UC Irvine
UC Irvine Previously Published Works

Title
Channel Estimation and Performance Analysis of One-Bit Massive MIMO Systems

Permalink
https://escholarship.org/uc/item/6hn3t8fr

Journal
IEEE Transactions on Signal Processing, 65(15)

ISSN
1053-587X

Authors
Li, Yongzhi
Tao, Cheng
Seco-Granados, Gonzalo
et al.

Publication Date
2017

DOI
10.1109/tsp.2017.2706179

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Channel Estimation and Performance Analysis of One-Bit Massive MIMO Systems

Yongzhi Li, Cheng Tao, *Member, IEEE,* Gonzalo Seco-Granados, *Senior Member, IEEE,*
Amine Mezghani, A. Lee Swindlehurst, *Fellow, IEEE* and Liu Liu, *Member, IEEE,*

arXiv:1609.07427v3 [cs.IT] 21 Mar 2017

***Abstract*—This paper considers channel estimation and system performance for the uplink of a single-cell massive multiple-input multiple-output (MIMO) system. Each receive antenna of the base station (BS) is assumed to be equipped with a pair of one-bit analog-to-digital converters (ADCs) to quantize the real and imaginary part of the received signal. We first propose an approach for channel estimation that is applicable for both flat and frequency-selective fading, based on the Bussgang decomposition that reformulates the nonlinear quantizer as a linear function with identical first- and second-order statistics. The resulting channel estimator outperforms previously proposed approaches across all SNRs. We then derive closed-form expressions for the achievable rate in flat fading channels assuming low SNR and a large number of users for the maximal ratio and zero forcing receivers that takes channel estimation error due to both noise and one-bit quantization into account. The closed-form expressions in turn allow us to obtain insight into important system design issues such as optimal resource allocation, maximal sum spectral efficiency, overall energy efficiency, and number of antennas. Numerical results are presented to verify our analytical results and demonstrate the benefit of optimizing system performance accordingly.**

***Index Terms*—massive MIMO, large-scale antenna systems, one-bit ADCs, channel estimation, power allocation**

I. Introduction

Massive multiple-input multiple-output (MIMO) technology is considered to be a key component for 5G wireless communications systems, and has recently attracted considerable research interest. The main characteristic of massive MIMO is a base station (BS) array equipped with many (perhaps a hundred or more) antennas, which provides unprecedented spatial degrees of freedom for simultaneously serving multiple user terminals on the same time-frequency channel. It has been shown that, with channel state information (CSI) available at the BS, relatively simple signal processing techniques such as maximum-ratio combining (MRC) or zero-forcing (ZF) can be employed to reduce the noise and interference at the terminals, and can lead to improvements not only in spectral efficiency, but in energy efficiency as well [1]–[5].

In most work on massive MIMO, perfect hardware implementations with infinite resolution analog-to-digital converters (ADCs) are assumed. There has been limited prior work on the impact of non-ideal hardware on massive MIMO systems including [6]–[8], which studied imperfections such as phase-drifts and additive distortion, and showed that a massive number of antennas can mitigate these effects. In terms of hardware, perhaps the most important issue at the BS for massive MIMO is the power consumption of the ADCs, which grows exponentially with the number of quantization bits [9], and also grows with increased sampling rates due to wider bandwidths. For example, commercially available ADCs with resolutions of 12 to 16 bits consume on the order of several watts [10]. For massive MIMO configurations employing large antenna arrays and many ADCs, the cost and power consumption will be prohibitive, and alternative approaches are needed.

The use of low resolution (1-3 bits) ADCs is a potential solution to this problem [11]–[18]. In this paper, we focus on the case of simple one-bit ADCs, which consist of a simple comparator and consume negligible power (a few milliwatts). One-bit ADCs do not require automatic gain control and linear amplifiers, and hence the corresponding radio frequency (RF) chains can be implemented with very low cost and power consumption [17], [18]. It was shown in [11] that the capacity maximizing transmit signals for one-bit ADCs operating in single-input single-output (SISO) channels are discrete, unlike the infinite resolution case where a Gaussian codebook is optimal. In addition, [11] showed that MIMO capacity is not severely reduced by the coarse quantization at low signal-to-noise ratios (SNRs); in particular, the power penalty due to one-bit quantization is approximately equal to only $\pi/2$ (1.96dB) in the low SNR region [12]. On the other hand, at high SNRs one-bit quantization can produce a large capacity loss [19], but there is reason to believe that massive MIMO systems will operate at relatively low SNRs for improved energy efficiency, exploiting array gain to overcome the resulting distortion. This will be especially true as systems move to higher (e.g., millimeter wave) frequencies. In either

The research was supported in part by Beijing Nova Programme (Grant No. xx2016023), the NSFC project under grant No.61471027, the Research Fund of National Mobile Communications Research Laboratory, Southeast University (No.2014D05, No. 2017D01), and Beijing Natural Science Foundation project under grant No.4152043. A. Swindlehurst was supported by the National Science Foundation under Grant ECCS-1547155, and by the Technische Universität München Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under grant agreement No. 291763, and by the European Union under the Marie Curie COFUND Program.

Y. Li, C. Tao, and L. Liu are with the Institute of Broadband Wireless Mobile Communications, Beijing Jiaotong University, Beijing, 100044, China (email: liyongzhi@bjtu.edu.cn; chtao@bjtu.edu.cn; liuliu@bjtu.edu.cn).

G. Seco-Granados is with the Telecommunications and Systems Engineering Department, Universitat Autònoma de Barcelona, Barcelona 08193, Spain (e-mail: gonzalo.seco@uab.es).

A. Mezghani and A. Swindlehurst are with the Center for Pervasive Communications and Computing, University of California, Irvine, CA 92697 USA (e-mail: amezghan@uci.edu; swindle@uci.edu). A. Swindlehurst is also a Hans Fischer Senior Fellow of the Institute for Advanced Study at the Technical University of Munich.

Corresponding authors: chtao@bjtu.edu.cn, liuliu@bjtu.edu.cn.

case, the availability of accurate BS-side CSI is indispensable for exploiting the full potential of a massive MIMO system, and an important open question is how to reliably estimate the channel and decode the data symbols under one-bit output quantization.

Several recent papers have investigated channel estimation in massive MIMO with one-bit ADCs [20]–[28]. A millimeter wave MIMO system with one-bit ADCs was considered in [24], which proposed a modified expectation-maximum (EM) channel estimator that exploits the sparsity of such channels. In [25] a near maximum likelihood (nML) channel estimator and detector were proposed, and the nML approach was shown to improve estimation accuracy and better support higher order constellations than the EM estimators using one-bit ADCs. However, the channel estimators and the computed rates obtained in [24], [25] rely on either the maximum-likelihood algorithm or on an iterative algorithm with high complexity, and their performance is difficult to theoretically quantify. More recently, [28] considered a low complexity channel estimator and the corresponding achievable rate for one-bit massive MIMO systems over frequency-selective channels, using a model in which the number of channel taps goes to infinity, and the quantization noise is essentially modeled as independent, identically distributed (i.i.d.) noise.

In this paper, we focus on channel estimation and uplink performance for massive MIMO systems with one-bit ADCs. In contrast to [28], we derive more general quantization noise models that are applied separately for data detection and channel estimation. One essential and unique aspect of our derivation is that the spatial correlation between the elements of the quantizer output is taken into account, calculated using the *arcsine law*. Our goal is to illustrate the impact of coarsely quantized ADCs, and to give an idea of the expected performance of massive MIMO systems with one-bit ADCs compared to conventional systems that assume infinite ADC resolution. Our specific contributions are summarized below.

- We focus on use of the Bussgang decomposition [29] to reformulate the nonlinear quantizer operation as a statistically equivalent linear system. Contrary to previous work, we perform a separate Bussgang decomposition for the pilot and data phase as well as for each channel realization, an approach that more accurately captures the full effect of the quantization. We derive an algorithm that we refer to as the Bussgang Linear Minimum Mean Squared Error (BLMMSE) channel estimator for both flat and frequency-selective channel models. We calculate the high-SNR channel estimation error floor achieved by the proposed approach under flat fading, and show via simulation that BLMMSE outperforms previously proposed methods.
- We derive a lower bound for the flat-fading case on the theoretical rate achievable in the uplink using MRC or ZF receivers based on the BLMMSE channel estimate, and we obtain a simple but tight closed-form approximation on the uplink rate assuming low SNR and a large number of users that accurately approximates our empirical observations. Similar work in [30], [31] relied on an additive quantization noise model [32], [33] to approximate the rate, but it assumed perfect rather than estimated CSI is available at the BS, which leads to an overly optimistic assessment.
- Using the closed-form expression for the achievable rate, we study the power efficiency of massive MIMO with one-bit ADCs and show that similar efficiency is obtained as in conventional massive MIMO. In particular, assuming M antennas, we show overall system performance remains unchanged if 1) for a fixed level of CSI accuracy (training data power independent of M), the transmit power of each user terminal is reduced proportionally to $1/M$, and 2) power during both training and data transmissions is reduced proportionally to $1/\sqrt{M}$.
- We propose an optimal resource allocation scheme to maximize the sum spectral efficiency of a one-bit massive MIMO system under a total power constraint. Numerical results indicate that the optimal training length in one-bit systems is no longer always equal to the number of users and the proposed resource allocation scheme notably improves performance compared to the case without power allocation.
- We show that to achieve similar performance, a one-bit massive MIMO system employing an MRC receiver will require approximately 2.2-2.3 times more antennas than a conventional system if the sum spectral efficiency for both systems is optimized by employing the optimal resource allocation scheme; for the ZF receiver, we show that to achieve the same goal, more and more antennas are needed as average transmit power increases.

A preliminary version of some of these results appeared in [34].

The rest of this paper is organized as follows. In the next section, we present the assumed system architecture and signal model. In Section III, we propose the BLMMSE channel estimator, and then based on the BLMMSE channel estimator, in Section IV we derive a simple closed-form expression for the lower bound on the achievable rate for MRC and ZF receivers in the low SNR region. Using the closed-form approximation, we then consider several system design issues related to resource allocation and the number of antennas in Section V. Simulation results are presented in Section VI and we conclude the paper in Section VII.

Notation: The following notation is used throughout the paper. Bold uppercase (lowercase) letters denote matrices (vectors); $(.)^*$, $(.)^T$, and $(.)^H$ denote complex conjugate, transpose, and Hermitian transpose operations, respectively; $||.||$ represents the 2-norm of a vector; tr$(.)$ represents the trace of a matrix; diag$\{\mathbf{X}\}$ denotes a diagonal matrix containing only the diagonal entries of $\mathbf{X}$; $\otimes$ represents the Kronecker product; $[\mathbf{X}]_{ij}$ denotes the (i,j)th entry of $\mathbf{X}$; $\mathbf{x} \sim \mathcal{CN}(\mathbf{a}, \mathbf{B})$ indicates that $\mathbf{x}$ is a complex Gaussian vector with mean $\mathbf{a}$ and covariance matrix $\mathbf{B}$; E$\{.\}$ and Var$\{.\}$ denote the expected value and variance of a random variable, respectively.

II. System Model

As depicted in Fig. 1, we consider a single-cell one-bit massive MIMO system with K single-antenna terminals and

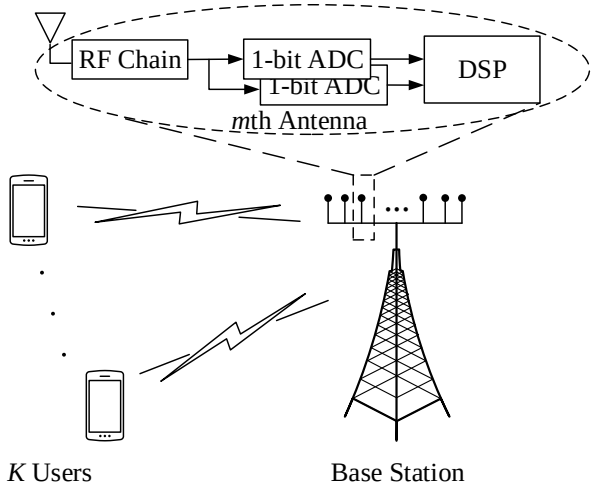

Fig. 1: One-bit massive MIMO system architecture.

an M-antenna BS, where each antenna is equipped with two one-bit ADCs and $M \gg K \gg 1$ is assumed. For the uplink, we assume all K users simultaneously transmit independent data symbols to the BS, so the received signal at the BS is

$$\mathbf{y} = \sqrt{\rho_{\mathrm{d}}}\mathbf{H}\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the $M \times 1$ additive white Gaussian noise vector, $\mathbf{H}$ is the $M \times K$ channel matrix, and $\mathbf{s}$ is a vector containing the signals transmitted by each user. We also define the vectorized channel $\underline{\mathbf{h}} = \mathrm{vec}(\mathbf{H})$ and we assume that $\underline{\mathbf{h}} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_{\underline{\mathbf{h}}})$, where $\mathbf{C}_{\underline{\mathbf{h}}}$ is the covariance matrix of $\underline{\mathbf{h}}$. We assume $\mathrm{E}\{|s_k|^2\} = 1$, and we will define the scale factor ρ_{d} to be the uplink SNR. Due to our assumption of one-bit quantization below and hence the lack of any signal dynamic range, we must assume that some type of power control is implemented that prevents a strong user from overwhelming other weaker users. For this reason, in our model we assume all users have the same level of large-scale fading/SNR ρ_{d}.

