

How Often Should CSI be Updated for Massive MIMO Systems with Massive Connectivity?

Ruichen Deng, Zhiyuan Jiang, Sheng Zhou, and Zhisheng Niu
Tsinghua National Laboratory for Information Science and Technology
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
Email: drc13@mails.tsinghua.edu.cn, {zhiyuan,sheng.zhou,niu zhs}@tsinghua.edu.cn

Abstract—Massive multiple-input multiple-output (MIMO) systems need to support massive connectivity for the application of the Internet of things (IoT). The overhead of channel state information (CSI) acquisition becomes a bottleneck in the system performance due to the increasing number of users. An intermittent estimation scheme is proposed to ease the burden of channel estimation and maximize the sum capacity. In the scheme, we exploit the temporal correlation of MIMO channels and analyze the influence of the age of CSI on the downlink transmission rate using linear precoders. We show the CSI updating interval should follow a quasi-periodic distribution and reach a trade-off between the accuracy of CSI estimation and the overhead of CSI acquisition by optimizing the CSI updating frequency of each user. Numerical results show that the proposed intermittent scheme provides significant capacity gains over the conventional continuous estimation scheme.

I. INTRODUCTION

By applying a large number of antenna elements, the massive multiple-input multiple-output (MIMO) system exploits the potentials of spatial multiplexing with tremendous degrees-of-freedom, and is believed to increase the spectral efficiency 10 times or more as well as improving the radiated energy efficiency in the order of 100 times [1]. Exploiting the advantages of massive MIMO technique requires the knowledge of channel state information (CSI) at the base station (BS) side. The canonical massive MIMO protocol is to operate in time-division-duplex (TDD) mode in order to acquire CSI through channel reciprocity [2]. However, the overhead of channel estimation is still unbearable for the application of the Internet of things (IoT) [3], where the systems need to serve a vast quantity of users within the limited channel coherence time.

Many prior works try to exploit the spatial correlation of channels to reduce the overhead of channel acquisition. The spatial correlation matrices of users are used to design inter-cell and intra-cell pilot reuse schemes [4] [5]. Ref. [6] assumes channel sparsity in spatial domain and estimates channels based on the compressive sensing theory. Mutual-information-optimal pilots for minimum mean square error (MMSE) estimation are proposed in [7] [8].

Another aspect of the channel statistics worth exploring is the temporal correlation. The correlation between the current channel and the previous channel becomes weaker with longer time elapsed [9]. Under the assumption of the Gilbert-Elliott channel model, the authors introduce the concept *the age of CSI* and study its impact on a general utility function [10]. A

more realistic model to characterize the temporal correlation of channels is the first-order Gaussian-Markov process [11]. Training techniques and pilot beam design for single user are discussed under this model [12] [13].

Our work focuses on the multiuser scenario. To ease the burden of CSI acquisition, we propose an intermittent estimation scheme. Different from the conventional scheme which continuously estimates channels of the whole users in every channel block, the proposed scheme exploits the temporal correlation and requires users to estimate their channels intermittently among blocks. The CSI updating interval of each user is shown to follow a quasi-periodic distribution and the CSI updating frequency is optimized based on the temporal correlation coefficient of each user. Hence, the scheme reaches a trade-off between the accuracy of the CSI and the overhead of the CSI acquisition to achieve the maximum sum capacity.

The rest of the paper is organized as follows. Section II describes the system model adopted in this paper. The influence of aged CSI on the downlink transmission rate is investigated in Section III. Then we propose the intermittent estimation scheme in Section IV. The numerical results are presented in Section V and conclusions are drawn in Section VI.

II. SYSTEM MODEL

We consider a massive MIMO system with M antenna elements serving K users. The processes of channel estimation and data transmission occur periodically in a block of L channel uses. The system operates in the calibrated time-division-duplex (TDD) mode so that channel reciprocity is utilized for channel estimation. More specifically, in the beginning of every block, a number of T channel uses is spent by the base station (BS) on estimating uplink channels, afterwards the BS obtains downlink channel vectors as the transpose of uplink channel vectors, and transmits precoded data in the remaining $L - T$ channel uses.

