One-Bit Quantization Design and Channel Estimation for Massive MIMO Systems

Feiyu Wang, Jun Fang, *Member, IEEE*, Hongbin Li, *Senior Member, IEEE*, Zhi Chen, and Shaoqian Li, *Fellow, IEEE*

***Abstract*—We consider the problem of channel estimation for uplink multiuser massive MIMO systems, where, in order to significantly reduce the hardware cost and power consumption, one-bit analog-to-digital converters (ADCs) are used at the base station to quantize the received signal. In this paper, we study the problem of optimal one-bit quantization design for channel estimation in one-bit massive MIMO systems. Our analysis reveals that, if the quantization thresholds are optimally devised, using one-bit ADCs can achieve an estimation error close to (with an increase by a factor of $\pi/2$) that of an ideal estimator that has access to the unquantized data. Since the optimal quantization thresholds are dependent on the unknown channel parameters, we introduce an adaptive quantization scheme in which the thresholds are adaptively adjusted, and a random quantization scheme that randomly generates a set of thresholds based on some statistical prior knowledge of the channel. Simulation results show that our proposed schemes presents a significant performance improvement over the conventional fixed quantization scheme that uses a fixed (typically zero) threshold.**

***Index Terms*—Massive MIMO systems, channel estimation, one-bit quantization design, Cramér-Rao bound (CRB), maximum likelihood (ML) estimator.**

I. Introduction

MASSIVE multiple-input multiple-output (MIMO), also known as large-scale or very-large MIMO, is a promising technology to meet the ever growing demands for higher throughput and better quality-of-service of next-generation wireless communication systems [1]–[3]. Massive MIMO systems are those that are equipped with a large number of antennas at the base station (BS) simultaneously serving a much smaller number of single-antenna users sharing the same time-frequency slot. By exploiting the asymptotic orthogonality among channel vectors associated with different users, massive MIMO systems can achieve almost perfect inter-user interference cancelation with a simple linear precoder and receive combiner [4], and thus have the potential to enhance the spectrum efficiency by orders of magnitude.

Despite the above benefits, massive MIMO systems pose new challenges for system design and hardware implementation. Due to the large number of antennas at the BS, the hardware cost and power consumption could become prohibitively high if we still employ expensive and power-hungry high-resolution analog-to-digital convertors (ADCs) [5]. To address this obstacle, recent studies (e.g. [6]–[14]) considered the use of low-resolution ADCs (e.g. 1–3 bits) for massive MIMO systems. It is known that the hardware complexity and power consumption grow exponentially with the resolution (i.e. the number of bits per sample) of the ADC. Therefore lowering the resolution of the ADC can effectively reduce the hardware cost and power consumption. In particular, for the extreme one-bit case, the ADC becomes a simple analog comparator. Also, automatic gain control (AGC) is no longer needed when one-bit ADCs are used, which further simplifies the hardware complexity.

Massive MIMO with low-resolution ADCs has attracted much attention over the past few years. Great efforts have been made to understand the effects of low-resolution ADCs on the performance of MIMO and massive MIMO systems. In [15], by assuming full knowledge of channel state information (CSI), the capacity at both finite and infinite signal-to-noise ratio (SNR) was derived for one-bit MIMO systems. For massive MIMO systems with low-resolution ADCs, the spectral efficiency and the uplink achievable rate were investigated in [6]–[8], [16] under different assumptions. These theoretical analyses suggest that the use of the low cost and low-resolution ADCs can still provide satisfactory achievable rates and spectral efficiency.

This paper focuses on the problem of uplink multiuser channel estimation for one-bit massive MIMO systems. Channel estimation is crucial to support multi-user MIMO operation in massive MIMO systems [17]–[24]. To reach the full potential of massive MIMO, accurate downlink CSI is required at the BS for precoding and other operations. Most literature on massive MIMO systems, e.g. [1], [4], [25], [26], assumes a time

Manuscript received February 28, 2018; revised June 15, 2018 and August 16, 2018; accepted September 11, 2018. Date of publication September 17, 2018; date of current version November 12, 2018. This work was supported in part by the National Science Foundation of China under Grants 61522104 and 61829103, in part by the National Science Foundation under Grants ECCS-1408182 and ECCS-1609393, and in part by the Science and Technology on Communication Networks Laboratory Fund. The review of this paper was coordinated by Prof. J.-C. Lin. *(Corresponding author: Jun Fang.)*

F. Wang is with the National Key Laboratory of Science and Technology on Communications, and the Center for Intelligent Networking and Communications, University of Electronic Science and Technology of China, Chengdu 611731, China, and also with the Science and Technology on Communication Networks Laboratory, Hebei 050081, China (e-mail: fywang.ee@outlook.com).

J. Fang, Z. Chen, and S. Li are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: JunFang@uestc.edu.cn; chenzhi@uestc.edu.cn; lsq@uestc.edu.cn).

H. Li is with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: Hongbin.Li@stevens.edu).

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.2018.2870580

0018-9545 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

division duplex (TDD) mode in which the downlink CSI can be immediately obtained from the uplink CSI by exploiting channel reciprocity. Nevertheless, channel estimation for massive MIMO systems with one-bit ADCs is challenging since the magnitude and phase information about the received signal are lost or severely distorted due to the coarse quantization. It was shown in [6] that one-bit massive MIMO systems require an excessively long training sequence (e.g. approximately 50 times the number of users) to achieve an acceptable performance. The work [9] showed that for one-bit massive MIMO systems, a least-squares channel estimation scheme and a maximum-ratio combining scheme are sufficient to support both multiuser operation and the use of high-order constellations. Nevertheless, a long training sequence is still a requirement. To alleviate this issue, a Bayes-optimal joint channel and data estimation scheme was proposed in [11], in which the estimated payload data are utilized to aid channel estimation. In [12], a maximum likelihood channel estimator, along with a near maximum likelihood detector, were proposed for uplink massive MIMO systems with one-bit ADCs. In addition to channel estimation, Bayesian optimal data detection methods based on low-resolution ADCs were developed for millimeter wave (mmWave) communication systems [13], [14].

In this paper, we study one-bit quantization design for uplink multiuser channel estimation in massive MIMO systems. Quantization design is an interesting and important issue but largely neglected by existing massive MIMO channel estimation studies. In fact, most massive MIMO channel estimation schemes, e.g. [6], [9], [11], [12], assume a fixed, typically zero, quantization threshold, which is a convenient choice but could be far away from the optimum. In this paper, we investigate the optimal design of quantization thresholds as well as the training sequences. Our theoretical analysis shows that using one-bit ADCs can achieve an estimation error close to that attained by using infinite-precision ADCs, provided the thresholds are optimally devised. Our analysis also reveals that the optimal quantization thresholds are functions of the unknown channel parameters. Motivated by this fact, we introduce an adaptive quantization (AQ) scheme, where the thresholds are dynamically adjusted in a way such that the thresholds converge to the optimal thresholds, and a random quantization (RQ) scheme which randomly generates a set of non-identical thresholds based on some statistical prior knowledge of the channel. The proposed AQ scheme is particularly suitable for slowly time-varying channels. Simulation results show that our proposed schemes achieve a substantial training overhead reduction for channel estimation. In particular, the AQ scheme, even with a moderate number of pilot symbols (about 5 times the number of users), can provide an achievable rate close to that of the perfect CSI case.

A. Past Related Works on Quantization Design

Note that one-bit quantization design has been extensively studied in the context of general estimation problems and distributed sensor networks. Specifically, [27] seems to be one of the earliest works to examine the impact of quantization thresholds on the estimation performance. It shows that the one-bit estimator only incurs a mild performance degradation (increases by a factor of $\pi/2$) if the threshold is optimally selected. This study was later extended to address the distributed estimation problem arising in wireless sensor networks, e.g. [28]–[30]. All these works, e.g. [27]–[30], consider estimation of a scalar deterministic parameter. In [31], estimation of vector parameters based on one-bit quantized data was considered. The work [31] is concerned with the low signal-to-noise ratio regime, in which case it was shown that the optimal thresholds are approximately equal to zero. Different from [27]–[31], the work [32] models the parameter to be estimated as a random variable. Unlike the deterministic case, the optimal threshold for the random parameter case does not have an analytical form, and needs to be found by gradient-based search algorithms [32]. In [33], the estimation performance was examined for a scalar parameter case when a multi-bit quantizer is used.

Our work can be considered as an extension of previous quantization design studies [27]–[32] to the one-bit massive MIMO channel estimation scenario. This extension, however, is not straightforward. For our problem, the quantization thresholds have to be jointly optimized along with the training sequence which is also an important factor in determining the estimation performance. To our best knowledge, this represents a first attempt to jointly optimize the quantization thresholds and training sequences in the one-bit channel estimation framework. Another contribution of this work is the development of an adaptive quantization scheme and a random quantization scheme which present a substantial performance improvement over the conventional fixed quantization scheme. Although the idea of adaptive quantization was originally introduced in [27], such an idea was more fully developed in our work, including examining the asymptotic behavior of the AQ scheme, and providing several possible ways of implementing AQ in one-bit massive MIMO systems.

We also noticed that a recent work [34] studied the impact of an unknown nonzero threshold on the channel estimation performance and obtained some interesting results. The threshold/offset considered in [34], however, is an unknown parameter out of control, instead of a variable that can be manually set or optimized to minimize the estimation errors, as in our work and other works, e.g. [27]–[30]. Additionally, in [34], the threshold is assumed fixed throughout the training process (since the offset effect remains constant over time), whereas in our work, the quantization thresholds used for different training symbols and different antennas are not necessarily the same and can be different. From the system design perspective, such a relaxation provides extra degrees of freedom for optimization, and has the potential to achieve better estimation performance.

B. Notations and Organization of the Paper

The following notations are adopted throughout this paper, where $(\cdot)^T$ and $(\cdot)^H$ represent the transpose and conjugate transpose, respectively. $\mathrm{tr}(\boldsymbol{A})$ denotes the trace of $\boldsymbol{A}$, and $\mathbf{A} \succeq \mathbf{0}$

means that the matrix is positive semidefinite. $\mathrm{vec}(\boldsymbol{A})$ denotes a vectorization operation which converts $\boldsymbol{A}$ into a column vector by stacking the columns of $\boldsymbol{A}$ on top of one another. The $m \times m$ identity matrix is denoted by $\boldsymbol{I}_m$. The symbols $\mathbb{R}^{n\times m}$ and $\mathbb{R}^n$ stand for the set of $n \times m$ matrices and the set of n-dimensional column vectors with real entries, respectively. $\mathbb{C}^{n\times m}$ and $\mathbb{C}^n$ denote the set of $n \times m$ matrices and the set of n-dimensional column vectors with complex entries, respectively.