The quantized signal obtained after the one-bit ADCs is represented as

$$\mathbf{r} = \mathcal{Q}(\mathbf{y}) = \mathcal{Q}\left(\sqrt{\rho_{\mathrm{d}}}\mathbf{H}\mathbf{s} + \mathbf{n}\right), \tag{2}$$

where $\mathcal{Q}(.)$ represents the one-bit quantization operation, which is applied separately to the real and imaginary part as $\mathcal{Q}(.) = \frac{1}{\sqrt{2}}\left(\mathrm{sign}\left(\Re\left(.\right)\right) + j\mathrm{sign}\left(\Im\left(.\right)\right)\right)$. Thus, the output set of the one-bit quantization is equivalent to the QPSK constellation points $\mathcal{R} = \frac{1}{\sqrt{2}}\{1+j, 1-j, -1+j, -1-j\}$.

III. Channel Estimation for One-bit MIMO

In a standard implementation, the CSI is estimated at the BS and then used to detect the data symbols transmitted from the K users. In the uplink transmission phase, we assume the coherence interval is divided into two parts: one dedicated to training and the other to data transmission. During training, all K users simultaneously transmit their pilot sequences of τ symbols each to the BS, which yields

$$\mathbf{Y}_{\mathrm{p}} = \sqrt{\rho_{\mathrm{p}}}\mathbf{H}\mathbf{\Phi}^T + \mathbf{N}_{\mathrm{p}}, \tag{3}$$

where $\mathbf{Y}_{\mathrm{p}} \in \mathbb{C}^{M\times\tau}$ is the received signal, ρ_{p} is the pilot transmit power, and $\mathbf{\Phi} \in \mathbb{C}^{\tau\times K}$ is the pilot matrix transmitted from the K users. We assume all pilot sequences are column-wise orthogonal, i.e., $\mathbf{\Phi}^T\mathbf{\Phi}^* = \tau\mathbf{I}_K$, which implies $\tau \geq K$. We further assume that both SNRs ρ_{d} and ρ_{p} are known at the BS via, for example, a low-rate control channel.

To match the matrix form of (3) to the vector form of (2), we vectorize the received signal as

$$\mathrm{vec}(\mathbf{Y}_{\mathrm{p}}) = \mathbf{y}_{\mathrm{p}} = \bar{\mathbf{\Phi}}\underline{\mathbf{h}} + \underline{\mathbf{n}}_{\mathrm{p}}, \tag{4}$$

where $\bar{\mathbf{\Phi}} = \left(\mathbf{\Phi} \otimes \sqrt{\rho_{\mathrm{p}}}\mathbf{I}_M\right)$ and $\underline{\mathbf{n}}_p = \mathrm{vec}(\mathbf{N}_p)$. After one-bit ADCs, the quantized signal can be expressed as

$$\mathbf{r}_{\mathrm{p}} = \mathcal{Q}(\mathbf{y}_{\mathrm{p}}), \tag{5}$$

where the ith element of $\mathbf{r}_{\mathrm{p}}$ takes values from the set $\mathcal{R}$.

A. Bussgang-Based Channel Estimator

The authors in [21], [22], [24], [25] have investigated various methods for channel estimation in one-bit systems that rely on either the maximum-likelihood algorithm or on iterative algorithms with relatively high complexity. Furthermore, the channel estimators obtained by these methods do not lend themselves to an analysis that provides insight on their performance.

To address these drawbacks, in this section we take a more fundamental approach and derive simple linear estimators whose performance can be analyzed in a straightforward way. These estimators are based on the so-called Bussgang decomposition [29], which finds a statistically equivalent (up to first and second moments) linear operator for any nonlinear function of a Gaussian signal. In particular, for the one-bit quantizer in (5), the Bussgang decomposition is written

$$\mathbf{r}_{\mathrm{p}} = \mathcal{Q}(\mathbf{y}_{\mathrm{p}}) = \mathbf{A}_{\mathrm{p}}\mathbf{y}_{\mathrm{p}} + \mathbf{q}_{\mathrm{p}}\ , \tag{6}$$

where $\mathbf{A}_{\mathrm{p}}$ is the linear operator and $\mathbf{q}_{\mathrm{p}}$ the statistically equivalent quantizer noise. The matrix $\mathbf{A}_{\mathrm{p}}$ is chosen to make $\mathbf{q}_{\mathrm{p}}$ uncorrelated with $\mathbf{y}_{\mathrm{p}}$ [29], [35], or equivalently, to minimize the power of the equivalent quantizer noise. This yields

$$\mathbf{A}_{\mathrm{p}} = \mathbf{C}^H_{\mathbf{y}_{\mathrm{p}}\mathbf{r}_{\mathrm{p}}}\mathbf{C}^{-1}_{\mathbf{y}_{\mathrm{p}}}, \tag{7}$$

where $\mathbf{C}_{\mathbf{y}_{\mathrm{p}}\mathbf{r}_{\mathrm{p}}}$ denotes the cross-correlation matrix between the received signal $\mathbf{y}_{\mathrm{p}}$ and the quantized signal $\mathbf{r}_{\mathrm{p}}$, and $\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}$ denotes the auto-correlation matrix of $\mathbf{y}_{\mathrm{p}}$. For one-bit quantization and Gaussian inputs, $\mathbf{C}_{\mathbf{y}_{\mathrm{p}}\mathbf{r}_{\mathrm{p}}}$ is given by [29][36, Ch.10]

$$\mathbf{C}_{\mathbf{y}_{\mathrm{p}}\mathbf{r}_{\mathrm{p}}} = \sqrt{\frac{2}{\pi}}\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\mathrm{diag}\left(\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\right)^{-\frac{1}{2}} \triangleq \sqrt{\frac{2}{\pi}}\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\mathbf{\Sigma}^{-\frac{1}{2}}_{\mathbf{y}_{\mathrm{p}}}. \tag{8}$$

where $\mathbf{\Sigma}_{\mathbf{y}_{\mathrm{p}}} = \mathrm{diag}\left(\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\right)$.

Using (4) and (6), we can express $\mathbf{r}_{\mathrm{p}}$ as

$$\mathbf{r}_{\mathrm{p}} = \mathcal{Q}(\mathbf{y}_{\mathrm{p}}) = \tilde{\mathbf{\Phi}}\underline{\mathbf{h}} + \tilde{\mathbf{n}}_{\mathrm{p}}, \tag{9}$$

where $\tilde{\mathbf{\Phi}} = \mathbf{A}_{\mathrm{p}}\bar{\mathbf{\Phi}} \in \mathbb{C}^{M\tau\times M\tau}$, $\tilde{\mathbf{n}}_{\mathrm{p}} = \mathbf{A}_{\mathrm{p}}\mathbf{n}_{\mathrm{p}} + \mathbf{q}_{\mathrm{p}} \in \mathbb{C}^{M\tau\times 1}$. For the sake of simplicity, we derive the subsequent formulas for the case of $\mathbf{C}_{\underline{\mathbf{h}}} = \mathbf{I}_{MK}$. We will show later in Section VI.A that they are readily modified to include a generic $\mathbf{C}_{\underline{\mathbf{h}}}$.

The matrix $\mathbf{A}_{\mathrm{p}}$ is given by substituting (8) into (7):

$$\mathbf{A}_p = \sqrt{\frac{2}{\pi}}\mathrm{diag}\left(\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\right)^{-\frac{1}{2}}$$

$$= \sqrt{\frac{2}{\pi}} \text{diag}\left(\left(\mathbf{\Phi}\mathbf{\Phi}^H \otimes \rho_{\rm p}\mathbf{I}_M\right) + \mathbf{I}_{M\tau}\right)^{-\frac{1}{2}}. \quad (10)$$

We can see from (10) that $\mathbf{A}_{\rm p}$ depends on the specific choice of pilot sequences $\mathbf{\Phi}$. In order to obtain a simple expression for $\mathbf{A}_{\rm p}$, we will consider pilot sequences composed of submatrices of the discrete Fourier transform (DFT) operator [37]. In particular, we define $\mathbf{\Phi}$ using K columns of the $\tau \times \tau$ DFT matrix, in which case $\mathbf{\Phi}$ has dimension $\tau \times K$, where $\tau \geq K$. The benefits of using DFT pilot sequences are: i) all the elements of the matrix have the same magnitude, which simplifies peak transmit power constraints, and ii) the diagonal terms of $\mathbf{\Phi}\mathbf{\Phi}^H$ are always equal to K, which results in a simple expression for $\mathbf{A}_{\rm p}$, as follows:

$$\mathbf{A}_{\rm p} = \sqrt{\frac{2}{\pi}\frac{1}{K\rho_{\rm p}+1}}\mathbf{I}_{M\tau} \triangleq \alpha_{\rm p}\mathbf{I}_{M\tau}. \quad (11)$$

Based on this statistically equivalent linear model, we can formulate the LMMSE estimator [38], which we refer to as Bussgang LMMSE (BLMMSE) channel estimator:

$$\hat{\underline{\mathbf{h}}}^{\rm BLM} = \mathbf{C}_{\underline{\mathbf{h}}\mathbf{r}_{\rm p}}\mathbf{C}_{\mathbf{r}_{\rm p}}^{-1}\mathbf{r}_{\rm p} = \left(\tilde{\mathbf{\Phi}}^H + \mathbf{C}_{\underline{\mathbf{h}}\mathbf{q}_{\rm p}}\right)\mathbf{C}_{\mathbf{r}_{\rm p}}^{-1}\mathbf{r}_{\rm p}, \quad (12)$$

where $\mathbf{C}_{\mathbf{h}\mathbf{r}_{\rm p}}$ is the cross-correlation matrix between $\mathbf{h}$ and $\mathbf{r}_{\rm p}$, $\mathbf{C}_{\mathbf{r}_{\rm p}}$ is the auto-correlation matrix of $\mathbf{r}_{\rm p}$.

The formula of (12) involves the auto-correlation function of the quantized signal $\mathbf{r}_{\rm p}$. It has been shown in [39] that for one-bit ADCs, the arcsin law can be used to obtain

$$\begin{aligned}\mathbf{C}_{\mathbf{r}_{\rm p}} =& \frac{2}{\pi}\left(\arcsin\left(\mathbf{\Sigma}_{\mathbf{y}_{\rm p}}^{-\frac{1}{2}}\Re\left(\mathbf{C}_{\mathbf{y}_{\rm p}}\right)\mathbf{\Sigma}_{\mathbf{y}_{\rm p}}^{-\frac{1}{2}}\right)\right.\\ &\left.+j\arcsin\left(\mathbf{\Sigma}_{\mathbf{y}_{\rm p}}^{-\frac{1}{2}}\Im\left(\mathbf{C}_{\mathbf{y}_{\rm p}}\right)\mathbf{\Sigma}_{\mathbf{y}_{\rm p}}^{-\frac{1}{2}}\right)\right). \quad (13)\end{aligned}$$

Moreover, using $\mathbf{A}_{\rm p}$ as in (10) according to the Bussgang theorem, the quantizer noise $\mathbf{q}_{\rm p}$ is not only uncorrelated with the received signal $\mathbf{y}_{\rm p}$, but also with the channel $\underline{\mathbf{h}}$ (see Appendix A). Therefore, we can simplify the BLMMSE channel estimator of (12) as

$$\hat{\underline{\mathbf{h}}}^{\rm BLM} = \tilde{\mathbf{\Phi}}^H\mathbf{C}_{\mathbf{r}_{\rm p}}^{-1}\mathbf{r}_{\rm p}. \quad (14)$$

Thus the covariance matrix of the BLMMSE channel estimate is given by

$$\mathbf{C}_{\hat{\underline{\mathbf{h}}}^{\rm BLM}} = \tilde{\mathbf{\Phi}}^H\mathbf{C}_{\mathbf{r}_{\rm p}}^{-1}\tilde{\mathbf{\Phi}}. \quad (15)$$

A similar LMMSE channel estimator is proposed in [28]. However, our proposed channel estimator in (14) is more general since the correlation between each element of the quantizer noise is taken into account by using the arcsine law. In fact, (15) can be reduced to the estimator derived in [28] if $\tau = K$ is assumed. When $\tau = K$, it is easy to see that $\mathbf{C}_{\mathbf{y}_{\rm p}} = (K\rho_{\rm p}+1)\mathbf{I}_{MK}$, and hence according to (13), $\mathbf{C}_{\mathbf{r}_{\rm p}} = \mathbf{I}_{MK}$. Therefore, when $\tau = K$ we can obtain the BLMMSE channel estimator of (14) as simply

$$\hat{\underline{\mathbf{h}}}^{\rm BLM} = \tilde{\mathbf{\Phi}}^H\mathbf{r}_{\rm p}. \quad (16)$$

We emphasize that, when $\tau = K$, there is no correlation between the quantizer noise $\mathbf{q}_{\rm p}$ and the normalized MSE for BLMMSE channel estimator is given by

$$\begin{aligned}\mathcal{M}^{\rm BLM} &= \frac{1}{MK}\mathrm{E}\left\{\left\|\tilde{\mathbf{\Phi}}^H\mathbf{r}_{\rm p} - \underline{\mathbf{h}}\right\|_2^2\right\}\\ &= \frac{1}{MK}\mathrm{tr}\left(\mathbf{I}_{MK} - \tilde{\mathbf{\Phi}}^H\tilde{\mathbf{\Phi}}\right)\\ &= 1 - \frac{2K\rho_{\rm p}}{\pi(K\rho_{\rm p}+1)}, \quad (17)\end{aligned}$$

and for high SNRs

$$\lim_{\rho_p\to\infty}\mathcal{M}^{\rm BLM} = 1 - \frac{2}{\pi} = -4.40\text{dB}. \quad (18)$$

The results in (17) and (18) are allied with the results in [28, Eq.(35)] by setting $p[l] = 1/L$ and $\beta_k P_k = \rho_{\rm p}$. In addition, the result in (18) implies that there exists an error floor for the channel estimate as the training power increases to infinity.