The uplink channel of the k -th user in the i -th block is

$$\mathbf{g}_k(i) = \sqrt{\beta_k} \mathbf{h}_k(i), \quad (1)$$

where $\sqrt{\beta_k}$ and $\mathbf{h}_k(i)$ denotes the large and small scale fading coefficient, respectively. The small scale fading coefficient satisfies the complex circularly-symmetric Gaussian distribution $\mathcal{CN}(0, \mathbf{I}_M)$. It stays constant in each block and

evolves between blocks according to the first-order stationary Gaussian-Markov process [11]

$$\mathbf{h}_k(i+1) = \rho_k \mathbf{h}_k(i) + \sqrt{1 - \rho_k^2} \mathbf{e}_k(i), \quad i \geq 0, \quad (2)$$

where ρ_k is the temporal correlation coefficient of channels between blocks, and $\mathbf{e}_k(i)$ is the innovation process distributed as $\mathcal{CN}(0, \mathbf{I}_M)$. We assume the initial channels $\mathbf{g}_k(0)$ and the innovation processes $\mathbf{e}_k(i)$ of each user are independent, so the channels of different users are independent in each block.

If the channel is updated in the i -th block, then the channel after n blocks can be easily derived by iterating (2), which is

$$\mathbf{h}_k(i+n) = \rho_k^n \mathbf{h}_k(i) + \sqrt{1 - \rho_k^{2n}} \mathbf{e}_{k,n}(i), \quad n \geq 1, \quad (3)$$

where $\mathbf{e}_{k,n}(i) \sim \mathcal{CN}(0, \mathbf{I}_M)$ is the equivalent innovation process in n blocks. *The age of CSI* [10] is defined as the number of blocks elapsed since last channel estimation in our work. We focus on the impact of the CSI updating frequency and only consider estimation errors caused by aging of CSI. As the age of CSI n increases, the temporal correlation between the estimated channel $\mathbf{h}_k(i)$ and the channel $\mathbf{h}_k(i+n)$ of ground truth declines exponentially as ρ_k^n , making the estimation less accurate.

III. TRANSMISSION RATE WITH AGED CSI

The BS needs to precode the user data before downlink transmission to support spatial multiplexing. The received signal of the k -th user is

$$\begin{aligned} y_k &= \sqrt{\epsilon_k} \mathbf{g}_k^T \mathbf{v}_k d_k + \sum_{i=1, i \neq k}^K \sqrt{\epsilon_i} \mathbf{g}_i^T \mathbf{v}_i d_i + \mathbf{n}_k \\ &= \sqrt{\epsilon_k \beta_k} \mathbf{h}_k^T \mathbf{v}_k d_k + \sum_{i=1, i \neq k}^K \sqrt{\epsilon_i \beta_i} \mathbf{h}_i^T \mathbf{v}_i d_i + \mathbf{n}_k, \end{aligned} \quad (4)$$

where ϵ_k , \mathbf{v}_k , d_k and \mathbf{n}_k denotes SNR at the transmitter, precoding vector, data symbol and normalized downlink noise, respectively. A lower bound of per user SINR at the receiver can be obtained by using the similar technique as [14, Theorem 1], which is denoted as

$$\gamma_k = \frac{\epsilon_k \beta_k |\mathbb{E} \mathbf{h}_k^T \mathbf{v}_k|^2}{1 + \epsilon_k \beta_k (\mathbb{E} |\mathbf{h}_k^T \mathbf{v}_k|^2 - |\mathbb{E} \mathbf{h}_k^T \mathbf{v}_k|^2) + \sum_{i \neq k} \epsilon_i \beta_i \mathbb{E} |\mathbf{h}_i^T \mathbf{v}_i|^2}. \quad (5)$$

And the downlink transmission rate is obtained:

$$R_k = \log(1 + \gamma_k). \quad (6)$$

We consider two kinds of linear precoding schemes, namely matched filter (MF) and zero forcing (ZF). Assume the estimated channel matrix is $\hat{\mathbf{H}}_k$, then the precoding matrix of matched filter is

$$\mathbf{V}_k^{\text{MF}} = \frac{1}{\sqrt{\eta_k^{\text{MF}}}} \hat{\mathbf{H}}_k^*, \quad (7)$$

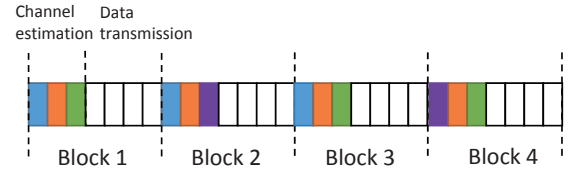


Fig. 1. An example of the intermittent estimation scheme.

where η_k^{MF} is the normalization coefficient. In the massive MIMO scenario, the value of η_k^{MF} tends to be constant due to the channel hardening effect, which is derived as

$$\eta_k^{\text{MF}} = \frac{1}{K} \mathbb{E} \text{tr} \{ \hat{\mathbf{H}}_k^T \hat{\mathbf{H}}_k^* \} = M. \quad (8)$$

