Channel Estimation for Large-Scale Multiple-Antenna Systems Using 1-Bit ADCs and Oversampling

ZHICHAO SHAO, (Student Member, IEEE), LUKAS T. N. LANDAU, (Member, IEEE), AND RODRIGO C. DE LAMARE, (Senior Member, IEEE)
Centre for Telecommunications Studies, Pontifical Catholic University of Rio de Janeiro, Rio de Janeiro 22451-900, Brazil

Corresponding author: Zhichao Shao (zhichao.shao@cetuc.puc-rio.br)

ABSTRACT Large-scale multiple-antenna systems with large bandwidth are fundamental for future wireless communications, where the base station employs a large antenna array. In this scenario, one problem faced is the large energy consumption as the number of receive antennas scales up. Recently, low-resolution analog-to-digital converters (ADCs) have attracted much attention. Specifically, 1-bit ADCs are suitable for such systems due to their low cost and low energy consumption. This paper considers uplink large-scale multiple-antenna systems with 1-bit ADCs on each receive antenna. We investigate the benefits of using oversampling for channel estimation in terms of the mean square error and symbol error rate performance. In particular, low-resolution aware channel estimators are developed based on the Bussgang decomposition for 1-bit oversampled systems and analytical bounds on the mean square error are also investigated. Numerical results are provided to illustrate the performance of the proposed channel estimation algorithms and the derived theoretical bounds.

INDEX TERMS Large-scale multiple-antenna systems, 1-bit quantization, oversampling, channel estimation, Cramér-Rao bound.

I. INTRODUCTION

Multi-user (MU) multiple-input multiple-output (MIMO) is currently being used in many wireless communication systems like long-term evolution (LTE), which allows for a small number of antennas at the base station [1]. However, in the last decade the number of wireless devices like mobiles, laptops and sensors, has experienced an explosive growth and current MU-MIMO systems cannot serve such a large number of users due to the limited bandwidth and increased multi-user interference (MUI). With large antenna arrays at the base station (BS), large-scale (or massive) MIMO can significantly increase the spectral efficiency, mitigate the propagation loss caused by channel fading, reduce the MUI and have many other advantages as compared to current systems [2], [3]. As such, large-scale MIMO is a key technique for future wireless communication systems, in which one favorable application is the large-scale millimeter-wave (mmWave) communication system [4]. However, many different

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Xu.

configurations and deployments need to be reconsidered. For example, by using current high-resolution (8-12 bits) analog-to-digital converters (ADCs) for each element of the antenna arrays at the BS, the hardware cost and the energy consumption may become prohibitively high since the dissipated power is exponentially scaled by the number of bits [5].

The high cost and energy consumption associated with high-resolution ADCs has motivated the use of low-cost and low-resolution ADCs for large-scale MIMO systems. As one extreme case, 1-bit ADCs can largely reduce the hardware cost and energy consumption of the receiver. Many recent works have studied this area. For instance, the works in [6]–[15] have studied massive MU-MIMO systems with coarsely quantized signals operating over frequency-flat, narrowband channels. The works in [6], [7] have investigated the uplink channel capacity by MU-MIMO systems with 1-bit ADCs at the BS and [8]–[10] have analyzed different precoding techniques for the downlink. Regarding channel estimation, the studies in [11]–[13] have proposed the Bussgang linear minimum mean squared error (BLMMSE), expectation-maximization (EM) based iterative

hard thresholding (IHT) and recursive least squares (RLS) adaptive channel estimators, respectively. In the context of the signal detection used in uplink 1-bit massive MU-MIMO systems, the work in [14] proposes the iterative detection and decoding (IDD) technique together with regular LDPC codes and [15] presents a low-complexity near maximum-likelihood-detection (near-MLD) algorithm called 1-bit sphere decoding.

Moreover, some prior works have investigated 1-bit ADCs used in wideband communication systems. The works in [16]–[19] have studied massive MU-MIMO systems with coarsely quantized signals that deploy orthogonal frequency-division multiplexing (OFDM) for wideband communications. Their results show that it is satisfactory to use 1-bit ADCs in wideband massive MU-MIMO-OFDM systems. Furthermore, the studies in [20]–[22] have discussed some key transceiver design challenges, including channel estimation, signal detection, achievable rates and precoding techniques, in millimeter-Wave (mmWave) massive MIMO systems, which are promising candidates for 5G cellular systems.

The previous works have considered quantized systems with sampling at the Nyquist rate. However, utilizing oversampling at the receiver can partially compensate for the information loss brought by the coarse quantization [23]. The work in [24] has proposed faster than symbol rate (FTSR) sampling in an uplink massive MIMO system with coarsely quantized signals in terms of the symbol error rate (SER). It shows that the FTSR sampling provides about 5dB signal-to-noise ratio (SNR) advantage in terms of SER and achievable rate with a linear zero forcing receiver. The work in [25] has analyzed the achievable rate for 1-bit oversampled systems over band-limited channels. To reduce the computational cost caused by the large number of samples due to oversampling, a sliding window based linear detection has been proposed in [26]. In addition to the conventional system models based on matched filtering and correlated noise samples, alternative receiver assumptions exist in literature such as in [27], where the authors consider a wideband receiver whose bandwidth scales proportionally with the oversampling factor and has the drawback of additional received noise and interference from neighboring frequency bands.

From the channel estimation point of view, the works in [11], [12] have proposed different channel estimation techniques for systems operating at the Nyquist rate. However, only few works have considered channel estimation in oversampled systems. The study in [28] considers time-of-arrival estimation for systems with 1-bit quantization and oversampling and proposes corresponding performance bounds. The study in [29] has proposed carrier phase estimation and given lower bounds on complex channel parameter estimation for 1-bit oversampled systems based on [30]. In the study in [24] the BLMMSE channel estimator is applied to the MIMO channel with 1-bit quantization and oversampling using the simplifying assumption of uncorrelated noise samples which

then yields performance degradation especially at low SNR and high oversampling factors.

In this work, low-resolution aware (LRA) channel estimators are developed for 1-bit oversampled large-scale MIMO systems in the uplink based on the Bussgang decomposition. Although the received signals are quantized to 1 bit, the computations after the 1-bit ADCs of all algorithms compared are performed at a higher resolution (8 bits or higher). The application of oversampling at the receiver can lead to significantly better performance. Unlike prior works we explicitly consider the correlation of the filtered noise, which is a main property of oversampled systems, and employ the Bussgang decomposition [31] to reformulate the nonlinear system into a statistically equivalent linear system. Based on this linear model, low-resolution aware least-squares (LS), linear minimum mean square error (LMMSE) and least-mean square (LMS) channel estimation algorithms are proposed for 1-bit oversampled systems and evaluate their computational costs. Moreover, an adaptive technique is devised to estimate the statistical quantities resulting from the Bussgang decomposition, which are required by channel estimators. We also examine the fundamental estimation limits by deriving a Bayesian framework and bounds on channel estimation for both non-oversampled and oversampled systems. In addition to the Bayesian Cramér-Rao bounds (CRBs), general CRBs is proposed for biased estimators due to the correlation between the signal and its quantization error. In summary, our work has the following contributions:

- The LRA-LS, LRA-LMMSE and LRA-LMS channel estimation algorithms are presented for the 1-bit large-scale MIMO systems in the uplink with oversampling.
- We obtain analytical expressions associated with the Bayesian CRBs for the oversampled systems and observe that the proposed bounds are very close to the results obtained from simulations at low SNR.
- An adaptive technique is proposed to estimate the auto-correlation of the channel vector, which is an essential part for the Bussgang decomposition in 1-bit systems.

Some preliminary results have been shown in [32] and [33]. However, as compared to [32], [33], this paper extends and refines the analysis of the correlation property of filtered noise and proposes a more practical adaptive channel estimator with lower computational cost. In the section of numerical results, the performance of the proposed LRA-LMMSE estimator is compared with its simplified version in [24]. Furthermore, a comparison of the performance of systems using ADCs with more bits is also shown in this paper.

The rest of this paper is organized as follows: Section II illustrates the system model and gives some statistical properties of 1-bit quantization. Section III derives the proposed oversampling based channel estimators and analyzes the computational complexity of the estimators. Section IV gives the upper bounds of the Bayesian CRBs and the general CRBs for 1-bit non-oversampled and oversampled MIMO systems.
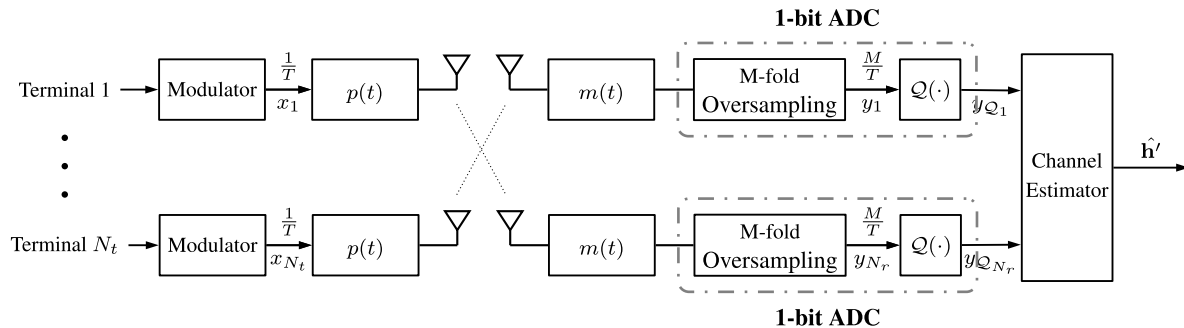
FIGURE 1. System model of 1-bit multi-user multiple-antenna system with oversampling at the receiver.

Section V compares the normalized mean square error (MSE) and SER performance of the proposed and existing channel estimators. Section VI concludes the paper.

Notation: The following notation is used throughout the paper. Matrices are in bold capital letters and vectors in bold lowercase. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix and $\mathbf{0}_n$ is the $n \times 1$ all-zero column vector. Additionally, diag($\mathbf{A}$) is a diagonal matrix only containing the diagonal elements of $\mathbf{A}$. The transpose, conjugate transpose and pseudoinverse of $\mathbf{A}$ are represented by $\mathbf{A}^T$, $\mathbf{A}^H$ and $\mathbf{A}^+$, respectively. $\mathbf{a}^*$ denotes the complex conjugate of $\mathbf{a}$ and $[\mathbf{a}]_k$ represents the kth element of vector $\mathbf{a}$. $(\cdot)^R$ and $(\cdot)^I$ get the real and imaginary part from the corresponding vector or matrix, respectively. $\otimes$ is the Kronecker product. Finally, vec($\mathbf{A}$) is the vectorized form of $\mathbf{A}$ obtained by stacking its columns and det($\mathbf{A}$) is the determinant function. $\mathbf{x} \sim \mathcal{CN}(\mathbf{a}, \mathbf{B})$ indicates that $\mathbf{x}$ is a complex Gaussian vector with mean $\mathbf{a}$ and covariance matrix $\mathbf{B}$. The expectation and covariance is denoted as $E\{\cdot\}$ and $Cov\{\cdot\}$, respectively.

