Statistical CSI Acquisition in the Nonstationary Massive MIMO Environment

Guoliang Wang, Wei Peng, *Senior Member, IEEE*, Dong Li, *Member, IEEE*, Tao Jiang, *Senior Member, IEEE*, and Fumiyuki Adachi, *Life Fellow, IEEE*

***Abstract*—This paper studies the statistical channel state information (S-CSI) acquisition problem in the nonstationary massive multiple-input multiple-output (MIMO) environment, where both the instantaneous and statistical channel states are time varying. First, we set up a hidden statistical channel state Markov model (HSCSM model). Then, the parameter of the HSCSM model is estimated through the observed sequence of received signals. Next, based on the HSCSM model and its estimated parameter, the S-CSI is obtained through a maximum *a-posteriori* decision process. Simulation results show that an accurate S-CSI acquisition can be achieved by the proposed approach in the nonstationary massive MIMO environment. In addition, the estimation accuracy rate of the proposed approach increases with the length of observation sequence as well as the number of antennas, where a tradeoff between them exists given a limited computing ability/storage space.**

***Index Terms*—Massive MIMO, non-stationary, statistical CSI acquisition, HSCSM-model.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO) employs hundreds or thousands of antennas at the base station (BS) to simultaneously serve a large number of users [1], [2]. Benefitting from the great spatial multiplexing/diversity gain, massive MIMO has many advantages. Firstly, the multiuser interference (MUI) almost disappears as the number of BS antennas increases to infinity, enabling power reduction and flexible network planning [3]. Secondly, considerable high spectrum and energy efficiency could be obtained by simple linear precoding in the downlink transmission or coherent combining in the uplink transmission. Thirdly, by exploiting the channel reciprocity [4] between uplink and downlink, the network capacity can be significantly improved by increasing the number of antennas without extra feedbacks.

It should be noted that, the reported benefits of massive MIMO are based on the assumption of accurate channel state information (CSI), which has to be obtained via effective CSI acquisition methods. E.g., precoding and coherent detection requires the instantaneous CSI (I-CSI) and user scheduling demands the statistical CSI (S-CSI) [5]–[8]. Generally speaking, the I-CSI acquisition is vulnerable to the additive random noise and is more difficult to be obtained than the S-CSI. Therefore, S-CSI based signal processing techniques are more preferable in practical massive MIMO systems.

To deal with the S-CSI acquisition problem, several approaches have been proposed in the massive MIMO environment and can be classified into two categories. The first category includes the I-CSI based algorithms [9], [10], by which the I-CSI is estimated through pilot transmission [11], and the S-CSI is then obtained via statistical calculation over the estimated I-CSI. The second category includes the expectation maximization (EM) based algorithms [12], [13], by which the S-CSI is "learned" in an iterative way. Namely, the expectation of the probability to receive the observed signals is calculated given a certain statistical channel state, then the statistical channel state can be obtained by maximizing the expectation. Comparatively speaking, the EM-based S-CSI acquisition is more spectrum efficient since it has lower requirements on the pilot sequence.

The existing S-CSI acquisition methods are grounded on the assumption that channels are wide-sense stationary [14], i.e., the statistical channel state should remain unchanged during the process of S-CSI acquisition. However, measurements on massive MIMO channels indicate the non-stationarity of channels [15], [16], which has been considered by recent researches on massive MIMO channel modelings [17], [18]. Actually, non-stationarity is a common phenomenon in real environments such as the high-speed railway and vehicle-to-vehicle communications [19], [20]. In non-stationary environments, both the instantaneous and statistical channel states are time-varying. As a result, the S-CSI can no longer be obtained via statistical averaging over the estimated I-CSI. At the same time, the EM-based S-CSI is not applicable as well. To the best of the authors' knowledge, research works concerning the S-CSI acquisition for non-stationary massive MIMO channels have not yet been reported.

Manuscript received December 23, 2017; revised March 20, 2018; accepted March 21, 2018. Date of publication April 20, 2018; date of current version August 13, 2018. This work was supported in part by National Science Foundation of China under Grants 61771214, 61771216, and 61501193, in part by the National Science Foundation for Distinguished Young Scholars of China under Grant 61325004, in part by the Innovative Project of Shenzhen city in China under Grant JCYJ20170307171931096, in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2018D10, and in part by the Macau Science and Technology Development Fund under Grant 052/2016/A. The review of this paper was coordinated by R. Dinis. *(Corresponding author: Wei Peng.)*

G. Wang, W. Peng, and T. Jiang are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: guoliangwang@hust.edu.cn; pengwei@hust.edu.cn; Tao.Jiang@ieee.org).

D. Li is with the School of Information Technology, Macau University of Science and Technology, Macau 999097, China (e-mail: dli@must.edu.mo).

F. Adachi is with the Research Organization of Electrical Communication, Tohoku University, Sendai 980-8579, Japan (e-mail: adachi@ecei.tohoku.ac.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2018.2828866

In this paper, we consider the non-stationary massive MIMO environment and propose an approach to solve the S-CSI acquisition problem. The contamination-free single-cell scenario is assumed for simplicity and an extreme case is considered, where the statistical channel state could vary in every time slot. Firstly, making use of the channel characteristics revealed by field experiments in [15], [16], we propose a hidden statistical channel state Markov model (HSCSM-model) to represent the probabilistic dependence of the observed receiving sequence on the statistical channel states. Then, the parameter of the HSCSM-model is estimated through the observed sequence of received signals. Next, given the HSCSM-model and its estimated parameter, the S-CSI is obtained through a maximum a-posteriori decision process.

The performance of the proposed S-CSI acquisition approach is verified by simulation results. It is indicated that accurate S-CSI can be obtained in the non-stationary massive MIMO environment, while the traditional EM-based S-CSI acquisition approach can not work at all. In addition, the estimation accuracy rate of the proposed approach increases with the length of the observation sequence as well as the number of antennas. Furthermore, the computational- complexity/storage-space of the proposed approach is evaluated. It is shown that, given a limited computing-ability/storage-capacity, there exists a trade-off between the affordable observation length and the number of antennas in terms of the estimation accuracy rate. In order to achieve the optimal performance, the limited computing/storage resource should be allocated to accommodate more antennas when the signal to noise power ratio (SNR) is low, while a longer observation should be allowed when the SNR is high.

The remainder of the paper is organized as follows. We present the massive MIMO system model and the non-stationary channel in Section II. Then, the HSCSM-model is proposed in Section III. In Section IV, we describe the HSCSM-model parameter estimation and the maximum a-posteriori decision process for the proposed S-CSI acquisition. Simulation results are shown in Section V. Finally, concluding remarks are given in Section VI and some detailed derivations are relegated to the Appendix.

Notations: Throughout this paper, the set of complex numbers is denoted by $\mathbb{C}$. The superscripts $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^{-1}$ represent the transpose, conjugate transpose and inverse operations, respectively. The notation v_{t-L+1}^{t} is used to represent a sequence $\{v(t-L+1), v(t-L+2), ..., v(t)\}$. $P(\cdot)$ and $p(\cdot)$ stand for the probability and the probability density function (PDF), respectively. In addition, a random vector $\mathbf{s}$ following the proper complex Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathbf{s} \sim N_{\mathbb{C}}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $N_{\mathbb{C}}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\det(\pi\boldsymbol{\Sigma})} e^{-(\mathbf{s}-\boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{s}-\boldsymbol{\mu})}$ [21] and $\det(\cdot)$ represents the determinant. Hereafter, $N_{\mathbb{C}}(\mathbf{s}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is simplified as $N_{\mathbb{C}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for conciseness.