For zero forcing precoding, the precoding matrix becomes

$$\mathbf{V}_k^{\text{ZF}} = \frac{1}{\sqrt{\eta_k^{\text{ZF}}}} \hat{\mathbf{H}}_k^* (\hat{\mathbf{H}}_k^T \hat{\mathbf{H}}_k^*)^{-1}, \quad (9)$$

where the normalization coefficient is

$$\eta_k^{\text{ZF}} = \frac{1}{K} \mathbb{E} \text{tr} \{ (\hat{\mathbf{H}}_k^T \hat{\mathbf{H}}_k^*)^{-1} \} = \frac{1}{M - K}. \quad (10)$$

Theorem 1. *If the CSI of the k -th user has an age of n blocks, then the SINR using MF precoding is given by*

$$\gamma_k^{\text{MF}}(n) = \frac{\epsilon_k \beta_k M \rho_k^{2n}}{1 + \sum_{i=1}^K \epsilon_i \beta_i}, \quad (11)$$

and the SINR using ZF precoding is given by

$$\gamma_k^{\text{ZF}}(n) = \frac{\epsilon_k \beta_k (M - K) \rho_k^{2n}}{1 + (1 - \rho_k^{2n}) \sum_{i=1}^K \epsilon_i \beta_i}. \quad (12)$$

Proof. See Appendix A. \square

From the above theorem, we observe the fact that for both precoding schemes, the SINR functions of the k -th user monotonically decreases as its age of CSI grows, and are not affected by the ages of CSI of other users. For the convenience of expressions, we will use the unified term $\gamma_k(n)$, $R_k(n)$ for the two precoding schemes in the rest of the paper.

IV. INTERMITTENT ESTIMATION SCHEME

A. Scheme Description

The channel estimation overhead increases linearly with the total number of users. When user number increases to a certain amount, the channel block can no longer support the estimation overhead of all users. So we propose an intermittent channel estimation scheme. In every block, the BS chooses a subset of users instead of the whole users to estimate their channels. Afterwards, the BS forms the precoder matrix using the latest CSI of each user and then transmits data to all users. The scheme reduces the overhead of channel estimation at the cost of a decline of the estimation accuracy.

An example of the intermittent estimation scheme is illustrated in Fig. 1. The system provides data service for total 4 users. The pilot pattern has a period of 4 blocks, and the system

selects 3 users from the whole user set for channel estimation in each block. In each period of the pilot pattern, the CSI of blue user is updated twice with an interval of 1 block and once with an interval of 2 blocks. So its average updating interval is $\frac{4}{3}$ blocks and CSI updating frequency equals $\frac{3}{4}$ per block. Similarly, the CSI updating frequencies of orange, green and purple users are 1, $\frac{3}{4}$ and $\frac{1}{2}$ per block, respectively.

For a given user set, we try to maximize the average user sum rate by choosing the proper user subset for channel estimation in every block. The optimization problem is formulated as

$$\begin{aligned} \max_{q_k(i), T} \quad & \lim_{I \rightarrow \infty} \frac{1}{I} \left(1 - \frac{T}{C}\right) \sum_{i=0}^{I-1} \sum_{k=1}^K R_{k,i} \\ \text{s.t.} \quad & q_k(i) \in \{0, 1\}, \quad k = 1, \dots, K \\ & \sum_k q_k(i) = T, \end{aligned} \quad (13)$$

where $q_k(i)$ indicates whether the k -th user is chosen to do channel estimation in the i -th block. The sum rate is multiplied by a discounting factor $1 - \frac{T}{C}$ where $\frac{T}{C}$ represents the ratio of estimation overhead.

Since the transmission rate of the k -th user is only affected by its own age of CSI, rearranging the order of CSI updating intervals has no influence on the rate performance of the k -th user. Denote p_k as the CSI updating frequency and $f_{k,n}$ as the distribution CSI updating interval. The CSI updating frequency is the reciprocal of the average CSI updating interval:

$$p_k = \frac{1}{\sum_n n f_{k,n}}. \quad (14)$$

We consider a new optimization problem as

$$\begin{aligned} \max_{T, p_k, f_{k,n}} \quad & \left(1 - \frac{T}{C}\right) \sum_{k=1}^K \mathbb{E}\{R_k\} \\ \text{s.t.} \quad & 0 \leq f_{k,n} \leq 1, \quad k = 1, \dots, K \\ & \sum_n f_{k,n} = 1, \quad k = 1, \dots, K \\ & \sum_k p_k = T, \end{aligned} \quad (15)$$

which is actually a relaxation of the original problem (13). The second constraint is relaxed from pilot length in every block equaling T to average pilot length equaling T . When the number of users goes to infinity, the pilot length tends to be the sum of the CSI updating frequencies of all users by the law of large numbers. Therefore, the relaxation is asymptotically tight.