B. Extension to Frequency Selective Fading with OFDM

Although for simplicity we focus on the flat fading case in this paper, we show here how to extend our channel estimation method to the frequency selective case, assuming the transmitter employs OFDM signaling. In particular, consider an OFDM system with $N_{\rm c}$ subcarriers, and denote the uplink OFDM symbol transmitted from the kth user as $\mathbf{x}_k^{\rm FD} \in \mathbb{C}^{N_{\rm c}\times 1}$. Before transmission, this vector is processed by a unitary IFFT operation $\mathbf{F}^H$, and then a cyclic prefix (CP) of length $N_{\rm cp}$ is added. Assume the CP length satisfies $L-1 \leq N_{\rm cp} \leq N_{\rm c}$, where L is the number of channel taps. After removing the CP, the $N_{\rm c}\times 1$ received time domain signal at the mth BS-antenna is given by

$$\begin{aligned}\mathbf{y}_m^{\rm TD} &= \sum_{k=1}^{K}\mathbf{G}_{mk}^{\rm TD}\mathbf{F}^H\mathbf{x}_k^{\rm FD} + \mathbf{n}_m^{\rm TD}\\ &= \sum_{k=1}^{K}\mathbf{\Phi}_k^{\rm TD}\mathbf{g}_{mk}^{\rm TD} + \mathbf{n}_m^{\rm TD} = \sum_{k=1}^{K}\mathbf{\Phi}_{k,L}^{\rm TD}\mathbf{h}_{mk}^{\rm TD} + \mathbf{n}_m^{\rm TD}, \quad (19)\end{aligned}$$

where the superscripts "TD" and "FD" refer to Time Domain and Frequency Domain, respectively. The matrix $\mathbf{G}_{mk}^{\rm TD} \in \mathbb{C}^{N_{\rm c}\times N_{\rm c}}$ is circulant and its first column is given by $\mathbf{g}_{mk}^{\rm TD} = [(\mathbf{h}_{mk}^{\rm TD})^T, 0, ..., 0]^T$, where $\mathbf{h}_{mk}^{\rm TD}$ is an $L\times 1$ column vector containing the L channel taps, and $\mathbf{n}_m^{\rm TD} \sim \mathcal{CN}(\mathbf{0},\mathbf{I})$ is additive white Gaussian noise. The matrix $\mathbf{\Phi}_k^{\rm TD} \in \mathbb{C}^{N_{\rm c}\times N_{\rm c}}$ is also circulant with first column given by $\phi_k^{\rm TD} = \mathbf{F}^H\mathbf{x}_k^{\rm FD}$. $\mathbf{\Phi}_{k,L}^{\rm TD}$ is a submatrix of $\mathbf{\Phi}_k^{\rm TD}$, corresponding to the first L columns of $\mathbf{\Phi}_k^{\rm TD}$. The second equation follows from the commutative property of circulant convolution. The third equation is due to the fact that there are only finite L channel taps.

After stacking the received time domain signal $\mathbf{y}_m^{\rm TD}$ for all M BS antennas, we have

$$\mathbf{y}^{\rm TD} = \bar{\mathbf{\Phi}}_L^{\rm TD}\mathbf{h}^{\rm TD} + \mathbf{n}^{\rm TD}, \quad (20)$$

where $\bar{\mathbf{\Phi}}_L^{\rm TD} = \mathbf{I}\otimes\mathbf{\Phi}_L^{\rm TD}$ with $\mathbf{\Phi}_L^{\rm TD} = [\mathbf{\Phi}_{1,L}^{\rm TD}, \mathbf{\Phi}_{2,L}^{\rm TD}, ..., \mathbf{\Phi}_{K,L}^{\rm TD}] \in \mathbb{C}^{N_{\rm c}\times LK}$ and $\mathbf{h}^{\rm TD} \in \mathbb{C}^{MKL\times 1}$ contains all channel taps between the M BS-antennas and K users. After one-bit quantization, the time domain quantized signal can be expressed as

$$\mathbf{r}^{\rm TD} = \mathcal{Q}(\mathbf{y}^{\rm TD}) = \mathcal{Q}(\bar{\mathbf{\Phi}}_L^{\rm TD}\mathbf{h}^{\rm TD} + \mathbf{n}^{\rm TD}), \quad (21)$$

and we see that, unlike a conventional system, OFDM cannot split the wideband channel into many parallel narrowband channel in a one-bit system.

Using the Bussgang decomposition, the non-linear quantization operation can be reformulated as

$$\begin{aligned}\mathbf{r}^{\mathrm{TD}} &= \mathbf{A}\mathbf{y}^{\mathrm{TD}} + \mathbf{q}^{\mathrm{TD}} \\ &= \mathbf{A}\bar{\mathbf{\Phi}}_L^{\mathrm{TD}}\mathbf{h}^{\mathrm{TD}} + \mathbf{A}\mathbf{n}^{\mathrm{TD}} + \mathbf{q}^{\mathrm{TD}},\end{aligned} \tag{22}$$

where the matrix $\mathbf{A}$ is chosen to make the quantizer noise $\mathbf{q}^{\mathrm{TD}}$ uncorrelated with $\mathbf{y}^{\mathrm{TD}}$. If the received time domain signal $\mathbf{y}^{\mathrm{TD}}$ is Gaussian, we have

$$\mathbf{A} = \sqrt{\frac{2}{\pi}}\mathrm{diag}\left(\mathbf{C}_{\mathbf{y}^{\mathrm{TD}}}\right)^{-\frac{1}{2}} . \tag{23}$$

Consequently, the BLMMSE channel estimator for the wideband OFDM case can be expressed as

$$\hat{\mathbf{h}}^{\mathrm{TD}} = \mathbf{C}_{\mathbf{h}^{\mathrm{TD}}}(\bar{\mathbf{\Phi}}_L^{\mathrm{TD}})^H\mathbf{A}^H\mathbf{C}_{\mathbf{r}^{\mathrm{TD}}}^{-1}\mathbf{r}^{\mathrm{TD}}, \tag{24}$$

where $\mathbf{C}_{\mathbf{h}^{\mathrm{TD}}}$ is the covariance matrix of $\mathbf{h}^{\mathrm{TD}}$, and the covariance matrix of $\mathbf{r}^{\mathrm{TD}}$ is obtained by using the arcsine law:

$$\begin{aligned}\mathbf{C}_{\mathbf{r}^{\mathrm{TD}}} =& \frac{2}{\pi}\left(\arcsin\left(\mathbf{\Sigma}_{\mathbf{y}^{\mathrm{TD}}}^{-\frac{1}{2}}\Re\left(\mathbf{C}_{\mathbf{y}^{\mathrm{TD}}}\right)\mathbf{\Sigma}_{\mathbf{y}^{\mathrm{TD}}}^{-\frac{1}{2}}\right)\right. \\ &\left. + j\arcsin\left(\mathbf{\Sigma}_{\mathbf{y}^{\mathrm{TD}}}^{-\frac{1}{2}}\Im\left(\mathbf{C}_{\mathbf{y}^{\mathrm{TD}}}\right)\mathbf{\Sigma}_{\mathbf{y}^{\mathrm{TD}}}^{-\frac{1}{2}}\right)\right),\end{aligned} \tag{25}$$

where $\mathbf{\Sigma}_{\mathbf{y}^{\mathrm{TD}}} = \mathrm{diag}\left(\mathbf{C}_{\mathbf{y}^{\mathrm{TD}}}\right)$. The covariance matrix of the quantizer noise $\mathbf{q}^{\mathrm{TD}}$ can be obtained by

$$\mathbf{C}_{\mathbf{q}^{\mathrm{TD}}} = \mathbf{C}_{\mathbf{r}^{\mathrm{TD}}} - \mathbf{A}\mathbf{C}_{\mathbf{y}^{\mathrm{TD}}}\mathbf{A}^H. \tag{26}$$

The above Bussgang-based channel estimators are more general than those derived in other work such as [28], since they take into account the fact that in general the covariance matrix of the quantizer noise $\mathbf{q}^{\mathrm{TD}}$ cannot be expressed as a diagonal matrix due to the arcsine law. This observation holds for any linear modulation scheme employed by the users, not just OFDM. While the derivations that follow will focus on the flat fading case, we can see from the above that they can be easily generalized to frequency-selective fading.

C. Low SNR Approximate BLMMSE Channel Estimate Covariance

As we can see from (13) and (14), it is difficult to obtain a general closed-form expression for the MSE of the BLMMSE channel estimator due to the 'arcsine' operation. However, it is expected that massive MIMO systems will operate at relatively low SNRs due to the availability of a large array gain [2]. Therefore in this subsection, we focus on deriving a low-SNR approximation for the covariance matrix of the BLMMSE channel estimator. According to (9), we can reformulate $\mathbf{C}_{\mathbf{r}_{\mathrm{p}}}$ as the following linear function,

$$\mathbf{C}_{\mathbf{r}_{\mathrm{p}}} = \tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^H + \mathbf{A}_{\mathrm{p}}\mathbf{A}_{\mathrm{p}}^H + \mathbf{C}_{\mathbf{q}_{\mathrm{p}}}, \tag{27}$$

where

$$\begin{aligned}\mathbf{C}_{\mathbf{q}_{\mathrm{p}}} &= \mathbf{C}_{\mathbf{r}_{\mathrm{p}}} - \mathbf{A}_{\mathrm{p}}\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\mathbf{A}_{\mathrm{p}}^H \\ &= \frac{2}{\pi}(\arcsin(\mathbf{X}) + j\arcsin(\mathbf{Y})) - \frac{2}{\pi}(\mathbf{X} + j\mathbf{Y}),\end{aligned} \tag{28}$$

and where we define

$$\mathbf{X} = \mathbf{\Sigma}_{\mathbf{y}_{\mathrm{p}}}^{-\frac{1}{2}}\Re\left(\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\right)\mathbf{\Sigma}_{\mathbf{y}_{\mathrm{p}}}^{-\frac{1}{2}} \tag{29}$$

$$\mathbf{Y} = \mathbf{\Sigma}_{\mathbf{y}_{\mathrm{p}}}^{-\frac{1}{2}}\Im\left(\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}\right)\mathbf{\Sigma}_{\mathbf{y}_{\mathrm{p}}}^{-\frac{1}{2}}. \tag{30}$$

We can see from (28) that the covariance matrix of the quantizer noise is in general not a diagonal matrix, which implies that there exists correlation between the quantization noise on each antenna. However, at low SNR or for large numbers of users, $\mathbf{C}_{\mathbf{y}_{\mathrm{p}}}$ is diagonally dominant and we can use the following approximation for applying the arcsine law:

$$\frac{2}{\pi}\arcsin(a) \cong \begin{cases} 1, & a = 1 \\ 2a/\pi, & a < 1. \end{cases} \tag{31}$$

Since the non-diagonal elements of $\mathbf{X}$ and $\mathbf{Y}$ are much smaller than 1 in the low SNR regime, we can approximate (28) as

$$\mathbf{C}_{\mathbf{q}_{\mathrm{p}}} \cong (1 - 2/\pi)\mathbf{I}_{M\tau}. \tag{32}$$

This implies that we can approximate the quantizer noise as uncorrelated noise with a variance of $1 - 2/\pi$ at low SNR. Substituting (32) and (27) into (15), we have

$$\begin{aligned}\mathbf{C}_{\hat{\underline{\mathbf{h}}}^{\mathrm{BLM}}} &\cong \tilde{\mathbf{\Phi}}^H\left(\tilde{\mathbf{\Phi}}\tilde{\mathbf{\Phi}}^H + (\alpha_{\mathrm{p}}^2 + 1 - 2/\pi)\mathbf{I}_{M\tau}\right)^{-1}\tilde{\mathbf{\Phi}} \\ &= (\alpha_{\mathrm{p}}^2\tau\rho_{\mathrm{p}} + \alpha_{\mathrm{p}}^2 + 1 - 2/\pi)^{-1}\alpha_{\mathrm{p}}^2\tau\rho_{\mathrm{p}}\mathbf{I}_{MK} \triangleq \sigma^2\mathbf{I}_{MK} ,\end{aligned} \tag{33}$$

where we have defined $\sigma^2 = (\alpha_{\mathrm{p}}^2\tau\rho_{\mathrm{p}} + \alpha_{\mathrm{p}}^2 + 1 - 2/\pi)^{-1}\alpha_{\mathrm{p}}^2\tau\rho_{\mathrm{p}}$. The equation on the second line holds due to the matrix inversion identity $(\mathbf{I} + \mathbf{A}\mathbf{B})^{-1}\mathbf{A} = \mathbf{A}(\mathbf{I} + \mathbf{B}\mathbf{A})^{-1}$. The result in (33) implies that in the low SNR regime, each element of the BLMMSE channel estimate is uncorrelated. In what follows, we will evaluate the uplink achievable rate by using the low SNR approximation in (33).