The rest of the paper is organized as follows. The system model and the problem of channel estimation using one-bit ADCs are discussed in Section II. In Section III, we develop a maximum likelihood estimator and carry out a Cramér-Rao bound analysis of the one-bit channel estimation problem. The optimal design of quantization thresholds and the pilot sequences is studied in Section IV. In Section V, we develop an adaptive quantization scheme and a random quantization scheme for practical threshold design. Simulation results are provided in Section VI, followed by concluding remarks in Section VII.

II. System Model and Problem Formulation

Consider a single-cell uplink multiuser massive MIMO system, where the BS equipped with M antennas serves K ($M \gg K$) single-antenna users simultaneously. The channel is assumed to be flat block fading, i.e. the channel remains constant over a certain amount of coherence time. The received signal at the BS can be expressed as

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{X} + \boldsymbol{W} \tag{1}$$

where $\boldsymbol{X} \in \mathbb{C}^{K\times L}$ is a training matrix and its row corresponds to each user's training sequence with L pilot symbols, $\boldsymbol{H} \in \mathbb{C}^{M\times K}$ denotes a deterministic channel parameter to be estimated, and $\boldsymbol{W} \in \mathbb{C}^{M\times L}$ represents the additive white Gaussian noise with its entries following a circularly symmetric complex Gaussian distribution with zero mean and variance $2\sigma_w^2$.

To reduce the hardware cost and power consumption, we consider a massive MIMO system which uses one-bit ADCs at the BS to quantize the received signal. Specifically, at each antenna, the real and imaginary components of the received signal are quantized separately using a pair of one-bit ADCs. Thus in total $2M$ one-bit ADCs are needed. The quantized output of the received signal, $\boldsymbol{B} \triangleq [b_{m,l}]$, can be written as

$$\boldsymbol{B} = \mathcal{Q}(\boldsymbol{Y}) \tag{2}$$

where $\mathcal{Q}(\boldsymbol{Y})$ is an element-wise operation performed on $\boldsymbol{Y}$, and for each element of $\boldsymbol{Y}$, $y_{m,l}$, we have

$$\mathcal{Q}(y_{m,l}) = \mathrm{sgn}(\Re(y_{m,l})) + j\mathrm{sgn}(\Im(y_{m,l})) \tag{3}$$

in which $\Re(y)$ and $\Im(y)$ denote the real and imaginary components of y, respectively, and the sign function $\mathrm{sgn}(\cdot)$ is defined as

$$\mathrm{sgn}(y) \triangleq \begin{cases} 1 & \text{if } y \geq 0 \\ -1 & \text{otherwise} \end{cases} \tag{4}$$

Therefore the quantized output belongs to the set

$$b_{m,l} \in \{1+j, -1+j, 1-j, -1-j\} \quad \forall m, l \tag{5}$$

Note that in (2), we implicitly assume a zero threshold for one-bit quantization. Nevertheless, using identically a zero threshold for all measurements is not necessarily optimal, and it is interesting to analyze the impact of the quantization thresholds on the channel estimation performance. To examine this problem, let $\boldsymbol{T} \triangleq [\tau_{m,l}]$ denote the thresholds used for one-bit quantization. The quantized output of the received signal, $\boldsymbol{B}$, is now given as

$$\boldsymbol{B} = \mathcal{Q}(\boldsymbol{Y} - \boldsymbol{T}) \tag{6}$$

To facilitate our analysis, we first convert (1) into a real-valued form as follows

$$\widetilde{\boldsymbol{Y}} = \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{H}} + \widetilde{\boldsymbol{W}} \tag{7}$$

where

$$\widetilde{\boldsymbol{Y}} \triangleq [\Re(\boldsymbol{Y})\ \Im(\boldsymbol{Y})]^T$$

$$\widetilde{\boldsymbol{H}} \triangleq [\Re(\boldsymbol{H})\ \Im(\boldsymbol{H})]^T$$

$$\widetilde{\boldsymbol{W}} \triangleq [\Re(\boldsymbol{W})\ \Im(\boldsymbol{W})]^T$$

and

$$\widetilde{\boldsymbol{A}} \triangleq \begin{bmatrix} \Re(\boldsymbol{X}) & \Im(\boldsymbol{X}) \\ -\Im(\boldsymbol{X}) & \Re(\boldsymbol{X}) \end{bmatrix}^T \tag{8}$$

Vectorizing the real-valued matrix $\widetilde{\boldsymbol{Y}}$, the received signal can be expressed as a real-valued vector form as

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{h} + \boldsymbol{w} \tag{9}$$

where $\boldsymbol{y} \triangleq \mathrm{vec}(\widetilde{\boldsymbol{Y}})$, $\boldsymbol{A} \triangleq \boldsymbol{I}_M \otimes \widetilde{\boldsymbol{A}}$, $\boldsymbol{h} \triangleq \mathrm{vec}(\widetilde{\boldsymbol{H}})$, and $\boldsymbol{w} \triangleq \mathrm{vec}(\widetilde{\boldsymbol{W}})$. Here $\otimes$ stands for the Kronecker product. It can be easily verified $\boldsymbol{y} \in \mathbb{R}^{2ML}$, $\boldsymbol{A} \in \mathbb{R}^{2ML\times 2MK}$, and $\boldsymbol{h} \in \mathbb{R}^{2MK}$. Accordingly, the one-bit quantized data can be written as

$$\boldsymbol{b} = \mathrm{sgn}(\boldsymbol{y} - \boldsymbol{\tau}) \tag{10}$$

where $\boldsymbol{\tau} \triangleq \mathrm{vec}([\Re(\boldsymbol{T})\ \Im(\boldsymbol{T})]^T)$ and $\boldsymbol{\tau} \in \mathbb{R}^{2ML}$. For simplicity, we define $N \triangleq 2ML$.

Our objective in this paper is to estimate the channel $\boldsymbol{h}$ based on the one-bit quantized data $\boldsymbol{b}$, examine the best achievable estimation performance and investigate the optimal thresholds $\boldsymbol{\tau}$ as well as the optimal training sequences $\boldsymbol{X}$. To this objective, in the following, we first develop a maximum likelihood (ML) estimator and carry out a Cramér-Rao bound (CRB) analysis. The optimal choice of the quantization thresholds as well as the training sequences is then studied based on the CRB matrix of the unknown channel parameter vector $\boldsymbol{h}$.

III. ML Estimator and CRB Analysis

A. ML Estimator

By combining (9) and (10), we have

$$b_n = \mathrm{sgn}(y_n - \tau_n) = \mathrm{sgn}(\boldsymbol{a}_n^T\boldsymbol{h} + w_n - \tau_n), \quad \forall n \tag{11}$$

where, by allowing a slight abuse of notation, we let b_n, y_n, τ_n, and w_n denote the nth entry of $\boldsymbol{b}$, $\boldsymbol{y}$, $\boldsymbol{\tau}$, and $\boldsymbol{w}$, respectively; and $\boldsymbol{a}_n^T$ denotes the nth row of $\boldsymbol{A}$. It is easy to derive that

$$P(b_n = 1; \boldsymbol{h}) = P(w_n \geq -(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n); \boldsymbol{h}) = F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n) \tag{12}$$

and

$$P(b_n = -1; \boldsymbol{h}) = P(w_n < -(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n); \boldsymbol{h}) = 1 - F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n) \tag{13}$$

where $F_w(\cdot)$ denotes the cumulative density function (CDF) of w_n, and w_n is a real-valued Gaussian random variable with zero-mean and variance σ_w^2. Therefore the probability mass function (PMF) of b_n is given by

$$p(b_n; \boldsymbol{h}) = [1 - F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)]^{(1-b_n)/2} \cdot [F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)]^{(1+b_n)/2} \tag{14}$$

Since $\{b_n\}$ are independent, the log-PMF or log-likelihood function can be written as

$$\begin{aligned}\mathcal{L}(\boldsymbol{h}) &\triangleq \log p(\boldsymbol{b}; \boldsymbol{h}, \boldsymbol{\tau}) \\ &= \sum_{n=1}^{N} \left\{ \frac{1-b_n}{2} \log[1 - F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)] \right. \\ &\quad \left. + \frac{1+b_n}{2} \log[F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)] \right\}\end{aligned} \tag{15}$$

The ML estimate of $\boldsymbol{h}$, therefore, is given as

$$\hat{\boldsymbol{h}} = \arg\max_{\boldsymbol{h}} \ \mathcal{L}(\boldsymbol{h}) \tag{16}$$

It can be proved that the log-likelihood function $\mathcal{L}(\boldsymbol{h})$ is a concave function (The proof of the concavity of $\mathcal{L}(\boldsymbol{h})$ is given in Appendix A). Hence computationally efficient gradient-based search algorithms can be used to find the global maximum.

In Algorithm 1, a summary of a gradient-based backtracking line search algorithm is provided to search for the optimal solution of (16). We see that at each iteration, the main computational task is to calculate $\mathcal{L}(\boldsymbol{h})$ and its first-order derivative $\nabla\mathcal{L}(\boldsymbol{h}) \triangleq \frac{\partial \mathcal{L}(\boldsymbol{h})}{\partial \boldsymbol{h}}$. To obtain $\mathcal{L}(\boldsymbol{h})$, we need to compute $\boldsymbol{A}\boldsymbol{h}$. Recalling that $\boldsymbol{A} = \boldsymbol{I}_M \otimes \tilde{\boldsymbol{A}}$, it can be readily verified that the number of flops required to calculate $\mathcal{L}(\boldsymbol{h})$ is of order $\mathcal{O}(KLM)$. Also, one can easily check that the number of flops required to compute $\nabla\mathcal{L}(\boldsymbol{h})$ is of order $\mathcal{O}(KLM)$. Thus the search algorithm has a computational complexity scaling as $\mathcal{O}(IKLM)$, where I denotes the number of iterations required for convergence. We see that the computational complexity of the gradient-based search algorithm grows linearly with the number of antennas M. As a comparison, a least squares channel estimation method given as follows

$$\hat{\boldsymbol{h}}_{\mathrm{LS}} = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{b} \tag{17}$$

has a computational complexity scaling as $\mathcal{O}(K^2LM)$. Since the number of iterations I is usually larger than the number of users K, the gradient-based ML estimator has a higher computational complexity than the least squares method. Nevertheless, our experiments suggest that with a good initialization point, the search algorithm is able to converge within tens of iterations, which makes its computational complexity comparable to the least squares method. The complexity of the ML estimator might be excessive for some practical massive MIMO systems. How to develop an efficient ML estimation algorithm is an important topic which has been studied in [12] and is also worthy of further investigation.