The relaxed problem can be solved in three steps. Firstly, we fix the value of the CSI updating frequency p_k and the pilot length T and try to optimize the distribution $f_{k,n}$ of CSI updating interval for each user. Then p_k is optimized under fixed value of T . Finally we find the optimal T by a one-dimensional search.

B. Optimizing the Distribution of CSI Updating Interval

To solve the relaxed problem (15), we firstly consider the rate performance of a single user, and optimize $f_{k,n}$ under fixed value of the CSI updating frequency p_k .

Assume the k -th user has done N_k times of channel estimations in a total of N blocks, and there are $F_{k,n}$ CSI updating intervals with block length of n . The total rate in a CSI updating interval with block length of n can be calculated as

$$G_k(n) \triangleq \sum_{i=0}^{n-1} R_k(i), \quad (16)$$

where $R_k(i) = \log(1 + \text{SINR}_k(i))$ is the transmission rate with a CSI age of i blocks. Obviously, $R_k(i)$ is a non-increasing function with respect to i . According to the definition of p_k and $f_{k,n}$, the limitations $\lim_{N \rightarrow \infty} \frac{N_k}{N} = p_k$ and $\lim_{N \rightarrow \infty} \frac{F_{k,n}}{N_k} = f_{k,n}$ both hold. Therefore, the average transmission rate of the k -th user is

$$\mathbb{E}\{R_k\} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_n F_{k,n} G_k(n) = p_k \sum_n f_{k,n} G_k(n). \quad (17)$$

We extend $G_k(n)$ to a piecewise function:

$$G_k(x) = \begin{cases} 0 & x = 0 \\ \sum_{i=0}^{x-1} R_k(i) & x \in \mathcal{N}_+ \\ (1-x + \lfloor x \rfloor)G_k(\lfloor x \rfloor) + (x - \lfloor x \rfloor)G_k(\lfloor x \rfloor + 1) & x \notin \mathcal{N}, x \geq 0 \end{cases} \quad (18)$$

The optimal distribution of CSI updating interval is obtained according to the following theorem.

Theorem 2. *The transmission rate of the k -th user achieves the maximum value*

$$S_k(p_k) = p_k G_k\left(\frac{1}{p_k}\right), \quad (19)$$

if the distribution of CSI updating interval is

$$f_{k,n} = \begin{cases} 1 - \frac{1}{p_k} + \lfloor \frac{1}{p_k} \rfloor & n = \lfloor \frac{1}{p_k} \rfloor \\ \frac{1}{p_k} - \lfloor \frac{1}{p_k} \rfloor & n = \lfloor \frac{1}{p_k} \rfloor + 1 \\ 0. & \text{else.} \end{cases} \quad (20)$$

Proof. See Appendix B. □

The theorem requires the CSI updating interval to be quasi-periodic. More specifically, if $\frac{1}{p_k}$ is an integer, the CSI should be updated every $\frac{1}{p_k}$ blocks. Otherwise the CSI should be updated every $\lfloor \frac{1}{p_k} \rfloor$ or $\lfloor \frac{1}{p_k} \rfloor + 1$ blocks with an average updating interval of $\frac{1}{p_k}$ blocks.

C. Optimizing the CSI Updating Frequency

After optimizing $f_{k,n}$, the transmission rate of the k -th user is $S_k(p_k)$. We need to allocate the total estimation resource

of T channel uses to all users by optimizing the CSI updating frequency:

$$\begin{aligned} \max_{p_k} \quad & \sum_{k=1}^K S_k(p_k) \\ \text{s.t.} \quad & 0 \leq p_k \leq 1, \quad k = 1, \dots, K \\ & \sum_k p_k = T. \end{aligned} \quad (21)$$

The function $S_k(p_k)$ is concave but non-differentiable, making the canonic convex analysis invalid. We consider the subdifferential of the convex function $-S_k(p_k)$, which is derived as

$$\partial(-S_k)(p_k) = \begin{cases} \nabla(-S_{k,n}) & \frac{1}{n+1} < p_k < \frac{1}{n} \\ [\nabla(-S_{k,n}), \nabla(-S_{k,n-1})] & p_k = \frac{1}{n} \end{cases} \quad (22)$$

where $S_{k,n}(p_k) = [G_k(n) - nR_k(n)]p_k + R_k(n)$, $n \in \mathcal{N}_+$ is an affine function, and its slope is $\nabla(-S_{k,n}) = nR_k(n) - G_k(n)$.