II. SYSTEM MODEL AND PROBLEM STATEMENT

In this paper, we consider a single-cell multi-user large-scale MIMO system with N_t single-antenna terminals and a BS with N_r receive antennas, where each receive antenna is equipped with two 1-bit ADCs (one for the in-phase component and the other for the quadrature-phase component) and $N_r \gg N_t$. The system model is depicted in Fig. 1. In the uplink, by assuming perfect synchronization the received oversampled signal $\mathbf{y} \in \mathbb{C}^{MN_rN \times 1}$ can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^{NN_t \times 1}$ contains independent identically distributed (i.i.d.) transmitted symbols from N_t terminals, each with block length N. The vector $\mathbf{x}$ is arranged as

$$\mathbf{x} = [x_{1,1} \quad \cdots \quad x_{N,1} \quad x_{1,2} \quad \cdots \quad x_{N,N_t}]^T, \tag{2}$$

where $x_{i,j}$ corresponds to the transmitted symbol of terminal j at time instant i. Each symbol has unit power so that $E[|x_{i,j}|^2] = 1$. The vector $\mathbf{n}$ represents the filtered oversampled noise expressed by

$$\mathbf{n} = (\mathbf{I}_{N_r} \otimes \mathbf{G})\mathbf{w} \tag{3}$$

with $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}_{3MN_rN}, \sigma_n^2 \mathbf{I}_{3MN_rN})$. Note that the noise samples are described such that each entry of $\mathbf{n}$ has the same statistical properties. Since in digital domain the receive filter has a length of $2MN + 1$ samples, $3MN$ unfiltered noise samples in the noise vector $\mathbf{w}$ need to be considered for the description of an interval of MN samples of the filtered noise $\mathbf{n}$. The matrix $\mathbf{G} \in \mathbb{R}^{MN \times 3MN}$ is a Toeplitz matrix that contains the coefficients of the matched filter $m(t)$ (operated in analog domain) at different time instants and is shown in (4), as shown at the bottom of the next page, where T is the symbol period and M denotes the oversampling rate. The equivalent channel matrix $\mathbf{H}$ is described as

$$\mathbf{H} = [\mathbf{I}_{N_r} \otimes \mathbf{Z}(\mathbf{I}_N \otimes \mathbf{u})](\mathbf{H}' \otimes \mathbf{I}_N), \tag{5}$$

where $\mathbf{H}' \in \mathbb{C}^{N_r \times N_t}$ is the channel matrix for non-oversampled systems and $\mathbf{u}$ is an oversampling vector with length M, which has the form

$$\mathbf{u} = [0 \quad \cdots \quad 0 \quad 1]^T. \tag{6}$$

The matrix $\mathbf{Z} \in \mathbb{R}^{MN \times MN}$ is a Toeplitz matrix that contains the coefficients of $z(t)$ at different time instants, where $z(t)$ is the convolution of the pulse shaping filter $p(t)$ and the matched filter $m(t)$ given by (7), as shown at the bottom of the next page.

In particular, $M = 1$ refers to the non-oversampling case.

Let $\mathcal{Q}(\cdot)$ represent the 1-bit quantization function, the resulting quantized signal $\mathbf{y}_\mathcal{Q}$ is given by

$$\mathbf{y}_\mathcal{Q} = \mathcal{Q}(\mathbf{y}) = \mathcal{Q}(\mathbf{y}^R) + j\mathcal{Q}(\mathbf{y}^I). \tag{8}$$

The real and imaginary parts of $\mathbf{y}$ are quantized element-wised to $\{\pm\frac{1}{\sqrt{2}}\}$ based on the sign. The factor $\frac{1}{\sqrt{2}}$ is to make the power of each quantized signal to be one.

Since quantization strongly changes the properties of signals, some statistical properties of quantization for Gaussian input signals will be shown. For 1-bit quantization and Gaussian inputs, the cross-correlation between the unquantized signal $\mathbf{s}$ with covariance matrix $\mathbf{C_s}$ and its 1-bit quantized signal $\mathbf{s}_\mathcal{Q}$ is described by [31]

$$\mathbf{C}_{\mathbf{s}_\mathcal{Q}\mathbf{s}} = \sqrt{\frac{2}{\pi}}\mathbf{K}\mathbf{C_s}, \text{ where } \mathbf{K} = \text{diag}(\mathbf{C_s})^{-\frac{1}{2}}. \tag{9}$$

Furthermore, the covariance matrix of the 1-bit quantized signal $\mathbf{s}_Q$ can be obtained through the arcsin law [34]

$$\mathbf{C}_{\mathbf{s}_Q} = \frac{2}{\pi}\left(\sin^{-1}(\mathbf{K}\mathbf{C}_{\mathbf{s}}^R\mathbf{K}) + j\sin^{-1}(\mathbf{K}\mathbf{C}_{\mathbf{s}}^I\mathbf{K})\right). \quad (10)$$

The problem we are interested in solving in this work is to cost-effectively estimate the channel parameters in $\mathbf{H}'$.

III. CHANNEL ESTIMATION FOR UPLINK 1-BIT OVERSAMPLED MIMO

In a standard uplink implementation, the channel state information (CSI) is estimated at the BS and then used to detect the data symbols transmitted from the N_t users. Each transmission block is divided into two sub-blocks: one for pilots and another for the data symbols. Pilots are either located at the beginning of each block or spread according to a desired pattern [35]. During the training phase, each terminal simultaneously transmits τ pilot symbols to the BS, which yields

$$\mathbf{y}_p = \mathbf{H}\mathbf{x}_p + \mathbf{n}_p. \quad (11)$$

Vectorizing (11) we get

$$\begin{aligned}\mathbf{y}_p &= (\mathbf{x}_p^T \otimes \mathbf{I}_{N_r})\text{vec}(\mathbf{H}) + \mathbf{n}_p \\ &= [\mathbf{x}_p^T \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}(\mathbf{I}_\tau \otimes \mathbf{u})]\text{vec}(\mathbf{H}' \otimes \mathbf{I}_\tau) + \mathbf{n}_p \\ &= \Phi_p\mathbf{h}' + \mathbf{n}_p, \end{aligned} \quad (12)$$

where $\mathbf{h}' = \text{vec}(\mathbf{H}')$ and the equivalent pilot matrix

$$\begin{aligned}\Phi_p &= [\mathbf{x}_p^T \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}(\mathbf{I}_\tau \otimes \mathbf{u})] \\ &\quad [\mathbf{I}_{N_t} \otimes (\mathbf{e}_1 \otimes \mathbf{I}_{N_r} \otimes \mathbf{e}_1 + \cdots + \mathbf{e}_\tau \otimes \mathbf{I}_{N_r} \otimes \mathbf{e}_\tau)].\end{aligned} \quad (13)$$

The vector $\mathbf{x}_p \in \mathbb{C}^{\tau N_t \times 1}$ contains the transmitted pilots and $\mathbf{e}_n \in \mathbb{R}^{\tau \times 1}$ represents a column vector with a one in the nth element and zeros elsewhere. After processing by 1-bit ADCs, the quantized signal can be expressed as

$$\mathbf{y}_{Q_p} = Q(\Phi_p\mathbf{h}' + \mathbf{n}_p) = \tilde{\Phi}_p\mathbf{h}' + \tilde{\mathbf{n}}_p, \quad (14)$$

where $\tilde{\Phi}_p = \mathbf{A}_p\Phi_p \in \mathbb{C}^{M\tau N_r \times N_t N_r}$ and $\tilde{\mathbf{n}}_p = \mathbf{A}_p\mathbf{n}_p + \mathbf{n}_q \in \mathbb{C}^{M\tau N_r \times 1}$. The vector $\mathbf{n}_q$ is the statistically equivalent

quantization noise[1] with covariance matrix $\mathbf{C}_{\mathbf{n}_q} = \mathbf{C}_{\mathbf{y}_{Q_p}} - \mathbf{A}_p\mathbf{C}_{\mathbf{y}_p}\mathbf{A}_p^H$. The matrix $\mathbf{A}_p \in \mathbb{R}^{M\tau N_r \times M\tau N_r}$ is the Bussgang-based linear operator chosen independently from $\mathbf{y}_p$ and is given by

$$\mathbf{A}_p = \mathbf{C}_{\mathbf{y}_p\mathbf{y}_{Q_p}}^H\mathbf{C}_{\mathbf{y}_p}^{-1} = \sqrt{\frac{2}{\pi}}\mathbf{K}, \quad (15)$$

where $\mathbf{C}_{\mathbf{y}_p\mathbf{y}_{Q_p}}$ denotes the cross-correlation matrix between the received signal $\mathbf{y}_p$ and its quantized signal $\mathbf{y}_{Q_p}$

$$\mathbf{C}_{\mathbf{y}_p\mathbf{y}_{Q_p}} = \sqrt{\frac{2}{\pi}}\mathbf{K}\mathbf{C}_{\mathbf{y}_p}, \quad \text{with} \quad \mathbf{K} = \text{diag}(\mathbf{C}_{\mathbf{y}_p})^{-\frac{1}{2}}. \quad (16)$$

The formulas of (15) and (16) involve the auto-correlation matrix $\mathbf{C}_{\mathbf{y}_p}$:

$$\mathbf{C}_{\mathbf{y}_p} = \Phi_p\mathbf{R}_{\mathbf{h}'}\Phi_p^H + \mathbf{C}_{\mathbf{n}_p}, \quad (17)$$

where $\mathbf{R}_{\mathbf{h}'} = E\{\mathbf{h}'\mathbf{h}'^H\}$.

A. NOISE COVARIANCE MATRIX $\mathbf{C}_{\mathbf{n}_p}$
With (3) the auto-correlation matrix $\mathbf{C}_{\mathbf{n}_p}$ in (17) is calculated as

$$\mathbf{C}_{\mathbf{n}_p} = \sigma_n^2(\mathbf{I}_{N_r} \otimes \mathbf{G}\mathbf{G}^H). \quad (18)$$

For non-oversampled system ($M = 1$), (18) is reduced to

$$\mathbf{C}_{\mathbf{n}_p} = \sigma_n^2\mathbf{I}_{\tau N_r}. \quad (19)$$

However, for oversampled system ($M \geq 2$) (18) cannot be further simplified due to the correlation of oversampled samples. The off-diagonal elements will appear in the matrix of $\mathbf{G}\mathbf{G}^H$. One example is shown in Fig. 2, where $m(t)$ is assumed to be a normalized root-raised cosine (RRC) filter with different roll-off factors, $M = 2$ and $\tau = 10$. It can be seen that the lower the roll-off factors the more off-diagonal elements appear in $\mathbf{G}\mathbf{G}^H$, which means that for systems with low roll-off factors it is important to consider $\mathbf{C}_{\mathbf{n}_p}$ as a full matrix rather than a simplified diagonal matrix as assumed in [24].

---

[1] In this paper, we assume the quantization noise $\mathbf{n}_q$ is Gaussian distributed with zero mean and covariance $\mathbf{C}_{\mathbf{n}_q}$.

$$\mathbf{G} = \begin{bmatrix} m(-NT) & m(-NT + \frac{1}{M}T) & \dots & m(NT) & 0 & \dots & 0 \\ 0 & m(-NT) & \dots & m(NT - \frac{1}{M}T) & m(NT) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m(-NT) & m(-NT + \frac{1}{M}T) & \dots & m(NT) \end{bmatrix} \quad (4)$$

$$\mathbf{Z} = \begin{bmatrix} z(0) & z(\frac{T}{M}) & \dots & z(NT - \frac{1}{M}T) \\ z(-\frac{T}{M}) & z(0) & \dots & z(NT - \frac{2}{M}T) \\ \vdots & \vdots & \ddots & \vdots \\ z(-NT + \frac{1}{M}T) & z(-NT + \frac{2}{M}T) & \dots & z(0) \end{bmatrix}. \quad (7)$$
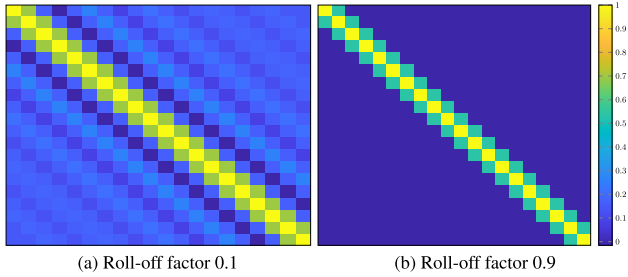
(a) Roll-off factor 0.1   (b) Roll-off factor 0.9

FIGURE 2. Matrix representation of $\mathbf{GG}^H$.

FIGURE 3. Illustration of the sliding window at each receive antenna when $l_{\text{win}} = 3$ and $M = 2$, where l_{win} is the window length representing the number of symbols sampled at the Nyquist (symbol) rate.

B. STANDARD LS CHANNEL ESTIMATOR

The work in [36] has proposed the standard LS estimator for 1-bit non-oversampled systems. Similar to this, this estimator is extended to oversampled systems, which can be computed according to

$$\hat{\mathbf{h}}'_{\text{Standard LS}} = \arg\min_{\bar{\mathbf{h}}'} ||\mathbf{y}_{\mathcal{Q}_p} - \Phi_p \bar{\mathbf{h}}'||^2$$
$$= (\Phi_p^H \Phi_p)^{-1} \Phi_p^H \mathbf{y}_{\mathcal{Q}_p}. \qquad (20)$$

The advantage of this estimator is that no a priori information is needed at the receiver. However, the issue with this estimator, when applied with 1-bit quantization, is that the channel estimate $\hat{\mathbf{h}}'$ scales with the amplitude associated with the quantizer, which then corresponds to a biased estimation.