II. System Model

Considering a time division duplexing (TDD) massive MIMO system in a single-cell scenario, there are N single-antenna users and the base station (BS) is equipped with M antennas. Taking the general assumption of channel reciprocity between uplink and downlink, CSI acquisition is performed in the uplink, where transmissions happen between N users and the BS simultaneously. The baseband received signal vector at the BS at time t is given as

$$\mathbf{y}(t) = \mathbf{H}(t)\mathbf{x}(t) + \mathbf{z}(t), \tag{1}$$

where the vector $\mathbf{x}(t) = [x_1(t), x_2(t), ..., x_N(t)]^T \in \mathbb{C}^{N\times 1}$ denotes the transmitted signals of N users, and the uplink channel matrix is $\mathbf{H}(t) = [\mathbf{h}_1(t), \mathbf{h}_2(t), ..., \mathbf{h}_N(t)] \in \mathbb{C}^{M\times N}$, where $\mathbf{h}_n(t) \in \mathbb{C}^{M\times 1}$ is the channel vector from the nth user to the BS, $\mathbf{z}(t) = [z_1(t), z_2(t), ..., z_M(t)]^T \in \mathbb{C}^{M\times 1}$ is the vector of complex additive white Gaussian noise with zero mean and element-wise variance σ_z^2.

The non-stationary property of massive MIMO channels is caused by the physical environment with limited but changing reflecting clusters [22], [23]. Assume that the elements of $\mathbf{H}(t)$ are independent and identically distributed random variables, which follow the Rayleigh distribution with zero mean and time-varying variance. To represent this non-stationarity, the finite set of statistical channel states [19] is denoted by $\mathbb{S} = \{S_1, ..., S_K\}$, where $S_i = (\mu_i, \sigma_i^2), i = 1, \ldots, K$ is the ith state with μ_i and σ_i^2 respectively being the mean value and variance, and K is the number of states in $\mathbb{S}$. It is assumed that the set of statistical states could be obtained through pilot-assisted training as the priori knowledge.

For a better understanding, the variables defined for the S-CSI acquisition are summarized in Table I.

III. Hidden Statistical Channel State Markov Model

Inspired by the continuous density hidden Markov model [24], [25], in this section, we propose a novel HSCSM-model. By the proposed model, for the first time, the probabilistic dependence of the observed receiving sequence on the statistical channel states is set up in massive MIMO systems. Denoting a bivariate stochastic process by $\{(q(t), \mathbf{y}(t))\}$, where $\{q(t)\}$ is the hidden process, and $\{\mathbf{y}(t)\}$ is the observable process. The hidden process is a discrete, finite-state, homogeneous Markov chain, and the observable process is conditioned on the hidden process. In the proposed HSCSM-model, the hidden process $\{q(t)\}$ represents the variation of statistical channel state over time, while the observable process is conditioned on $\{q(t)\}$ through the massive MIMO channel. Namely, at each time slot, a new statistical channel state is entered with a state transition probability given the previous statistical channel state, then a new receiving signal vector is observed, which depends on the current statistical channel state, as shown in Fig. 1.

The parameter of the proposed HSCSM-model consists of three components: initial state probability vector $\boldsymbol{\eta}$, state transition probability matrix $\mathbf{G}$ and probability density matrix $\mathbf{B}$, denoted by $\lambda = (\boldsymbol{\eta}, \mathbf{G}, \mathbf{B})$. Assume that the length of the observation sequence is L (including the current observation and previous $L-1$ observations) and let $q_{t-L+1}^{t} = \{q(t-L+1), q(t-L+2), ..., q(t)\}$ denote the hidden sequence of statistical channel states from time $t-L+1$ to time t. The sequence of observations is denoted by $\mathbf{y}_{t-L+1}^{t} =$

TABLE I
VARIABLES FOR THE S-CSI ACQUISITION

Variables	**Descriptions**
$\mathbf{H}(t) \in \mathbb{C}^{M\times N}$	Channel matrix
$\mathbf{x}(t) \in \mathbb{C}^{N\times 1}$	Transmitted signal vector
$\mathbf{Z}(t) \in \mathbb{C}^{M\times 1}$	Noise vector
$\mathbf{y}(t) \in \mathbb{C}^{M\times 1}$	Received signal vector
$\mathbb{S} = \{S_1, ..., S_K\}$	The set of statistical channel states, where S_i is the i^{th} state
$\{q(t)\}$	Hidden process of statistical channel state
$\mathbf{y}_{t-L+1}^{t}$	Sequence of the received signal vector from time $t-L+1$ to time t
q_{t-L+1}^{t}	Sequence of the statistical channel states from time $t-L+1$ to time t
$p(\mathbf{y}(t-l)\|q(t-l)=S_i)$	Probability density of the observation vector $\mathbf{y}(t-l)$ given S_i
$\eta_{S_i}(t)$	Probability of the event that the initial state is S_i
$g_{ij}(t)$	Transition probability from statistical channel state S_i to S_j
$\boldsymbol{\mu}_{S_i}(t)$	Vector of mean values given S_i
$\boldsymbol{\Sigma}_{S_i}(t)$	Covariance matrix given S_i
$\lambda(t) = (\boldsymbol{\eta}(t), \mathbf{G}(t), \boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t))$	HSCSM-model parameter
$\hat{\lambda}(t) = \left(\hat{\boldsymbol{\eta}}(t), \hat{\mathbf{G}}(t), \hat{\boldsymbol{\mu}}(t), \hat{\boldsymbol{\Sigma}}(t)\right)$	Estimated HSCSM-model parameter
$P(q_{t-L+1}^{t}\|\lambda(t))$	Conditional probability of hidden sequence q_{t-L+1}^{t} given $\lambda(t)$
$P(\mathbf{y}_{t-L+1}^{t}\|\lambda(t))$	Conditional probability density of $\mathbf{y}_{t-L+1}^{t}$ given $\lambda(t)$
$P(q(t-l)=S_i\|\mathbf{y}_{t-L+1}^{t}, \lambda(t))$	Probability of being in state S_i at time $t-l$
$P(q(t-l)=S_i, q(t-l+1)=S_j\|\mathbf{y}_{t-L+1}^{t}, \lambda(t))$	Probability of being in state S_i at time $t-l$ and state S_j at time $t-l+1$
$\bar{\alpha}_{t-l}(S_i)$	Scaled forward variable
$\bar{\beta}_{t-l}(S_i)$	Scaled backward variable
$p(q(t)\|\mathbf{y}_{t-L+1}^{t}, \hat{\lambda}(t))$	Probability of hidden state given $(\mathbf{y}_{t-L+1}^{t}, \hat{\lambda}(t))$

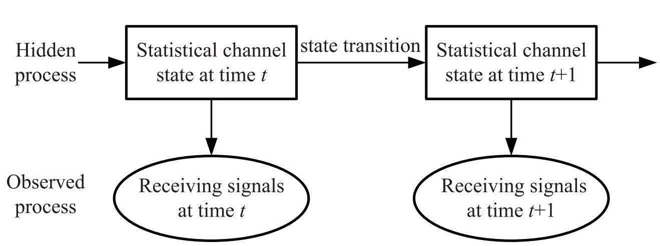

Fig. 1. The proposed HSCSM-model.

$\{\mathbf{y}(t-L+1), \mathbf{y}(t-L+2), ..., \mathbf{y}(t)\}$. In the following, the calculation of $\boldsymbol{\eta}(t)$, $\mathbf{G}(t)$, and $\mathbf{B}(t)$ will be presented, respectively.

Calculation of $\boldsymbol{\eta}(t)$*:* $\boldsymbol{\eta}(t)$ is the initial state probability vector, $\boldsymbol{\eta}(t) = [\eta_{S_1}(t), \eta_{S_2}(t), ..., \eta_{S_K}(t)]^T \in \mathbb{R}^{K\times 1}$, its ith element represents the probability that the initial state is S_i at time

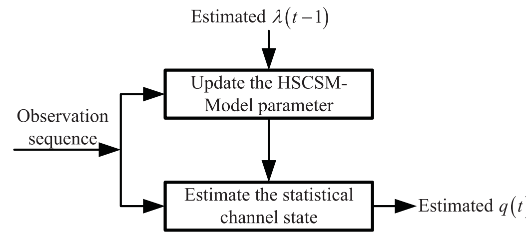

Fig. 2. Statistical channel state estimation at time t.