IV. Achievable Rate Analysis in the One-Bit MIMO Uplink

A. Data Transmission

In the data transmission stage, we assume the K users simultaneously transmit their data symbols, represented as the vector $\mathbf{s}$, to the BS. After one-bit quantization, the signal at the BS can be expressed as

$$\begin{aligned}\mathbf{r}_d &= \mathcal{Q}(\mathbf{y}_{\mathrm{d}}) = \mathcal{Q}(\sqrt{\rho_{\mathrm{d}}}\mathbf{H}\mathbf{s} + \mathbf{n}_d) \\ &= \sqrt{\rho_{\mathrm{d}}}\mathbf{A}_{\mathrm{d}}\mathbf{H}\mathbf{s} + \mathbf{A}_{\mathrm{d}}\mathbf{n}_{\mathrm{d}} + \mathbf{q}_{\mathrm{d}},\end{aligned} \tag{34}$$

where the same definitions as in the previous sections apply, but replacing the subscript 'p' with 'd', since the power ρ_{d} during data transmission may be different than during training. Again, according to the Bussgang decomposition and assuming a Gaussian input, we have

$$\mathbf{A}_{\mathrm{d}} = \sqrt{\frac{2}{\pi}}\mathrm{diag}\left(\mathbf{C}_{\mathbf{y}_{\mathrm{d}}}\right)^{-\frac{1}{2}} = \sqrt{\frac{2}{\pi}}\mathrm{diag}\left(\rho_{\mathrm{d}}\mathbf{H}\mathbf{H}^H + \mathbf{I}_M\right)^{-\frac{1}{2}} . \tag{35}$$

Note that, in contrast to the model of [28], in which the quantizer noise can still be correlated with the desired signal since the same Bussgang decomposition is employed for different channel realizations, the Bussgang decomposition in (35) is employed for each individual channel realization. This

approach ensures that the quantizer noise is uncorrelated with the desired signal.

As can be seen in (35), the covariance matrix $\mathbf{C}_{\mathbf{y}_\mathrm{d}}$ of the quantizer input (and hence, the matrix $\mathbf{H}\mathbf{H}^H$) must be known at the BS in order to implement the Bussgang decomposition. In practice, however, we can use the same technique provided in [40] to reconstruct the covariance matrix of $\mathbf{C}_{\mathbf{y}_\mathrm{d}}$ using the measurements of the quantizer output. In addition, relying on channel hardening for $K \gg 1$ in massive MIMO systems and for i.i.d. unit-variance channel coefficients, we can approximate the matrix $\mathbf{A}_\mathrm{d}$ as

$$\mathbf{A}_\mathrm{d} \cong \sqrt{\frac{2}{\pi}}\sqrt{\frac{1}{1+K\rho_\mathrm{d}}}\mathbf{I}_M = \alpha_\mathrm{d}\mathbf{I}_M, \tag{36}$$

without requiring perfect CSI. This approximate gain matrix is assumed without derivation in other previous work such as [28].

Next we assume the BS uses the BLMMSE channel estimate to compute a linear receiver to detect the data symbols transmitted from the K users. The linear receiver attempts to separate the quantized signal into K streams by multiplying the signal by the matrix $\mathbf{W}^T$ as follows:

$$\begin{aligned}\hat{\mathbf{s}} &= \mathbf{W}^T\mathbf{r}_\mathrm{d} \\ &= \sqrt{\rho_\mathrm{d}}\mathbf{W}^T\mathbf{A}_\mathrm{d}(\hat{\mathbf{H}}\mathbf{s} + \boldsymbol{\mathcal{E}}\mathbf{s}) + \mathbf{W}^T\mathbf{A}_\mathrm{d}\mathbf{n}_\mathrm{d} + \mathbf{W}^T\mathbf{q}_\mathrm{d}\ ,\end{aligned} \tag{37}$$

where $\hat{\mathbf{H}} = \mathrm{unvec}(\hat{\underline{\mathbf{h}}}^{\mathrm{BLM}})$ is the estimated channel matrix (unvec is the inverse of the vec operator in Eq. (4)) and $\boldsymbol{\mathcal{E}} = \mathbf{H} - \hat{\mathbf{H}}$ denotes the channel estimation error. The kth element of $\hat{\mathbf{s}}$ is then used to decode the signal transmitted from the kth user:

$$\begin{aligned}\hat{s}_k =& \sqrt{\rho_\mathrm{d}}\mathbf{w}_k^T\mathbf{A}_\mathrm{d}\hat{\mathbf{h}}_k s_k + \sqrt{\rho_\mathrm{d}}\mathbf{w}_k^T\sum\nolimits_{i\neq k}^{K}\mathbf{A}_\mathrm{d}\hat{\mathbf{h}}_i s_i \\ &+ \sqrt{\rho_\mathrm{d}}\mathbf{w}_k^T\sum\nolimits_{i=1}^{K}\mathbf{A}_\mathrm{d}\boldsymbol{\varepsilon}_i s_i + \mathbf{w}_k^T\mathbf{A}_\mathrm{d}\mathbf{n}_\mathrm{d} + \mathbf{w}_k^T\mathbf{q}_\mathrm{d},\end{aligned} \tag{38}$$

where $\mathbf{w}_k$, $\hat{\mathbf{h}}_k$ and $\boldsymbol{\varepsilon}_k$ are the kth columns of $\mathbf{W}$, $\hat{\mathbf{H}}$ and $\boldsymbol{\mathcal{E}}$, respectively.

The last four terms in (38) respectively correspond to user interference, channel estimation error, AWGN noise and quantizer noise. In our analysis, we will consider the performance of the common MRC and ZF receivers, defined by

$$\mathbf{W}_{\mathrm{MRC}}^T = \hat{\mathbf{H}}^H \tag{39}$$

$$\mathbf{W}_{\mathrm{ZF}}^T = \left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^H, \tag{40}$$

respectively.

B. Uplink Achievable Rate Approximation at Low SNR

Although prior work has obtained expressions for the mutual information or the achievable rate of one-bit systems using the joint probability distribution of the transmitted and received symbols [21], [23], [24], this approach does not result in easily computable or insightful expressions. To overcome this drawback, in this section we provide a simple closed-form expression for an approximation of the achievable rate for both MRC and ZF processing in the low SNR region. Using the same reasoning as in Section III-C, the covariance matrix of $\mathbf{q}_\mathrm{d}$ can be expressed as

$$\mathbf{C}_{\mathbf{q}_\mathrm{d}} = \mathbf{C}_{\mathbf{r}_\mathrm{d}} - \mathbf{A}_\mathrm{d}\mathbf{C}_{\mathbf{y}_\mathrm{d}}\mathbf{A}_\mathrm{d}^H, \tag{41}$$

where $\mathbf{C}_{\mathbf{r}_\mathrm{d}}$ is the covariance matrix of $\mathbf{r}_\mathrm{d}$ and can be obtained using the arcsine law in (13). Note that, again, the covariance matrix of (41) is in general not a diagonal matrix, which implies that there exists some correlations among the elements of $\mathbf{q}_\mathrm{d}$. For the special case where $\mathbf{C}_{\mathbf{y}_\mathrm{d}} = \rho_\mathrm{d}\mathbf{H}\mathbf{H}^H + \mathbf{I}_M$ is diagonally dominant due to low SNR or for large K with i.i.d. channels, then similar to the pilot phase, the approximation (32) can be used.

Furthermore, while the quantizer noise $\mathbf{q}_\mathrm{d}$ is non-Gaussian, we can obtain a lower bound on the achievable rate by making the worst-case assumption [41], [42] that in fact it is Gaussian with the same covariance matrix in (41). Using this approach and (38), the ergodic achievable rate of the one-bit MIMO uplink is lower bounded by (42) shown on the next page. In order to obtain a closed-form expression for the achievable rate, we rewrite the detected signal in (38) as a known mean gain (which only depends on the channel distribution instead of the instantaneous channel) times the desired symbol plus an uncorrelated effective noise, as follows:

$$\hat{s}_k = \mathrm{E}\left\{\sqrt{\rho_\mathrm{d}}\mathbf{w}_k^T\mathbf{A}_\mathrm{d}\mathbf{h}_k\right\}s_k + \tilde{n}_{\mathrm{d},k}, \tag{43}$$

where $\tilde{n}_{\mathrm{d},k}$ is the effective noise given by

$$\begin{aligned}\tilde{n}_{\mathrm{d},k} =& \left(\sqrt{\rho_\mathrm{d}}\mathbf{w}_k^T\mathbf{A}_\mathrm{d}\mathbf{h}_k - \mathrm{E}\left\{\sqrt{\rho_\mathrm{d}}\mathbf{w}_k^T\mathbf{A}_\mathrm{d}\mathbf{h}_k\right\}\right)s_k \\ &+ \sqrt{\rho_\mathrm{d}}\mathbf{w}_k^T\sum_{i\neq k}^{K}\mathbf{A}_\mathrm{d}\mathbf{h}_i s_i + \mathbf{w}_k^T\mathbf{A}_\mathrm{d}\mathbf{n}_\mathrm{d} + \mathbf{w}_k^T\mathbf{q}_\mathrm{d}.\end{aligned} \tag{44}$$

Lemma 1: In a massive MIMO system with one-bit quantization and $\mathbf{A}_\mathrm{d} = \alpha_\mathrm{d}\mathbf{I}_M$, the uplink achievable rate for the kth user at low SNR can be approximated by

$$R_k = \log_2\left(1 + \frac{\rho_\mathrm{d}\alpha_\mathrm{d}^2\left|\mathrm{E}\left\{\mathbf{w}_k^T\mathbf{h}_k\right\}\right|^2}{\rho_\mathrm{d}\alpha_\mathrm{d}^2\mathrm{Var}\left(\mathbf{w}_k^T\mathbf{h}_k\right) + \mathrm{UI}_k + \mathrm{AQN}_k}\right), \tag{45}$$

where

$$\mathrm{UI}_k = \rho_\mathrm{d}\alpha_\mathrm{d}^2\sum_{i\neq k}^{K}\mathrm{E}\left\{\left|\mathbf{w}_k^T\mathbf{h}_i\right|^2\right\} \tag{46}$$

$$\mathrm{AQN}_k = \left(\alpha_\mathrm{d}^2 + 1 - \frac{2}{\pi}\right)\mathrm{E}\left\{\left\|\mathbf{w}_k^T\right\|^2\right\}. \tag{47}$$

Proof: See Appendix B. ■

The result in (45) is obtained by approximating the effective noise as Gaussian. In a massive MIMO system, the effective noise is a sum of a very large number of independent zero-mean terms, and thus we expect via the central limit theorem that the approximation will be asymptotically tight to the lower bound of (42) in M. In Section V, it will be shown that the gap between the achievable rate approximation given by (45) and the lower bound of the ergodic achievable rate given in (42) is small, which implies that our resulting closed-form expression is an excellent predictor of the system performance.

Based on Lemma 1, we derive in the theorems below closed-form expressions for the lower bound on the achievable rate

$$\tilde{R}_k = \mathrm{E}\left\{\log_2\left(1+\frac{\rho_{\mathrm{d}}\left|\mathbf{w}_k^T\mathbf{A}_{\mathrm{d}}\hat{\mathbf{h}}_k\right|^2}{\rho_{\mathrm{d}}\sum_{i\neq k}^{K}\left|\mathbf{w}_k^T\mathbf{A}_{\mathrm{d}}\hat{\mathbf{h}}_i\right|^2+\rho_{\mathrm{d}}\sum_{i=1}^{K}\left|\mathbf{w}_k^T\mathbf{A}_{\mathrm{d}}\boldsymbol{\varepsilon}_i\right|^2+\left\|\mathbf{w}_k^T\mathbf{A}_{\mathrm{d}}\right\|^2+\mathbf{w}_k^T\mathbf{C}_{\mathbf{q}_{\mathrm{d}}}\mathbf{w}_k^*}\right)\right\} \tag{42}$$

for the MRC and ZF receivers.