Algorithm 1: Backtracking Line Search algorithm.

Input: $\boldsymbol{A}$, $\boldsymbol{b}$, $\boldsymbol{\tau}$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$; $i_{\max}$, $\tilde{i}_{\max}$, and $\epsilon > 0$.
Output: $\hat{\boldsymbol{h}}$.

```
Initialize h^(0);
for i = 0 to i_max do
    Calculate L(h^(i)) and ∇L(h^(i));
    Initialize the step size ρ = 1 and ĩ = 1;
    Update h^(i+1) = h^(i) + ρ∇L(h^(i));
    while L(h^(i+1)) < L(h^(i)) + ρα‖∇L(h^(i))‖^2 and
    ĩ < ĩ_max do
        Update the step size ρ = βρ;
        Update h^(i+1) = h^(i) + ρ∇L(h^(i));
        ĩ = ĩ + 1;
    end while
    if ‖∇L(h^(i+1))‖^2 < ε then
        Set: ĥ = h^(i+1);
        Break;
    end if
end for
```

B. CRB

We now carry out a CRB analysis of the one-bit channel estimation problem (16). The CRB results help understand the effect of different system parameters, including quantization thresholds as well as training sequences, on the estimation performance. We first summarize our derived CRB results in the following theorem.

Theorem 1: The Fisher information matrix (FIM) for the estimation problem (16) is given as

$$\boldsymbol{J}(\boldsymbol{h}) = \sum_{n=1}^{N} g(\tau_n, \boldsymbol{a}_n) \boldsymbol{a}_n \boldsymbol{a}_n^T \tag{18}$$

where $g(\tau_n, \boldsymbol{a}_n)$ is defined as

$$g(\tau_n, \boldsymbol{a}_n) \triangleq \frac{f_w^2(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)}{F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)(1 - F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n))} \tag{19}$$

in which $f_w(\cdot)$ denotes the probability density function (PDF) of w_n. Accordingly, the CRB matrix for the estimation problem

(16) is given by

$$\mathrm{CRB}(\boldsymbol{h}) = \boldsymbol{J}^{-1}(\boldsymbol{h}) = \left(\sum_{n=1}^{N} g(\tau_n, \boldsymbol{a}_n)\boldsymbol{a}_n\boldsymbol{a}_n^T\right)^{-1} \tag{20}$$

Proof: See Appendix B. ■

As is well known, the CRB places a lower bound on the estimation error of any unbiased estimator [35] and is asymptotically attained by the ML estimator. More specifically, the covariance matrix of any unbiased estimate satisfies: $\mathrm{cov}(\hat{\boldsymbol{h}}) - \mathrm{CRB}(\boldsymbol{h}) \succeq \boldsymbol{0}$. Also, the variance of each component is bounded by the corresponding diagonal element of $\mathrm{CRB}(\boldsymbol{h})$, i.e., $\mathrm{var}(\hat{h}_i) \geq [\mathrm{CRB}(\boldsymbol{h})]_{ii}$.

We observe from (20) that the CRB matrix of $\boldsymbol{h}$ depends on the quantization thresholds $\boldsymbol{\tau}$ as well as the matrix $\boldsymbol{A}$ which is constructed from training sequences $\boldsymbol{X}$ (cf. (8)). Naturally, we wish to optimize $\boldsymbol{\tau}$ and $\boldsymbol{A}$ (i.e. $\boldsymbol{X}$) by minimizing the trace of the CRB matrix, i.e. the overall estimation error asymptotically achieved by the ML estimator. The optimization therefore can be formulated as follows

$$\begin{aligned} \min_{\boldsymbol{X},\boldsymbol{\tau}} \quad & \mathrm{tr}\{\mathrm{CRB}(\boldsymbol{h})\} = \mathrm{tr}\left\{\left(\sum_{n=1}^{N} g(\tau_n, \boldsymbol{a}_n)\boldsymbol{a}_n\boldsymbol{a}_n^T\right)^{-1}\right\} \\ \text{s.t.} \quad & \boldsymbol{A} = \boldsymbol{I}_M \otimes \widetilde{\boldsymbol{A}} \\ & \widetilde{\boldsymbol{A}} = \begin{bmatrix} \Re(\boldsymbol{X}) & \Im(\boldsymbol{X}) \\ -\Im(\boldsymbol{X}) & \Re(\boldsymbol{X}) \end{bmatrix}^T \\ & \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P \end{aligned} \tag{21}$$

where $\mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P$ is a transmit power constraint imposed on the pilot signals. Such an optimization is examined in the following section.

IV. Optimal Design and Performance Analysis

A. Optimal Quantization Thresholds and Pilot Sequences

Before proceeding, we first examine the property of the function $g(\tau_n, \boldsymbol{a}_n)$ defined in (19). Clearly, from Fig. 1, we see that $g(\tau_n, \boldsymbol{a}_n)$ is a positive and symmetric function attaining its maximum when $\tau_n = \boldsymbol{a}_n^T\boldsymbol{h}$ (a rigorous proof for this result is also available in [32]). Hence the optimal quantization thresholds conditional on $\boldsymbol{A}$ are given by

$$\tau_n^\star = \boldsymbol{a}_n^T\boldsymbol{h}, \quad \forall n \in \{1, \ldots, N\} \tag{22}$$

The result (22) comes by noting that

$$\sum_{n=1}^{N} g_n(\tau_n^\star, \boldsymbol{a}_n)\boldsymbol{a}_n\boldsymbol{a}_n^T - \sum_{n=1}^{N} g_n(\tau_n, \boldsymbol{a}_n)\boldsymbol{a}_n\boldsymbol{a}_n^T \succeq \boldsymbol{0} \tag{23}$$

and resorting to the convexity of $\mathrm{tr}(\boldsymbol{P}^{-1})$ over the set of positive definite matrix, i.e. for any $\boldsymbol{P} \succ \boldsymbol{0}$, $\boldsymbol{Q} \succ \boldsymbol{0}$, and $\boldsymbol{P} - \boldsymbol{Q} \succeq \boldsymbol{0}$, the following inequality $\mathrm{tr}(\boldsymbol{P}^{-1}) \leq \mathrm{tr}(\boldsymbol{Q}^{-1})$ holds (see [36]).

It can be observed that the optimal choice of the quantization threshold τ_n is dependent on the unknown channel $\boldsymbol{h}$. To facilitate our analysis, we, for the time being, suppose $\boldsymbol{h}$ is known.

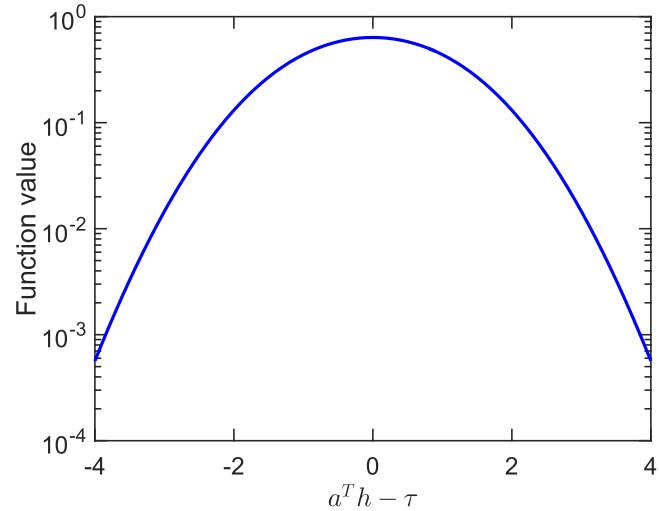


Fig. 1. The function value of $g(\tau_n, \boldsymbol{a}_n)$ vs. $(\boldsymbol{a}_n^T\boldsymbol{h} - \tau_n)$, where $\sigma_w^2 = 1$.

Substituting (22) into (21) and noting that

$$g(\tau_n^\star, \boldsymbol{a}_n) = \frac{f_w^2(0)}{F_w(0)(1 - F_w(0))} = \frac{2}{\pi\sigma_w^2} \quad \forall n \tag{24}$$

the optimization (21) reduces to

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & \mathrm{tr}\left\{\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\right\} \\ \text{s.t.} \quad & \boldsymbol{A} = \boldsymbol{I}_M \otimes \widetilde{\boldsymbol{A}} \\ & \widetilde{\boldsymbol{A}} = \begin{bmatrix} \Re(\boldsymbol{X}) & \Im(\boldsymbol{X}) \\ -\Im(\boldsymbol{X}) & \Re(\boldsymbol{X}) \end{bmatrix}^T \\ & \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P \end{aligned} \tag{25}$$

By exploiting the block diagonal structure of $\boldsymbol{A}$, the above optimization problem can be further simplified as

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & \mathrm{tr}\left\{\left(\widetilde{\boldsymbol{A}}^T\widetilde{\boldsymbol{A}}\right)^{-1}\right\} \\ \text{s.t.} \quad & \widetilde{\boldsymbol{A}} = \begin{bmatrix} \Re(\boldsymbol{X}) & \Im(\boldsymbol{X}) \\ -\Im(\boldsymbol{X}) & \Re(\boldsymbol{X}) \end{bmatrix}^T \\ & \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P \end{aligned} \tag{26}$$

It can be shown that the above optimization problem is essentially equivalent to the following optimization

$$\begin{aligned} \min_{\boldsymbol{X}} \quad & \mathrm{tr}\left\{\left(\boldsymbol{X}\boldsymbol{X}^H\right)^{-1}\right\} \\ \text{s.t.} \quad & \mathrm{tr}(\boldsymbol{X}\boldsymbol{X}^H) \leq P \end{aligned} \tag{27}$$

whose optimal solution has been studied in [37] and given as

$$\boldsymbol{X}\boldsymbol{X}^H = (P/K)\boldsymbol{I}_K \tag{28}$$

This result means that, when thresholds are optimally selected, pilot sequences associated with different users should be orthogonal to each other in order to minimize the channel estimation error.