C. LRA-LS CHANNEL ESTIMATOR

Based on the Bussgang decomposition, the LS estimate is proposed for the linear equivalent system model in (14). The LRA-LS channel estimator is obtained by solving the following optimization problem:

$$\hat{\mathbf{h}}'_{\text{LRA-LS}} = \arg\min_{\bar{\mathbf{h}}'} ||\mathbf{y}_{\mathcal{Q}_p} - \tilde{\Phi}_p \bar{\mathbf{h}}'||^2$$
$$= (\tilde{\Phi}_p^H \tilde{\Phi}_p)^{-1} \tilde{\Phi}_p^H \mathbf{y}_{\mathcal{Q}_p}. \qquad (21)$$

Compared to the standard LS channel estimator, the proposed estimator has taken $\mathbf{R}_{\mathbf{h}'}$ into consideration in order to obtain the linear operator $\mathbf{A}_p$.

D. LRA-LMMSE CHANNEL ESTIMATOR

The LMMSE channel estimator has the advantage of superior MSE performance to that of the LS channel estimator. Based on the statistically equivalent linear model in (14), the oversampling based LRA-LMMSE channel estimator is proposed. The optimal filter is given by

$$\mathbf{W}_{\text{LMMSE}} = \arg\min_{\mathbf{W}} E\{||\mathbf{h}' - \mathbf{W}\mathbf{y}_{\mathcal{Q}_p}||^2\}$$
$$= \mathbf{R}_{\mathbf{h}'} \tilde{\Phi}^H \mathbf{C}_{\mathbf{y}_{\mathcal{Q}_p}}^{-1}, \qquad (22)$$

where

$$\mathbf{C}_{\mathbf{y}_{\mathcal{Q}_p}} = \frac{2}{\pi} \left(\sin^{-1}(\mathbf{K}\mathbf{C}_{\mathbf{y}_p}^R \mathbf{K}) + j\sin^{-1}(\mathbf{K}\mathbf{C}_{\mathbf{y}_p}^I \mathbf{K}) \right). \qquad (23)$$

The resulting LRA-LMMSE channel estimator is then

$$\hat{\mathbf{h}}'_{\text{LRA-LMMSE}} = \mathbf{R}_{\mathbf{h}'} \tilde{\Phi}^H \mathbf{C}_{\mathbf{y}_{\mathcal{Q}_p}}^{-1} \mathbf{y}_{\mathcal{Q}_p}. \qquad (24)$$
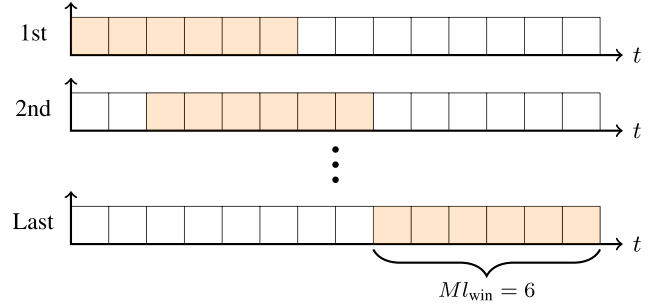
Proof: See Appendix A. ∎

Note that when $M = 1$, (24) reduces to the same as that of the BLMMSE channel estimator in [11].

E. LRA-LMS CHANNEL ESTIMATOR

LMS is the most widely used adaptive algorithm and has been adopted in various applications like system identification and channel equalization. In addition, LMS has robust performance and a low cost of implementation. Based on the linear equivalent model in (14), an LRA-LMS channel estimator for 1-bit oversampled systems is devised.

Since for large-scale MIMO with $N_r \gg N_t$, in order to reduce the computational complexity the multiplications and divisions involving large matrices, whose dimensions contain N_r elements, need to be avoided. For this reason, we concentrate on the channel from N_t users to only one receive antenna n_r and the received quantized signal is modelled as

$$\mathbf{y}_{\mathcal{Q}_p}^{n_r} = \tilde{\Phi}_p^{n_r} \mathbf{h}'^{n_r} + \tilde{\mathbf{n}}_p^{n_r}, \qquad (25)$$

where $\mathbf{y}_{\mathcal{Q}_p}^{n_r} = [y_{\mathcal{Q}_p}^{n_r}(1), y_{\mathcal{Q}_p}^{n_r}(2), \ldots, y_{\mathcal{Q}_p}^{n_r}(M\tau)]^T$ and $\mathbf{h}'^{n_r} \in \mathbb{C}^{N_t \times 1}$ is the n_rth row of $\mathbf{H}'$. Different from $\tilde{\Phi}_p$ in (14), $\tilde{\Phi}_p^{n_r} \in \mathbb{C}^{M\tau \times N_t}$ is an equivalent pilot matrix to the n_rth receive antenna. The sliding window based technique [26] (shown in Fig. 3) is applied, which combines the adjacent symbol-rate-sampled symbols together to estimate the instantaneous channel parameters, since in oversampled systems the interference from adjacent symbol-rate-sampled symbols should be considered. The first window contains the first Ml_{win} oversampled samples and the second contains the next Ml_{win} samples until the last window. Note that only one symbol-rate-sampled symbol (or M oversampled samples) is shifted for the subsequent window.

Based on (25), the received signal at the nth window can be expressed as

$$\mathbf{y}_{\mathcal{Q}_p}^{n_r}(n) = \tilde{\Phi}_p^{n_r}(n)\mathbf{h}'^{n_r} + \tilde{\mathbf{n}}_p^{n_r}(n), \qquad (26)$$

where $\mathbf{y}_{\mathcal{Q}_p}^{n_r}(n) = [y_{\mathcal{Q}_p}^{n_r}(M(n-1)+1), \ldots, y_{\mathcal{Q}_p}^{n_r}(M(n-1)+Ml_{\text{win}})]^T$ and $\tilde{\Phi}_p^{n_r}(n) = \mathbf{A}_p^{n_r}(n)\Phi_p^{n_r}(n) \in \mathbb{C}^{Ml_{\text{win}} \times N_t}$ contains the transmit pilot sequences in the nth window.

The optimization problem that leads to the proposed LRA-LMS channel estimation algorithm can be stated as

$$\hat{\mathbf{h}}'^{n_r}_{\text{LRA-LMS}}(n) = \arg\min_{\bar{\mathbf{h}}'^{n_r}(n)} \sum_{n=1}^{\tau-l_{\text{win}}+1} ||\mathbf{y}^{n_r}_{\mathcal{Q}_p}(n) - \tilde{\Phi}^{n_r}_p(n)\bar{\mathbf{h}}'^{n_r}(n)||^2,$$

$$(27)$$

where $\bar{\mathbf{h}}'^{n_r}(n)$ is the instantaneous estimate of $\mathbf{h}'^{n_r}$ in the nth window.

Taking the partial derivative of the objective function in (27) with respect to $\bar{\mathbf{h}}'^{n_r}(n)^H$, we obtain

$$\frac{\partial \sum_{n=1}^{\tau-l_{\text{win}}+1} ||\mathbf{y}^{n_r}_{\mathcal{Q}_p}(n) - \tilde{\Phi}^{n_r}_p(n)\bar{\mathbf{h}}'^{n_r}(n)||^2}{\partial \bar{\mathbf{h}}'^{n_r}(n)^H}$$

$$= \sum_{n=1}^{\tau-l_{\text{win}}+1} -\tilde{\Phi}^{n_r}_p(n)^H(\mathbf{y}^{n_r}_{\mathcal{Q}_p}(n) - \tilde{\Phi}^{n_r}_p(n)\bar{\mathbf{h}}'^{n_r}(n))$$

$$= \sum_{n=1}^{\tau-l_{\text{win}}+1} -\tilde{\Phi}^{n_r}_p(n)^H \mathbf{e}^{n_r}(n). \quad (28)$$

The recursion of the proposed LRA-LMS algorithm is

$$\bar{\mathbf{h}}'^{n_r}(n+1) = \bar{\mathbf{h}}'^{n_r}(n) + \mu\tilde{\Phi}^{n_r}_p(n)^H\mathbf{e}^{n_r}(n),$$

$$n = 1, \ldots, \tau - l_{\text{win}} + 1, \quad (29)$$

where the constant step size μ fulfills

$$0 < \mu < \frac{2}{\gamma_{\max}}. \quad (30)$$

$\gamma_{\max}$ is the largest eigenvalue of $\mathbf{C}_{\tilde{\Phi}^{n_r}_p(n)}$, which is $E\{\tilde{\Phi}^{n_r}_p(n)\tilde{\Phi}^{n_r}_p(n)^H\}$.   *Proof:* See Appendix B. ∎

The proposed adaptive channel estimator is summarized in Algorithm 1, where $\mathbf{x}_p(n) \in \mathbb{C}^{l_{\text{win}}N_t \times 1}$ contains the pilot symbols in the nth window. Both $\mathbf{e}'_n \in \mathbb{R}^{l_{\text{win}}\times 1}$ and $\mathbf{e}''_n \in \mathbb{R}^{N_r \times 1}$ represent all-zero column vectors except that the nth elements are ones. Fig. 4 shows the convergence performance of the proposed LRA-LMS channel estimator for each receive antenna. The proposed estimator achieves its steady state after $\tau = 40$.
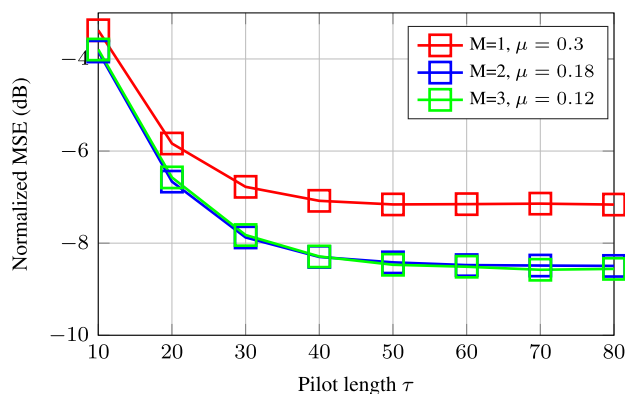

FIGURE 4. Convergence of the LRA-LMS channel estimator with $N_t = 8$ and $N_r = 64$ at SNR = 20dB.

Algorithm 1 Proposed LRA-LMS Channel Estimator

1: **Parameters:**
   μ: forgetting factor
2: **Initialization:**
   $\mathbf{h}'^{n_r}(1) = \mathbf{0}_{N_t \times 1}$
3: **Iteration:**

4: **for** $n_r = 1 : N_r$ **do**
5:    **for** $n = 1 : \tau - l_{\text{win}} + 1$ **do**
6:

      $\Phi^{n_r}_p(n) = [\mathbf{x}^T_p(n) \otimes \mathbf{Z}(\mathbf{I}_{l_{\text{win}}} \otimes \mathbf{u})]$
       $[\mathbf{I}_{N_t} \otimes (\mathbf{e}'_1 \otimes \mathbf{e}'_1 + \cdots + \mathbf{e}'_{l_{\text{win}}} \otimes \mathbf{e}'_{l_{\text{win}}})];$

7:       $\mathbf{C}^{n_r}_{\mathbf{y}_p}(n) = \Phi^{n_r}_p(n)\Phi^{n_r}_p(n)^H + \sigma^2_n\mathbf{G}\mathbf{G}^H;$
8:       $\mathbf{A}^{n_r}_p(n) = \sqrt{\frac{2}{\pi}}\text{diag}(\mathbf{C}^{n_r}_{\mathbf{y}_p}(n))^{-\frac{1}{2}};$
9:       $\tilde{\Phi}^{n_r}_p(n) = \mathbf{A}^{n_r}_p(n)\Phi^{n_r}_p(n);$
10:     $\mathbf{e}^{n_r}(n) = \mathbf{y}^{n_r}_{\mathcal{Q}_p}(n) - \tilde{\Phi}^{n_r}_p(n)\mathbf{h}'^{n_r}(n);$
11:     $\mathbf{h}'^{n_r}(n+1) = \mathbf{h}'^{n_r}(n) + \mu\tilde{\Phi}^{n_r}_p(n)^H\mathbf{e}^{n_r}(n);$
12:   **end for**
13: **end for**

F. COMPLEXITY ANALYSIS

The computational complexities of the proposed channel estimators are compared in this subsection. For the sake of simplification and a fair comparison among the estimators, we assume $\mathbf{R}_{\mathbf{h}'}$ is an identity matrix. Table 1 shows the total required complex additions/subtractions and multiplications/divisions for obtaining the channel estimate $\hat{\mathbf{h}}'$. More intuitively, Fig. 5 shows the total number of complex operations, which is a sum of complex additions and multiplications, as a function of the number of receive antennas N_r. Compared to other channel estimators, the LRA-LMS channel estimator consumes the lowest computational cost since there are no matrix inversions or large matrix multiplications in the algorithm. The comparisons in terms of MSE performance are shown in the simulations section.
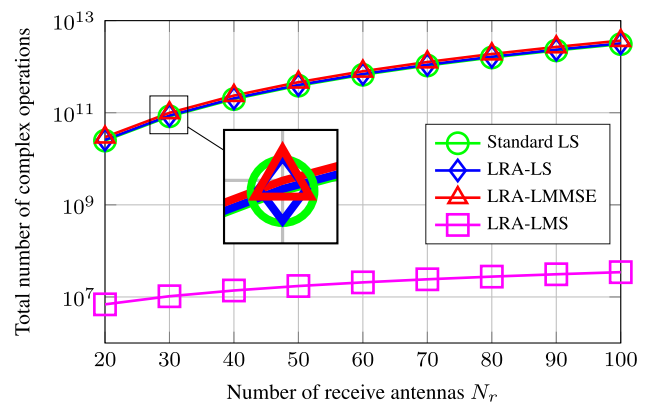

FIGURE 5. Computational complexity comparison between different channel estimators in an oversampled system $M = 3$ with $\tau = 20$, $l_{\text{win}} = 3$ and $N_t = 8$.