$t-L+1$, denoted by

$$\eta_{S_i}(t) = P(q(t-L+1) = S_i). \tag{2}$$

Note that, $\eta_{S_i}(t) \geq 0$ and $\sum_{i=1}^{K} \eta_{S_i}(t) = 1$.

Calculation of $\mathbf{G}(t)$*:* $\mathbf{G}(t) \in \mathbb{R}^{K\times K}$ is the matrix of state transition probability, its (i,j)th element represents the state transition probability from state S_i at time $t-l$ to state S_j at time $t-l+1$, given as

$$g_{ij}(t) = P(q(t-l+1) = S_j \,|\, q(t-l) = S_i). \tag{3}$$

It is clear that $\sum_{j=1}^{K} g_{ij}(t) = 1$.

Calculation of $\mathbf{B}(t)$*:* $\mathbf{B}(t) \in \mathbb{R}^{K\times L}$ is the probability density matrix. Its (i,l)th element represents the probability density of the observation vector $\mathbf{y}(t-l)$ in the presence of statistical channel state S_i, given as

$$b_{il}(t) = p(\mathbf{y}(t-l) \,|\, q(t-l) = S_i), \tag{4}$$

where $p(\mathbf{y}(t-l) \,|\, q(t-l) = S_i)$ is a multi-variate complex Gaussian density with the mean vector $\boldsymbol{\mu}_{S_i}(t)$ and covariance matrix $\boldsymbol{\Sigma}_{S_i}(t)$. According to (1), $p(\mathbf{y}(t-l) \,|\, q(t-l) = S_i)$ is given as

$$\begin{aligned} &p(\mathbf{y}(t-l) \,|\, q(t-l) = S_i) \\ &= \frac{1}{\det(\pi \boldsymbol{\Sigma}_{S_i}(t))} e^{-\left(\mathbf{y}(t-l)-\boldsymbol{\mu}_{S_i}(t)\right)^H \left(\boldsymbol{\Sigma}_{S_i}(t)\right)^{-1} \left(\mathbf{y}(t-l)-\boldsymbol{\mu}_{S_i}(t)\right)}. \end{aligned} \tag{5}$$

Note that, to get the derivation in (5), a unit transmitting power is assumed for each user for simplicity. However, an extension to the case with un-equal power allocation is straight forward. Let $\boldsymbol{\mu}(t) = \{\boldsymbol{\mu}_{S_1}(t), \boldsymbol{\mu}_{S_2}(t), ..., \boldsymbol{\mu}_{S_K}(t)\}$ and $\boldsymbol{\Sigma}(t) = \{\boldsymbol{\Sigma}_{S_1}(t), \boldsymbol{\Sigma}_{S_2}(t), ..., \boldsymbol{\Sigma}_{S_K}(t)\}$. The elements of $\mathbf{B}(t)$ can be calculated by the elements of $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$. Therefore, the parameter of the HSCSM-model at time t is represented by $\lambda(t) = (\boldsymbol{\eta}(t), \mathbf{G}(t), \boldsymbol{\mu}(t), \boldsymbol{\Sigma}(t))$ hereafter.

IV. STATISTICAL CHANNEL STATE INFORMATION ACQUISITION

In this section, we propose a new S-CSI acquisition approach based on the proposed HSCSM-model. Specifically, at each time slot, the newly observed receiving signal vector will be used to update the HSCSM-model parameter, then, the statistical channel state is estimated through the maximum a-posteriori decision process, as shown in Fig. 2.

A. HSCSM-Model Parameter Estimation

The probability density of $\mathbf{y}_{t-L+1}^{t}$ conditioned on $\lambda(t)$ is given as

$$\begin{aligned} &p\left(\mathbf{y}_{t-L+1}^{t}|\lambda(t)\right) \\ &= \sum_{q_{t-L+1}^{t}} p\left(\mathbf{y}_{t-L+1}^{t}|q_{t-L+1}^{t},\lambda(t)\right) P\left(q_{t-L+1}^{t}|\lambda(t)\right), \end{aligned} \tag{6}$$

where the conditional probability of the hidden sequence q_{t-L+1}^{t} is given as

$$P\left(q_{t-L+1}^{t}\,|\lambda(t)\right) = \prod_{l=0}^{L-1} g_{q(t-l-1)q(t-l)}(t), \tag{7}$$

with $g_{q(t-L)q(t-L+1)}(t) = \eta_{q(t-L+1)}(t)$. Elements of $\mathbf{y}_{t-L+1}^{t}$ are conditionally independent, and their distributions depend on the corresponding statistical channel states q_{t-L+1}^{t}. For any non-negative integer L, we have

$$p\left(\mathbf{y}_{t-L+1}^{t}\,\middle|q_{t-L+1}^{t},\lambda(t)\right) = \prod_{l=0}^{L-1} p\left(\mathbf{y}(t-l)\,|q(t-l)\right). \tag{8}$$

Substituting (7) and (8) into (6), the conditional probability density in (6) is given as

$$\begin{aligned} &p\left(\mathbf{y}_{t-L+1}^{t}|\lambda(t)\right) \\ &= \sum_{q_{t-L+1}^{t}} \prod_{l=0}^{L-1} g_{q(t-l-1)q(t-l)}(t) p\left(\mathbf{y}(t-l)|q(t-l)\right). \end{aligned} \tag{9}$$

Note that, a more efficient calculation of (9) can be found in the Appendix. The HSCSM-model parameter estimation is then obtained by maximizing the likelihood given an observation sequence $\mathbf{y}_{t-L+1}^{t}$, given as

$$\hat{\lambda}(t) = \arg\max_{\lambda(t)} p\left(\mathbf{y}_{t-L+1}^{t}\,|\lambda(t)\right), \tag{10}$$

where $\hat{}$ denotes the optimal parameter that maximizes the likelihood. The maximization problem in (10) is solved by an extended Baum-Welch method [26], which iteratively generates the HSCSM-model parameter with nondecreasing likelihood [27]. Each iteration starts with the current parameter $\lambda^{n}(t)$ and estimates a new parameter $\lambda^{n+1}(t)$ by maximizing the auxiliary function, given as

$$\lambda^{n+1}(t) = \arg\max_{\lambda(t)} Q\left(\lambda^{n}(t), \lambda(t)\right), \tag{11}$$

where n is the index of iteration and

$$\begin{aligned} &Q\left(\lambda^{n}(t), \lambda(t)\right) \\ &= \sum_{q_{t-L+1}^{t}} p\left(q_{t-L+1}^{t}\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda^{n}(t)\right) \log p\left(q_{t-L+1}^{t},\mathbf{y}_{t-L+1}^{t}\,|\lambda(t)\right). \end{aligned} \tag{12}$$

The iteration is terminated when a convergence criterion is satisfied, e.g.,

$$p\left(\mathbf{y}_{t-L+1}^{t}\,|\lambda^{n+1}(t)\right) - p\left(\mathbf{y}_{t-L+1}^{t}\,|\lambda^{n}(t)\right) < \varepsilon, \tag{13}$$

where ε is a predefined threshold.

Let $P\left(q(t-l)=S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)$ be the probability of being in state S_i at time $t-l$ given $\lambda(t)$. Defining $P\left(q(t-l)=S_i, q(t-l+1)=S_j\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)$ as the probability of being in state S_i at time $t-l$ and state S_j at time $t-l+1$ given $\left(\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)$. The calculation of $P\left(q(t-l)=S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)$ and $P\left(q(t-l)=S_i, q(t-l+1)=S_j\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)$ relies on the scaled forward and backward variables, which can be found in the Appendix.