The Lagrange function of the problem (21) is

$$L(\mathbf{p}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \nu) = \sum_k [-S_k(p_k) - \lambda_k p_k + \mu_k (p_k - 1)] - \nu (\sum_k p_k - T), \quad (23)$$

where $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_K]$, $\boldsymbol{\mu} = [\mu_1, \dots, \mu_K]$, ν are the Lagrange multipliers. It can be easily verified that there exists a strictly feasible point for (21), thus Slater's condition is satisfied. Therefore strong duality is hold for the prime problem and dual problem. Let \mathbf{p}^* , $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$, ν^* be any primal and dual points, then they satisfy KKT optimality conditions, which include normal Lagrange inclusion:

$$\nu^* + \lambda_k^* - \mu_k^* \in \partial(-S_k)(p_k^*). \quad (24)$$

Theorem 3. *The value of an optimal ν^* of the problem (21) lies in the candidate set*

$$\mathcal{C} = \cup_{k=1}^K \{\nabla(-S_{k,1}), \dots, \nabla(-S_{k,\Phi_k})\}, \quad (25)$$

where $\Phi_k = \max \left\{ n \mid \nabla(-S_{k,n}) \geq \min_k \{\nabla(-S_{k, \lfloor \frac{K}{T} \rfloor})\} \right\}$.

Proof. See Appendix C. \square

The subdifferential of $-S_k(p_k)$ has some kind of "monotonicity": if $p_k \leq p'_k$, then $\forall t_k \in \partial(-S_k)(p_k)$ and $\forall t'_k \in \partial(-S_k)(p'_k)$ we have $t_k \leq t'_k$. Therefore, we apply a binary search algorithm to find the optimal ν^* and \mathbf{p}^* , which is described in Algorithm 1.

D. Optimizing the Pilot Length

The optimization problem (15) is finally solved by finding the maximum sum rate provided by the solution of (21) for each value of $T \in \{1, \dots, K\}$.

When $T < K$, we apply a greedy algorithm to construct the pilot pattern and obtain the sum rate, which greedily fills the pilot pattern according to the descending order of the updating frequencies.

When $T = K$, the scheme degrades into conventional one, which estimates the channels of the whole users in every block. The sum rate can be calculated as $(1 - \frac{K}{C}) \sum_{k=1}^K R_k(0)$.

Algorithm 1 Binary Search for optimal ν^*

Input: $\mathcal{C} = \cup_{k=1}^K \{\nabla(-S_{k,1}), \dots, \nabla(-S_{k,\Phi_k})\}$, $\Phi = \sum_{k=1}^K \Phi_k$
Output: \mathbf{p}^*

- 1: Sort the elements of \mathcal{C} in descending order and get $\{\nabla(-S_{k_1, n_1}), \dots, \nabla(-S_{k_\Phi, n_\Phi})\}$.
- 2: Initialize $a = 1, b = \Phi$.
- 3: **repeat**
- 4: $t \leftarrow \lfloor \frac{a+b}{2} \rfloor$
- 5: **for** $i \in \{1, 2, \dots, K\} \setminus \{t\}$ **do**
- 6: $p(i) \leftarrow 0$
- 7: $j \leftarrow i + 1$
- 8: **while** $j \leq \Phi$ **do**
- 9: **if** $k_j = k_i$ **then**
- 10: $p(i) \leftarrow \frac{1}{n_j}$
- 11: **Break**
- 12: **end if**
- 13: $j \leftarrow j + 1$
- 14: **end while**
- 15: **end for**
- 16: $p(t) \leftarrow T - \sum_{i \neq t} p(i)$
- 17: **if** $p(t) < \frac{1}{1+n_t}$ **then**
- 18: $a \leftarrow i + 1$
- 19: **else if** $p(t) > \frac{1}{n_t}$ **then**
- 20: $b \leftarrow i - 1$
- 21: **else**
- 22: **Break**
- 23: **end if**
- 24: **until** $a > b$

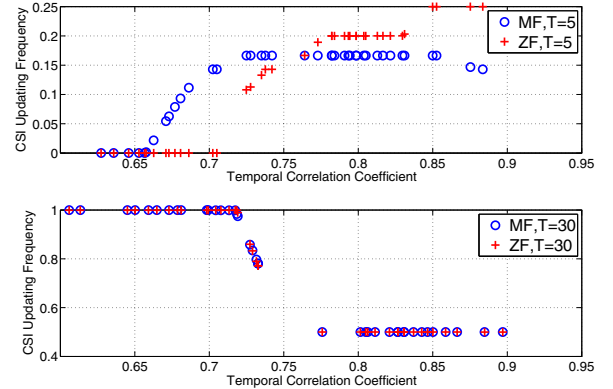


Fig. 2. CSI updating frequency vs. temporal correlation coefficient under different pilot length. $M = 64, K = 40, C = 50$, per user SNR is 10 dB.