TABLE 1. Computational complexity of different channel estimators.

	Complex Additions/Subtractions	Complex Multiplications/Divisions
Standard LS	$N_r^3 N_t^2 (N_t + M\tau^3 + 2M\tau)$ $-N_r^2 N_t (M\tau + N_t) - 2N_r N_t$ $+M\tau^2 (M\tau - 1)$	$N_r^3 N_t^2 (N_t + 2M\tau + \tau^3 M)$ $+N_r^2 \tau [\tau N_t^2 + (\tau^2 + 1)MN_t + \tau^2 + (1+M)\tau]$ $+2N_r N_t + M\tau^2 (1 + \tau)$
LRA-LS	$N_r^3 N_t [N_t^2 + (M\tau^3 + 2M\tau)N_t + 2M^2\tau^2]$ $-N_r^2 N_t (2M\tau + N_t) - 2N_r N_t$ $+M\tau^2 (M\tau - M - 1 + 3M^2\tau)$	$N_r^3 N_t [N_t^2 + (M\tau^3 + 2M\tau)N_t + 2M^2\tau^2]$ $+N_r^2 \tau [\tau N_t^2 + (\tau^2 + 1)MN_t + \tau^2 + (1 + M + 2M^2)\tau]$ $+2N_r (M\tau + N_t) + 3M^3\tau^3 + M\tau^2 (1 + \tau)$
LRA-LMMSE	$N_r^3 [M\tau^3 N_t^2 + 3M^2\tau^2 N_t + M^3\tau^3]$ $-2N_r^2 M\tau N_t - N_r (M\tau + N_t)$ $+M\tau^2 (M\tau - 1 + 3M^2\tau - M)$	$N_r^3 [M\tau^3 N_t^2 + 3M^2\tau^2 N_t + M^3\tau^3]$ $+N_r^2 \tau [\tau N_t^2 + (\tau^2 + 1)MN_t + \tau^2 + (1 + \tau + M + 5M^2)\tau]$ $+3M\tau N_r + 3M^3\tau^3 + M\tau^2 (1 + \tau)$
LRA-LMS	$N_r (\tau - l_{\text{win}} + 1)[l_{\text{win}}^2 M(2MN_t - M - 1)$ $+l_{\text{win}}^3 M(3M^2 + M + N_t^2)]$	$N_r (\tau - l_{\text{win}} + 1)[N_t + 2l_{\text{win}} M(1 + N_t)$ $+l_{\text{win}}^2 (1 + N_t^2 + 2M + 2M^2 N_t + 2M^2)$ $+l_{\text{win}}^3 (N_t^2 M + 1 + MN_t + M + 3M^3)]$

G. ESTIMATION OF $\mathbf{R}_{\mathbf{h}'}$

In practical environments, there is no prior information about $\mathbf{R}_{\mathbf{h}'}$ at the receiver. In this subsection, an adaptive technique is proposed to recursively estimate $\mathbf{R}_{\mathbf{h}'}$ as

$$\hat{\mathbf{R}}_{\mathbf{h}'}(n+1) = \lambda \hat{\mathbf{R}}_{\mathbf{h}'}(n) + (1 - \lambda)\hat{\mathbf{h}}'(n)\hat{\mathbf{h}}'(n)^H, \quad n = 1, \ldots, \tau, \tag{31}$$

where λ is the forgetting factor and $\hat{\mathbf{h}}'(n)$ is the channel estimate at the Nyquist time instant n. Consider the system model

$$\begin{aligned} \mathbf{y}_{\mathcal{Q}}(n) &= \mathcal{Q}(\mathbf{Hx}(n) + \mathbf{n}(n)) \\ &= \mathcal{Q}((\mathbf{x}'^T_p(n) \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}'\mathbf{u})\mathbf{h}' + \mathbf{n}(n)), \end{aligned} \tag{32}$$

where $\mathbf{y}_{\mathcal{Q}}(n)$ and $\mathbf{n}(n)$ are column vectors with size $MN_r \times 1$. Different from $\mathbf{x}_p(n)$ in Algorithm 1, $\mathbf{x}'_p(n) \in \mathbb{C}^{N_t \times 1}$ contains pilot symbols from N_t terminals at time instant n. $\mathbf{Z}' \in \mathbb{R}^{M \times M}$ is a simplified version of $\mathbf{Z}$ with $N = 1$. The instantaneous estimate of $\mathbf{h}'$ is calculated as

$$\hat{\mathbf{h}}'(n) = (\mathbf{x}'^T_p(n) \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}'\mathbf{u})^+ \mathbf{y}_{\mathcal{Q}}(n), \tag{33}$$

where the initial guess of $\hat{\mathbf{R}}_{\mathbf{h}'}(1)$ is an identity matrix by assuming channel parameters are uncorrelated and each has unit power.

IV. CRAMÉR-RAO BOUNDS

Unlike the works in [29], [30], which have proposed the CRBs for the unbiased estimators, the existing CRBs are extended suitable for the biased estimators. Two different types of CRBs are proposed depending on whether the prior information $\mathbf{R}_{\mathbf{h}'}$ is known at the receiver, namely Bayesian CRB with known $\mathbf{R}_{\mathbf{h}'}$ and general CRB with estimated $\mathbf{R}_{\mathbf{h}'}$.

A. BAYESIAN CRAMÉR-RAO BOUNDS

Bayesian bounds on the fundamental limits of estimation are derived for non-oversampled and oversampled systems. Without loss of generality, we extend (12) considering the whole system and not just the pilots, and rewrite the complex-valued model in the following real-valued form

$$\begin{bmatrix} \mathbf{y}^R \\ \mathbf{y}^I \end{bmatrix} = \begin{bmatrix} \Phi^R & -\Phi^I \\ \Phi^I & \Phi^R \end{bmatrix} \begin{bmatrix} \mathbf{h}'^R \\ \mathbf{h}'^I \end{bmatrix} + \begin{bmatrix} \mathbf{n}^R \\ \mathbf{n}^I \end{bmatrix}. \tag{34}$$

Let $\tilde{\mathbf{h}}' = [\mathbf{h}'^R; \mathbf{h}'^I]$ be the unknown parameter vector, since the real and imaginary parts are independent, the Bayesian information matrix (BIM) [37] for the quantized signal is defined as

$$\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}(\tilde{\mathbf{h}}') = \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^R}(\tilde{\mathbf{h}}') + \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^I}(\tilde{\mathbf{h}}'), \tag{35}$$

where

$$[\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^{R/I}}(\tilde{\mathbf{h}}')]_{ij} \triangleq E_{\mathbf{y}_{\mathcal{Q}}^{R/I}, \tilde{\mathbf{h}}'} \left\{ \frac{\partial \ln p(\mathbf{y}_{\mathcal{Q}}^{R/I}, \tilde{\mathbf{h}}')}{\partial [\tilde{\mathbf{h}}']_i} \frac{\partial \ln p(\mathbf{y}_{\mathcal{Q}}^{R/I}, \tilde{\mathbf{h}}')}{\partial [\tilde{\mathbf{h}}']_j} \right\} \tag{36}$$

with $[\tilde{\mathbf{h}}']_i$ and $[\tilde{\mathbf{h}}']_j$ being the elements of $\tilde{\mathbf{h}}'$. The expression in (36) can be divided into two parts:

$$[\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^{R/I}}(\tilde{\mathbf{h}}')]_{ij} = [\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^{R/I}}^D(\tilde{\mathbf{h}}')]_{ij} + [\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^{R/I}}^P(\tilde{\mathbf{h}}')]_{ij}, \tag{37}$$

where

$$[\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^{R/I}}^D(\tilde{\mathbf{h}}')]_{ij} \triangleq E_{\mathbf{y}_{\mathcal{Q}}^{R/I}|\tilde{\mathbf{h}}'} \left\{ \frac{\partial \ln p(\mathbf{y}_{\mathcal{Q}}^{R/I} \mid \tilde{\mathbf{h}}')}{\partial [\tilde{\mathbf{h}}']_i} \frac{\partial \ln p(\mathbf{y}_{\mathcal{Q}}^{R/I} \mid \tilde{\mathbf{h}}')}{\partial [\tilde{\mathbf{h}}']_j} \right\} \tag{38}$$

$$[\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}^{R/I}}^P(\tilde{\mathbf{h}}')]_{ij} \triangleq E_{\tilde{\mathbf{h}}'} \left\{ \frac{\partial \ln p(\tilde{\mathbf{h}}')}{\partial [\tilde{\mathbf{h}}']_i} \frac{\partial \ln p(\tilde{\mathbf{h}}')}{\partial [\tilde{\mathbf{h}}']_j} \right\}. \tag{39}$$

To transform the real-valued $\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}(\tilde{\mathbf{h}}')$ back to the complex domain $\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}(\mathbf{h}')$, $\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}(\tilde{\mathbf{h}}')$ is defined with the following structure:

$$\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}(\tilde{\mathbf{h}}') = \begin{bmatrix} \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{RR}(\tilde{\mathbf{h}}') & \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{RI}(\tilde{\mathbf{h}}') \\ \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{IR}(\tilde{\mathbf{h}}') & \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{II}(\tilde{\mathbf{h}}') \end{bmatrix} \tag{40}$$

and apply the chain rule to get:

$$\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}(\mathbf{h}') = \frac{1}{4}(\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{RR}(\tilde{\mathbf{h}}') + \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{II}(\tilde{\mathbf{h}}')) + \frac{j}{4}(\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{RI}(\tilde{\mathbf{h}}') - \mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{IR}(\tilde{\mathbf{h}}')), \tag{41}$$

where $\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{RR}(\tilde{\mathbf{h}}')$, $\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{RI}(\tilde{\mathbf{h}}')$, $\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{IR}(\tilde{\mathbf{h}}')$ and $\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{II}(\tilde{\mathbf{h}}')$ have the same dimensions $N_r N_t \times N_r N_t$. The variance of the estimator $\hat{\mathbf{h}}'(\mathbf{y}_{\mathcal{Q}})$ is lower bounded by

$$Var\{\hat{h}'_i(\mathbf{y}_{\mathcal{Q}})\} \geq [\mathbf{J}_{\mathbf{y}_{\mathcal{Q}}}^{-1}(\mathbf{h}')]_{ii}. \tag{42}$$