The HSCSM-model parameter obtained during the iterative maximization of (12) is given as

$$\widetilde{\eta}_{S_i}(t) = P\left(q(t-L+1) = S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right), \tag{14a}$$

$$\widetilde{g}_{ij}(t) = \frac{\sum_{l=1}^{L-1} P\left(q(t-l)=S_i, q(t-l+1)=S_j\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)}{\sum_{l=1}^{L-1} P\left(q(t-l)=S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)}, \tag{14b}$$

$$\widetilde{\boldsymbol{\mu}}_{S_i}(t) = \frac{\sum_{l=0}^{L-1} P\left(q(t-l) = S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)\mathbf{y}(t-l)}{\sum_{l=0}^{L-1} P\left(q(t-l) = S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)}, \tag{14c}$$

and

$$\widetilde{\boldsymbol{\Sigma}}_{S_i}(t) = \frac{\sum_{l=0}^{L-1} P\left(q(t-l)=S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)\mathbf{W}(t-l)}{\sum_{l=0}^{L-1} P\left(q(t-l) = S_i\,\middle|\mathbf{y}_{t-L+1}^{t},\lambda(t)\right)}, \tag{14d}$$

where $\widetilde{}$ denotes the temporary value during the iteration, $\mathbf{W}(t-l) = \mathbf{I}_M \odot \left((\mathbf{y}(t-l)-\widetilde{\boldsymbol{\mu}}_{S_i})(\mathbf{y}(t-l)-\widetilde{\boldsymbol{\mu}}_{S_i})^{H}\right)$ with $\mathbf{I}_M$ denoting the $M\times M$ identity matrix and $\odot$ representing the element-wise matrix multiplication.

B. S-CSI Acquisition

The statistical channel state at time t is obtained by maximizing the likelihood of $q(t)$ conditioned on the observation sequence $\mathbf{y}_{t-L+1}^{t}$ and the estimated HSCSM-model parameter $\hat{\lambda}(t)$, given as

$$\hat{q}(t) = \arg\max_{q(t)\in\mathbb{S}} \left\{p\left(q(t)\,\middle|\mathbf{y}_{t-L+1}^{t},\hat{\lambda}(t)\right)\right\}. \tag{15}$$

To solve the maximization problem in (15), we use the maximum a-posteriori decision process, given as

$$\begin{aligned} &p\left(q(t-l)\,\middle|\mathbf{y}(t-l),\hat{\lambda}(t)\right) \\ &= \frac{\hat{\eta}_{q(t-l)}(t)\,p\left(\mathbf{y}(t-l)\,|q(t-l)\right)}{\sum_{q(t-l)\in\mathbb{S}} \hat{\eta}_{q(t-l)}(t)\,p\left(\mathbf{y}(t-l)\,|q(t-l)\right)} \end{aligned} \tag{16}$$

when $l = L-1$, and

$$\begin{aligned} &p\left(q(t-l)\,\middle|\mathbf{y}_{t-L+1}^{t-l},\hat{\lambda}(t)\right) \\ &= \frac{p\left(q(t-l)\,\middle|\mathbf{y}_{t-L+1}^{t-l-1},\hat{\lambda}(t)\right)p\left(\mathbf{y}(t-l)\,|q(t-l)\right)}{\sum_{q(t-l)\in\mathbb{S}} p\left(q(t-l)\,\middle|\mathbf{y}_{t-L+1}^{t-l-1},\hat{\lambda}(t)\right)p\left(\mathbf{y}(t-l)\,|q(t-l)\right)} \end{aligned} \tag{17}$$

Algorithm 1: Statistical Channel State Estimator.

Input: $\mathbf{y}_{t-L+1}^{t}=\{y(t-L+1),\ldots,y(t)\}$ and $\mathbb{S}=\{S_1,\ldots,S_K\}$
Output: Estimated statistical channel state $\hat{q}\,(t)$
$n=0$;
Initialize $\lambda^{n}\,(t)=(\boldsymbol{\eta}^{n}\,(t),\mathbf{G}^{n}\,(t),\boldsymbol{\mu}^{n}\,(t),\boldsymbol{\Sigma}^{n}\,(t))$;
Obtain $\mathrm{B}^{n}\,(t)$ following $p(\mathbf{y}(t-l)|q(t-l)=S_i)\sim N_{\mathbb{C}}\left(\boldsymbol{\mu}_{S_i}^{n}\,(t)\,,\boldsymbol{\Sigma}_{S_i}^{n}\,(t)\right)$;
repeat
  Calculate $\bar{\alpha}_{t-l}(S_i)$ and $\bar{\beta}_{t-l}(S_i)$ following (27) and (30);
  Calculate $p\left(\mathbf{y}_{t-L+1}^{t}|\lambda^{n}\,(t)\right)$ following (34);
  Update the HSCSM-model parameter:
  Calculate $P\left(q(t-l)=S_i\,|\mathbf{y}_{t-L+1}^{t},\lambda^{n}\,(t)\right)$ and $P\left(q(t-l)=S_i,q(t-l+1=S_j)=S_i\,|\mathbf{y}_{t-L+1}^{t},\lambda^{n}\,(t)\right)$ following (31) and (32);
  Calculate $\lambda^{n+1}(t)$ following (14);
  Obtain $\mathrm{B}^{n+1}(t)$ following $p(\mathbf{y}(t-l)|q(t-l)=S_i)\sim N_{\mathbb{C}}\left(\boldsymbol{\mu}_{S_i}^{n+1}\,(t)\,,\boldsymbol{\Sigma}_{S_i}^{n+1}\,(t)\right)$;
  $n=n+1$;
until $p\left(\mathbf{y}_{t-L+1}^{t}|\lambda^{n}\,(t)\right)-p\left(\mathbf{y}_{t-L+1}^{t}|\lambda^{n-1}(t)\right)<\varepsilon$;
Obtain the statistical channel state:
Estimate the statistical channel state $\hat{q}\,(t)$ following (16)–(18);

when $l=L-2,L-3,\ldots,0$, where

$$p\left(q\,(t-l)\,\middle|\mathbf{y}_{t-L+1}^{t-l-1},\hat{\lambda}\,(t)\right)=\sum_{q(t-l-1)\in\mathbb{S}}\hat{g}_{q(t-l-1)q(t-l)}\,(t)\,p\left(q\,(t-l-1)\,\middle|\mathbf{y}_{t-L+1}^{t-l-1},\hat{\lambda}\,(t)\right).\tag{18}$$

In summary, the necessary steps for the proposed S-CSI acquisition approach are listed in Algorithm 1.

V. Simulation Results

In this section, we present some representative simulation results to evaluate the performance of the proposed S-CSI acquisition approach. Due to fact that the field experiments on massive MIMO channels are still at an exploratory stage and the raw data of channel sounding are difficult to be obtained, the parameters used in the simulations are randomly selected, as shown in Table II. Note that, Setup 1 and 2 both have three statistical channel states, but with different initial state probability vectors and transition matrices, while Setup 3 has four statistical channel states. For each statistical channel state S_i, the mean vector and covariance matrix of the (i,l)th element of the probability density matrix $\mathbf{B}\,(t)$ are $\boldsymbol{\mu}_{S_i}\,(t)=\mathbf{0}$ and $\boldsymbol{\Sigma}_{S_i}\,(t)=(\sum_{n=1}^{N}\mathrm{E}\{|x_n\,(t)|^2\}\sigma_i^2+\sigma_z^2)\mathbf{I}_M$, where $\mathrm{E}\{|x_n\,(t)|^2\}$ represents the transmitting power of the nth user. Since the proposed S-CSI acquisition approach only requires the second-order statistical knowledge of the transmitted signals, binary phase shift keying (BPSK) modulated signals with unit transmitting power is used for simplicity. In addition, the threshold for (13) is set as $\varepsilon=10^{-6}$.