B. Performance Analysis

We now investigate the estimation performance when the optimal thresholds are employed, and its comparison with the performance attained by a conventional massive MIMO system which assumes infinite-precision ADCs. Substituting the optimal thresholds (22) into the CRB matrix (20), we have

$$\mathrm{CRB}_{\mathrm{OQ}}(\boldsymbol{h}) = \frac{\pi\sigma_w^2}{2}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1} \tag{29}$$

where for clarity, we use the subscript OQ to represent the estimation scheme using optimal quantization thresholds. On the other hand, when the unquantized observations $\boldsymbol{y}$ are available, it can be readily verified that the CRB matrix is given as

$$\mathrm{CRB}_{\mathrm{NQ}}(\boldsymbol{h}) = \sigma_w^2\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1} \tag{30}$$

where we use the subscript NQ to represent the scheme which has access to the unquantized observations. Comparing (29) with (30), we can see that if optimal thresholds are employed, then using one-bit ADCs for channel estimation incurs only a mild performance loss relative to using infinite-precision ADCs, with the CRB increasing by only a factor of $\pi/2$, i.e.

$$\mathrm{CRB}_{\mathrm{OQ}}(\boldsymbol{h}) = \frac{\pi}{2}\mathrm{CRB}_{\mathrm{NQ}}(\boldsymbol{h}) \tag{31}$$

We also take a glimpse of the estimation performance as the thresholds deviate from their optimal values. For simplicity, let $\tau_n = \tau_n^\star + \delta = \boldsymbol{a}_n^T\boldsymbol{h} + \delta, \forall n$, in which case the CRB matrix is given by

$$\mathrm{CRB}_{\mathrm{Q}}(\boldsymbol{h}) = \frac{F_w(\delta)(1-F_w(\delta))}{f_w^2(\delta)}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1} \tag{32}$$

Since $(F_w(\delta)(1-F_w(\delta)))/f_w^2(\delta)$ is the reciprocal of $g(\tau_n, \boldsymbol{a}_n)$, from Fig. 1, we know that the function value $(F_w(\delta)(1-F_w(\delta)))/f_w^2(\delta)$ grows exponentially as $|\delta|$ increases. This indicates that a deviation of the thresholds from their optimal values results in a substantial performance loss.

In summary, the above results have important implications for the design of one-bit massive MIMO systems. It points out that a careful choice of quantization thresholds can help improve the estimation performance significantly, and help achieve an estimation accuracy close to an ideal estimator which has access to the raw observations $\boldsymbol{y}$. Note that the $\pi/2$ estimation performance loss result was also reported in previous studies, e.g. [27], [28]. Nevertheless, this result was obtained in a relatively simple scenario in which estimation of a scalar deterministic parameter was considered.

A problem with the above result is that the optimal thresholds $\boldsymbol{\tau}$ are functions of $\boldsymbol{h}$, as described in (22). Since $\boldsymbol{h}$ is unknown and to be estimated, the optimal thresholds $\boldsymbol{\tau}$ are also unknown. To address this difficulty, we, in the following, introduce an AQ scheme by which the thresholds are dynamically adjusted from one iteration to another, and a RQ scheme which randomly generates a set of nonidentical thresholds based on some statistical prior knowledge of the channel.

Algorithm 2: Adaptive Quantization Scheme.

1. Select an initial quantization threshold $\boldsymbol{\tau}^{(0)}$ and the maximum number of iterations $k_{\max}$.
2. At iteration $k = 1, 2, \ldots$: Based on $\boldsymbol{y}^{(k)}$ and $\boldsymbol{\tau}^{(k)}$, calculate the new binary data $\boldsymbol{b}^{(k)} = \mathrm{sgn}(\boldsymbol{y}^{(k)} - \boldsymbol{\tau}^{(k)})$.
3. Compute a new estimate of $\boldsymbol{h}$, $\hat{\boldsymbol{h}}^{(k)}$, via (33).
4. Calculate new thresholds according to $\boldsymbol{\tau}^{(k+1)} = \boldsymbol{A}\hat{\boldsymbol{h}}^{(k)}$.
5. Go to Step 2 if $k < k_{\max}$.

V. Practical Threshold Design Strategies

A. Adaptive Quantization

One strategy to overcome the above difficulty is to use an iterative algorithm in which the thresholds are iteratively refined based on the previous estimate of $\boldsymbol{h}$.

Suppose we have $\boldsymbol{y}^{(k)} = \boldsymbol{A}\boldsymbol{h} + \boldsymbol{w}^{(k)}$ at each iteration. At iteration k, we use the current quantization thresholds $\boldsymbol{\tau}^{(k)}$ to generate the one-bit observation data $\boldsymbol{b}^{(k)}$. Then a new estimate $\hat{\boldsymbol{h}}^{(k)}$ is obtained from the ML estimator (16). This estimate is then plugged in (22) to obtain updated quantization thresholds, i.e. $\boldsymbol{\tau}^{(k+1)} = \boldsymbol{A}\hat{\boldsymbol{h}}^{(k)}$, for subsequent iteration. When computing the ML estimate $\hat{\boldsymbol{h}}^{(k)}$, not only the quantized data from the current iteration but also from all previous iterations can be used. Since the data are independent across different iterations, the ML estimator (16) can be easily adapted as follows

$$\hat{\boldsymbol{h}}^{(k)} = \arg\max_{\boldsymbol{h}} \tilde{\mathcal{L}}^{(k)}(\boldsymbol{h}) \tag{33}$$

where

$$\tilde{\mathcal{L}}^{(k)}(\boldsymbol{h}) \triangleq \sum_{i=1}^{k} \log p(\boldsymbol{b}^{(i)}; \boldsymbol{h}, \boldsymbol{\tau}^{(i)}) \tag{34}$$

For clarity, we summarize the AQ scheme as Algorithm 2.

We have the following result regarding the convergence of the proposed AQ scheme.

Theorem 2: As the iterative process evolves, the estimated thresholds of the AQ scheme converge to the optimal thresholds, i.e.

$$\boldsymbol{\tau}^{(k)} \xrightarrow{k\to\infty} \boldsymbol{\tau}^\star \tag{35}$$

Proof: See Appendix C. ■

Theorem 2 indicates that as the iterative process evolves, the quantization thresholds of the AQ scheme will converge to the optimal values. In Theorem 2, we assume that the observations $\{\boldsymbol{y}^{(k)}\}$ are independent across iterations. For the case where the observations $\{\boldsymbol{y}^{(k)}\}$ remain unchanged across the iterative process, i.e. $\boldsymbol{y}^{(k)} = \boldsymbol{y}, \forall k$, our experimental results suggest that the quantization thresholds of the AQ scheme will come close to the optimal values, although its convergence cannot be rigorously proved.

In the following, we discuss several ways of implementing the AQ scheme in practical systems. Clearly, the AQ scheme

can be employed to estimate channels that are unchanged or slowly time-varying across a number of consecutive frames, in which case the quantization thresholds can be updated at each frame based on the channel estimate obtained from the previous frame. Note that the channel may indeed remain static across multiple frames in practice. For example, for the scenario where the relative speeds between the mobile terminals and the base station are slow, say, 2 meters per second, the channel coherence time could be up to tens of milliseconds, more precisely, about 60 milliseconds if the carrier frequency is set to 1GHz, according to the Clarke's model [38]. Suppose the time duration of each frame is 10 milliseconds which is a typical value for practical LTE systems. Then the channel will remain unaltered across 6 consecutive frames.

From the above discussion, we see that our AQ scheme can be easily implemented for slowly time-varying channels. On the other hand, for the scenario where the channel varies across different frames, i.e. the channel only keeps constant over each transmission frame, there are two different ways to implement the AQ scheme. The first option is to devise a structured pilot matrix $\boldsymbol{X}$ that is amiable for adaptive processing. Suppose the length of the training sequence, L, is an integer multiple of the number of users K, i.e. $L = dK$, in which d is an integer. In this case, the pilot matrix $\boldsymbol{X}$ can be devised as a concatenation of a series of identity matrix, i.e.

$$\boldsymbol{X} = \mathbf{1}_d^T \otimes \boldsymbol{I}_K \tag{36}$$

where $\mathbf{1}_d$ denotes a d-dimensional column vector with its entries equal to one. Clearly, such a pilot matrix indicates that at each time instant, only one user is activated to transmit a pilot symbol. Besides, each user, say user k, is activated periodically (every K time instants) for pilot symbol transmission. This pilot structure allows the channel estimation to be decomposed into K independent and parallel tasks, with each user's channel estimated separately. We now discuss how to implement the AQ for each individual task. After receiving a pilot symbol from each user, say, the kth user, we can obtain a new estimate of the kth user's channel, based on which we can update the quantization thresholds for the kth user. The refined thresholds are then used for quantization when the next pilot symbol from the kth user arrives. For this scheme, the only issue is to make sure that for each user, the quantization thresholds are calculated and updated before a new pilot symbol from this user arrives, which is possible since there is a time interval between the current and next pilot symbol transmissions. We see that with this carefully devised pilot matrix, the AQ scheme can be easily implemented.

Another way of implementing the AQ scheme, without resorting to the assumption of a static channel across multiple frames, is to use a number of sample-and-hold (S/H) circuits to sample the analog received signals and to store their values for subsequent offline processing. Specifically, each antenna/RF chain is followed by $2L$ S/H circuits which are equally divided into two groups, with one group of S/H circuits storing the real components of the received signal, and the other group storing the imaginary components of the received signal (see Fig. 2).

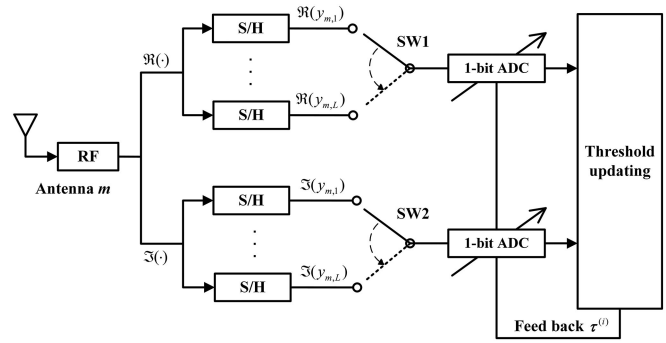


Fig. 2. An off-line implementation of the AQ scheme.

In each group, the L S/H circuits are used to store the L pilot symbols in a sequential manner, with the lth S/H circuit holding the lth received pilot symbol, i.e. $\Re(y_{m,l})$ (or $\Im(y_{m,l})$). Also, to avoid using a one-bit ADC for each S/H circuit, a switch can be used to connect a single one-bit ADC with multiple S/H circuits. Once the analog signals $\boldsymbol{y}$ have been stored, the AQ scheme can be implemented in an offline manner. Clearly, this offline approach can be implemented on a single frame basis, and thus no longer requires that the channel keeps constant across multiple transmission frames. Although using S/H circuits makes the system more complex, hopefully, S/H circuits are standard, elementary analog devices with low cost and low power consumption. Also, the number of S/H circuits per antenna does not have to be $2L$ and can be reduced if needed. Specifically, we can use $2L_1$ S/H circuits to sample and hold the received signal of the first L_1 pilot symbols, with the remaining $L - L_1$ pilot signals simply quantized using a fixed threshold. The proposed AQ scheme can be easily adapted to handle such a case. In doing this way, a tradeoff between the complexity and the performance can be achieved.