1) BIM FOR NON-OVERSAMPLED SYSTEMS

For non-oversampled systems, i.e, $M = 1$, the covariance matrix of the equivalent noise vector $\mathbf{n}$ is $\mathbf{C_n} = \sigma_n^2 \mathbf{I}_{NN_r}$. With the independence of the real and imaginary parts, the log-likelihood function can be expressed as

$$\ln p(\mathbf{y}_Q \mid \tilde{\mathbf{h}}') = \sum_{k=1}^{NN_r} [\ln p([\mathbf{y}_Q^R]_k \mid \tilde{\mathbf{h}}') + \ln p([\mathbf{y}_Q^I]_k \mid \tilde{\mathbf{h}}')] \quad (43)$$

with

$$p([\mathbf{y}_Q^R]_k = \pm\frac{1}{\sqrt{2}} \mid \tilde{\mathbf{h}}') = Q\left(\mp \frac{[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k}{\sigma_n/\sqrt{2}}\right) \quad (44)$$

$$p([\mathbf{y}_Q^I]_k = \pm\frac{1}{\sqrt{2}} \mid \tilde{\mathbf{h}}') = Q\left(\mp \frac{[\Phi^I \mathbf{h}'^R + \Phi^R \mathbf{h}'^I]_k}{\sigma_n/\sqrt{2}}\right) \quad (45)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-\frac{u^2}{2})du$. Inserting (43) into (38), we obtain

$$[\mathbf{J}_{\mathbf{y}_Q}^D(\tilde{\mathbf{h}}')]_{ij} = -E\left\{\frac{\partial^2 \ln p(\mathbf{y}_Q \mid \tilde{\mathbf{h}}')}{\partial[\tilde{\mathbf{h}}']_i \partial[\tilde{\mathbf{h}}']_j}\right\}$$
$$= [\mathbf{J}_{\mathbf{y}_Q^R}^D(\tilde{\mathbf{h}}')]_{ij} + [\mathbf{J}_{\mathbf{y}_Q^I}^D(\tilde{\mathbf{h}}')]_{ij}. \quad (46)$$

With the derivative of the $Q(x)$ function, the real part in (38) $[\mathbf{J}_{\mathbf{y}_Q^R}^D(\tilde{\mathbf{h}}')]_{ij}$ is given by

$$[\mathbf{J}_{\mathbf{y}_Q^R}^D(\tilde{\mathbf{h}}')]_{ij}$$
$$= \sum_{k=1}^{NN_r} -E\left\{\frac{\partial^2 \ln p([\mathbf{y}_Q^R]_k \mid \tilde{\mathbf{h}}')}{\partial[\tilde{\mathbf{h}}']_i \partial[\tilde{\mathbf{h}}']_j}\right\} = \frac{1}{\pi\sigma_n^2}$$
$$\times \sum_{k=1}^{NN_r} \frac{\exp(-\frac{[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k^2}{\sigma_n^2/2}) \frac{\partial[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k}{\partial[\tilde{\mathbf{h}}']_i} \frac{\partial[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k}{\partial[\tilde{\mathbf{h}}']_j}}{Q\left(\frac{[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k}{\sigma_n/\sqrt{2}}\right) Q\left(-\frac{[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k}{\sigma_n/\sqrt{2}}\right)}. \quad (47)$$

The derivation for the imaginary part $[\mathbf{J}_{\mathbf{y}_Q^I}^D(\tilde{\mathbf{h}}')]_{ij}$ is analogous.

By assuming that $\tilde{\mathbf{h}}'$ is Gaussian distributed with zero mean and covariance matrix $\mathbf{C}_{\tilde{\mathbf{h}}'} = \frac{1}{2}\mathbf{I}_2 \otimes \mathbf{C}_{\mathbf{h}'}$, $\ln p(\tilde{\mathbf{h}}')$ yields

$$\ln p(\tilde{\mathbf{h}}') = -\frac{1}{2}N_r N_t \ln[(2\pi)^{2N_r N_t} \det(\mathbf{C}_{\tilde{\mathbf{h}}'})] - \frac{1}{2}\tilde{\mathbf{h}}'^T \mathbf{C}_{\tilde{\mathbf{h}}'}^{-1}\tilde{\mathbf{h}}'. \quad (48)$$

Substituting (48) into (39), we obtain

$$\mathbf{J}_{\mathbf{y}_Q}^P(\tilde{\mathbf{h}}') = 2\mathbf{J}_{\mathbf{y}_Q^{R/I}}^P(\tilde{\mathbf{h}}') = 2\mathbf{C}_{\tilde{\mathbf{h}}'}^{-1}. \quad (49)$$

Finally, the resulting BIM is the summation of (46) and (49) as described by

$$\mathbf{J}_{\mathbf{y}_Q}(\tilde{\mathbf{h}}') = \mathbf{J}_{\mathbf{y}_Q}^D(\tilde{\mathbf{h}}') + \mathbf{J}_{\mathbf{y}_Q}^P(\tilde{\mathbf{h}}'). \quad (50)$$

2) BIM FOR OVERSAMPLED SYSTEMS

When $M \geq 2$ the equivalent noise vector $\mathbf{n}$ consists of colored Gaussian noise samples. Computing $p(\mathbf{y}_Q^{R/I} \mid \tilde{\mathbf{h}}')$ requires the orthant probabilities, which are not available or too difficult to compute. The authors in [28], [30] have introduced a lower bounding technique on the Fisher information for real-valued system. To employ this lower bounding technique in the complex-valued system, the work of [29] has come out. The lower bound of $\mathbf{J}_{\mathbf{y}_Q^{R/I}}^D(\tilde{\mathbf{h}}')$ is calculated based on the first and second order moments as

$$\mathbf{J}_{\mathbf{y}_Q^{R/I}}^D(\tilde{\mathbf{h}}') \geq \left(\frac{\partial\mu_{\mathbf{y}_Q^{R/I}}}{\partial\tilde{\mathbf{h}}'}\right)^T \mathbf{C}_{\mathbf{y}_Q^{R/I}}^{-1}\left(\frac{\partial\mu_{\mathbf{y}_Q^{R/I}}}{\partial\tilde{\mathbf{h}}'}\right) = \tilde{\mathbf{J}}_{\mathbf{y}_Q^{R/I}}^D(\tilde{\mathbf{h}}'). \quad (51)$$

Since the lower-bounding technique is identical for the real and the imaginary parts, only the derivation of $\tilde{\mathbf{J}}_{\mathbf{y}_Q^R}^D(\tilde{\mathbf{h}}')$ is presented. The mean value of the kth received symbol is

$$[\mu_{\mathbf{y}_Q^R}]_k = \frac{1}{\sqrt{2}}p([\mathbf{y}_Q]_k = +1 \mid \tilde{\mathbf{h}}') - \frac{1}{\sqrt{2}}p([\mathbf{y}_Q]_k = -1 \mid \tilde{\mathbf{h}}')$$
$$= \frac{1}{\sqrt{2}}\left[1 - 2Q\left(\frac{[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k}{\sqrt{[\mathbf{C_n}]_{kk}/2}}\right)\right]. \quad (52)$$

The partial derivative of (52) with respect to $[\tilde{\mathbf{h}}']_i$ is

$$\frac{\partial[\mu_{\mathbf{y}_Q^R}]_k}{\partial[\tilde{\mathbf{h}}']_i} = \frac{2\exp\left(-\frac{[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k^2}{[\mathbf{C_n}]_{kk}}\right)\frac{\partial[\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k}{\partial[\tilde{\mathbf{h}}']_i}}{\sqrt{2\pi[\mathbf{C_n}]_{kk}}}. \quad (53)$$

The diagonal elements of the covariance matrix are given by

$$[\mathbf{C}_{\mathbf{y}_Q^R}]_{kk} = \frac{1}{2} - [\mu_{\mathbf{y}_Q^R}]_k^2, \quad (54)$$

while the off-diagonal elements are calculated as

$$[\mathbf{C}_{\mathbf{y}_Q^R}]_{kn} = p(z_k > 0, z_n > 0) + p(z_k \leq 0, z_n \leq 0)$$
$$- \frac{1}{2} - [\mu_{\mathbf{y}_Q^R}]_k[\mu_{\mathbf{y}_Q^R}]_n, \quad (55)$$

where $[z_k, z_n]^T$ is a bi-variate Gaussian random vector

$$\begin{bmatrix} z_k \\ z_n \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} [\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_k \\ [\Phi^R \mathbf{h}'^R - \Phi^I \mathbf{h}'^I]_n \end{bmatrix}, \frac{1}{2}\begin{bmatrix} [\mathbf{C_n}]_{kk} & [\mathbf{C_n}]_{kn} \\ [\mathbf{C_n}]_{nk} & [\mathbf{C_n}]_{nn} \end{bmatrix}\right).$$

The lower bound for the imaginary part is derived in the same way. With the calculations above the lower bound of the BIM is obtained as

$$\mathbf{J}_{\mathbf{y}_Q}(\tilde{\mathbf{h}}') \geq \tilde{\mathbf{J}}_{\mathbf{y}_Q}^D(\tilde{\mathbf{h}}') + \mathbf{J}_{\mathbf{y}_Q}^P(\tilde{\mathbf{h}}'), \quad (56)$$

where the equality holds for $M = 1$, as shown in [30] for the real valued CRB and in [29] for the complex valued CRB. Based on (42), the inverse of this BIM lower bound will result in an upper bound of the actual Bayesian CRB for oversampled systems.

B. GENERAL CRAMÉR-RAO BOUNDS

When $\mathbf{R}_{\mathbf{h}'}$ is unknown and needs to be estimated at the receiver, the Bayesian CRBs will not be applicable. The general CRBs are derived for the proposed channel estimators with estimated $\mathbf{R}_{\mathbf{h}'}$.

Lemma 1: The proposed LRA channel estimators with combination of estimated $\hat{\mathbf{R}}_{\mathbf{h}'}$ are biased channel estimators.

Proof: See Appendix C. ∎

Since the proposed LRA channel estimators are biased, while calculating the CRBs, they should apply as

$$Cov\{\hat{\mathbf{h}}'^{R}_{\text{bias}}\} \geq \frac{\partial E\{\hat{\mathbf{h}}'^{R}_{\text{bias}}\}}{\partial \mathbf{h}'^{R}} \left(\mathbf{J}^{D^{RR}}_{\mathbf{y}_Q}(\mathbf{h}'^{R})^{-1} \frac{\partial E\{\hat{\mathbf{h}}'^{R}_{\text{bias}}\}}{\partial \mathbf{h}'^{R}}\right)^{T} \quad (57)$$

$$Cov\{\hat{\mathbf{h}}'^{I}_{\text{bias}}\} \geq \frac{\partial E\{\hat{\mathbf{h}}'^{I}_{\text{bias}}\}}{\partial \mathbf{h}'^{I}} \left(\mathbf{J}^{D^{II}}_{\mathbf{y}_Q}(\mathbf{h}'^{I})^{-1} \frac{\partial E\{\hat{\mathbf{h}}'^{I}_{\text{bias}}\}}{\partial \mathbf{h}'^{I}}\right)^{T}, \quad (58)$$

where $\mathbf{J}^{D^{RR}}_{\mathbf{y}_Q}(\mathbf{h}'^{R})$ and $\mathbf{J}^{D^{II}}_{\mathbf{y}_Q}(\mathbf{h}'^{I})$ are defined by

$$[\mathbf{J}^{D^{RR}}_{\mathbf{y}_Q}(\mathbf{h}'^{R})]_{ij} \triangleq E\left\{\frac{\partial \ln p(\mathbf{y}_Q \mid \mathbf{h}'^{R})}{\partial [\mathbf{h}'^{R}]_i} \frac{\partial \ln p(\mathbf{y}_Q \mid \mathbf{h}'^{R})}{\partial [\mathbf{h}'^{R}]_j}\right\} \quad (59)$$

$$[\mathbf{J}^{D^{II}}_{\mathbf{y}_Q}(\mathbf{h}'^{I})]_{ij} \triangleq E\left\{\frac{\partial \ln p(\mathbf{y}_Q \mid \mathbf{h}'^{I})}{\partial [\mathbf{h}'^{I}]_i} \frac{\partial \ln p(\mathbf{y}_Q \mid \mathbf{h}'^{I})}{\partial [\mathbf{h}'^{I}]_j}\right\}, \quad (60)$$

which are the upper left and lower right part of the $\mathbf{J}^{D}_{\mathbf{y}_Q}(\tilde{\mathbf{h}}')$ (similar as (40)), respectively.