TABLE II
Parameters for the Simulations

Setup	Statistical channel states	Initial state probability vector $\boldsymbol{\eta}$	Transition matrix $\mathbf{G}$
1	$S_1=(0,0.5)$ $S_2=(0,2.0)$ $S_3=(0,3.5)$	$[1,0,0]$	$\begin{bmatrix}0.5420 & 0.3685 & 0.0895\\0.6496 & 0.3246 & 0.0258\\0.6407 & 0.2933 & 0.0660\end{bmatrix}$
2	$S_1=(0,0.5)$ $S_2=(0,2.0)$ $S_3=(0,3.5)$	$[0,0,1]$	$\begin{bmatrix}0.5696 & 0.3016 & 0.1288\\0.4685 & 0.1295 & 0.4020\\0.3260 & 0.5305 & 0.1435\end{bmatrix}$
3	$S_1=(0,0.5)$ $S_2=(0,2.0)$ $S_3=(0,3.5)$ $S_4=(0,5.0)$	$[1,0,0,0]$	$\begin{bmatrix}0.3460 & 0.2878 & 0.2406 & 0.1256\\0.2639 & 0.3280 & 0.1454 & 0.2627\\0.2940 & 0.1423 & 0.3237 & 0.2400\\0.1626 & 0.2414 & 0.2896 & 0.3064\end{bmatrix}$

In the Monte Carlo simulations, the following steps are carried out.

Step 1: Generate a random hidden sequence q_{t-L+1}^{t} using the initial state probability vector and the state transition probability matrix.

Step 2: Generate an observation sequence $\mathbf{y}_{t-L+1}^{t}$ conditioned on the hidden sequence and the HSCSM-model.

Step 3: Estimate the parameter of the HSCSM-model from the observation sequence.

Step 4: Estimate the statistical channel state based on the observation sequence and the estimated HSCSM-model parameter.

The performance indicators are the mean square error (MSE) and the accuracy rate P_c, respectively given as

$$\mathrm{MSE}=\frac{1}{\zeta}\sum_{j=1}^{\zeta}\left|\hat{\sigma}_{ij}^2\,(t)-\sigma_{ij}^2\,(t)\right|^2,\tag{19}$$

and

$$P_c=\frac{\sum_{j=1}^{\zeta}\mathrm{Sign}\left(|\hat{q}_j\,(t)-q_j\,(t)|\right)}{\zeta},\tag{20}$$

where $\zeta=10000$ denotes the number of independent Monte Carlo simulation trials, $\hat{q}_j\,(t)$ is the estimated statistical channel state in the jth Monte Carlo trial, and $\hat{\sigma}_{ij}^2\,(t)$ denotes its variance. The function $\mathrm{Sign}\,(\cdot)$ is defined as

$$\mathrm{Sign}\,(x)=\begin{cases}1, & x=0,\\0, & x\neq 0.\end{cases}\tag{21}$$

A. Feasibility of the Proposed Approach

In this subsection, the feasibility of the proposed approach is testified. Specifically, we compare the MSE performance of the proposed statistical channel state estimator and an EM-based estimator in [13], where the number of users $N=16$, the number of BS antennas $M=128$ and the length of observation sequence $L=10$.

The results are shown in Fig. 3, it can be observed that: 1) The EM-based estimator cannot work for all the three Setups in the

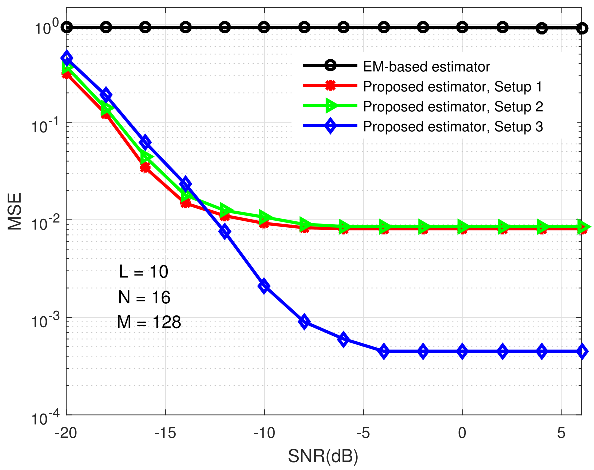

Fig. 3. The MSE performance of the proposed statistical channel state estimator and the EM-based estimator.

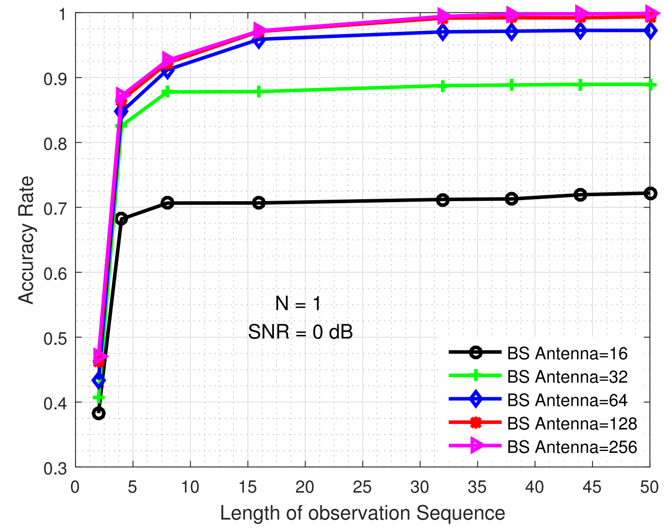

Fig. 4. Effect of the observation length on the estimation accuracy rate.

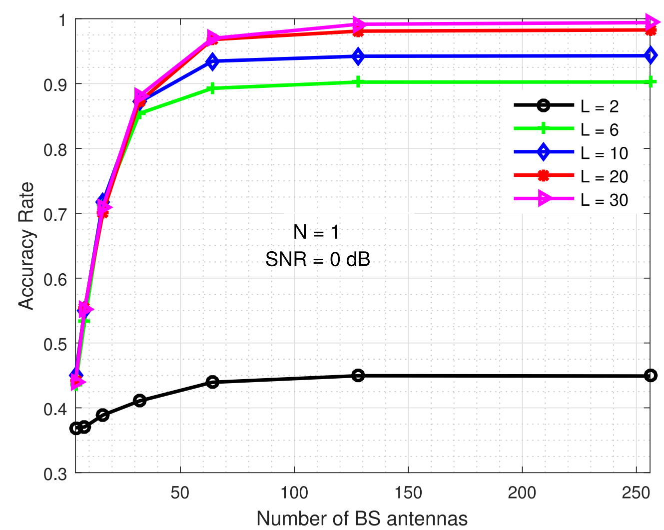

Fig. 5. Effect of the number of BS antennas on the estimation accuracy rate.

non-stationary environment, even when the SNR is increased to a high level. In contrast, our proposed approach can obtain the S-CSI with a much lower MSE. 2) For Setups 1 and 2, the proposed statistical channel state estimator has similar performance in terms of the MSE. It is indicated that, given the same set of statistical states, the performance of the proposed approach is not sensitive to the initial state probability or the transition probability. 3) Compared to Setup 1 or 2, a better MSE performance can be obtained by the proposed approach with Setup 3 when SNR > -12dB. This is due to the fact that, more statistical states will result in a larger "capacity" of the HSCSM-model, thus improve the performance [28]. Furthermore, it is noticed that a crosspoint exists between the MSE curve of Setup 3 and those of Setups 1 and 2 when SNR ≈ -12dB. I.e., in the region when SNR < -12dB, the MSE performance with Setup 3 is worse than that of Setup 1 or 2. The explanation is that, in the low SNR region, the sampled data is severely contaminated and the model parameter is estimated through the "bad" samples. As a result, erroneous model parameter is used and wrong decision is made, where more statistical states lead to a higher error probability.

The above observations show that, accurate S-CSI can be obtained by the proposed approach in the simulated non-stationary massive MIMO environment, especially when the number of statistical states is large. In the following simulations, we use only Setup 1 for simplicity since it represents the worst-case performance among the three setups.

B. Accuracy Rate of the Proposed Approach

In the next, the accuracy rate of the proposed S-CSI acquisition approach is testified, where single-user and multi-user scenarios are considered, respectively.