V. NUMERICAL RESULTS

In this section we evaluate the performance of the proposed estimation scheme.

We assume the same path loss coefficient for each user. The BS splits the total power equally to each user in order to ensure fairness. So the receive SNR of each user is the same. The temporal correlation coefficient of each user is uniformly distributed in the interval $[0.6, 0.9]$. The results are obtained by averaging the performance of 50 simulation runs.

Fig. 2 shows the allocation of estimation resource to users with different temporal correlation coefficients in a single si-

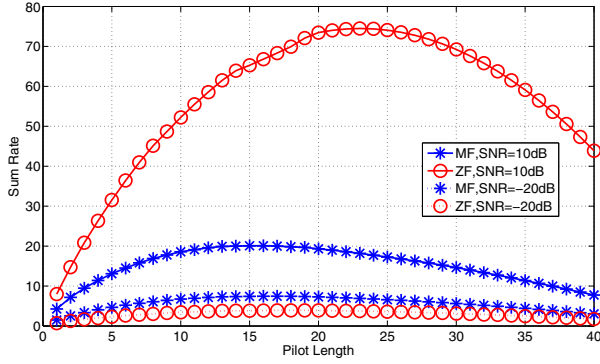


Fig. 3. Sum rate vs. pilot length. $M = 64, K = 40, C = 50$.

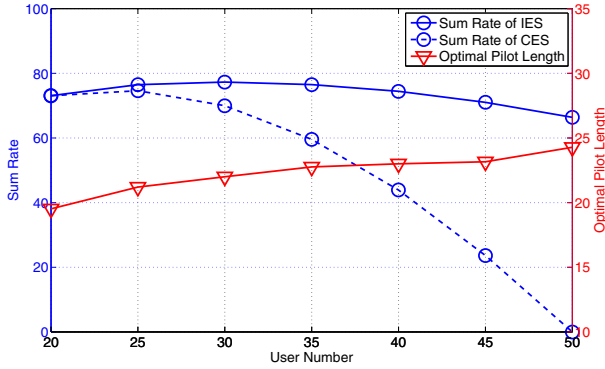


Fig. 4. The sum rate and optimal pilot length of the proposed scheme using ZF precoding. $M = 64, C = 50$, per user SNR is 10 dB.

mulation run. When the total available pilot length is extremely low ($T = 5$), the proposed scheme allocates zero estimation resource to users with temporal correlation coefficients below a certain threshold to leave the limited resource to users with better temporal correlation. The threshold of ZF precoder is higher than the threshold of the MF precoder, because it suffers more severely from the aged CSI. On the other hand, when the available pilot length is sufficiently high ($T = 30$), both MF and ZF precoder estimate allocation the maximum estimation resource to the users with poor temporal correlation by setting the CSI updating frequency to 1.

The relationship between sum rate and pilot length is depicted in Fig. 3. The sum rate of ZF outperforms that of MF in high SNR region and loses its advantage in low SNR region. The optimal pilot length of ZF is larger than that of MF under the same receive SNR. The sum rate of the conventional continuous estimation scheme is achieved at the point where pilot length is $K = 40$, and our proposed scheme improves the performance by about 74% using ZF precoder when the per user receive SNR is 10 dB.

Finally we investigate the influence of user number on the proposed scheme using ZF precoding, which is shown in Fig. 4. The results using MF precoding are omitted due to similarity. As the user number grows, the sum rate of the proposed scheme (IES) experiences the process of first

increasing and then declining, which means the system first enjoys user diversity and then becomes overwhelmed by the excessive users. The performance gap between our scheme and the continuous estimation scheme (CES) increases with the growing number of users. When user number equals the length of channel block, CES fails to transmit any data while our scheme still obtains about 86% of the peak sum rate. On the other hand, the optimal pilot length increases all the time, but its growth rate is much lower than the growth rate of user number.

VI. CONCLUSION

In this paper, we investigate the influence of the age of CSI on the channel capacity of massive MIMO systems and derive closed-form expressions for transmission rate under two linear precoders, namely matched filter and zero forcing. The intermittent estimation scheme is proposed to reduce the overhead of CSI acquisition. We show the CSI updating intervals should follow a quasi-periodic distribution and obtain the optimal CSI updating frequency for each user to maximize sum capacity in the limited block length. The numerical results shows a great performance gain of the proposed scheme compared to the conventional continuous estimation scheme.