B. Random Quantization

We propose a random quantization (RQ) scheme that does not involve any iterative procedure and is simple to implement. The idea is to randomly generate a set of non-identical thresholds based on the statistical knowledge of $\boldsymbol{h}$, with the hope that some of the thresholds are close to the unknown optimal thresholds. For example, suppose each entry of $\boldsymbol{h}$ follows a Gaussian distribution with zero mean and variance σ_h^2. Note that different entries of $\boldsymbol{h}$ may have different variances due to the reason that they may correspond to different users. Nevertheless, we assume the same variance for all entries for simplicity. We randomly generate N different realizations of $\boldsymbol{h}$, denoted as $\{\tilde{\boldsymbol{h}}_n\}$, following this known distribution. The N quantization thresholds are then devised according to the optimality condition (22)

$$\tau_n = \boldsymbol{a}_n^T \tilde{\boldsymbol{h}}_n, \quad \forall n \in \{1, \ldots, N\} \tag{37}$$

Our simulation results suggest that this RQ scheme can achieve a considerable performance improvement over the conventional fixed quantization scheme which uses a fixed (typically zero)

threshold. The reason is that the thresholds produced by (37) are more likely to be close to their optimal values.

Remarks: Both the RQ scheme and the well-known dithering technique [39], [40] employ random thresholds. However, they use different rules to generate the thresholds. For the RQ scheme, the thresholds are randomly generated according to the optimality condition (37). While for the dithering technique, it is usually assumed that the dithering signal is an i.i.d. random process independent of the raw signal, and follows some distribution (e.g. a Gaussian distribution) with an appropriate power level [41], i.e.

$$\tau_n \sim \mathcal{N}(0, \sigma^2), \quad \forall n \in \{1, \ldots, N\}$$

We see that the thresholds $\{\tau_n\}$ generated by these two schemes generally have different distributions.

Note that the RQ scheme only needs to run the gradient-based search algorithm once to obtain a ML estimate of the channel, whereas the AQ scheme requires performing the search algorithms multiple times in order to search for the optimal thresholds. This means the AQ scheme has a complexity multiple times higher than the RQ scheme. Although the AQ scheme presents a significant performance improvement over the RQ scheme, this indeed comes at the expense of a much higher computational complexity.

C. Estimation of the Noise Variance and the Channel Variance

For the RQ scheme, we assume that the variance of the channel is known *a priori*. Also, in the ML estimator, the knowledge of the noise variance is assumed available. In the following, we discuss how to estimate the noise variance and the channel variance based on one-bit quantized data. Notice that the estimation of the noise variance and the estimation of the channel variance can be conducted in two separate stages. In practice, the noise variance can be estimated from the received signals collected during an idle period when users do not transmit signals. In this case, the signal model (9) becomes

$$\boldsymbol{y} = \boldsymbol{w} \tag{38}$$

and the objective now is to estimate the noise variance σ_w^2 based on quantized data

$$\boldsymbol{b} = \text{sgn}(\boldsymbol{y} - \boldsymbol{\tau}) = \text{sgn}(\boldsymbol{w} - \boldsymbol{\tau}) \tag{39}$$

Such a problem has been studied in previous works, e.g. [42], and it was shown that σ_w^2 can be accurately estimated from one-bit quantized data $\boldsymbol{b}$ via a ML estimator. To see this, write $w_n = \sigma_w v_n$, where v_n denotes a random Gaussian variable with zero mean and unit variance. The corresponding probability mass function of b_n can be written as

$$p(b_n; \sigma_w) = (1 - F_v(\tau_n/\sigma_w))^{\frac{1+b_n}{2}} F_v(\tau_n/\sigma_w)^{\frac{1-b_n}{2}} \tag{40}$$

where $F_v(x)$ denotes the CDF of v_n. The log-likelihood function of σ_w can therefore be expressed as

$$\begin{aligned}\mathcal{L}(\sigma_w) &= \log[p(b_1, \ldots, b_N; \sigma_w)] \\ &= \sum_{n=1}^{N} \Big\{ \frac{1+b_n}{2} \log[1 - F_v(\tau_n/\sigma_w)] \\ &\quad + \frac{1-b_n}{2} \log[F_v(\tau_n/\sigma_w)] \Big\}\end{aligned} \tag{41}$$

from which a ML estimate of σ_w can be obtained. Specifically, if a common threshold is used, i.e. $\tau_n = \tau, \forall n$, then the ML estimate of σ_w has a closed form solution given as

$$\hat{\sigma}_w = \frac{\tau}{F_v^{-1}\left(\left(1 - \left(\sum_{n=1}^{N} b_n\right)/N\right)/2\right)}$$

where $F_v^{-1}(\cdot)$ denotes the inverse of the CDF.

After an estimate of the noise variance is obtained, we are now ready to estimate the channel variance. For simplicity, we consider a single-user case. The extension to the multi-user case is straightforward by letting users successively transmit their pilot symbols. To estimate the channel variance, let the user transmit a single pilot symbol $x = 1$. The received signal model (9) can be simplified as

$$\boldsymbol{y} = \boldsymbol{h} + \boldsymbol{w}$$

Note that here $\boldsymbol{h}$ is the channel vector associated with that user. Assume $h_n \sim \mathcal{N}(0, \sigma_h^2)$. Our objective is to estimate the channel variance σ_h^2 based on the one-bit quantized data

$$\boldsymbol{b} = \text{sgn}(\boldsymbol{y} - \boldsymbol{\tau})$$

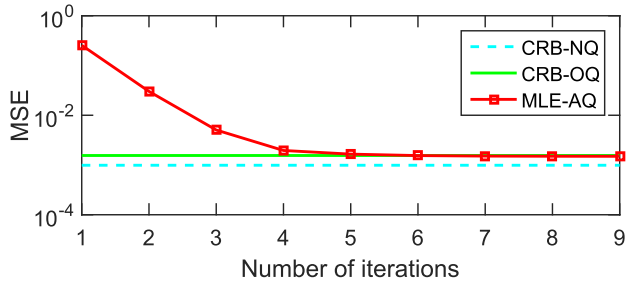
Since $\boldsymbol{h}$ and $\boldsymbol{w}$ are mutually independent, clearly we have $y_n = h_n + w_n \sim \mathcal{N}(0, \sigma_y^2)$, where $\sigma_y^2 = \sigma_h^2 + \sigma_w^2$. Similar to the way of estimating the noise variance, the variance σ_y^2 can be estimated from one-bit quantized data $\boldsymbol{b}$. The channel variance σ_h^2 can then be obtained as $\hat{\sigma}_h^2 = \hat{\sigma}_y^2 - \hat{\sigma}_w^2$.

VI. Simulation Results

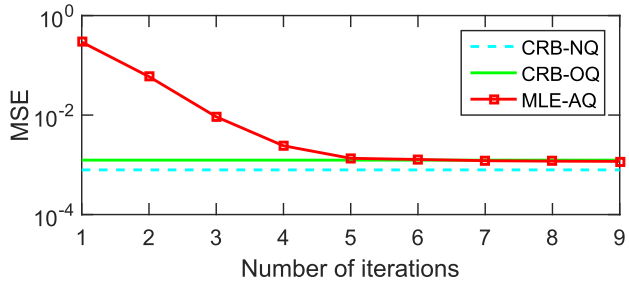
We now carry out experiments to corroborate our theoretical analysis and to illustrate the performance of our proposed one-bit quantization schemes, i.e. the AQ and the RQ schemes. We compare our schemes with the conventional fixed quantization (FQ) scheme which employs a fixed zero threshold for one-bit quantization, and a no-quantization scheme (referred to as NQ) which uses the original unquantized data for channel estimation. For the NQ scheme, it can be easily verified that its ML estimate is given by

$$\hat{\boldsymbol{h}} = (\boldsymbol{A}^T \boldsymbol{A})^{-1} \boldsymbol{A}^T \boldsymbol{y} \tag{42}$$

and its associated CRB is given by (30). For other schemes such as the RQ and the FQ, although a close-form expression is not available, the ML estimate can be obtained by solving the convex optimization (16). In our simulations, we assume independent and identically distributed (i.i.d.) rayleigh fading channels, i.e. all elements of $\boldsymbol{H}$ follow a circularly symmetric



(a) $K = 8$, $L = 32$ and SNR $= 15$ dB.



(b) $K = 16$, $L = 40$ and SNR $= 15$ dB.

Fig. 3. MSEs of the AQ scheme as a function of the number of iterations. Here, CRB-OQ represents the CRB attained by using optimal quantization thresholds.

complex Gaussian distribution with zero mean and unit variance. Training sequences $\boldsymbol{X}$ which satisfy (28) are randomly generated. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{P}{KL\sigma_w^2} \tag{43}$$

We first examine the estimation performance of our proposed AQ scheme which adaptively adjusts the thresholds based on the previous estimate of the channel. To facilitate our comparison, we assume that observations remain unchanged during the iterative process. Fig. 3 plots the mean-squared errors (MSEs) vs. the number of iterations for the AQ scheme, where we set $M = 64$, $K = 8$, $L = 32$ for Fig. 3(a) and $M = 64$, $K = 16$, $L = 40$ for Fig. 3(b). The SNR is set to 15 dB. The MSE is calculated as

$$\text{MSE} = \frac{1}{KM}\|\boldsymbol{H} - \hat{\boldsymbol{H}}\|_F^2 \tag{44}$$

To better illustrate the effectiveness of the AQ scheme, we also include the CRB results in Fig. 3, in which the CRB-OQ and the CRB-NQ are respectively given by (29) and (30). From Fig. 3, we see that our proposed AQ scheme approaches the CRB-OQ within only a few (say, 5) iterations, and achieves performance close to the CRB-NQ. This result demonstrates the effectiveness of the AQ scheme in searching for the optimal thresholds. In the rest of our simulations, we set the maximum number of iterations, $k_{\max}$, equal to 5 for the AQ scheme.

We now compare the estimation performance of different schemes. Fig. 4 plots the MSEs of respective schemes as a function of the number of pilot symbols, L, where we set $M = 64$, $K = 8$ and SNR $= 15$ dB. The corresponding CRBs

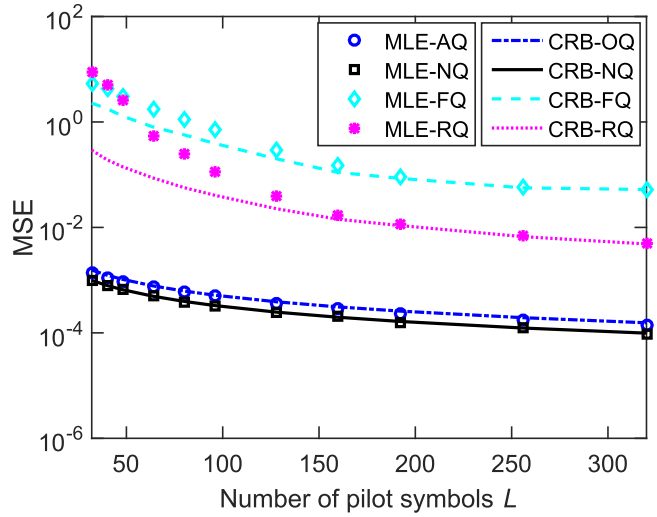


Fig. 4. MSEs vs. number of pilot symbols, where $K = 8$ and SNR $= 15$ dB. Here, CRB-OQ represents the CRB attained by using optimal quantization thresholds.