1) Single-User Scenario: Firstly, the effect of the observation length on the estimation accuracy rate is studied and the results are shown in Fig. 4, where SNR $= 0$ dB, and the number of BS antennas is set as $M = \{16, 32, 64, 128, 256\}$. It can be observed that: *a)* When the length of observation sequence increases, the accuracy rate increases as expected. Thus, to obtain accurate S-CSI, an adequate long observation sequence should be available. This is natural due to the fact that the proposed HSCSM-model is a statistical one, where a longer observation sequence can help to obtain a better parameter estimation and thus yield more accurate S-CSI. *b)* The effect of the accuracy rate improvement by increasing the length of observation sequence is marginal when $L \geq 30$. This is because that, there exists a uniform capacity convergence for the deterministic statistical model [28], [29]. Namely, the improvement associated with increasing the observation length becomes saturated when the convergence is obtained. Therefore, in Fig. 4, the accuracy rate is dominated by the noise level when $L \geq 30$.

Then, the effect of the number of BS antennas on the estimation accuracy rate is studied and the results are shown in Fig. 5,

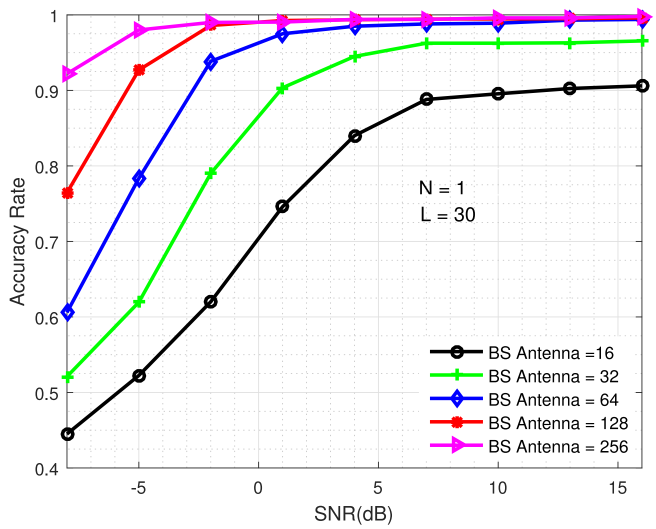

Fig. 6. Effect of the SNR on the estimation accuracy rate.

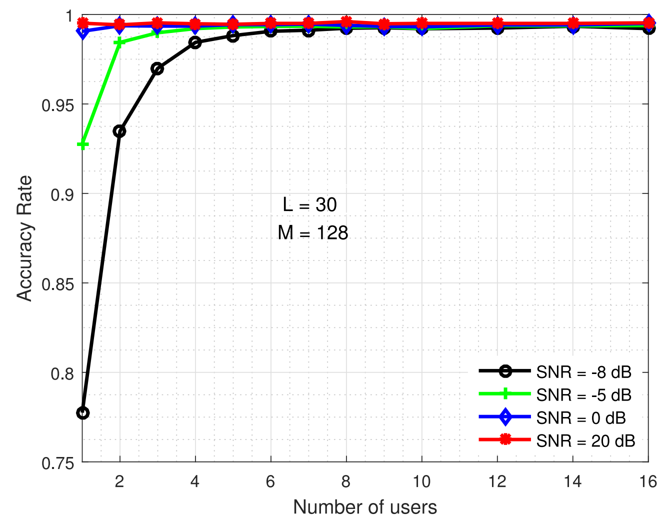

Fig. 7. Effect of the number of users on the estimation accuracy rate.

where SNR $= 0$ dB, and the length of observation sequence is set as $L = \{2, 6, 10, 20, 30\}$. Note that, increasing the number of BS antennas will increase the number of elements following the same distribution in one observation. It can be observed that, the number of BS antennas affects the accuracy rate in a similar way as the observation length does. Namely, increasing the number of BS antennas can improve the estimation accuracy rate significantly when M is small, and such an improvement becomes marginal when M is large, e.g., a steady-state accuracy rate is achieved when $M \geq 128$.

Next, the effect of the SNR on the accuracy rate is studied and the results are shown in Fig. 6, where $L = 30$ is used and the number of BS antennas is set as $M = \{16, 32, 64, 128, 256\}$. As we can see, given an adequate long observation sequence, the estimation accuracy rate improves as the SNR increases. In addition, as the number of BS antennas increases, the corresponding SNR of the steady-state accuracy rate decreases. When the number of BS antennas is 128, the steady-state accuracy rate $(P_c \geq 0.99)$ corresponds to SNR $= 1$ dB.

In summary, in order to obtain accurate S-CSI $(P_c \geq 0.99)$ in the single-user scenario, the length of observation sequence should be longer than 30, the number of BS antennas should be more than 128 and the SNR should be greater than 1 dB.

2) Multi-User Scenario: The estimation accuracy rate of the proposed approach in the multi-user scenario is shown in Fig. 7, where the number of BS antennas $M = 128$, the observation length L equals to 30, and SNR $= \{-8\text{ dB}, -5\text{ dB}, 0\text{ dB}, 20\text{ dB}\}$. Note that, according to (1), by increasing the number of users, the SNR is increased by N times. Namely, the multi-user scenario is equivalent to the single-user scenario with N-fold SNR, this is confirmed by the results in Fig. 7. Specifically, when SNR $= -8$ dB, accurate S-CSI $(P_c \geq 0.99)$ can be obtained when $N \geq 8$, and while SNR $= -5$ dB, accurate S-CSI is obtained when $N \geq 4$.

TABLE III
COMPUTATIONAL COMPLEXITY OF STEPS IN ALGORITHM 1

Steps	Computational complexity
1-3	$O(LMK)$
5	$O(LK^2)$
6	$O(LK^2)$
8-9	$O(LK^2 + LK + K)$
10-11	$O(LMK)$
14	$O(LK^2)$
Total	$O(LMK + LK^2 + LK + K)$

Therefore, the over all requirements for accurate S-CSI $(P_c \geq 0.99)$ acquisition in both single-user and multi-user scenarios is $L \geq 30$, $M \geq 128$ and $10 \log N +$ SNR ≥ 1 dB.

C. Computational Complexity of the Proposed Approach

Last but not least, the computational complexity of the proposed statistical channel state estimator is analyzed and evaluated by the number of complex multiplications, which can be obtained by summing up the computational complexity of each step in Algorithm 1, as listed in Table III. Note that, due to the diagonal property of the covariance matrix, the complexity of calculating $\mathbf{B}(t)$ inversion is $O(M)$ instead of $O(M^3)$. Thus, the total computational complexity of the proposed statistical channel state estimator is

$$N_{\text{complexity}} = O\left(LMK + LK^2 + LK + K\right). \quad (22)$$

Discussion: The computational complexity in (22) is primarily dominated by $O(LMK)$, while the storage space consumed by the L-length, M-dimensional observation sequence is dominated by $O(LMN_s)$, where N_s is the storage space used for one receiving signal. Therefore, it is clear that, given a limited computing-ability/storage-capacity which is determined by the hardware capability, there exists a tradeoff between the affordable observation length L and the number of BS antennas M.

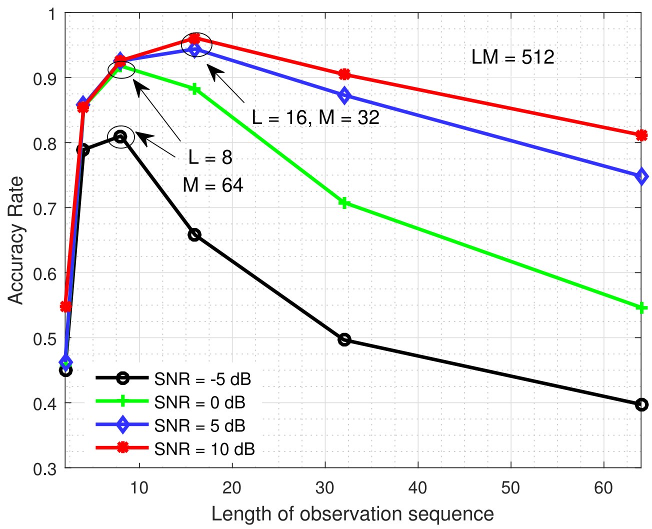

Fig. 8. Tradeoff between the affordable observation length and the number of BS antennas.