Theorem 1: For the MRC receiver with CSI estimated by the BLMMSE channel estimator, the achievable rate of the kth user in a one-bit massive MIMO uplink at low SNR can be approximated by

$$\begin{aligned} R_{\mathrm{MRC},k} &= \log_2\left(1+\frac{\rho_{\mathrm{d}}\alpha_{\mathrm{d}}^2 M\sigma^2}{\rho_d\alpha_{\mathrm{d}}^2 K+\alpha_{\mathrm{d}}^2+(1-2/\pi)}\right) \\ &= \log_2\left(1+\rho_{\mathrm{d}}\alpha_{\mathrm{d}}^2 M\sigma^2\right). \end{aligned} \tag{48}$$

Proof: See Appendix C. ■

Theorem 2: For the ZF receiver with CSI estimated by the BLMMSE channel estimator, the achievable rate of the kth user in a one-bit massive MIMO uplink at low SNR can be approximated by

$$R_{\mathrm{ZF},k} = \log_2\left(1+\frac{\rho_{\mathrm{d}}\alpha_{\mathrm{d}}^2\sigma^2(M-K)}{\rho_{\mathrm{d}}\alpha_{\mathrm{d}}^2 K\eta+\alpha_{\mathrm{d}}^2+(1-2/\pi)}\right), \tag{49}$$

where $\eta = (1-\sigma^2)$.

Proof: See Appendix D. ■

V. One-Bit Massive MIMO System Design

The simple approximation for the achievable rate derived in the previous section provides us with a tool for easily quantifying the impact of system design decisions. In this section, we study design issues surrounding the length of the training sequence, the power allocated for training and data transmission, and the number of BS antennas. Our performance metric will be the sum spectral efficiency, defined by

$$\mathcal{S}_{\mathbb{A}} = \frac{T-\tau}{T}\sum_{k=1}^{K} R_{\mathbb{A},k}, \tag{50}$$

where T represents the length of the coherence interval, during which the channel satisfies the block fading model and stays constant. The notation $\mathbb{A} \in \{\mathtt{MRC}, \mathtt{ZF}\}$ indicates that we will perform the analysis for both the MRC and ZF receivers.

A. Power Efficiency in One-Bit Massive MIMO

In this section, we study the power efficiency achieved by one-bit massive MIMO systems, where an increase in the number of antennas can be traded for reduced transmit power at the user terminals. We will consider two cases: i) the training power ρ_{p} (and hence the channel estimation accuracy) is fixed, but user transmit power decreases as $1/M$; and ii) the training power ρ_{p} and data transmission power ρ_{d} are equal and scale as $1/\sqrt{M}$.

1) Case I: In the first case, we assume ρ_{p} is fixed and independent of M, while $\rho_{\mathrm{d}} = E_{\mathrm{u}}/M^c$ for a given c, where E_{u} is fixed independent of M. We will find the largest value for c such that scaling down the users' power by $1/M^c$ results in no change in spectral efficiency as $M \to \infty$. Substituting $\rho_{\mathrm{d}} = E_{\mathrm{u}}/M^c$ into (48) and (49) and assuming M increases to infinity, we can readily see that choosing $c = 1$ will result in the spectral efficiency converging to a fixed value. This implies that, when the channel estimation accuracy is fixed, the transmit power of each user can be reduced proportionally by $1/M$ for both the MRC and ZF receivers while maintaining a given sum spectral efficiency. Moreover, the asymptotic performance for MRC and ZF is the same and is given by

$$\lim_{M\to\infty}\mathcal{S}_{\mathbb{A}}\big|_{\rho_{\mathrm{d}}=\frac{E_{\mathrm{u}}}{M}} = \frac{T-\tau}{T}K\log_2\left(1+\frac{2}{\pi}\sigma^2 E_{\mathrm{u}}\right). \tag{51}$$

2) Case II: For the second case, we assume the training and data transmission power are reduced at the same rate: $\rho_{\mathrm{p}} = \rho_{\mathrm{d}} = E_{\mathrm{u}}/M^c$, where again E_{u} is fixed independent of M. Substituting $\rho_{\mathrm{p}} = \rho_{\mathrm{d}} = E_{\mathrm{u}}/M^c$ into (48) and (49) and assuming M increases to infinity, the value of $c = 1/2$ can be seen to provide constant performance. Thus, we cannot reduce the user transmit power as aggressively as in the first case where the channel estimation accuracy is fixed. The asymptotic performance for MRC and ZF is again the same in this case, but with a different asymptotic value:

$$\lim_{M\to\infty}\mathcal{S}_{\mathbb{A}}\big|_{\rho_{\mathrm{d}}=\rho_{\mathrm{p}}=\frac{E_{\mathrm{u}}}{\sqrt{M}}} = \frac{T-\tau}{T}K\log_2\left(1+\frac{4}{\pi^2}\tau E_{\mathrm{u}}^2\right). \tag{52}$$

Note that both of the spectral efficiency expressions in (51) and (52) are equivalent to that of K SISO channels with transmit power $2\sigma^2 E_{\mathrm{u}}/\pi$ and $4\tau E_{\mathrm{u}}^2/\pi^2$, respectively, without interference. Thus, even though one-bit ADCs are deployed at the BS, the spectral efficiency increases proportionally to the number of users K.

B. Resource Allocation in One-Bit Massive MIMO System

It has been proved in [41] that for conventional MIMO systems with infinite precision ADCs, the optimal training length is always $\tau = K$. However, due to the quantizer noise, we will see that this result does not hold for one-bit massive MIMO systems. Considerable gains in spectral efficiency can be obtained by proper resource allocation. Thus in this subsection, we assume the users can vary the training power and the data transmission power and study the optimal resource allocation scheme that jointly selects the length of the training sequence, and the power allocated to training and data transmission with the goal of maximizing the sum spectral efficiency.

Let ρ be the average transmit power and $P = \rho T$ be the total power budget for the users in one coherence interval, which satisfies the constraint $\tau\rho_{\mathrm{p}} + (T-\tau)\rho_{\mathrm{d}} \leq P$. Then, following the approach of [43], the optimization problem can be formulated as

$$\begin{aligned} \underset{\rho_{\mathrm{p}},\rho_{\mathrm{d}},\tau}{\text{maximize}} \quad & \mathcal{S}_{\mathbb{A}} \\ \text{subject to} \quad & \tau\rho_{\mathrm{p}} + (T-\tau)\rho_{\mathrm{d}} \leq P \\ & K \leq \tau \leq T \end{aligned} \tag{53}$$

$$\rho_{\rm p} \geq 0, \rho_{\rm d} \geq 0. \tag{54}$$

For any power allocation in which the users do not employ the full energy budget, the users could increase their training power (and, thus, increase the channel estimation accuracy) without causing any inter-user interference in the data transmission phase, and hence in turn improve their rate. Therefore, we can replace the inequality constraint on the total energy budget with an equality constraint, i.e., $\tau\rho_{\rm p} + (T-\tau)\rho_{\rm d} = P$. To facilitate the presentation, let $\gamma \in (0,1)$ denote the fraction of the total energy budget that is devoted to pilot training, so that $\gamma P = \tau\rho_{\rm p}$ and $(1-\gamma)P = (T-\tau)\rho_{\rm d}$. The optimization problem in (53) is then equivalent to

$$\begin{aligned} \underset{\gamma,\tau}{\text{maximize}} \quad & \mathcal{S}_{\rm A}\big|_{\rho_{\rm p}=\frac{\gamma P}{\tau}, \rho_{\rm d}=\frac{(1-\gamma)P}{T-\tau}} \\ \text{subject to} \quad & 0 < \gamma < 1, \quad K \leq \tau \leq T. \end{aligned} \tag{55}$$

Lemma 2: For both the MRC and ZF receivers in one-bit massive MIMO, the optimal training length τ^* that maximizes the sum spectral efficiency is not always equal to the number of users.

Proof: See Appendix E. ■

Although we cannot obtain a closed-form expression for τ^*, we can numerically evaluate τ^* using a simple search algorithm since there are only a few parameters in problem (55). As we will show in the numerical results, unlike conventional MIMO systems, the optimal training duration depends on various system parameters such as the coherence interval T and the total energy budget P.

C. How Many More Antennas are Needed for One-Bit Massive MIMO?

In this subsection, we compare the performance of one-bit and conventional massive MIMO with infinite resolution ADCs in terms of the number of antennas deployed at the BS. In particular, we wish to answer the question of how many more antennas a one-bit massive MIMO system would need to achieve the same spectral efficiency of a conventional massive MIMO implementation. For this analysis, we denote the number of antennas in the one-bit and conventional massive MIMO systems as $M_{\rm one}$ and $M_{\rm conv}$, respectively, and show the lower bound on the uplink achievable rate for both one-bit and conventional MIMO in Table I, where we define $C(x) = \frac{T-\tau}{T} K \log_2(1+x)$.

For the special case of $\tau = K$ and $\rho_{\rm d} = \rho_{\rm p}$, it was shown in [28] that 2.5 times more antennas are needed in one-bit systems to ensure the same rate as the conventional system with MRC, and also for ZF at low SNR. This can be easily verified using our results as well. However, this result will not hold in general for the optimal values of $\tau, \rho_{\rm d}$ and $\rho_{\rm p}$ resulting from the optimization in (55). In fact, we can pose a complementary optimization problem in which we attempt to minimize the ratio $\kappa = M_{\rm one}/M_{\rm conv}$ required for both systems to achieve the same spectral efficiency, as follows:

$$\begin{aligned} \underset{\gamma,\tau,\kappa}{\text{minimize}} \quad & \kappa \\ \text{subject to} \quad & \mathcal{S}_{\rm A}^{\rm one} = \mathcal{S}_{\rm A}^{\rm conv}, \\ & 0 < \gamma < 1, \quad K \leq \tau \leq T. \end{aligned} \tag{56}$$

TABLE I: Lower bound on individual achievable rates for conventional and one-bit MIMO systems

	Conv. MIMO [2]	One-bit MIMO
MRC	$C\left(\frac{\rho_{\rm d}\tau\rho_{\rm p}M_{\rm conv}}{(1+K\rho_{\rm d})(1+\tau\rho_{\rm p})}\right)$	$C\left(\rho_{\rm d}\alpha_{\rm d}^2 M_{\rm one}\sigma^2\right)$
ZF	$C\left(\frac{\rho_{\rm d}\tau\rho_{\rm p}(M_{\rm conv}-K)}{K\rho_{\rm d}+\tau\rho_{\rm p}+1}\right)$	$C\left(\frac{\rho_{\rm d}\alpha_{\rm d}^2\sigma^2(M_{\rm one}-K)}{\rho_{\rm d}\alpha_{\rm d}^2 K\eta+\alpha_{\rm d}^2+(1-2/\pi)}\right)$

where $\mathcal{S}_{\rm A}^{\rm conv}$ is the maximum spectral efficiency achieved for the conventional system by optimizing $\rho_{\rm p}$ and $\rho_{\rm d}$ with $\tau = K$ for fixed $M_{\rm conv}$. Since the problem in (56) only has a few parameters, we can use a simple search algorithm for the optimization.

Although no closed-form expression for the optimal κ can be obtained, we will show in the simulations that less than 2.5 times more antennas are needed for the MRC receiver, and also for the ZF receiver at low SNR, if the training length τ, training power $\rho_{\rm p}$ and data transmission power $\rho_{\rm d}$ are all optimized.

VI. Numerical Results

The simulation results presented here consider an uplink single-cell one-bit massive MIMO system with a coherence interval of $T = 200$ symbols. Unless otherwise indicated, we assume $\rho_{\rm p} = \rho_{\rm d} = {\rm SNR}$.

A. Channel Estimation Performance

In this subsection, we evaluate the performance of the BLMMSE channel estimator proposed in Section III-A compared with the LS channel estimator of [21] and the near maximum-likelihood channel estimator of [25]. Note that although the nML channel estimator proposed in [25] focused on estimating the channel vector between the K users and one receive antenna, we can define the nML estimator for the entire channel for all M receive antennas and K users using logic similar to [25] as follows:

$$\hat{\mathbf{h}}^{\rm nML} = \underset{\substack{\acute{\mathbf{h}}_{\rm R} \in \mathbb{R}^{2MK\times 1} \\ \|\acute{\mathbf{h}}_{\rm R}\|^2 \leq K}}{\arg\max} \sum_{i=1}^{2M\tau} \log\left(F\left(\sqrt{2}\bar{\boldsymbol{\varphi}}_{\rm R}^{(i)}\acute{\mathbf{h}}_{\rm R}\right)\right), \tag{57}$$

where $F(x)$ is the cumulative distribution function (CDF) of the standard normal distribution, and $\bar{\boldsymbol{\varphi}}_{\rm R}^{(i)} = \sqrt{2} r_{\rm R,p}^{(i)} \bar{\boldsymbol{\Phi}}_{\rm R}^{(i)} \cdot r_{\rm R,p}^{(i)}$ and $\bar{\boldsymbol{\Phi}}_{\rm R}^{(i)}$ are respectively the ith element of $\mathbf{r}_{\rm R,p}$ and the ith row of $\bar{\boldsymbol{\Phi}}_{\rm R}$:

$$\mathbf{r}_{\rm R,p} = [\Re(\mathbf{r}_{\rm p}) \quad \Im(\mathbf{r}_{\rm p})]^T \tag{58}$$

$$\bar{\boldsymbol{\Phi}}_{\rm R} = \begin{bmatrix} \Re\left(\bar{\boldsymbol{\Phi}}\right) & -\Im\left(\bar{\boldsymbol{\Phi}}\right) \\ \Im\left(\bar{\boldsymbol{\Phi}}\right) & \Re\left(\bar{\boldsymbol{\Phi}}\right) \end{bmatrix}. \tag{59}$$