APPENDIX A

PROOF OF THEOREM 1

According to the Gaussian-Markov process, we know the correlation between the estimated channel and the channel of ground truth reads

$$\mathbf{h}_k = \rho_k^n \hat{\mathbf{h}}_k + \sqrt{1 - \rho_k^{2n}} \mathbf{e}_k. \quad (26)$$

When the MF precoding scheme is applied, the random variable $\hat{\mathbf{h}}_k^H \mathbf{v}_k$ satisfies the chi square distribution with $2M$ degrees of freedom by a scale factor of $\frac{1}{2\eta_k^{MF}}$, so we can obtain the expectation of effective channel, the variation of effective channel and the interference as

$$\epsilon_k \beta_k |\mathbb{E} \mathbf{h}_k^T \mathbf{v}_k|^2 = \epsilon_k \beta_k M \rho_k^{2n}, \quad (27)$$

$$\epsilon_k \beta_k (\mathbb{E} |\mathbf{h}_k^T \mathbf{v}_k|^2 - |\mathbb{E} \mathbf{h}_k^T \mathbf{v}_k|^2) = \epsilon_k \beta_k, \quad (28)$$

$$\sum_{i \neq k} \epsilon_i \beta_i \mathbb{E} |\mathbf{h}_k^T \mathbf{v}_i|^2 = \sum_{i \neq k} \epsilon_i \beta_i, \quad (29)$$

respectively. We get the expression for SINR under MF precoding scheme by substituting (27), (28) and (29) into (5).

Similarly, we calculate the expectation of effective channel, the variation of effective channel and the interference under ZF precoding scheme as

$$\epsilon_k \beta_k |\mathbb{E} \mathbf{h}_k^T \mathbf{v}_k|^2 = \epsilon_k \beta_k (M - K) \rho_k^{2n}, \quad (30)$$

$$\epsilon_k \beta_k (\mathbb{E} |\mathbf{h}_k^T \mathbf{v}_k|^2 - |\mathbb{E} \mathbf{h}_k^T \mathbf{v}_k|^2) = \epsilon_k \beta_k (1 - \rho_k^{2n}), \quad (31)$$

$$\sum_{i \neq k} \epsilon_i \beta_i \mathbb{E} |\mathbf{h}_k^T \mathbf{v}_i|^2 = (1 - \rho_k^{2n}) \sum_{i \neq k} \epsilon_i \beta_i, \quad (32)$$

respectively. The SINR is derived by substituting (30), (31) and (32) into (5).

APPENDIX B
PROOF OF THEOREM 2

Firstly, we show that the function $G_k(x)$ is concave in its domain of definition $\mathcal{D} = \{x \in \mathcal{R} | x \geq 0\}$.

Define $G_{k,n}(x) = R_k(n)(x - n) + G_k(n)$, $n \in \mathcal{N}$, which are a series of linear functions with respect to x . For $x \geq n(n \geq 1)$, we have

$$G_{k,n-1}(x) \geq G_{k,n}(x), \quad (33)$$

due to the slope of the left linear function is no less than that of the right one, i.e., $R_k(n-1) \geq R_k(n)$ and the two function have an intersection at the point $(n, G_k(n))$.

By induction, we get

$$G_{k,n'}(x) \geq G_{k,n}(x), \quad \text{for } n' < n, x \geq n. \quad (34)$$

Similarly, we also get

$$G_{k,n'}(x) \geq G_{k,n}(x), \quad \text{for } n' > n, x \leq n+1. \quad (35)$$

Therefore, $\min_{n' \in \mathcal{N}} \{G_{k,n'}(x)\}$ actually equals $G_{k,n}(x)$ in the interval $[n, n+1]$, which also equals the value of function $G_k(x)$ in the interval $[n, n+1]$. As a result, $G_k(x)$ can be restated as

$$G_k(x) = \min_{n' \in \mathcal{N}} \{G_{k,n'}(x)\}, \quad (36)$$

which is the minimum of a series of affine functions and thus concave.

Then, by *Jensen's inequality* [15], the transmission rate of the k -th user satisfies the inequality

$$\mathbb{E}\{R_k\} = p_k \sum_n f_{k,n} G_k(n) \leq p_k G_k \left(\sum_n f_{k,n} \right) = p_k G \left(\frac{1}{p_k} \right), \quad (37)$$

where the equality holds when $\{f_{k,n}\}$ satisfies (20).

APPENDIX C
PROOF OF THEOREM 3

Firstly, we show that the value of an optimal ν^* equals $\nabla(-S_{k,n}) = nR_k(n) - G_k(n)$, where $k \in \{1, \dots, K\}$, $n \in \mathcal{N}_+$.