V. NUMERICAL RESULTS

The simulation results presented here consider an uplink single-cell 1-bit large-scale MIMO system with $N_t = 8$ and $N_r = 64$. The modulation scheme is quadrature phase-shift keying (QPSK). The $m(t)$ and $p(t)$ filters are normalized RRC filters with a roll-off factor of 0.8. The channel is assumed to experience block fading and the pilots are column-wise orthogonal with length 20. The SNR is defined as $10\log(\frac{N_t}{\sigma_n^2})$. The normalized MSE and SER performance plots are obtained by taking the average of 300 channel matrices, noise and symbol vectors.

For the LRA-LMS channel estimator, the window length l_{win} is chosen as three to ensure low computational complexity. The step size μ is optimized according to the oversampling factor and SNR. In the simulation, μ varies between 0.05 and 0.3. While recovering the transmitted symbols from the received quantized signal, the sliding-window based LMMSE detector [26] with window length equal to three ($l_{\text{win}} = 3$) and the estimate of the channel obtained by the proposed algorithms is applied in the system for obtaining both high accuracy and low computational cost.

The performance of the channel estimators is evaluated based on the channel model simulated in [38]. The channel for user n_t is assumed Rayleigh distributed

$$\mathbf{h}'_{n_t} = \mathbf{R}^{\frac{1}{2}}_{r,n_t}\mathbf{h}'_{w,n_t}, \quad (61)$$

where $\mathbf{R}_{r,n_t}$ denotes the receive correlation matrix with the following form

$$\mathbf{R}_{r,n_t} = \begin{bmatrix} 1 & \rho_{n_t} & \cdots & \rho_{n_t}^{(N_r-1)} \\ \rho_{n_t}^{*} & 1 & \cdots & \rho_{n_t}^{(N_r-2)} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n_t}^{*(N_r-1)} & \rho_{n_t}^{*(N_r-2)} & \cdots & 1 \end{bmatrix}. \quad (62)$$

ρ_{n_t} is the correlation index of neighboring antennas. ($|\rho_{n_t}| = 0$ represents an uncorrelated scenario and $|\rho_{n_t}| = 1$ implies a fully correlated scenario.) The elements of $\mathbf{h}'_{w,n_t}$ are i.i.d. complex Gaussian random variables with zero mean and unit variance. All users are assumed to experience the same value of $|\rho_{n_t}| = |\rho|$ but different phases uniformly distributed over 2π. The overall channel model is summarized as

$$\mathbf{H}' = [\mathbf{h}'_1, \mathbf{h}'_2, \cdots, \mathbf{h}'_{n_t}] \quad (63)$$

and $\mathbf{R}_{\mathbf{h}'}$ is calculated as

$$\mathbf{R}_{\mathbf{h}'} = \begin{bmatrix} \mathbf{R}_{r,1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{r,2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{R}_{r,n_t} \end{bmatrix}. \quad (64)$$

A. $\mathbf{R}_{\mathbf{h}'}$ IS KNOWN AT THE RECEIVER

In this subsection, we evaluate the performance of the proposed LRA channel estimators with known $\mathbf{R}_{\mathbf{h}'}$ at the receiver. Fig. 6a and Fig. 6b compare the normalized MSE of the various channel estimators as a function of SNR in uncorrelated ($|\rho| = 0$) and correlated channel ($|\rho| = 0.75$), respectively. There is a 2dB performance gain of the oversampled systems as compared to the non-oversampled systems for the LRA-LMMSE channel estimator at low SNR, whereas a much larger gain at high SNR. In both channels the LRA-LMMSE achieves the best MSE performance at the cost of high computational cost.

In contrast, the LRA-LMS estimates the channel matrix $\mathbf{H}'$ row by row. This approach can largely reduce the computational cost (shown in Fig. 5). Note that this separation into several rows may overlook the correlation of receive antennas. More specifically, the proposed LRA-LMS treats $\mathbf{R}_{r,n_t}$ as an identity matrix. As an amendment, the resulting estimated channel matrix $\hat{\mathbf{h}}'_{\text{LRA-LMS}}$ needs to be multiplied with the square root of the receive correlation matrix $\mathbf{R}^{\frac{1}{2}}_{r,n_t}$, which can be derived from $\mathbf{R}_{\mathbf{h}'}$ in (64). From the results, it can be seen that in both channels the LRA-LMS approaches the performance of the LRA-LMMSE at low SNR (≤ 5 dB), whereas at high SNR this performance gap becomes large.

The Bayesian CRBs illustrated in Section IV-A are also depicted in Fig. 6. Note that for the oversampled systems ($M \geq 2$) the upper bounds of Bayesian CRBs are higher than the actual Bayesian CRBs, since they are derived from the lower bounds of Bayesian information. The black lines represent the standard LMMSE performance for the systems with unquantized signals, which can be treated as lower bounds for the systems with 1-bit quantized signals.
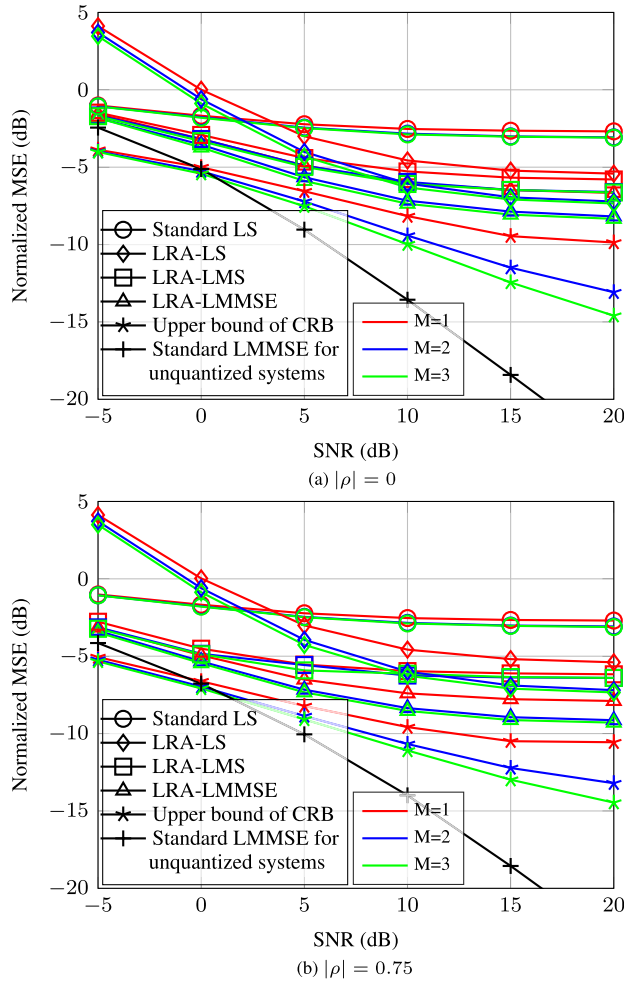
(a) $|\rho| = 0$

(b) $|\rho| = 0.75$

FIGURE 6. Normalized MSE comparisons of different channel estimators with known $\mathbf{R}_{\mathbf{h}'}$.
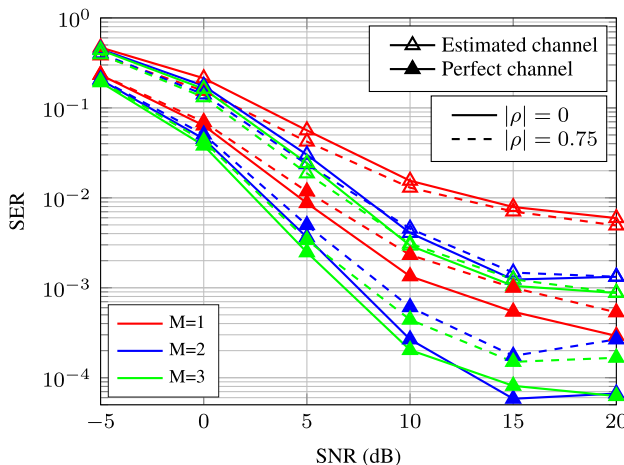

FIGURE 7. SER comparisons of different oversampling factors for the LRA-LMMSE channel estimator with known $\mathbf{R}_{\mathbf{h}'}$.

The LMMSE detector with sliding-window based SER performance of the system with the LRA-LMMSE estimated and perfect channel matrix are illustrated in Fig. 7, where the oversampled systems obviously outperform the
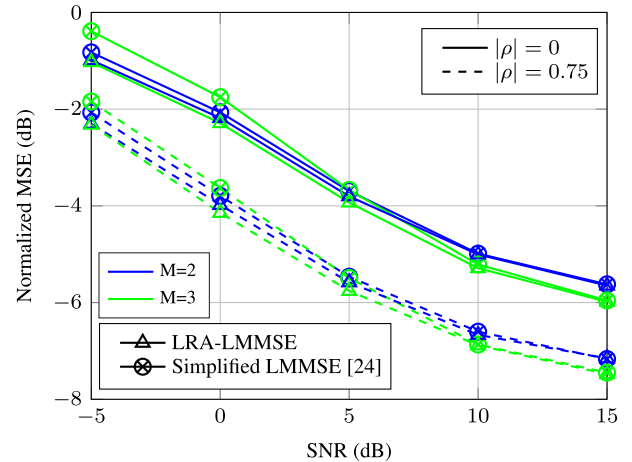

FIGURE 8. Normalized MSE comparisons between LRA-LMMSE and simplified LMMSE [24].

non-oversampled systems. As described in III-A, Fig. 8 shows the MSE comparisons between LRA-LMMSE and simplified LMMSE [24] channel estimator in the system with $\tau = 10$ and roll-off factor 0.1. We emphasize again that in our work, the correlation of filtered noise is taken into account, and hence $\mathbf{C}_{\mathbf{n}_p}$ is not a diagonal matrix in over-sampled systems. It can be seen that at low SNR (≤ 10 dB) the performance of simplified LMMSE [24] is worse than the proposed LRA-LMMSE, although they converge together at high SNR (> 10 dB). Another observation is that at low SNR the simplified LMMSE estimator with $M = 3$ performs worse than that with $M = 2$, which shows that the assumption in [24] is inaccurate.

B. $\mathbf{R}_{\mathbf{h}'}$ IS UNKNOWN AT THE RECEIVER

Practically, $\mathbf{R}_{\mathbf{h}'}$ is not known at the receiver. Fig. 9 shows the MSE performance of the LRA channel estimators by using the proposed adaptive recursion to estimate $\mathbf{R}_{\mathbf{h}'}$, where λ is set to 0.99. It can be seen that the performance remains almost the same as Fig. 6a, which shows that the proposed estimation of $\mathbf{R}_{\mathbf{h}'}$ works well under uncorrelated channel.

While analyzing the general CRBs proposed in (57) and (58), instead of directly calculating the gradient of the expected value with respect to the channel vector $\frac{\partial E\{\hat{\mathbf{h}}_{\text{bias}}^{R/I}\}}{\partial \mathbf{h}'^{R/I}}$, this gradient is numerically evaluated, since there is an adaptive estimation technique inside the channel estimator, which makes the calculation more difficult. As one example, Fig. 10 shows the normalized MSE performance of the LRA-LS channel estimator with estimated $\hat{\mathbf{R}}_{\mathbf{h}'}$ in (31) for estimating the first N_r elements[2] of $\mathbf{h}'^{R}$ and its corresponding numerically calculated general CRBs under uncorrelated channels ($|\rho| = 0$). More specifically, each element of the gradient vector $\frac{\partial E\{\hat{\mathbf{h}}_{\text{bias}}^{R/I}\}}{\partial \mathbf{h}'^{R/I}}$ is calculated with the following steps:

---

[2]For the sake of simplicity, only first N_r elements are considered, since for the large-scale MIMO there are $N_t N_r$ elements in $\mathbf{h}'^{R}$, which will cost much time for calculating the general CRBs.
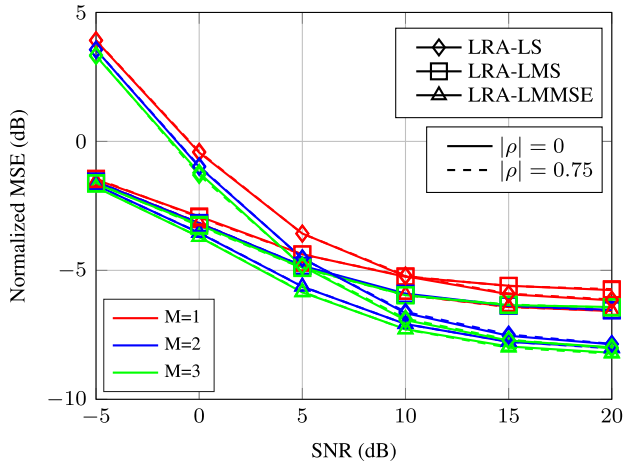
FIGURE 9. Normalized MSE comparisons of different channel estimators with adaptively estimated $\hat{\mathbf{R}}_{\mathbf{h}'}$.
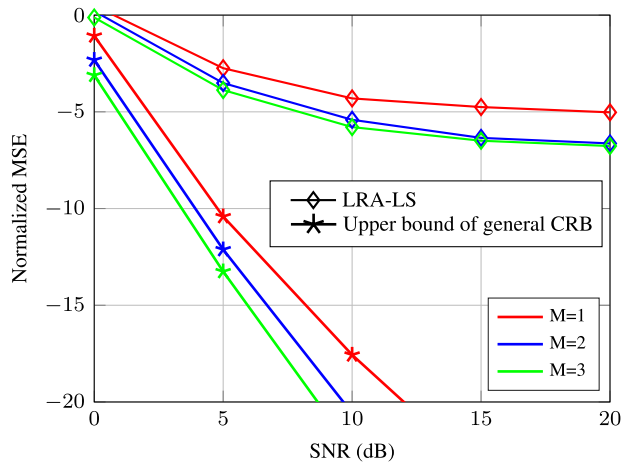

FIGURE 10. Normalized MSE comparisons of different oversampling factors for the LRA-LS channel estimator with estimated $\hat{\mathbf{R}}_{\mathbf{h}'}$.