The performance of the proposed statistical channel state estimator is then studied given a limitation of $LM = 512$, where $L = \{2, 4, 8, 16, 32, 64\}$ and $M = \{256, 128, 64, 32, 16, 8\}$ accordingly, the results are shown in Fig. 8. It is observed that, as expected, the tradeoff between the observation length and the number of BS antennas does exist. In order to achieve the optimal performance, $\{L = 8, M = 64\}$ is chosen for SNR $= -5$ dB, and 0 dB, while $\{L = 16, M = 32\}$ is selected when SNR $= 5$ dB, and 10 dB. It can be inferred that, the limited resource should be allocated to accommodate more BS antennas when the SNR is low, while a longer observation sequence should be allowed when the SNR is high. The reason is that, in the low SNR region, the credibility of each receiving signal is low due to the noise perturbation, and it can be improved by using more BS antennas. While in the high SNR region, each observation is credible and thereby a longer observation sequence can yield a better performance. In other words, in some cases, we have to abandon the data belonging to part of the BS antennas in order to allow a longer observation sequence for a better performance .

VI. Conclusion

In this paper, an S-CSI acquisition approach has been proposed for the non-stationary massive MIMO channel. Firstly, a HSCSM-model has been set up to represent the probabilistic dependence of the observed receiving sequence on the statistical channel states. Then, the parameter of the HSCSM-model has been estimated given the observed sequence of receiving signals. In the next, the statistical channel state has been estimated based on the HSCSM-model and its estimated parameter. The effectiveness of the proposed S-CSI acquisition approach has been verified by simulation results. In addition, a tradeoff between the affordable observation length and the number of BS antennas has been found when the computing/storage resource is limited. Specifically, in order to achieve the optimal performance, the limited computing/storage resource should be allocated to accommodate more antennas when the SNR is low, while a longer observation sequence should be allowed when the SNR is high.

In the following, we start from defining the forward and backward variables, then we show how their scaled versions are calculated, finally, we give the derivation of the probability $P\left(q\left(t-l\right)=S_i\left|\mathbf{y}_{t-L+1}^{t}, \lambda\left(t\right)\right.\right)$ and the probability $P\left(q\left(t-l\right)=S_i, q\left(t-l+1\right)=S_j\left|\mathbf{y}_{t-L+1}^{t}, \lambda\left(t\right)\right.\right)$.

The forward variable $\alpha_{t-l}\left(S_i\right)$ denotes the probability density of partial observation sequence $\mathbf{y}_{t-L+1}^{t-l}$ and the hidden state $q\left(t-l\right) = S_i$ given the condition of the HSCSM-model parameter $\lambda\left(t\right)$, i.e.,

$$\alpha_{t-l}\left(S_i\right) = p\left(\mathbf{y}_{t-L+1}^{t-l}, q\left(t-l\right) = S_i\left|\lambda\left(t\right)\right.\right). \tag{23}$$

For $l = L-1, L-2, ..., 0$, $\alpha_{t-l}\left(S_i\right)$ can be obtained consequently, given as

$$\alpha_{t-l}\left(S_i\right) = \eta_{S_i}\left(t\right) p\left(\mathbf{y}\left(t-l\right)\left|S_i\right.\right) \tag{24a}$$

when $l = L-1$, and

$$\alpha_{t-l}\left(S_i\right) = \left(\sum_{j=1}^{K} \alpha_{t-l-1}\left(S_j\right) g_{ji}\left(t\right)\right) p\left(\mathbf{y}\left(t-l\right)\left|S_i\right.\right) \tag{24b}$$

when $l = L-2, L-3, ..., 0$.

Scaling is used to make sure that the calculation of (23) will not exceed the calculating limit of computer simulations, given as (without scaling, the calculation of (23) may occasionally exceeds the precision limit)

$$\bar{\alpha}_{t-l}\left(S_i\right) = \frac{\alpha_{t-l}\left(S_i\right)}{\prod_{d=l}^{L-1} c_{t-d}}, \tag{25}$$

where

$$c_{t-d} = \sum_{i=1}^{K} \eta_{S_i}\left(t\right) p\left(\mathbf{y}\left(t-d\right)\left|S_i\right.\right) \tag{26a}$$

when $d = L-1$, and

$$c_{t-d} = \sum_{i=1}^{K} \left(\sum_{j=1}^{K} \bar{\alpha}_{t-d-1}\left(S_j\right) g_{ji}\left(t\right)\right) p\left(\mathbf{y}\left(t-d\right)\left|S_i\right.\right) \tag{26b}$$

when $d = L-2, L-3, ..., 0$.

Substituting (24) into (25), the scaled forward variable can be obtained as

$$\bar{\alpha}_{t-l}\left(S_i\right) = \frac{\eta_{S_i}\left(t\right) p\left(\mathbf{y}\left(t-l\right)\left|S_i\right.\right)}{c_{t-l}} \tag{27a}$$

when $l = L-1$, and

$$\bar{\alpha}_{t-l}\left(S_i\right) = \frac{\left(\sum_{j=1}^{K} \bar{\alpha}_{t-l-1}\left(S_j\right) g_{ji}\left(t\right)\right) p\left(\mathbf{y}\left(t-l\right)\left|S_i\right.\right)}{c_{t-l}} \tag{27b}$$

when $l = L-2, L-3, ..., 0$.

The backward variables can be obtained in a similar manner. It denotes the probability density of partial observations $\mathbf{y}_{t-l+1}^{t}$ given the hidden state $q(t-l) = S_i$ and the HSCSM-model parameter $\lambda(t)$, i.e.,

$$\beta_{t-l}(S_i) = p\left(\mathbf{y}_{t-l+1}^{t} | q(t-l) = S_i, \lambda(t)\right). \tag{28}$$

For $l = 0, 1, ..., L-1$, $\beta_{t-l}(S_i)$ can be obtained reversely, given as

$$\beta_{t-l}(S_j) = 1 \tag{29a}$$

when $l = 0$, and

$$\beta_{t-l}(S_j) = \sum_{i=1}^{K} \beta_{t-l+1}(S_i)\, g_{ji}(t)\, p(\mathbf{y}(t-l+1) | S_i) \tag{29b}$$

when $l = 1, 2, ..., L-1$.

Similarly, the scaled backward variable is calculated as

$$\bar{\beta}_{t-l}(S_j) = 1 \tag{30a}$$

when $l = 0$, and

$$\bar{\beta}_{t-l}(S_j) = \frac{\sum_{i=1}^{K} \bar{\beta}_{t-l+1}(S_i)\, g_{ji}(t)\, p(\mathbf{y}(t-l+1) | S_i)}{c_{t-l}} \tag{30b}$$

when $l = 1, 2, .., L-1$.

