alternative pilot structure to mitigate pilot contamination in massive MIMO," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3366–3370.

[28] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[29] W. Wang, K. Zhang, X. Chen, and G. Wei, "Fast sparse channel estimation for many access MIMO system: On the benefit of block sparsity," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 116–122.

[30] S. Narayanan, S. K. Sahoo, and A. Makur, "Greedy pursuits assisted basis pursuit for compressive sensing," in *Proc. 23rd Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 694–698.

[31] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[32] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[33] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, pp. 301–321, May 2009.

[34] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[35] Z. Kun, W. Wang, X. Chen, and G. Wei, "A novel sparse channel estimation method for multiuser MIMO systems," in *Communications, Signal Processing, and Systems*. Singapore: Springer, 2019, pp. 755–765.

[36] A. M. Tulino and S. Verdíš, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004, doi: 10.1561/0100000001.

**KUN ZHANG** received the B.S. degree from the Department of Electronic Engineering, University of Science and Technology of China, Hefei, China, in 2013, where he is currently pursuing the Ph.D. degree in communication and information engineering. His current research interests include compression sensing and channel estimation in many access MIMO systems.

**WENDI WANG** (Graduate Student Member, IEEE) received the B.S. degree from the Department of Electronic Engineering, University of Science and Technology of China, Hefei, China, in 2016, where she is currently pursuing the Ph.D. degree in communication and information engineering. Her current research interests include millimeter wave communication, beamforming, and wireless channel estimation.

**HUARUI YIN** (Member, IEEE) received the bachelor's and Ph.D. degrees in electronic engineering and information science from the University of Science and Technology of China, Hefei, Anhui, China, in 1996 and 2006, respectively. Since 1999, he has been with the Department as an Associate Professor. His research interests include digital signal processing, low-complexity receiver design, and throughput analysis of wireless networks.