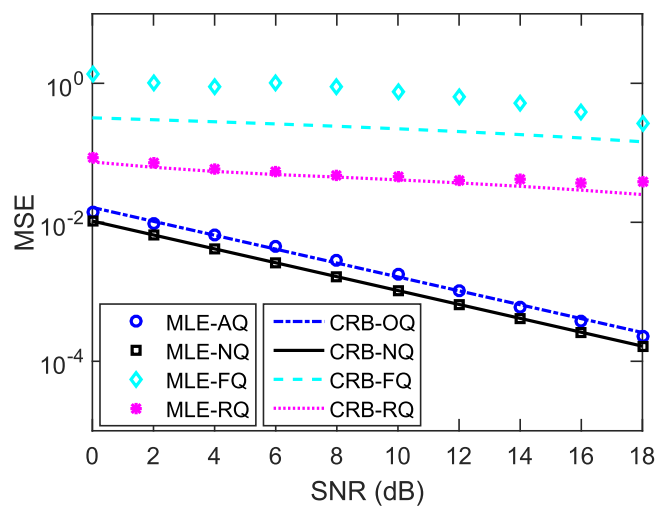


Fig. 5. MSEs vs. SNR(dB), where $K = 8$ and $L = 96$.

of these schemes are also included. Note that the CRBs for the FQ and the RQ schemes can be obtained by substituting the thresholds into (20). Results are averaged over 10^3 independent runs, with the channel and the pilot sequences randomly generated for each run. From Fig. 4, we can see that the proposed AQ scheme outperforms the FQ and RQ schemes by a big margin. This result corroborates our analysis that carefully devised quantization thresholds helps achieve a substantial performance improvement. In particular, the AQ scheme needs less than 30 pilot symbols to achieve a decent estimation accuracy with a MSE of 0.01, while the FQ and RQ schemes require a much larger number of pilot symbols to attain a same estimation accuracy. We also observe that the RQ scheme presents a clear performance advantage over the FQ scheme. The RQ requires about 100 symbols to achieve a MSE of 0.1, whereas the FQ needs about 250 pilot symbols to reach a same estimation accuracy. In Fig. 5, we plot the MSEs of respective schemes

TABLE I
AVERAGE RUN TIME (SEC) AND MSEs OF RESPECTIVE SCHEMES

Average Run Time	$L = 80$	$L = 96$	$L = 128$
AQ	60.64	71.68	94.88
FQ	49.12	51.20	58.40
RQ	41.60	44.96	49.92
MSE	$L = 80$	$L = 96$	$L = 128$
AQ	0.0007	0.0006	0.0004
FQ	1.0916	0.7271	0.2873
RQ	0.2459	0.1123	0.0399

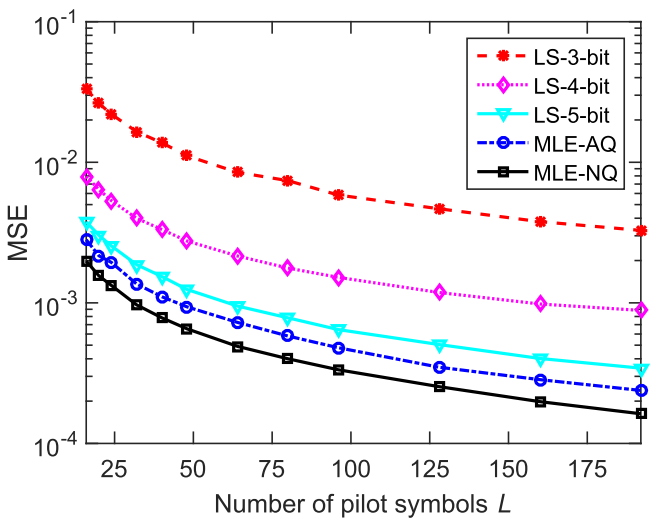


Fig. 6. MSEs of the AQ and the multi-bit uniform quantization scheme as a function of the number of pilot symbols.

under different SNRs, where we set $K = 8$ and $L = 96$. Similar conclusions can be made from Fig. 5.

Note that the performance advantage of the AQ scheme over the RQ and FQ schemes comes at the expense of a higher computational complexity. To show this, in Table I, the average run times of respective schemes are provided, where we set $M = 16$, $K = 8$ and SNR $= 15$ dB. We see that the average run time of the AQ scheme is about 1.5 times that of the RQ and the FQ schemes. Note that although the AQ scheme requires to perform the ML search algorithms multiple times (5 times for our experiments), the complexity for the last few searches can be substantially reduced since the previous estimate can be used as a good initialization point for subsequent searches.

We also conduct experiments to compare our proposed one-bit AQ scheme with a q-bit uniform quantization scheme, where we set $q = 3$, $q = 4$ and $q = 5$ in our experiments. For the multi-bit uniform quantization scheme, the quantization errors can be deemed as independent and identically distributed random noise, and a least squares (LS) method can be used to estimate the channel. Fig. 6 plots MSEs of the AQ scheme and the multi-bit uniform quantization scheme vs. the number of pilot symbols, where we set $K = 8$ and SNR $= 15$ dB. Our results show that the proposed AQ scheme achieves a higher estimation accuracy than the multi-bit uniform quantization scheme. This result is not unexpected since the proposed AQ scheme, due to its more prudent choice of the quantization thresholds, can achieve an estimation accuracy close to that of the ideal estimator which has access to the raw observations.

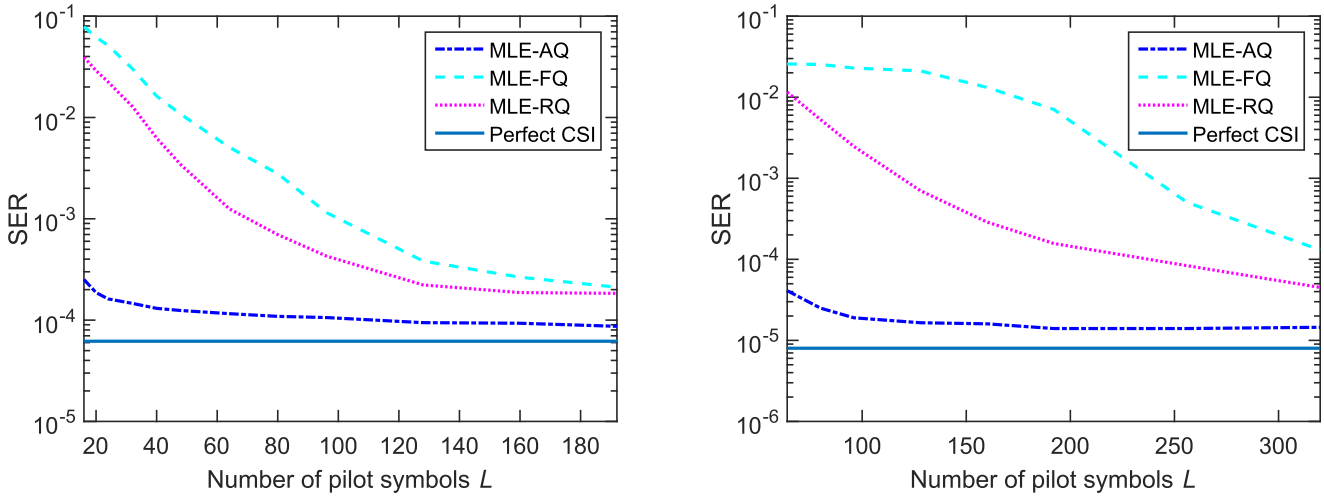
Next, we examine the effect of channel estimation accuracy on the symbol error rate (SER) performance. For each scheme, after the channel is estimated, a near maximum likelihood detector [12] developed for one-bit massive MIMO is adopted for symbol detection. For a fair comparison, in the symbol detection stage, the quantization thresholds are all set equal to zero, as assumed in [12]. In our experiments, QPSK symbols are transmitted by all users. Fig. 7 plots the SERs of respective schemes vs. the number of pilot symbols, where we set $K = 8$, $M = 64$, SNR $= 5$ dB in Fig. 7(a), and $K = 16$, $M = 128$, SNR $= 5$ dB in Fig. 7(b). Results are averaged over all K users. The SER performance obtained by assuming perfect channel knowledge is also included. It can be seen that the SER performance improves as the number of pilot symbols increases, which is expected since a more accurate channel estimate can be obtained when more pilot symbols are available for channel estimation. We also observe that the AQ scheme, using a moderate number of pilot symbols, can achieve SER performance close to that attained by assuming perfect channel knowledge. Moreover, the SER results, again, demonstrate the superiority of the RQ over the FQ scheme. In order to attain a same SER, say, 10^{-3}, the RQ requires about 60 and 120 pilot symbols for the two cases, whereas the FQ requires about 100 and 240 pilot symbols, respectively.