Then, $P\left(q(t-l) = S_i \left| \mathbf{y}_{t-L+1}^{t}, \lambda(t)\right.\right)$ is obtained as

$$P\left(q(t-l) = S_i \left| \mathbf{y}_{t-L+1}^{t}, \lambda(t)\right.\right) = \frac{\bar{\alpha}_{t-l}(S_i)\, \bar{\beta}_{t-l}(S_i)}{\sum_{i=1}^{K} \bar{\alpha}_{t-l}(S_i)\, \bar{\beta}_{t-l}(S_i)}, \tag{31}$$

and $P\left(q(t-l) = S_i, q(t-l+1) = S_j \left| \mathbf{y}_{t-L+1}^{t}, \lambda(t)\right.\right)$ is calculated as

$$P\left(q(t-l) = S_i, q(t-l+1) = S_j \left| \mathbf{y}_{t-L+1}^{t}, \lambda(t)\right.\right)$$
$$= \frac{\bar{\alpha}_{t-l}(S_i)\, g_{ij}(t)\, p(\mathbf{y}(t-l+1) | S_j)\, \bar{\beta}_{t-l+1}(S_j)}{\sum_{i=1}^{K}\sum_{j=1}^{K} \bar{\alpha}_{t-l}(S_i)\, g_{ij}(t)\, p(\mathbf{y}(t-l+1) | S_j)\, \bar{\beta}_{t-l+1}(S_j)}. \tag{32}$$

As aforementioned, the conditional probability density of $\mathbf{y}_{t-L+1}^{t}$ in (9) can be calculated more efficiently, i.e.,

$$p\left(\mathbf{y}_{t-L+1}^{t} | \lambda(t)\right) = \sum_{i=1}^{K} \alpha_t(S_i). \tag{33}$$

Substituting (25) into (33), the probability in (9) is obtained as

$$p\left(\mathbf{y}_{t-L+1}^{t} | \lambda\right) = \prod_{d=0}^{L-1} c_{t-d}. \tag{34}$$

References

[1] T. L. Marzetta, “Noncooperative cellular wireless with unlimited number of base station antennas,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] V. Jungnickel *et al.*, “The role of small cells, coordinated multipoint, and massive MIMO in 5G,” *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 44–51, May 2014.

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

[4] 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.

[5] 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.

[6] S. Wang, Y. Li, M. Zhao, and J. Wang, “Energy-efficient and low-complexity uplink transceiver for massive spatial modulation MIMO,” *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4617–4632, Oct. 2015.

[7] T. E. Bogale and L. B. Le, “Pilot optimization and channel estimation for multiuser massive MIMO systems,” in *Proc. Annu. Conf. Inf. Sci. Syst.*, Princeton, NJ, USA, Mar. 2014, pp. 1–6.

[8] 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 Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3352–3366, Jun. 2015.

[9] C. Sun, X. Gao, S. Jin, M. Matthaiou, Z. Ding, and C. Xiao, “Beam division multiple access transmission for massive MIMO communications,” *IEEE Trans. Commun.*, vol. 63, no. 6, pp. 2170–2184, Jun. 2015.

[10] Z. Zheng, C. Zhu, B. Jiang, W. Zhong, and X. Gao, “Statistical channel state information acquisition for massive MIMO communications,” in *Proc. Int. Conf. Wireless Commun. Signal Process.*, Nanjing, China, Oct. 2015, pp. 1–5.

[11] L. Tong, B. M. Sadler, and M. Dong, “Pilot-assisted wireless transmissions: General model, design criteria, and signal processing,” *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[12] J. P. Vila and P. Schniter, “Expectation-maximization Gaussian-mixture approximate message passing,” *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[13] C.-K. Wen, S. Jin, K.-K. Wong, J.-C. Chen, and P. Ting, “Channel estimation for massive MIMO using Gaussian-mixture Bayesian learning,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1356–1368, Mar. 2015.

[14] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: Benefits and challenges,” *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[15] S. Payami and F. Tufvesson, “Channel measurements and analysis for very large array systems at 2.6 GHz,” in *Proc. Eur. Conf. Antennas Propag.*, Prague, Czech Republic, Mar. 2012, pp. 433–437.

[16] X. Gao, F. Tufvesson, O. Edfors, and F. Rusek, “Measured propagation characteristics for very-large MIMO at 2.6 GHz,” in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 2012, pp. 295–299.

[17] S. Wu *et al.*, “A non-stationary wideband channel model for massive MIMO communication systems,” *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1434–1446, Mar. 2015.

[18] S. Wu, C.-X. Wang, H.-M. Aggoune, M. Alwakeel, and Y. He, “A nonstationary 3-D wideband twin-cluster model for 5G massive MIMO channels,” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1207–1218, Jun. 2014.

[19] S. Lin, Z. Zhong, L. Cai, and Y. Luo, “Finite state Markov modelling for high speed railway wireless communication channel,” in *Proc. IEEE Global Commun. Conf.*, Anaheim, CA, USA, Dec. 2012, pp. 5421–5426.

[20] A. Ghazal *et al.*, “A non-stationary IMT-Advanced MIMO channel model for high-mobility wireless communication systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2057–2068, Apr. 2017.

[21] A. Lapidoth, *A Foundation in Digital Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[22] F. Babich and G. Lombardi, “A Markov model for the mobile propagation channel,” *IEEE Trans. Veh. Technol.*, vol. 48, no. 1, pp. 63–73, Jan. 2000.

[23] T. Zwick, C. Fischer, and W. Wiesbeck, “A stochastic multipath channel model including path directions for indoor environments,” *IEEE J. Sel. Areas Commun.*, vol. 20, no. 6, pp. 1178–1192, Aug. 2002.

[24] M. A. Mohamed and P. Gader, “Generalized hidden Markov models. II. Application to handwritten word recognition,” *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 1, pp. 82–94, Feb. 2000.

[25] L. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition,” *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[26] M. A. Mohamed and P. Gader, “Generalized hidden Markov models-part I: Theoretical frameworks,” *IEEE Trans. Fuzzy Syst.*, vol. 8, no. 1, pp. 67–81, Feb. 2000.

[27] L. E. Baum and G. R. Sell, “Growth functions for transformations on manifolds,” *Pacific J. Math.*, vol. 27, no. 2, pp. 211–227, Jan. 1968.

[28] V. N. Vapnik, *Statistical Learning Theory*. New York, NY, USA: Wiley, 1998.

[29] A. Y. Ng, *Machine Learning Yearning*, 2016. [Online]. Available: http://www.mlyearning.org

Guoliang Wang is currently working toward the Master’s degree with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China.

Wei Peng (M’07–SM’12) received the Ph.D. degree in wireless communications from the University of Hong Kong, Hong Kong, in 2007. She was with Tohoku University, Sendai, Japan, as a Postdoctoral Research Fellow from 2008 to 2009 and as an Assistant Professor from 2009 to 2013. Since 2013, she has been with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, as an Associate Professor. She held a visiting position in Columbia University, New York City, NY, USA. Her research interests include signal processing for large-scale wireless communication systems, online learning based network optimization, and space-ground integrated network.

Dong Li received the Ph.D. degree in electronics and communication engineering from Sun Yat-sen University, Guangzhou, China, in 2010. Since 2010, he has been with the Faculty of Information Technology, Macau University of Science and Technology, Macau, China, where he is currently an Associate Professor. He held research and visiting positions with the Institute for Infocomm Research, Singapore, and The Chinese University of Hong Kong, Shenzhen, China. His research interests include signal processing and machine learning for wireless communications, and Internet of Things.

Tao Jiang (M’06–SM’10) received the Ph.D. degree in information and communication engineering from Huazhong University of Science and Technology, Wuhan, China, in 2004. He is currently a Distinguished Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology. From 2004 to 2007, he worked in some universities including Brunel University and University of Michigan-Dearborn. He has authored or coauthored about 300 technical papers in major journals and conferences and 9 books/chapters in areas of communications and networks. He served or is serving as an Associate Editor of some technical journals in communications, including in IEEE Transactions on Signal Processing, IEEE Communications Surveys and Tutorials, IEEE Transactions on Vehicular Technology, and IEEE Internet of Things Journal.

Fumiyuki Adachi (M’79–SM’90–F’02–LF’16) received the B.S. and Dr. Eng. degrees in electrical engineering from Tohoku University, Sendai, Japan, in 1973 and 1984, respectively. In April 1973, he joined the NTT company and conducted research on digital cellular mobile communications. From 1992 to 1999, he was with NTT DOCOMO, where he led a research group on wideband/broadband wireless access for 3G and beyond. Since 2000, he has been with Tohoku University, where he was a full Professor with the Department of Communications Engineering until March 2016. He is currently a Specially Appointed Professor with the Research Organization of Electrical Communication. His research interests include the area of wireless signal processing and networking.