Figure 2 compares the MSE of the various channel estimators as a function of SNR for a case with $M = 16$, $K = 4$ and $\tau = 20$. Note that we also include the performance of a similar channel estimator proposed in [28], in which the quantizer noise is modeled as uncorrelated additive noise with a covariance matrix $\mathbf{C}_{\mathbf{q}_{\rm p}} = (1-2/\pi)\mathbf{I}_{M\tau}$. We emphasize again that in our work, the correlation between the elements of the quantization noise vector is taken into account using the

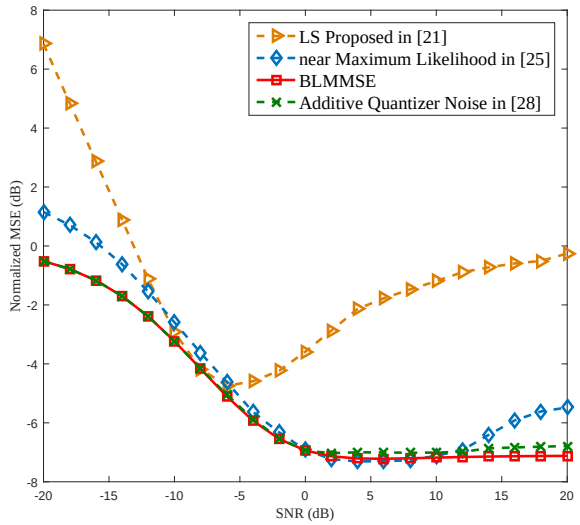

Fig. 2: MSE of channel estimators versus SNR with $M = 16, K = 4$ and $\tau = 20$. Least Squares estimator is from [21] and nML estimator is from [25].

arcsine law, and hence $\mathbf{C}_{\mathbf{q}_\mathrm{p}}$ is not in general a diagonal matrix. We see that our proposed BLMMSE approach outperforms the other previously proposed approaches. We also see that at low SNR, BLMMSE and the method based on uncorrelated quantization noise achieves the same performance, which verifies the observation that the approximation of (32) is reasonable at low SNR. However, with the increase of SNR, a small performance gap can be seen between these two curves, indicating that not considering the correlation between the quantizer noise in one-bit systems may cause performance loss and the correlation should be taken into account.

A larger gap will result in cases where the quantizer noise is spatially correlated, since the analysis of [28] did not take this possibility into account. This will occur for example if the channel or the additive noise is itself spatially correlated. For example, take the simple case depicted in Fig. 3 for $M = 16, K = 1$ and $\tau = 2$, which shows the MSE performance for a case with a spatially correlated channel where $\mathbf{C}_{\underline{\mathbf{h}}}$ is non-diagonal. In this case, the BLMMSE channel estimator is given by

$$\hat{\underline{\mathbf{h}}}^{\mathtt{BLM}} = \mathbf{C}_{\underline{\mathbf{h}}}(\mathbf{A}_\mathrm{p}\bar{\mathbf{\Phi}})^H \mathbf{C}_{\mathbf{r}_\mathrm{p}}^{-1}\mathbf{r}_\mathrm{p}, \tag{60}$$

where, following the same step as in (7), the matrix $\mathbf{A}_\mathrm{p}$ is

$$\mathbf{A}_\mathrm{p} = \sqrt{\frac{2}{\pi}}\mathrm{diag}\left(\bar{\mathbf{\Phi}}\mathbf{C}_{\underline{\mathbf{h}}}\bar{\mathbf{\Phi}}^H + \mathbf{I}_{M\tau}\right)^{-\frac{1}{2}}. \tag{61}$$

For this example, we consider a typical urban channel model as described in [44], where the power angle spectrum of the channel is modeled by a Laplacian distribution with an angle spread of $10°$. The covariance matrix $\mathbf{C}_{\underline{\mathbf{h}}}$ can then be obtained according to [45, Eq. (2)]. We can see that the MSE performance gap grows to over 1 dB, indicating that the spatial correlation of the quantizer noise has an impact on performance and should be taken into account.

B. Validation of Achievable Rate Results

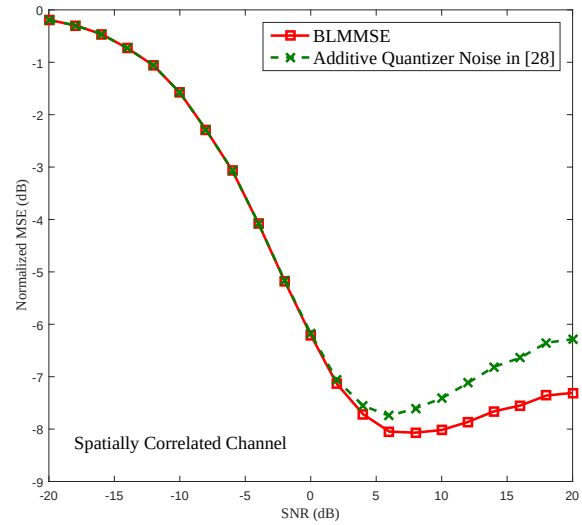

Fig. 3: MSE of channel estimators versus SNR with $M = 16, K = 1$ and $\tau = 2$ over a spatially correlated channel.

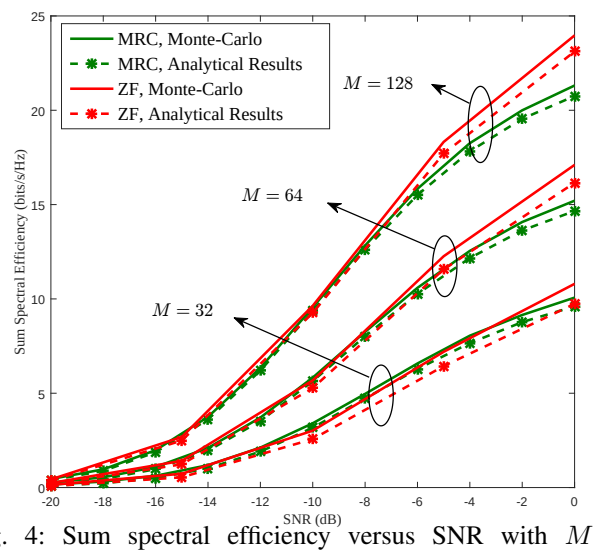

Fig. 4: Sum spectral efficiency versus SNR with $M = \{32, 64, 128\}$ and $K = \tau = 8$ for MRC and ZF receivers.

Here we evaluate the validity of the lower bounds on the achievable rate for the MRC and ZF receivers derived in Theorems 1 and 2 compared with the ergodic rate given in (42). Fig. 4 shows the sum spectral efficiency versus SNR with $K = \tau = 8$ for different numbers of transmit antennas $M = \{32, 64, 128\}$. The dashed lines represent the sum spectral efficiencies obtained using the closed-form expressions in (48) and (49) for the MRC and ZF receivers, respectively, while the solid lines represent the ergodic sum spectral efficiencies obtained from (42). For both the MRC and ZF receiver, the gap between the approximation and the lower bound of the ergodic rate is small. For example, with $M = 128$ and $\mathrm{SNR} = -10$dB, the sum spectral efficiency gap is 0.19 bits/s/Hz and 0.38 bits/s/Hz for the MRC and ZF receivers, respectively. This implies that the approximation on the achievable rate given in (45) is a good predictor of the performance of one-bit massive MIMO systems. Thus, in the following plots we will show only the approximation when evaluating performance.

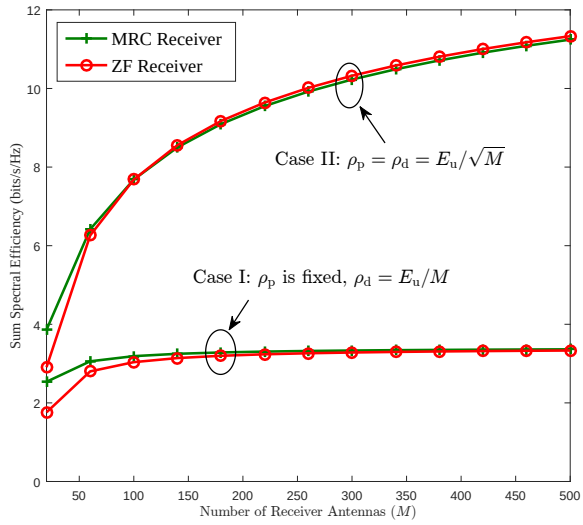

Fig. 5: Sum spectral efficiency versus number of BS antennas M for MRC and ZF receivers with $\rho_{\rm p} = 0$dB, $\rho_{\rm d} = E_{\rm u}/M$ in Case I, and $\rho_{\rm p} = \rho_{\rm d} = E_{\rm u}/\sqrt{M}$ in Case II.

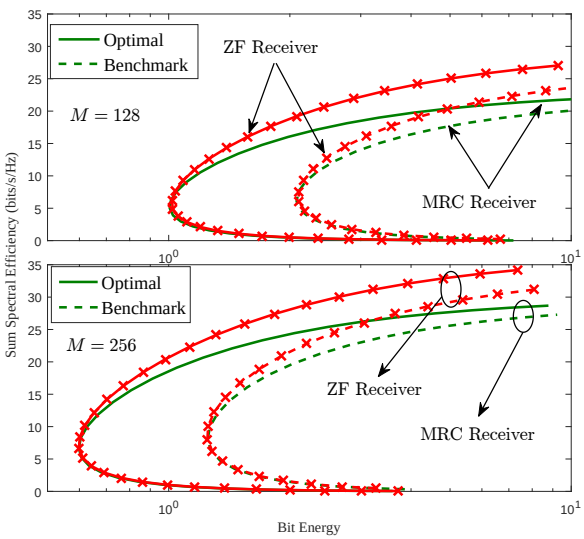

Fig. 6: Bit energy versus sum spectral efficiency with and without resource allocation for $M = \{128, 256\}$ and $K = 8$.

C. One-Bit Massive MIMO Power Efficiency

This example considers the power efficiency of using large antenna arrays in one-bit massive MIMO for the two cases considered in Section V-A. Fig. 5 shows the sum spectral efficiency versus the number of receive antennas with $K = \tau = 8$ for the MRC and ZF receivers for Cases I and II. In Case I, we assume $\rho_{\rm p} = 10$dB is fixed and $\rho_{\rm d} = E_{\rm u}/M$, while in Case II we choose $\rho_{\rm p} = \rho_{\rm d} = E_{\rm u}/\sqrt{M}$, where $E_{\rm u} = 0$dB. As predicted by the analysis, in Case I the sum spectral efficiency converges to the same constant value for both the MRC and ZF receivers. In Case II where $\rho_{\rm p} = \rho_{\rm d} = E_{\rm u}/\sqrt{M}$, the sum spectral efficiency also converges to a constant value for both the MRC and ZF receivers, although the constant is only reached for very large M.

D. Resource Allocation

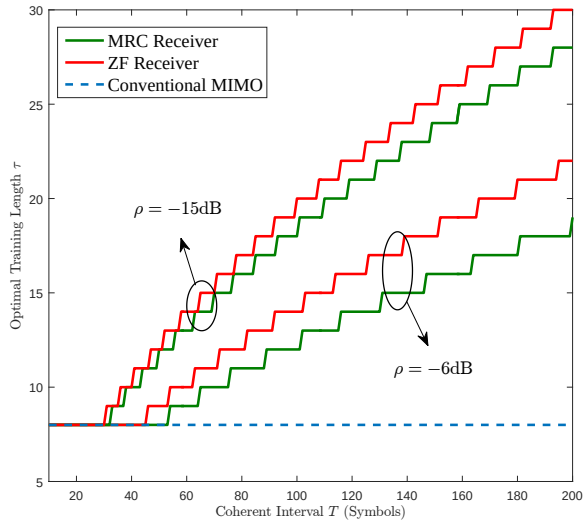

Fig. 7: Optimal training length versus the coherence interval with $M = 128$ and average transmit power $\rho = \{-15, -6\}$dB for the MRC and ZF receivers.

We now investigate the benefit of our proposed optimal resource allocation scheme that adjusts the training length, training power, and data transmission power. In order to illustrate the benefit achieved by our proposed allocation scheme, we define the bit energy as the total transmit power expended divided by the sum spectral efficiency, or energy consumed per transmitted bit:

$$\zeta_{\rm A} = \frac{\tau\rho_{\rm p} + (T-\tau)\rho_{\rm d}}{\mathcal{S}_{\rm A}}. \tag{62}$$

Fig. 6 shows the sum spectral efficiency versus the bit energy with and without optimal power allocation for $M = \{128, 256\}$ and for the MRC and ZF receivers. The 'Benchmark' curves correspond to choosing $\tau = K$ and $\rho_{\rm p} = \rho_{\rm d}$, while the 'Optimal' curves are obtained using the optimal resource allocation of (55). Different points on the curves correspond to different values of total available power. The benefit of an optimal power allocation is very evident in all cases. For example, to achieve a sum spectral efficiency of 15 bits/s/Hz with $M = 128$, the optimal resource allocation can reduce the bit energy by a factor of 1.9 for both the MRC and ZF receivers compared to the benchmark case. The improvement in bit energy achieved by increasing the number of antennas is also apparent. For a sum spectral efficiency of 15 bits/s/Hz and using the optimal resource allocation, we can reduce the bit energy by a factor of about 2.2 for the MRC and ZF receivers, by doubling the number of antennas from 128 to 256.