In Fig. 8, the achievable rates of respective schemes vs. the number of pilot symbols are depicted, where we set $K = 8$, $M = 64$, SNR $= 5$ dB in Fig. 8(a), and $K = 16$, $M = 128$, SNR $= 5$ dB, in Fig. 8(b), respectively. A tight lower bound on the achievable rate for the kth user is calculated as [43]

$$R_k \triangleq \log_2\left(1 + \frac{|E\left[s_k^*(t)\hat{s}_k(t)\right]|^2}{E\left[|\hat{s}_k(t)|^2\right] - |E\left[s_k^*(t)\hat{s}_k(t)\right]|^2}\right) \tag{45}$$

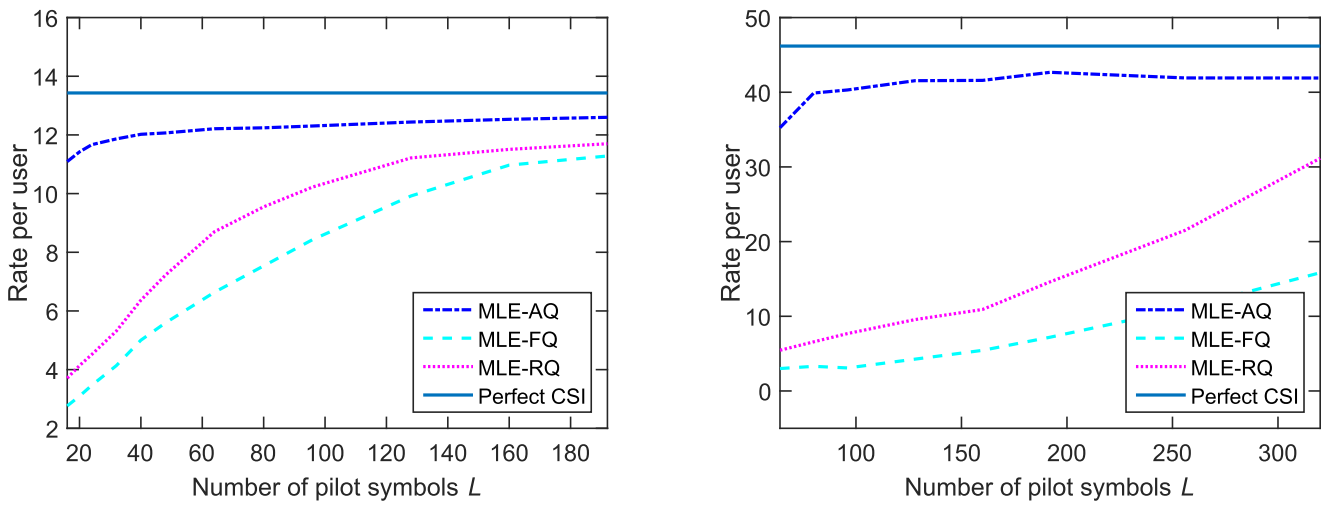
where $s_k(t)$ is the transmit symbol of the kth user at time t, $()^*$ denotes the conjugate, and $\hat{s}_k(t)$ is the estimated symbol of $s_k(t)$, which is obtained via the near maximum likelihood detector by using the channel estimated by respective schemes. The achievable rate we plotted is averaged over all K users. It can be seen that, even with a moderate number of pilot symbols (about 5 times the number of users), the AQ scheme can provide an achievable rate close to that of the perfect CSI case, whereas the achievable rates attained by the other two schemes are far below the level of the AQ scheme.

To illustrate the difference between the RQ scheme and the dithering scheme. We consider a scenario where the elements of $\boldsymbol{h}$ follow a circularly symmetric complex Gaussian distribution with a nonzero mean ($\mu_h = 5$) and unit variance. For the dithering scheme, the dithering signal is assumed to be an i.i.d. complex Gaussian process independent of the raw signal with zero mean and unit variance. Fig. 9 plots MSEs of the RQ and the dither scheme as a function of the number of



(a) $K = 8$, $M = 64$ and SNR $= 5$dB. (b) $K = 16$, $M = 128$ and SNR $= 5$dB

Fig. 7. SERs vs. number of pilot symbols.



(a) $K = 8$, $M = 64$ and SNR $= 5$dB. (b) $K = 16$, $M = 128$ and SNR $= 5$dB

Fig. 8. Achievable rates vs. number of pilot symbols.

pilot symbols, where we set $K = 8$ and SNR $= 15$ dB. Results are averaged over 10^3 independent runs. This result shows that the RQ presents a clear performance advantage over the dithering technique, which implies that the thresholds used by the RQ scheme are in average better than those of the dithering technique.

VII. Conclusion

We studied the problem of one-bit quantization design and channel estimation for uplink multiuser massive MIMO systems. Specifically, based on the derived CRB matrix, we examined the impact of quantization thresholds on the channel estimation performance. Our theoretical analysis revealed that using one-bit ADCs can achieve an estimation error close to that attained by using infinite-precision ADCs, given that the quantization thresholds are optimally set. Our analysis also suggested that the optimal quantization thresholds are dependent on the unknown channel parameters. We developed two practical quantization design schemes, namely, an adaptive quantization scheme which adaptively adjusts the thresholds, and a random quantization scheme which randomly generates a set of thresholds based on some statistical prior knowledge of the channel. Simulation results showed that the proposed quantization schemes achieved a significant performance improvement over the fixed quantization scheme that uses a fixed (typically zero) quantization threshold, and thus can help substantially reduce the training overhead in order to attain a same estimation accuracy target.

Appendix A
Proof of Concavity of the Log-Likelihood Function (15)

It can be easily verified that $f_w(\boldsymbol{a}_n^T\boldsymbol{h} - \tau_n)$ is log-concave in $\boldsymbol{h}$ since the Hessian matrix of $\log f_w(\boldsymbol{a}_n^T\boldsymbol{h} - \tau_n)$, which is

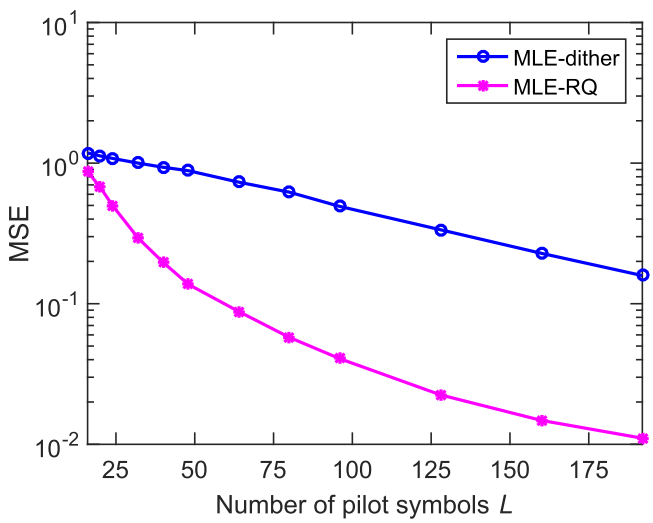


Fig. 9. MSEs of the RQ scheme and the dithering scheme as a function of the number of pilot symbols.

given by

$$\frac{\partial^2 \log f_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)}{\partial \boldsymbol{h} \partial \boldsymbol{h}^T} = -\frac{\boldsymbol{a}_n \boldsymbol{a}_n^T}{\sigma_w^2} \tag{46}$$

is negative semidefinite. Consequently the corresponding cumulative density function (CDF) and complementary CDF (CCDF), which are integrals of the log-concave function $f_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)$ over convex sets $(-\infty, \tau_n)$ and (τ_n, ∞) respectively, are also log-concave, and their logarithms are concave. Since summation preserves concavity, $\mathcal{L}(\boldsymbol{h})$ is a concave function of $\boldsymbol{h}$.

Appendix B
Proof of Theorem 1

Define a new variable $z_n \triangleq \boldsymbol{a}_n^T \boldsymbol{h}$ and define

$$l(z_n) \triangleq \frac{1-b_n}{2} \log[1 - F_w(z_n - \tau_n)] + \frac{1+b_n}{2} \log[F_w(z_n - \tau_n)]. \tag{47}$$

The first and second-order derivative of $\mathcal{L}(\boldsymbol{h})$ are given by

$$\frac{\partial \mathcal{L}(\boldsymbol{h})}{\partial \boldsymbol{h}} = \sum_{n=1}^{N} \frac{\partial l(z_n)}{\partial z_n} \frac{\partial z_n}{\partial \boldsymbol{h}} = \sum_{n=1}^{N} \frac{\partial l(z_n)}{\partial z_n} \boldsymbol{a}_n \tag{48}$$

and

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{h})}{\partial \boldsymbol{h} \partial \boldsymbol{h}^T} = \sum_{n=1}^{N} \boldsymbol{a}_n \frac{\partial^2 l(z_n)}{\partial z_n^2} \frac{\partial z_n}{\partial \boldsymbol{h}^T} = \sum_{n=1}^{N} \frac{\partial^2 l(z_n)}{\partial z_n^2} \boldsymbol{a}_n \boldsymbol{a}_n^T. \tag{49}$$

where

$$\frac{\partial l(z_n)}{\partial z_n} = \frac{1-b_n}{2} \frac{f_w(z_n - \tau_n)}{F_w(z_n - \tau_n) - 1} + \frac{1+b_n}{2} \frac{f_w(z_n - \tau_n)}{F_w(z_n - \tau_n)} \tag{50}$$

and

$$\frac{\partial^2 l(z_n)}{\partial z_n^2} = \frac{1-b_n}{2} \left[\frac{f_w'(z_n - \tau_n)}{F_w(z_n - \tau_n) - 1} - \frac{f_w^2(z_n - \tau_n)}{(F_w(z_n - \tau_n) - 1)^2} \right] + \frac{1+b_n}{2} \cdot \left[\frac{f_w'(z_n - \tau_n)}{F_w(z_n - \tau_n)} - \frac{f_w^2(z_n - \tau_n)}{F_w^2(z_n - \tau_n)} \right] \tag{51}$$

where $f_w(x)$ denotes the probability density function (PDF) of w_n, and $f_w'(x) \triangleq \frac{\partial f_w(x)}{\partial x}$.

Therefore, the Fisher information matrix (FIM) of the estimation problem is given as

$$J(\boldsymbol{h}) = -E\left[\frac{\partial^2 \mathcal{L}(\boldsymbol{h})}{\partial \boldsymbol{h} \partial \boldsymbol{h}^T}\right] = -\sum_{n=1}^{N} E_{b_n}\left[\frac{\partial^2 l(z_n)}{\partial z_n^2}\right] \boldsymbol{a}_n \boldsymbol{a}_n^T \overset{(a)}{=} \sum_{n=1}^{N} \frac{f_w^2(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)}{F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n)(1 - F_w(\boldsymbol{a}_n^T \boldsymbol{h} - \tau_n))} \boldsymbol{a}_n \boldsymbol{a}_n^T \tag{52}$$

where $E_{b_n}[\cdot]$ denotes the expectation with respect to the distribution of b_n, and (a) follows from the fact that b_n is a binary random variable with $P(b_n = 1 | \tau_n, z_n) = F_w(z_n - \tau_n)$ and $P(b_n = -1 | \tau_n, z_n) = 1 - F_w(z_n - \tau_n)$. This completes the proof.