- increasing a small value δ (e.g. 0.1) in the corresponding element of $\mathbf{h}'^{R/I}$
- estimating the channel $\hat{\mathbf{h}}'^{R/I}_{\text{bias}}$ with different transmit symbols and noises (e.g. 1000 different realizations)
- calculating the mean value of all estimates $E\{\hat{\mathbf{h}}'^{R/I}_{\text{bias}}\}$, which will be divided by δ.

These steps are repeated until all the elements in $\frac{\partial E\{\hat{\mathbf{h}}'^{R/I}_{\text{bias}}\}}{\partial \mathbf{h}'^{R/I}}$ are obtained.

C. 1-BIT OR B-Bit ADC?

In this subsection, the channel estimation performance of the 1-bit oversampled system is compared with the b-bit non-oversampled systems. In Fig. 11 the LRA-LMMSE channel estimator for a system with 2 or 3 bits is based on the work in [6]. It can be seen that a system with 2 or 3 bits has better MSE performance than the 1-bit system especially at high SNR. However, the advantages of 1-bit ADCs is that they do not require automatic gain control (AGC) and linear amplifiers, and hence the corresponding radio frequency chains can be implemented with very low cost and power consumption
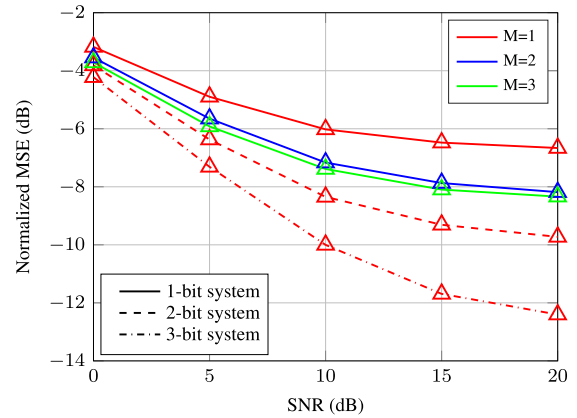

FIGURE 11. Normalized MSE comparisons of LRA-LMMSE channel estimator with known $\mathbf{R}_{\mathbf{h}'}$ under uncorrelated channel ($|\rho| = 0$).
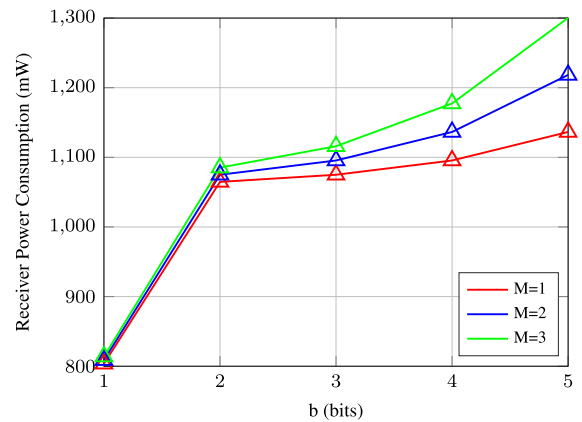

FIGURE 12. Receiver power consumption as a function of the quantization bits b.

(a few milliwatts) [7], [11], [39]. As one example, Fig. 12 shows the total receiver power consumption as a function of the quantization bits b. The calculation of receiver power consumption is based on the work in [40]

$$P_{\text{total}} = P_{\text{BB}} + P_{\text{LO}} + N_r(P_{\text{LNA}} + P_{\text{H}} + 2P_{\text{M}}) + 2N_r(cP_{\text{AGC}} + P_{\text{ADC}}), \quad (65)$$

where P_{BB}, P_{LO}, P_{LNA}, P_{H}, P_{M} and P_{AGC} denote the power consumption in the baseband processor, local oscillator (LO), low noise amplifier (LNA), $\frac{\pi}{2}$ hybrid and LO buffer, Mixer and AGC, respectively. c is chosen as 0 for the 1-bit system and 1 for b-bit systems. The power consumption of different hardware components is given as $P_{\text{BB}} = 200$ mW, $P_{\text{LO}} = 22.5$ mW, $P_{\text{LNA}} = 5.4$ mW, $P_{\text{H}} = 3$ mW, $P_{\text{AGC}} = 2$ mW and $P_{\text{M}} = 0.3$ mW. The P_{ADC} is calculated as

$$P_{\text{ADC}} = \text{FOM}_w \times Mf_n \times 2^b, \quad (66)$$

where FOM_w is 200 fJ/conversion-step at 50 MHz bandwidth and f_n is 100 MHz. From the results, it can be seen that the 1-bit system consumes much less power than the 2-bit and 3-bit systems in both non-oversampled and oversampled systems. Indeed, the 1-bit oversampled systems have largely improved the estimation performance and allows the

estimator to approach the performance of the 2-bit system at low SNR.

VI. CONCLUSION

In this work, oversampling based low-resolution aware channel estimators have been proposed for uplink single-cell large-scale MIMO systems with 1-bit ADCs employed at the receiver. The Bussgang decomposition is used to derive linear channel estimators based on different criteria. With oversampling in such systems, it is observed that we can achieve obvious advantage compared to the non-oversampled system in terms of the normalized MSE. Moreover, the LMS adaptive technique used for channel estimation can largely reduce the computational cost and has almost the same accuracy as the LRA-LMMSE channel estimator at low SNR, which is important to ensure low computational complexity and for hardware implementation. In addition, we have also derived Bayesian and general CRBs on MSE, which give theoretical limits on the performance of the channel estimators. Furthermore, we have proposed an adaptive technique to estimate the auto-correlation of channel vector, which is important for practical use. In general, the 1-bit ADCs have the advantage of energy saving. Our proposed oversampling based channel estimation, especially the LRA-LMS estimator, increases the accuracy of estimation while maintaining low computational cost, which is important for future low cost and low latency wireless systems.

APPENDIX A
PROOF OF (23)

Recall the optimization problem

$$\mathbf{W}_{\text{LMMSE}} = \arg\min_{\mathbf{W}} E\{||\mathbf{h}' - \mathbf{W}\mathbf{y}_{\mathcal{Q}_p}||^2\}. \tag{67}$$

Taking the partial derivative with respect to $\mathbf{W}^H$, we obtain

$$\frac{\partial E\{||\mathbf{h}' - \mathbf{W}\mathbf{y}_{\mathcal{Q}_p}||^2\}}{\partial \mathbf{W}^H} = -E\{\mathbf{h}'\mathbf{y}_{\mathcal{Q}_p}^H\} + \mathbf{W}E\{\mathbf{y}_{\mathcal{Q}_p}\mathbf{y}_{\mathcal{Q}_p}^H\}. \tag{68}$$

Inserting (14) into (68), the LMMSE filter is

$$\mathbf{W}_{\text{LMMSE}} = E\{\mathbf{h}'\mathbf{y}_{\mathcal{Q}_p}^H\}E\{\mathbf{y}_{\mathcal{Q}_p}\mathbf{y}_{\mathcal{Q}_p}^H\}^{-1}$$
$$= (E\{\mathbf{h}'\mathbf{h}'^H\}\tilde{\Phi}_p^H + E\{\mathbf{h}'\tilde{\mathbf{n}}_p^H\})\mathbf{C}_{\mathbf{y}_{\mathcal{Q}_p}}^{-1}. \tag{69}$$

Since $\mathbf{h}'$ is uncorrelated with $\mathbf{n}_p$ and $\mathbf{n}_q$ [11], we have

$$E\{\mathbf{h}'\tilde{\mathbf{n}}_p^H\} = E\{\mathbf{h}'(\mathbf{A}_p\mathbf{n}_p + \mathbf{n}_q)^H\} = \mathbf{0}. \tag{70}$$

The resulting LRA-LMMSE channel estimator is

$$\hat{\mathbf{h}}'_{\text{LRA-LMMSE}} = \mathbf{R}_{\mathbf{h}'}\tilde{\Phi}_p^H\mathbf{C}_{\mathbf{y}_{\mathcal{Q}_p}}^{-1}\mathbf{y}_{\mathcal{Q}_p}. \tag{71}$$

APPENDIX B
PROOF OF (29)

Defining $\epsilon(n) = \bar{\mathbf{h}}'^{n_r}(n) - \mathbf{h}'^{n_r}$ and inserting it into (29), we obtain

$$\epsilon(n+1) = \epsilon(n) + \mu\tilde{\Phi}_p^{n_r}(n)^H(\mathbf{y}_{\mathcal{Q}_p}^{n_r}(n) - \tilde{\Phi}_p^{n_r}(n)\bar{\mathbf{h}}'^{n_r}(n))$$
$$= \epsilon(n) + \mu\tilde{\Phi}_p^{n_r}(n)^H\mathbf{y}_{\mathcal{Q}_p}^{n_r}(n)$$

$$-\mu\tilde{\Phi}_p^{n_r}(n)^H\tilde{\Phi}_p^{n_r}(n)(\epsilon(n) + \mathbf{h}'^{n_r})$$
$$= (\mathbf{I} - \mu\tilde{\Phi}_p^{n_r}(n)^H\tilde{\Phi}_p^{n_r}(n))\epsilon(n)$$
$$+\mu\tilde{\Phi}_p^{n_r}(n)^H(\mathbf{y}_{\mathcal{Q}_p}^{n_r}(n) - \tilde{\Phi}_p^{n_r}(n)\mathbf{h}'^{n_r}). \tag{72}$$

Taking the expected value from $\epsilon(n+1)$, we have

$$E\{\epsilon(n+1)\} = (\mathbf{I} - \mu E\{\tilde{\Phi}_p^{n_r}(n)^H\tilde{\Phi}_p^{n_r}(n)\})E\{\epsilon(n)\}. \tag{73}$$

With the eigenvalue decomposition $E\{\tilde{\Phi}_p^{n_r}(n)^H\tilde{\Phi}_p^{n_r}(n)\} = \mathbf{Q}\Gamma\mathbf{Q}^H$, (73) can be written as

$$\mathbf{Q}^H E\{\epsilon(n+1)\} = \mathbf{Q}^H(\mathbf{I} - \mu\mathbf{Q}\Gamma\mathbf{Q}^H)E\{\epsilon(n)\}$$
$$= (\mathbf{I} - \mu\Gamma)\mathbf{Q}^H E\{\epsilon(n)\}, \tag{74}$$

where $\mathbf{Q}$ is an unitary matrix and Γ is a diagonal matrix, whose diagonal entries are the eigenvalues of $E\{\tilde{\Phi}_p^{n_r}(n)^H\tilde{\Phi}_p^{n_r}(n)\}$. With $\mathbf{u}(n) = \mathbf{Q}^H E\{\epsilon(n)\}$, (74) is then

$$\mathbf{u}(n+1) = (\mathbf{I} - \mu\Gamma)\mathbf{u}(n). \tag{75}$$

Decoupling the matrix form into individual elements we get

$$u_{n_t}(n+1) = (1 - \mu\gamma_{n_t})u_{n_t}(n)$$
$$= (1 - \mu\gamma_{n_t})^{\tau - l_{\text{win}}+1}u_{n_t}(1), \quad n_t = 1, \ldots, N_t. \tag{76}$$

In order for the LRA-LMS to converge, we must have

$$|1 - \mu\gamma_{n_t}| < 1. \tag{77}$$

The stability condition is then given by

$$0 < \mu < \frac{2}{\gamma_{\text{max}}}, \tag{78}$$

where γ_{max} is the largest eigenvalue of $E\{\tilde{\Phi}_p^{n_r}(n)^H\tilde{\Phi}_p^{n_r}(n)\}$.