Fig. 7 shows the optimal training duration versus the length of the coherence interval for $M = 128$, $K = 8$ and average transmit power $\rho = \{-15, -6\}$dB for conventional and one-bit massive MIMO systems. We can see that the optimal training length is always equal to the number of users for conventional massive MIMO systems, while it depends on the coherence interval and the total power budget for one-bit MIMO systems. This is because a larger proportion of the coherence interval devoted to training is required in one-

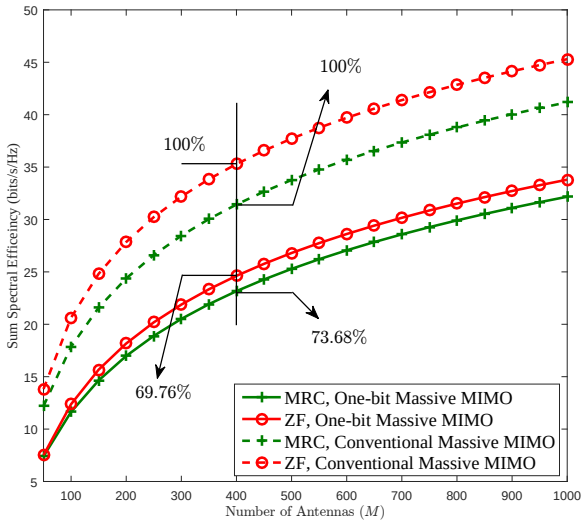

Fig. 8: Comparison of the sum spectral efficiency versus number of receive antennas for one-bit and conventional massive MIMO systems with average transmit power $\rho = -10$dB.

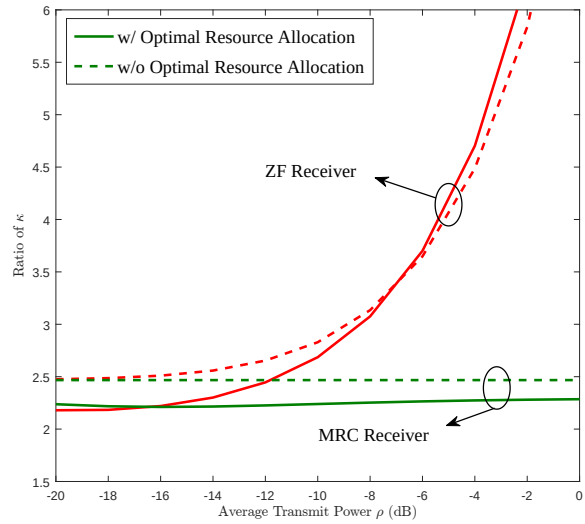

Fig. 9: The ratio of κ versus the average transmit power ρ with $K = 8$ for the MRC and ZF receivers.

bit systems to combat the quantization noise. In addition, we observe that the optimal training length for the MRC receiver is smaller than that for the ZF receiver, implying that the ZF receiver demands a higher quality channel estimate than MRC in order to reduce the interuser interference, and hence improve the sum spectral efficiency.

E. Number of Antennas for One-Bit and Conventional Massive MIMO

In this example we compare the sum spectral efficiencies between one-bit and conventional massive MIMO systems. Fig. 8 illustrates the sum spectral efficiency versus the number of receive antennas for the MRC and ZF receivers with an average transmit power $\rho = -10$dB. Since we are more interested in comparing the maximum sum spectral efficiencies of both one-bit and conventional systems, each curve is obtained by adjusting the training length, the training power and data transmission power to maximize the sum spectral efficiency, as in problem (55). The curves for 'Conventional massive MIMO' are obtained using the formulas in Table I. Compared with the conventional system, the rate loss of the one-bit system is not as severe as might be imagined. For example, with $M = 400$, the one-bit system can still achieve a sum spectral efficiency of 23.2 bits/s/Hz and 24.6 bits/s/Hz for the MRC and ZF receivers, respectively, which amounts to 73.68% and 69.76% of the sum spectral efficiency of the conventional system. This is a remarkably high value for such a coarsely quantized signal that only retains sign information about the received signals. The figure also verifies the increase in the number of antennas required for the one-bit system with MRC to achieve performance equivalent to a conventional massive MIMO system; the one-bit system requires about 480 antennas, or approximately $480/215 = 2.23$ times more antennas than for a conventional system to achieve a spectral efficiency of 25 bits/s/Hz.

This relationship is further illustrated in Fig. 9 which shows the ratio of $\kappa = M_{\text{one}}/M_{\text{conv}}$ needed for the two types of systems to achieve equivalent performance. The curves labeled 'w/o Optimal Resource Allocation' are obtained assuming $\tau = K$ and $\rho_{\text{p}} = \rho_{\text{d}} = \rho$, while the curves labeled 'w/ Optimal Resource Allocation' are obtained by solving problem (56). We can see that the ratio is constant at 2.5 for MRC and also at low SNR for ZF for the case without resource allocation, which verifies the conclusion in [28]. However, for the case with an optimal resource allocation, the ratio is around 2.2-2.3, which implies that fewer antennas are needed for the one-bit system if its performance is optimized. In addition, we see that as the average transmit power ρ increases, the number of antennas required for a one-bit system to have equivalent performance with the ZF receiver grows without bound, since the conventional ZF receiver is theoretically able to obtain a better and better channel estimate that allows it to ultimately eliminate all inter-user interference.

VII. Conclusions

This paper has investigated channel estimation and overall system performance for the single-cell, flat Rayleigh fading massive MIMO uplink when one-bit ADCs are employed at the BS. We used the Bussgang decomposition to derive a new channel estimator based on the LMMSE criteria, and showed that the resulting BLMMSE estimator provides the lowest MSE among various competing algorithms. However, even the BLMMSE estimator has a high-SNR error floor due to the one-bit quantization. We derived simple closed-form approximations for the massive MIMO uplink achievable rate for low SNR and a large number of users assuming MRC and ZF receivers that employ the BLMMSE channel estimate. We then used the approximation to study the sum spectral efficiency and energy efficiency of the one-bit massive MIMO uplink. Our results show that massive MIMO still yields similar gains in energy efficiency when one-bit quantizers are employed, and we developed an optimization problem that when solved yields significant gains in spectral efficiency by properly selecting the training length, training power and data

transmission power. We showed that for an MRC receiver with optimal resource allocation, approximately 2.2-2.3 times more antennas are required in a one-bit massive MIMO system to achieve the same spectral efficiency as a conventional system with full-precision ADCs. However, significantly more antennas are required in a one-bit system for the ZF receiver at high SNR. Finally, we presented a number of simulation results that validate our analysis and illustrate the potential performance of massive MIMO systems with one-bit ADCs.

Appendix A

For a given $\mathbf{y}_{\rm p}$, the covariance matrix between the quantizer noise $\mathbf{q}_{\rm p}$ and the channel vector $\underline{\mathbf{h}}$ can be expressed as

$$\mathrm{E}\left\{\mathbf{q}_{\rm p}\underline{\mathbf{h}}^H\right\} = \mathrm{E}_{\mathbf{y}_{\rm p}}\left\{\mathrm{E}\left\{\mathbf{q}_{\rm p}\underline{\mathbf{h}}^H|\mathbf{y}_{\rm p}\right\}\right\}. \tag{63}$$

Since the quantizer noise $\mathbf{q}_{\rm p} = \mathbf{r}_{\rm p} - \mathbf{A}_{\rm p}\mathbf{y}_{\rm p}$ is fixed for a given $\mathbf{y}_{\rm p}$, we can remove $\mathbf{q}_{\rm p}$ from the inner expectation of (63) to obtain

$$\mathrm{E}_{\mathbf{y}_{\rm p}}\left\{\mathrm{E}\left\{\mathbf{q}_{\rm p}\underline{\mathbf{h}}^H|\mathbf{y}_{\rm p}\right\}\right\} = \mathrm{E}_{\mathbf{y}_{\rm p}}\left\{\mathbf{q}_{\rm p}\mathrm{E}\left\{\underline{\mathbf{h}}^H|\mathbf{y}_{\rm p}\right\}\right\}. \tag{64}$$

According to [38], the value of $\mathrm{E}\left\{\underline{\mathbf{h}}^H|\mathbf{y}_{\rm p}\right\}$ is the linear MMSE estimate of $\underline{\mathbf{h}}$, leading to

$$\mathrm{E}_{\mathbf{y}_{\rm p}}\left\{\mathrm{E}\left\{\mathbf{q}_{\rm p}\underline{\mathbf{h}}^H|\mathbf{y}_{\rm p}\right\}\right\} = \mathrm{E}_{\mathbf{y}_{\rm p}}\left\{\mathbf{q}_{\rm p}\mathbf{y}_{\rm p}^H\mathbf{C}_{\mathbf{y}_{\rm p}}^{-1}\mathbf{C}_{\mathbf{y}_{\rm p}\underline{\mathbf{h}}}\right\}. \tag{65}$$

Choosing $\mathbf{A}_{\rm p}$ according to (7), the quantizer noise $\mathbf{q}_{\rm p}$ is uncorrelated with $\mathbf{y}_{\rm p}$, and hence we have

$$\mathrm{E}\left\{\mathbf{q}_{\rm p}\underline{\mathbf{h}}^H\right\} = \mathrm{E}_{\mathbf{y}_{\rm p}}\left\{\mathbf{q}_{\rm p}\mathbf{y}_{\rm p}^H\mathbf{C}_{\mathbf{y}_{\rm p}}^{-1}\mathbf{C}_{\mathbf{y}_{\rm p}\underline{\mathbf{h}}}\right\} = \mathbf{0}, \tag{66}$$

which implies that the quantizer noise $\mathbf{q}_{\rm p}$ is uncorrelated with the channel $\underline{\mathbf{h}}$.

Appendix B

We follow the approach of [46] and only exploit knowledge of the average effective channel $\mathrm{E}\left\{\sqrt{\rho_{\rm d}}\mathbf{w}_k^T\mathbf{A}_{\rm d}\mathbf{h}_k\right\}$ in the detection. Then, according to [41], the lower bound of the achievable rate in (45) is obtained by treating the uncorrelated inter-user interference and the quantizer noise as independent Gaussian noise, which is a worst-case assumption when computing the mutual information [41]. Therefore the variance of the effective noise is

$$\begin{aligned}\mathrm{E}\{|\tilde{n}_{{\rm d},k}|^2\} =& \mathrm{Var}\left\{\left(\mathbf{w}_k^T\mathbf{A}_{\rm d}\mathbf{h}_k\right)\right\} + \rho_{\rm d}\sum_{i\neq k}^{K}\mathrm{E}\left\{\left|\mathbf{w}_k^T\mathbf{A}_{\rm d}\mathbf{h}_i\right|^2\right\} \\ &+ \mathrm{E}\left\{\left\|\mathbf{w}_k^T\mathbf{A}_{\rm d}\right\|^2\right\} + \mathrm{E}\left\{\mathbf{w}_k^T\mathbf{C}_{\mathbf{q}_{\rm d}}\mathbf{w}_k^T\right\},\end{aligned} \tag{67}$$

where the expectation operation is taken with respect to the channel realizations. By using the same result in (32) at low SNR, we can approximate the quantizer noise as

$$\mathrm{E}\left\{\mathbf{w}_k^T\mathbf{C}_{\mathbf{q}_{\rm d}}\mathbf{w}_k^T\right\} = \left(1-\frac{2}{\pi}\right)\mathrm{E}\left\{\left\|\mathbf{w}_k^T\right\|^2\right\}. \tag{68}$$

Substituting $\mathbf{A}_{\rm d} = \alpha_{\rm d}\mathbf{I}_K$ and combining (67) and (68), we arrive at Lemma 1.

Appendix C

From (45), we need to compute $\mathrm{E}\left\{\mathbf{w}_k^T\mathbf{h}_k\right\}$, $\mathrm{Var}\left(\mathbf{w}_k^T\mathbf{h}_k\right)$, UI_k and AQN_k. Note that, although the channel vector $\mathbf{h}_k$ is Gaussian, the BLMMSE channel estimate $\hat{\mathbf{h}}_k$ is not Gaussian due to the quantizer noise. However, we can approximate $\hat{\mathbf{h}}_k$ as Gaussian using Cramér's central limit theorem [47].