Suppose an optimal ν^* does not satisfy the statement. Divide the user set $\mathcal{K} = \{1, 2, \dots, K\}$ into three subsets: $\mathcal{K}_0 = \{k | p_k = 0\}$, $\mathcal{K}_1 = \{k | p_k = 1\}$ and $\mathcal{K}_2 = \{k | 0 < p_k < 1\}$. From KKT conditions, we observe the following fact:

- 1) For $k_0 \in \mathcal{K}_0$, $\nu^* < \nabla(-S_{k_0,n}) \quad \forall n \in \mathcal{N}_+$.
- 2) For $k_1 \in \mathcal{K}_1$, $\nu^* > \nabla(-S_{k_1,1})$.
- 3) For $k_2 \in \mathcal{K}_2$, there exists $n_{k_2} > 1$ so that $\nabla(-S_{k_2,n_{k_2}}) < \nu^* < \nabla(-S_{k_2,n_{k_2}-1})$.

Let $\hat{\nu} = \max\{\nabla(-S_{k_1,1}), \nabla(-S_{k_2,n_{k_2}}) | k_1 \in \mathcal{K}_1, k_2 \in \mathcal{K}_2\}$ and replace ν^* with $\hat{\nu}$, the KKT conditions still hold. Therefore we get a new optimal $\hat{\nu}$ which satisfies the statement.

Then we derive an upper bound of Φ_k of the candidate set for each k . Assume the k_M -th user is assigned with the maximum updating frequency. Then $p_{k_M} \geq \frac{T}{K}$ and the subgradient of $-S_{k_M}(p_{k_M})$ is no less than

$\nabla(-S_{k_M, \lfloor \frac{K}{T} \rfloor}(p_{k_M}))$. As a result, $\nu^* \geq \nabla(-S_{k_M, \lfloor \frac{K}{T} \rfloor}(p_{k_M}))$ and hence $\nu^* \geq \min_k \{\nabla(-S_{k, \lfloor \frac{K}{T} \rfloor}(p_k))\}$. Therefore $\Phi_k = \max\{n | \nabla(-S_{k,n}(p_k)) \geq \min_k \{\nabla(-S_{k, \lfloor \frac{K}{T} \rfloor}(p_k))\}\}$. As a result, the value of ν^* lies in the set \mathcal{C} as described in the theorem.

ACKNOWLEDGMENT

This work is sponsored in part by the Nature Science Foundation of China (No.61701275, No. 91638204, No. 61571265, No. 61621091), and Hitachi R&D Headquarter.

REFERENCES

- [1] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.
- [2] E. Bjrnson, E. G. Larsson, and T. L. Marzetta, "Massive MIMO: ten myths and one critical question," *IEEE Communications Magazine*, vol. 54, no. 2, pp. 114–123, February 2016.
- [3] S. Y. Lien, K. C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 66–74, April 2011.
- [4] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, February 2013.
- [5] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive mimo transmission over spatially correlated rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3352–3366, June 2015.
- [6] S. L. H. Nguyen and A. Ghayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems," in *IEEE Wireless Communications and Networking Conference (WCNC)*, April 2013, pp. 2890–2895.
- [7] Z. Jiang, A. F. Molisch, G. Caire, and Z. Niu, "Achievable rates of FDD massive MIMO systems with spatial channel correlation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2868–2882, May 2015.
- [8] R. Deng, S. Zhou, and Z. Niu, "Scalable non-orthogonal pilot design for massive MIMO systems with massive connectivity," in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–6.
- [9] A. Abdi and M. Kaveh, "A space-time correlation model for multi-element antenna systems in mobile fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 3, pp. 550–560, Apr 2002.
- [10] M. Costa, S. Valentin, and A. Ephremides, "On the age of channel state information for non-reciprocal wireless links," in *2015 IEEE International Symposium on Information Theory (ISIT)*, June 2015, pp. 2356–2360.
- [11] M. Dong, L. Tong, and B. M. Sadler, "Optimal insertion of pilot symbols for transmissions over time-varying flat fading channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 5, pp. 1403–1418, May 2004.
- [12] J. Choi, D. J. Love, and P. Bidigare, "Downlink training techniques for FDD massive MIMO systems: Open-loop and closed-loop training with memory," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 802–814, Oct 2014.
- [13] S. Noh, M. D. Zoltowski, Y. Sung, and D. J. Love, "Pilot beam pattern design for channel estimation in massive MIMO systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 787–801, Oct 2014.
- [14] J. Jose, A. Ashikhmin, T. L. Marzetta, and S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems," *IEEE Transactions on Wireless Communications*, vol. 10, no. 8, pp. 2640–2651, August 2011.
- [15] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.