APPENDIX C
PROOF OF LEMMA 1

The biasness of the adaptive estimator $\hat{\mathbf{R}}_{\mathbf{h}'}$ is firstly examined. The expected value of $\hat{\mathbf{h}}'(n)$ in (33) is

$$E\{\hat{\mathbf{h}}'(n)\} = E\{(\mathbf{x}_p'^T(n) \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}'\mathbf{u})^+\mathbf{y}_{\mathcal{Q}}(n)\}. \tag{79}$$

From the Bussgang theorem (32) can be decomposed as

$$\mathbf{y}_{\mathcal{Q}}(n) = \mathcal{Q}\left((\mathbf{x}_p'^T(n) \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}'\mathbf{u})\mathbf{h}' + \mathbf{n}(n)\right)$$
$$= \mathbf{A}'_p(n)((\mathbf{x}_p'^T(n) \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}'\mathbf{u})\mathbf{h}' + \mathbf{n}(n)) + \mathbf{n}_q(n), \tag{80}$$

where $\mathbf{A}'_p(n)$ is the linear operator and $\mathbf{n}_q(n)$ is the statistically equivalent quantizer noise. Substituting (80) into (79) and with $\Phi'(n) = (\mathbf{x}^T(n) \otimes \mathbf{I}_{N_r} \otimes \mathbf{Z}'\mathbf{u})$, we obtain

$$E\{\hat{\mathbf{h}}'(n)\} = E\{\Phi'(n)^+(\mathbf{A}'_p(n)(\Phi'(n)\mathbf{h}' + \mathbf{n}(n)) + \mathbf{n}_q(n))\}$$
$$= E\{\Phi'(n)^+\mathbf{A}'_p(n)\Phi'(n)\mathbf{h}'\}$$
$$+E\{\Phi'(n)^+\mathbf{A}'_p(n)\mathbf{n}(n)\} + E\{\Phi'(n)^+\mathbf{n}_q(n)\}. \tag{81}$$

Since $\Phi'(n)$ and $\mathbf{n}(n)$ are uncorrelated and $E\{\mathbf{n}(n)\} = \mathbf{0}$, we have

$$E\{\Phi'(n)^+\mathbf{A}'_p(n)\mathbf{n}(n)\} = \mathbf{0}. \tag{82}$$

Similarly,

$$E\{\Phi'(n)^+ \mathbf{n}_q(n)\} = \mathbf{0}. \quad (83)$$
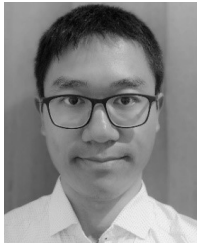
Equation (81) can be further simplified as

$$E\{\hat{\mathbf{h}}'(n)\} = E\{\Phi'(n)^+ \mathbf{A}'_p(n)\Phi'(n)\}\mathbf{h}'. \quad (84)$$

The matrix $\mathbf{A}'_p(n)$ depends on $\mathbf{R_{h'}}$ such that the expectation in (84) can be different from the identity matrix especially for channels without normalization, which verifies that (33) has an unknown bias [37]. With the analysis above, it is concluded that the adaptive estimator $\hat{\mathbf{R}}_{\mathbf{h'}}$ is also biased, which shows that the estimation procedures together with the proposed LRA channel estimators are biased.

REFERENCES

[1] E. Dahlman, S. Parkvall, J. Skold, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband*. New York, NY, USA: Academic, 2007.

[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] F. Rusek, D. Persson, B. Kiong Lau, E. G. Larsson, T. L. Marzetta, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[4] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[5] R. H. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 4, pp. 539–550, Apr. 1999.

[6] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "Throughput analysis of massive MIMO uplink with low-resolution ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 4038–4051, Jun. 2017.

[7] J. Mo and R. W. Heath, Jr., "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.

[8] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4624–4634, Sep. 2017.

[9] L. T. N. Landau and R. C. de Lamare, "Branch-and-Bound precoding for multiuser MIMO systems with 1-Bit quantization," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 770–773, Dec. 2017.

[10] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE precoder for massive MIMO using 1-bit quantization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3381–3385.

[11] Y. Li, C. Tao, G. Seco-Granados, A. Mezghani, A. L. Swindlehurst, and L. Liu, "Channel estimation and performance analysis of one-bit massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 15, pp. 4075–4089, Aug. 2017.

[12] C. Stockle, J. Munir, A. Mezghani, and J. A. Nossek, "Channel estimation in massive MIMO systems using 1-bit quantization," in *Proc. IEEE 17th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2016.

[13] Z. Shao, L. T. N. Landau, and R. C. Lamare, "Adaptive RLS channel estimation and SIC for large-scale antenna systems with 1-bit ADCs," in *Proc. 22nd Int. ITG Workshop Smart Antennas*, Mar. 2018, pp. 1–4.

[14] Z. Shao, R. C. de Lamare, and L. T. N. Landau, "Iterative detection and decoding for large-scale multiple-antenna systems with 1-Bit ADCs," *IEEE Wireless Commun. Lett.*, vol. 7, no. 3, pp. 476–479, Jun. 2018.

[15] Y.-S. Jeon, N. Lee, S.-N. Hong, and R. W. Heath, Jr., "One-bit sphere decoding for uplink massive MIMO systems with One-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4509–4521, Jul. 2018.

[16] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, Jun. 2016.

[17] C. Mollen, J. Choi, E. G. Larsson, and R. W. Heath, Jr., "One-bit ADCs in wideband massive MIMO systems with OFDM transmission," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3386–3390.

[18] C. Mollen, J. Choi, E. G. Larsson, and R. W. Heath, Jr., "Uplink performance of wideband massive MIMO with one-bit ADCs," *IEEE Trans. Wireless Commun.*, vol. 16, no. 1, pp. 87–100, Jan. 2017.

[19] S. Jacobsson, G. Durisi, M. Coldrey, and C. Studer, "Massive MU-MIMO-OFDM downlink with one-bit DACs and linear precoding," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[20] J. Zhang, L. Dai, X. Li, Y. Liu, and L. Hanzo, "On low-resolution ADCs in practical 5G millimeter-wave massive MIMO systems," *IEEE Commun. Mag.*, vol. 56, no. 7, pp. 205–211, Jul. 2018.

[21] J. Mo, P. Schniter, N. G. Prelcic, and R. W. Heath, Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization," in *Proc. 48th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2014, pp. 957–961.

[22] J. Mo and R. W. Heath, Jr., "High SNR capacity of millimeter wave MIMO systems with one-bit quantization," in *Proc. Inf. Theory Appl. Workshop (ITA)*, Feb. 2014, pp. 1–5.

[23] L. Landau and G. Fettweis, "On reconstructable ask-sequences for receivers employing 1-bit quantization and oversampling," in *Proc. IEEE Int. Conf. Ultra-WideBand (ICUWB)*, Sep. 2014, pp. 180–184.

[24] A. B. Ucuncu and A. O. Yilmaz, "Oversampling in one-bit quantized massive MIMO systems and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 7952–7964, Dec. 2018.

[25] L. T. N. Landau, M. Dörpinghaus, and G. P. Fettweis, "1-bit quantization and oversampling at the receiver: Communication over bandlimited channels with noise," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1007–1010, May 2017.

[26] Z. Shao, L. T. N. Landau, and R. C. de Lamare, "Sliding window based linear signal detection using 1-Bit quantization and oversampling for large-scale multiple-antenna systems," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Jun. 2018, pp. 183–187.

[27] S. Krone and G. Fettweis, "Capacity of communications channels with 1-bit quantization and oversampling at the receiver," in *Proc. 35th IEEE Sarnoff Symp.*, May 2012.

[28] M. S. Stein, "Performance analysis for time-of-arrival estimation with oversampled low-complexity 1-bit A/D conversion," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 4491–4495.

[29] M. Schluter, M. Dorpinghaus, and G. P. Fettweis, "Bounds on channel parameter estimation with 1-Bit quantization and oversampling," in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2018.

[30] M. Stein, A. Mezghani, and J. A. Nossek, "A lower bound for the Fisher information measure," *IEEE Signal Process. Lett.*, vol. 21, no. 7, pp. 796–799, Jul. 2014.

[31] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 216, Mar. 1952, p. 14.

[32] Z. Shao, L. T. N. Landau, and R. C. de Lamare, "Channel estimation using 1-Bit quantization and oversampling for large-scale multiple-antenna systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 4669–4673.

[33] Z. Shao, L. T. N. Landau, and R. C. de Lamare, "Oversampling based channel estimation for 1-bit large-scale multiple-antenna systems," in *Proc. 23rd Int. ITG Workshop Smart Antennas*, Apr. 2019, pp. 1–5.

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

[35] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," *IEEE Trans. Broadcast.*, vol. 48, no. 3, pp. 223–229, Sep. 2002.

[36] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," 2014, *arXiv:1404.7736*. [Online]. Available: http://arxiv.org/abs/1404.7736

[37] K. L. B. H. L. Van Trees and Z. Tian, *Detection Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory, Detection Estimation and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2013.

[38] B. Clerckx, G. Kim, and S. Kim, "Correlated fading in broadcast MIMO channels: Curse or blessing?" in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Nov. 2008, pp. 1–5.

[39] J. Singh, S. Ponnuru, and U. Madhow, "Multi-gigabit communication: The ADC bottleneck1," in *Proc. IEEE Int. Conf. Ultra-Wideband*, Sep. 2009, pp. 22–27.

[40] Y. Xiong, Z. Zhang, N. Wei, B. Li, and Y. Chen, "Performance analysis of uplink massive MIMO systems with variable-resolution ADCs using MMSE and MRC detection," *Trans. Emerg. Telecommun. Technol.*, vol. 30, no. 5, p. e3549, May 2019.

LUKAS T. N. LANDAU (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Ilmenau University of Technology, Germany, in 2009 and 2011, respectively, and the Ph.D. degree in electrical engineering from Technische Universität Dresden, Germany, in 2016.

Since July 2016, he has been with the Center for Studies in Telecommunications, Pontifical Catholic University of Rio de Janeiro, Brazil, where he is currently an Assistant Professor. His research interests include communications and signal processing. He serves as an Associate Editor for *EURASIP Journal on Wireless Communications*.

RODRIGO C. DE LAMARE (Senior Member, IEEE) was born in Rio de Janeiro, Brazil, in 1975. He received the Diploma degree in electronic engineering from the Federal University of Rio de Janeiro, in 1998, and the M.Sc. and Ph.D. degrees in electrical engineering from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), in 2001 and 2004, respectively.

Since January 2006, he has been with the Communications Group, Department of Electronic Engineering, University of York, U.K. Since April 2013, he has been a Professor with PUC-Rio. He is currently a Professor with the University of York. His research interests include communications and signal processing, areas in which he has published over 400 articles in international journals and conferences. He is an Elected Member of the IEEE Signal Processing Theory and Method Technical Committee. He serves as an Editor for the IEEE Transactions on Communications and the IEEE Wireless Communications Letters.

ZHICHAO SHAO (Student Member, IEEE) received the B.Sc. degree in information engineering from Xidian University, China, in 2012, and the Dipl.-Ing. or M.Sc. degree in electrical engineering from Technische Universität Dresden, Germany, in 2016. He is currently pursuing Ph.D. degree in electrical engineering with the Pontifical Catholic University of Rio de Janeiro, Brazil.

He is also with the Center for Studies in Telecommunications under the supervision of Prof. R. C. de Lamare. His current research interests include communications and signal processing. He is a Reviewer of the IEEE Wireless Communications Letters, *EURASIP Journal on Wireless Communications*, *Transactions on Emerging Telecommunications Technologies*, and several technical conferences.

• • •