For the MRC receiver $\mathbf{W}_{\rm MRC}^T = \hat{\mathbf{H}}^H$, we have

$$\mathbf{w}_k^T\mathbf{h}_k = \hat{\mathbf{h}}_k^H\mathbf{h}_k = \left\|\hat{\mathbf{h}}_k\right\|^2 + \hat{\mathbf{h}}_k^H\boldsymbol{\varepsilon}_k. \tag{69}$$

Therefore,

$$\mathrm{E}\left\{\hat{\mathbf{h}}_k^H\mathbf{h}_k\right\} = \mathrm{E}\left\{\left\|\hat{\mathbf{h}}_k\right\|^2\right\} = M\sigma^2. \tag{70}$$

The variance of $\mathbf{w}_k^T\mathbf{h}_k$ is given by

$$\begin{aligned}\mathrm{Var}\left(\mathbf{w}_k^T\mathbf{h}_k\right) &= \mathrm{E}\left\{\left\|\hat{\mathbf{h}}_k^H\mathbf{h}_k\right\|^2\right\} - M^2\sigma^4 \\ &= \mathrm{E}\left\{\left\|\hat{\mathbf{h}}_k\right\|^4\right\} + \mathrm{E}\left\{\left\|\hat{\mathbf{h}}_k^H\boldsymbol{\varepsilon}_k\right\|^2\right\} - M^2\sigma^4.\end{aligned} \tag{71}$$

Since $\hat{\mathbf{h}}_k$ is approximately Gaussian with variance of each element $M\sigma^2$, we obtain

$$\begin{aligned}\mathrm{Var}\left(\mathbf{w}_k^T\mathbf{h}_k\right) &= \sigma^4M(M+1) + \sigma^2(1-\sigma^2)M - M^2\sigma^4 \\ &= M\sigma^2.\end{aligned} \tag{72}$$

For $i \neq k$ we have

$$\mathrm{UI}_k = \rho_{\rm d}\alpha_{\rm d}^2\sum_{i\neq k}^{K}\mathrm{E}\left\{\left|\hat{\mathbf{h}}_k^H\mathbf{h}_i\right|^2\right\} = (K-1)\rho_{\rm d}\alpha_{\rm d}^2M\sigma^2. \tag{73}$$

Similarly, we obtain

$$\mathrm{AQN}_k = (\alpha_{\rm d}^2 + 1 - 2/\pi)M\sigma^2. \tag{74}$$

Substituting (70), (72), (73) and (74) into (45), Theorem 1 is obtained.

Appendix D

For the ZF receiver $\mathbf{W}_{\rm ZF}^T = \left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^H$, we have

$$\mathbf{W}_{\rm ZF}^T\mathbf{H} = \mathbf{W}_{\rm ZF}^T(\hat{\mathbf{H}} + \mathcal{E}) = \mathbf{I}_K + \mathbf{W}_{\rm ZF}^T\mathcal{E}. \tag{75}$$

Therefore,

$$\mathbf{w}_{{\rm ZF},k}^T\mathbf{h}_k = 1 + \mathbf{w}_{{\rm ZF},k}^T\boldsymbol{\varepsilon}_k. \tag{76}$$

Similar to the derivation of the MRC receiver, we need to compute $\mathrm{E}\left\{\mathbf{w}_k^T\mathbf{h}_k\right\}$, $\mathrm{Var}\left(\mathbf{w}_k^T\mathbf{h}_k\right)$, UI_k and AQN_k.

For the ZF receiver, we have

$$\mathrm{E}\left\{\mathbf{w}_k^T\mathbf{h}_k\right\} = 1 + \mathrm{E}\left\{\mathbf{w}_{{\rm ZF},k}^T\boldsymbol{\varepsilon}_k\right\} = 1. \tag{77}$$

The variance of $\mathbf{w}_k^T\mathbf{h}_k$ is given by

$$\begin{aligned}\mathrm{Var}\left(\mathbf{w}_k^T\mathbf{h}_k\right) &= \mathrm{E}\left\{\left\|\mathbf{w}_{{\rm ZF},k}^T\boldsymbol{\varepsilon}_k\right\|^2\right\} \\ &= (1-\sigma^2)\mathrm{E}\left\{\left\|\mathbf{w}_{{\rm ZF},k}^T\right\|^2\right\} \\ &= (1-\sigma^2)\mathrm{E}\left\{\left[\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\right]_{k,k}\right\}.\end{aligned} \tag{78}$$

Since $\hat{\mathbf{H}}$ is approximately Gaussian, $\hat{\mathbf{H}}^H\hat{\mathbf{H}}$ is a $K\times K$ central Wishart matrix with M degrees of freedom, Thus,

$$\mathrm{Var}\left(\mathbf{w}_k^T\mathbf{h}_k\right)=\frac{(1-\sigma^2)}{\sigma^2(M-K)}. \tag{79}$$

From (76), for $i\neq k$ we have

$$\begin{aligned}\mathrm{UI}_k&=\rho_{\mathrm{d}}\alpha_{\mathrm{d}}^2\sum_{i\neq k}^{K}\mathrm{E}\left\{\left|\hat{\mathbf{h}}_k^H\boldsymbol{\varepsilon}_i\right|^2\right\}\\&=\rho_{\mathrm{d}}\alpha_{\mathrm{d}}^2\sum_{i\neq k}^{K}(1-\sigma^2)\mathrm{E}\left\{\left[\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\right]_{k,k}\right\}\\&=(K-1)\rho_{\mathrm{d}}\alpha_{\mathrm{d}}^2\frac{(1-\sigma^2)}{\sigma^2(M-K)}.\end{aligned} \tag{80}$$

Similarly,

$$\mathrm{AQN}_k=\frac{\alpha_{\mathrm{d}}^2+1-2/\pi}{\sigma^2(M-K)}. \tag{81}$$

Substituting (77), (79), (80) and (81) into (45), Theorem 2 is obtained.

Appendix E

First we rewrite the sum spectral efficiency of (48) and (49) for the MRC and ZF receivers as a function with respect to γ and τ:

$$\mathcal{S}_{\mathbb{A}}(\gamma,\tau)=\frac{T-\tau}{T}K\log_2\left(1+\frac{a_1\tau}{a_2\tau^2+a_3\tau+a_4}\right), \tag{82}$$

where we define

$$\begin{aligned}&a_1=4MP^2(\gamma-\gamma^2)\qquad a_2=\pi^2+2\pi P\gamma\\&a_3=\pi(KP(\pi-2)\gamma-KP(1-\gamma)(\pi+2P\gamma)-(\pi+2P\gamma)T)\\&a_4=\pi(K^2P^2(\pi-2)(-1+\gamma)\gamma-KP(\pi-2)\gamma T)\end{aligned}$$

for $\mathbb{A}=\mathtt{MRC}$, and

$$\begin{aligned}&a_1=4(M-K)P^2(\gamma-\gamma^2)\qquad a_2=\pi^2+2\pi P\gamma\\&a_3=-KP(2\pi(\gamma+P(\gamma-\gamma^2))+4P(\gamma-\gamma^2)+\pi^2(2\gamma-1))\\&\quad-(\pi^2+2\pi P\gamma)T\\&a_4=\pi(K^2P^2(\pi-2)(-1+\gamma)\gamma-KP(\pi-2)\gamma T)\end{aligned}$$

for $\mathbb{A}=\mathtt{ZF}$.

Then we denote $\{\gamma^*,\tau^*\}$ to be the solution of (55), such that $\gamma^*P=\tau^*\rho_{\mathrm{p}}^*$ is the optimal power for training, and $(1-\gamma^*)P=(T-\tau^*)\rho_{\mathrm{d}}^*$ is the optimal amount for data transmission. Next we choose $\bar{\tau}=K,\bar{\rho}_p=\gamma^*P/\bar{\tau}$ and $\bar{\rho}_d=(1-\gamma^*)P/(T-\bar{\tau})$. Clearly, the function in (82) is not a monotonic function with respect to τ with a given γ^*. That is to say, it is difficult to compare the values of $\mathcal{S}(\gamma^*,\tau^*)$ and $\mathcal{S}(\gamma^*,\bar{\tau})$. Therefore, we conclude that the optimal training length is not always equal to the number of users for one-bit systems.

References

[1] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transactions on Wireless Communications*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[2] H. Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[3] E. Larsson, O. Edfors, F. Tufvesson, and T. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 186–195, February 2014.

[4] F. Rusek, D. Persson, B. K. Lau, E. Larsson, T. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.

[5] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.

[6] E. Björnson, J. Hoydis, M. Kountouris, and M. Debbah, "Massive MIMO systems with non-ideal hardware: Energy efficiency, estimation, and capacity limits," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 7112–7139, Nov 2014.

[7] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4353–4368, Aug 2015.

[8] X. Zhang, M. Matthaiou, M. Coldrey, and E. Björnson, "Impact of residual transmit RF impairments on training-based MIMO systems," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2899–2911, Aug 2015.

[9] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 4, pp. 539–550, Apr 1999.

[10] "Texas instruments ADC products." [Online]. Available: http://www.ti.com/lsds/ti/data-converters/analog-to-digital-converter-products.page

[11] A. Mezghani and J. Nossek, "Analysis of Rayleigh-fading channels with 1-bit quantized output," in *IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 260–264.

[12] J. A. Nossek and M. T. Ivrlač, "Capacity and coding for quantized MIMO systems," in *Proceedings of the international conference on Wireless communications and mobile computing*. ACM, 2006, pp. 1387–1392.

[13] A. Mezghani, M.-S. Khoufi, and J. Nossek, "A modified MMSE receiver for quantized MIMO systems," in *Proc. ITG/IEEE Workshop on Smart Antennas (WSA)*, Feb 2007.

[14] A. Mezghani, M. Rouatbi, and J. Nossek, "An iterative receiver for quantized MIMO systems," in *IEEE Mediterranean Electrotechnical Conference (MELECON)*, March 2012, pp. 1049–1052.

[15] A. Mezghani and J. Nossek, "Efficient reconstruction of sparse vectors from quantized observations," in *International ITG Workshop on Smart Antennas (WSA)*, March 2012, pp. 193–200.

[16] J. Singh, O. Dabeer, and U. Madhow, "On the limits of communication with low-precision analog-to-digital conversion at the receiver," *IEEE Transactions on Communications*, vol. 57, no. 12, pp. 3629–3639, December 2009.

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

[18] J. Singh, S. Ponnuru, and U. Madhow, "Multi-gigabit communication: The ADC bottleneck," in *IEEE International Conference on Ultra-Wideband (ICUWB)*, Sept 2009, pp. 22–27.

[19] J. Mo and R. W. Heath Jr., "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Information Theory and Applications Workshop (ITA), 2014*, Feb 2014.

[20] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 983–997, April 2016.

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

[22] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *IEEE International Conference on Communication Workshop (ICCW), 2015*, June 2015, pp. 1304–1309.

[23] ——, "Throughput analysis of massive MIMO uplink with low-resolution ADCs." [Online]. Available: https://arxiv.org/abs/1602.01139

[24] J. Mo, P. Schniter, N. González-Prelcic, and R. W. Heath Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *48th Asilomar Conference on Signals, Systems and Computers, 2014*, Nov 2014, pp. 957–961.

[25] J. Choi, J. Mo, and R. W. Heath Jr., "Near maximum-likelihood detector and channel estimator for uplink multiuser massive MIMO systems with

one-bit ADCs," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 2005–2018, May 2016.

[26] C. Mollén, J. Choi, E. G. Larsson, and R. W. Heath Jr., "Performance of linear receivers for wideband massive MIMO with one-bit ADCs," in *International ITG Workshop on Smart Antennas (WSA)*, March 2016.

[27] ——, "One-bit ADCs in wideband massive MIMO systems with OFDM transmission," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3386–3390.

[28] ——, "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Transactions on Wireless Communications*, vol. 16, no. 1, pp. 87–100, 2017.

[29] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," MIT Research Lab. Electronics, Tech. Rep. 216, 1952.

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

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

[32] O. Orhan, E. Erkip, and S. Rangan, "Low power analog-to-digital conversion in millimeter wave systems: Impact of resolution and bandwidth on performance," in *Information Theory and Applications Workshop (ITA)*, Feb 2015, pp. 191–198.

[33] Q. Bai, A. Mezghani, and J. A. Nossek, "On the optimization of ADC resolution in multi-antenna systems," in *Proceedings of the Tenth International Symposium on Wireless Communication Systems (ISWCS),*, Aug 2013.

[34] Y. Li, C. Tao, L. Liu, G. Seco-Granados, and A. L. Swindlehurst, "Channel estimation and uplink achievable rates in one-bit massive MIMO systems," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2016.

[35] A. Mezghani and J. A. Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012.

[36] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*. Tata McGraw-Hill Education, 2002.

[37] M. Biguesh and A. B. Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback," *EURASIP Journal on Applied Signal Processing*, vol. 2004, pp. 1330–1339, 2004.

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

[39] G. Jacovitti and A. Neri, "Estimation of the autocorrelation function of complex Gaussian stationary processes by amplitude clipped signals," *IEEE Transactions on Information Theory*, vol. 40, no. 1, pp. 239–245, Jan 1994.

[40] O. Bar-Shalom and A. Weiss, "DOA estimation using one-bit quantized measurements," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 3, pp. 868–884, 2002.

[41] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, April 2003.

[42] S. Diggavi and T. Cover, "The worst additive noise under a covariance constraint," *IEEE Transactions on Information Theory*, vol. 47, no. 7, pp. 3072–3081, Nov 2001.

[43] H. Q. Ngo, M. Matthaiou, and E. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 605–608, Dec 2014.

[44] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Transactions on Vehicular Technology*, vol. 49, no. 2, pp. 437–447, Mar 2000.

[45] L. You, X. Gao, X. G. Xia, N. Ma, and Y. Peng, "Pilot reuse for massive MIMO transmission over spatially correlated Rayleigh fading channels," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3352–3366, June 2015.

[46] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Transactions on Information Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[47] H. Cramér, *Random Variables and Probability Distributions*. Cambridge University Press, 2004, vol. 36.