Appendix C
Proof of Theorem 2

From the estimation theory, we know that the ML estimate is consistent if the following regularity conditions are satisfied [35]: 1. the observations are i.i.d; 2. the first-order and second-order derivatives of the log-likelihood function are well defined; 3. the condition

$$E\left[\frac{\partial \log p(\boldsymbol{b}^{(k)}; \boldsymbol{h}, \boldsymbol{\tau}^{(k)})}{\partial \boldsymbol{h}}\right] = \boldsymbol{0} \tag{53}$$

holds. It can be easily verified that, for our problem, the first and second conditions hold valid. Now we check whether the third condition is satisfied. We have

$$E\left[\frac{\partial \log p(\boldsymbol{b}^{(k)}; \boldsymbol{h}, \boldsymbol{\tau}^{(k)})}{\partial \boldsymbol{h}}\right] = \sum_{n=1}^{N} E\left[\frac{\partial l(z_n)}{\partial z_n}\right] \boldsymbol{a}_n \tag{54}$$

where $z_n \triangleq \boldsymbol{a}_n^T \boldsymbol{h}$ and

$$l(z_n) \triangleq \frac{1-b_n^{(k)}}{2} \log[1 - F_w(z_n - \tau_n^{(k)})] + \frac{1+b_n^{(k)}}{2} \log[F_w(z_n - \tau_n^{(k)})]. \tag{55}$$

Due to the fact that $b_n^{(k)}$ is a binary random variable with $P(b_n^{(k)} = 1|\tau_n^{(k)}, z_n) = F_w(z_n - \tau_n^{(k)})$ and $P(b_n^{(k)} = -1|\tau_n^{(k)}, z_n) = 1 - F_w(z_n - \tau_n^{(k)})$, and

$$\frac{\partial l(z_n)}{\partial z_n} = \frac{1 - b_n^{(k)}}{2} \frac{f_w(z_n - \tau_n^{(k)})}{F_w(z_n - \tau_n^{(k)}) - 1} + \frac{1 + b_n^{(k)}}{2} \frac{f_w(z_n - \tau_n^{(k)})}{F_w(z_n - \tau_n^{(k)})}, \tag{56}$$

we have $E[\partial l(z_n)/\partial z_n] = 0$ so that the condition (53) holds.

Consequently, according to the asymptotic properties of the ML estimator, for any small $\epsilon > 0$ and $\varepsilon > 0$, we can find a sufficiently large m such that

$$P(\|\hat{\boldsymbol{h}}^{(k-1)} - \boldsymbol{h}\|_2 < \epsilon) > 1 - \varepsilon, \quad k \geq m \tag{57}$$

Since $\boldsymbol{\tau}^{(k)} = \boldsymbol{A}\hat{\boldsymbol{h}}^{(k-1)}$ and $\boldsymbol{\tau}^\star = \boldsymbol{A}\boldsymbol{h}$, we have

$$P(\|\boldsymbol{\tau}^{(k)} - \boldsymbol{\tau}^\star\|_2 < \rho\epsilon) > 1 - \varepsilon, \quad k \geq m \tag{58}$$

where $\rho \triangleq \|\boldsymbol{A}\|_2$ denotes the largest singular value, which means that we have $\boldsymbol{\tau}^{(k)} \overset{k\to\infty}{\longrightarrow} \boldsymbol{\tau}^\star$. The proof is completed here.

References

[1] F. Rusek *et al.*, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[2] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[3] S. Chen, S. Sun, Q. Gao, and X. Su, "Adaptive beamforming in TDD-based mobile communication systems: State of the art and 5G research directions," *IEEE Wireless Commun.*, vol. 23, no. 6, pp. 81–87, Dec. 2016.

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

[5] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 36–43, May 2014.

[6] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," unpublished paper, 2014. [Online]. Available: https://arxiv.org/abs/1404.7736

[7] L. Fan, S. Jin, C.-K. Wen, and H. Zhang, "Uplink achievable rate for massive MIMO systems with low-resolution ADC," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2186–2189, Dec. 2015.

[8] J. Zhang, L. Dai, S. Sun, and Z. Wang, "On the spectral efficiency of massive MIMO systems with low-resolution ADCs," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 842–845, May 2016.

[9] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *Proc. IEEE Int. Conf. Commun. Workshop*, London, U.K., 2015, pp. 1304–1309.

[10] S. Wang, Y. Li, and J. Wang, "Multiuser detection in massive spatial modulation MIMO with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2156–2168, Apr. 2015.

[11] C.-K. Wen, C.-J. Wang, S. Jin, K.-K. Wong, and P. Ting, "Bayes-optimal joint channel-and-data estimation for massive MIMO with low-precision ADCs," *IEEE Trans. Signal Process.*, vol. 64, no. 10, pp. 2541–2556, May 2016.

[12] J. Choi, J. Mo, and R. W. Heath, "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with one-bit ADCs," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[13] H. Wang, C.-K. Wen, and S. Jin, "Bayesian optimal data detector for mmWave OFDM system with low-resolution ADC," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1962–1979, Sep. 2017.

[14] H. He, C.-K. Wen, and S. Jin, "Bayesian optimal data detector for hybrid mmWave MIMO-OFDM systems with low-resolution ADCs," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 469–483, Jun. 2018.

[15] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5498–5512, Oct. 2015.

[16] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.

[17] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing–The large-scale array regime," *IEEE Trans. Inf. Theory*, vol. 59, no. 10, pp. 6441–6463, Oct. 2013.

[18] 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 J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 802–814, Oct. 2014.

[19] 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.

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

[21] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1935–1947, Mar. 2017.

[22] 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.

[23] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, Jun. 2018.

[24] D. Fan, F. Gao, G. Wang, Z. Zhong, and A. Nallanathan, "Angle domain signal processing-aided channel estimation for indoor 60-GHz TDD/FDD massive MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1948–1961, Sep. 2017.

[25] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 264–273, Feb. 2013.

[26] R. R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 773–786, Oct. 2014.

[27] H. C. Papadopoulos, G. W. Wornell, and A. V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.

[28] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, Mar. 2006.

[29] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks–Part II: Unknown probability density function," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2784–2796, Jul. 2006.

[30] J. Fang and H. Li, "Distributed adaptive quantization for wireless sensor networks: From delta modulation to maximum likelihood," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5246–5257, Oct. 2008.

[31] A. Mezghani, F. Antreich, and J. A. Nossek, "Multiple parameter estimation with quantized channel output," in *Proc. IEEE Int. ITG Workshop Smart Antennas*, Bremen, Germany, 2010, pp. 143–150.

[32] G. Zeitler, G. Kramer, and A. C. Singer, "Bayesian parameter estimation using single-bit dithered quantization," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2713–2726, Jun. 2012.

[33] R. Karlsson and F. Gustafsson, "Particle filtering for quantized sensor information," in *Proc. IEEE Eur. Signal Process. Conf.*, Antalya, Turkey, 2005, pp. 1–4.

[34] M. Stein, S. Bar, J. A. Nossek, and J. Tabrikian, "Performance analysis for pilot-based 1-bit channel estimation with unknown quantization threshold," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Shanghai, China, 2016, pp. 4353–4357.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing, Volume I: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

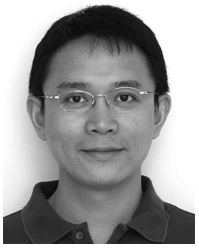
[36] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge University Press, 2004.
[37] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.
[38] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge University Press, 2005.
[39] L. G. Roberts, "Picture coding using pseudo-random noise," *IRE Trans. Inf. Theory*, vol. 8, no. 2, pp. 145–154, Feb. 1962.
[40] L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Trans. Commun. Technol.*, vol. 12, no. 4, pp. 162–165, Dec. 1964.
[41] R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, May 1993.
[42] J. Fang and H. Li, "Adaptive distributed estimation of signal power from one-bit quantized data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 4, pp. 1893–1905, Oct. 2010.
[43] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

Feiyu Wang received the B.Sc. degree from Zhejiang Ocean University, Zhoushan, China, in 2011, and the M.Sc. degree from Ningbo University, Ningbo, China, in 2014. Since August 2014, he has been working toward the Ph.D. degree with the University of Electronic Science and Technology of China, Chengdu, China. His current research interests include compressed sensing, tensor analysis, and statistical signal processing.

Jun Fang (M'08) received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 1998 and 2001, respectively, and the Ph.D. degree from the National University of Singapore, Singapore, in 2006, all in electrical engineering.

During 2006, he was a Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. From January 2007 to December 2010, he was a Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA. Since 2011, he has been with the University of Electronic of Science and Technology of China, Chengdu, China. His research interests include compressed sensing and sparse theory, massive MIMO/mmWave communications, and statistical inference.

Dr. Fang was the recipient the IEEE Jack Neubauer Memorial Award in 2013 for the best systems paper published in the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY. He serves as a Senior Associate Editor for the IEEE SIGNAL PROCESSING LETTERS.



Hongbin Li (M'99–SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, USA, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, the University of Florida, Gainesville, FL, USA. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, where he became a Professor in 2010. He was a Summer Visiting Faculty Member with the Air Force Research Laboratory in the summers of 2003, 2004, and 2009. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li was the recipient of the IEEE Jack Neubauer Memorial Award from the IEEE Vehicular Technology Society in 2013, the Outstanding Paper Award from the IEEE AFICON Conference in 2011, the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for excellence in research in 2001 from the Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He has been a member of the IEEE SPS Signal Processing Theory and Methods Technical Committee (TC) and the IEEE SPS Sensor Array and Multichannel TC, an Associate Editor for *Signal Processing* (Elsevier), the IEEE TRANSACTIONS ON SIGNAL PROCESSING, the IEEE SIGNAL PROCESSING LETTERS, and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, as well as a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities, including serving as the General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing Workshop, Hoboken, NJ, USA, June 17–20, 2012. He is a Member of Tau Beta Pi and Phi Kappa Phi.

Zhi Chen received the B. Eng, M. Eng., and Ph.D. degrees in electrical engineering from the University of Electronic Science and Technology of China, Chengdu, China, (UESTC) in 1997, 2000, and 2006, respectively. In April 2006, he joined the National Key Laboratory on Communications, UESTC, where he is currently a Professor. During 2010–2011, he was a Visiting Scholar with the University of California, Riverside, CA, USA. His current research interests include relay and cooperative communications, multiuser beamforming in cellular networks, interference coordination and cancellation, and THz communication. He was a Reviewer for various international journals and conferences, including the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, theIEEE TRANSACTIONS ON SIGNAL PROCESSING, etc.

Shaoqian Li (F'16) received the B.S.E. degree in communication technology from the Northwest Institute of Telecommunication, Xidian University, Xi'an, China, in 1982, and the M.S.E. degree in communication system from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 1984. He is currently a Professor, a Ph.D. supervisor, and the Director of the National Key Lab of Communication, UESTC, and a member of the National High Technology R&D Program (863 Program) Communications Group. His research interests include wireless communication theory, anti-interference technology for wireless communications, spread-spectrum and frequency hopping technology, and mobile and personal communications.