Study of Channel Estimation Algorithms for Large-Scale Multiple-Antenna Systems using 1-Bit ADCs and Oversampling

Zhichao Shao, Lukas T. N. Landau and Rodrigo C. de Lamare

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.

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 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], [7], [8], [9], [10], [11], [12], [13], [14], [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], [9], [10] have analyzed different precoding techniques for the downlink. Regarding channel estimation, the studies in [11], [12], [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], [17], [18], [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], [21], [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 (LMSE) 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-LMSE 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 autocorrelation 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. 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. \( I_n \) denotes the \( n \times n \) identity matrix and \( 0_n \) is the \( n \times 1 \) all-zero column vector. Additionally, \( \text{diag}(A) \) is a diagonal matrix only containing the diagonal elements of \( A \). The transpose, conjugate transpose and pseudoinverse of \( A \) are represented by \( A^T \), \( A^H \) and \( A^+ \), respectively. \( a^* \) denotes the complex conjugate of \( a \) and \( |a|_k \) represents the \( k \)th element of vector \( a \), \( (\cdot)^H \) and \( (\cdot)^T \) get the real and imaginary part from the corresponding vector or matrix, respectively. \( \otimes \) is the Kronecker product. Finally, \( \text{vec}(A) \) is the vectorized form of \( A \) obtained by stacking its columns and \( \text{det}(A) \) is the determinant function. \( x \sim \mathcal{CN}(\mu, \Sigma) \) indicates that \( x \) is a complex Gaussian vector with mean \( \mu \) and covariance matrix \( \Sigma \). The expectation and covariance is denoted as \( E[\cdot] \) and \( \text{Cov}[\cdot] \), 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 \( y \in \mathbb{C}^{MN_rN_t \times 1} \) can be expressed as

\[
y = Hx + n,
\]

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

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

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 \( n \) represents the filtered oversampled noise expressed by

\[
n = (I_{N_r} \otimes G)w
\]

with \( w \sim \mathcal{CN}(0_{3MN_tN_r}, \sigma_n^2 I_{3MN_tN_r}) \). Note that the noise samples are described such that each entry of \( n \) 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 $w$ need to be considered for the description of an interval of $MN$ samples of the filtered noise $n$. The matrix $G \in \mathbb{R}^{MN \times 3MN}$ 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), where $T$ is the symbol period and $M$ denotes the oversampling rate. The equivalent channel matrix $H$ is described as

$$H = [I_{N_r} \otimes Z(I_N \otimes u)](H' \otimes I_N), \tag{5}$$

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

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

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

$$Z = \begin{bmatrix}
  z(0) & z(\frac{T}{2}) & \cdots & z(NM - \frac{T}{2}) \\
  z(-\frac{T}{2}) & z(0) & \cdots & z(NM - \frac{T}{2}) \\
  \vdots & \vdots & \ddots & \vdots \\
  z(-NT + \frac{T}{2}) & z(-NT + \frac{T}{2}) & \cdots & z(0)
\end{bmatrix}. \tag{7}$$

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

Let $Q(\cdot)$ represent the 1-bit quantization function, the resulting quantized signal $y_q$ is given by

$$y_q = Q(y) = Q(y^R) + jQ(y^I). \tag{8}$$

The real and imaginary parts of $y$ are quantized element-wised to $\{\pm \frac{1}{2}\}$ based on the sign. The factor $\frac{1}{2}$ 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 $s$ with covariance matrix $C_s$ and its 1-bit quantized signal $s_Q$ is described by \cite{31}

$$C_{s,s_Q} = \sqrt{\frac{2}{\pi}}K C_s, \text{ where } K = \text{diag}(C_s)^{-\frac{1}{2}}. \tag{9}$$

Furthermore, the covariance matrix of the 1-bit quantized signal $s_Q$ can be obtained through the arcsin law \cite{34}

$$C_{s,s_Q} = \frac{2}{\pi} (\text{sin}^{-1}(KC_s^R K) + j\text{sin}^{-1}(KC_s^I K)). \tag{10}$$

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

### 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 \cite{35}. During the training phase, each terminal simultaneously transmits $\tau$ pilot symbols to the BS, which yields

$$y_p = Hx_p + n_p. \tag{11}$$

Vectorizing (11) we get

$$y_p = (x_p^T \otimes I_{N_r}) \text{vec}(H) + n_p$$

$$= [x_p^T \otimes I_{N_r} \otimes Z(I_r \otimes u)] \text{vec}(H' \otimes I_r) + n_p$$

$$= \Phi_p h' + n_p, \tag{12}$$

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

$$\Phi_p = [x_p^T \otimes I_{N_r} \otimes Z(I_r \otimes u)]$$

$$[I_{N_r} \otimes (e_1 \otimes I_{N_r} \otimes e_1 + \cdots + e_r \otimes I_{N_r} \otimes e_r)]. \tag{13}$$

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

\[ y_{Q_p} = Q(\Phi_p h' + n_p) = \tilde{\Phi}_p h' + \tilde{n}_p, \]

(14)

where \( \tilde{\Phi}_p = A_p \Phi_p \in \mathbb{C}^{M \tau N_r \times N_t N_r} \) and \( \tilde{n}_p = A_p n_p + n_q \in \mathbb{C}^{M \tau N_r \times 1} \). The vector \( n_q \) is the statistically equivalent quantization noise\(^1\) with covariance matrix \( C_{n_q} = C_{y_{Q_p}} - A_p C_y A_p^H \). The matrix \( A_p \in \mathbb{R}^{M \tau N_r \times M \tau N_r} \) is the Bussgang-based linear operator chosen independently from \( y_p \) and is given by

\[ A_p = C_{y_p y_{Q_p}}^H C_{y_p}^{-1} = \sqrt{\frac{2}{\pi}} K, \]

(15)

where \( C_{y_p y_{Q_p}} \) denotes the cross-correlation matrix between the received signal \( y_p \) and its quantized signal \( y_{Q_p} \)

\[ C_{y_p y_{Q_p}} = \sqrt{\frac{2}{\pi}} K C_{y_p}, \quad \text{with} \quad K = \text{diag}(C_{y_p})^{-\frac{1}{2}}. \]

(16)

The formulas of (15) and (16) involve the auto-correlation matrix \( C_{y_p} \):

\[ C_{y_p} = \tilde{\Phi}_p R_{h} \tilde{\Phi}_p^H + C_{n_p}, \]

(17)

where \( R_{h} = E\{h' h'^H\} \). We note that more sophisticated parameter estimation techniques can be adapted for use with 1-bit quantization [36], [37], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49], [50], [51], [52], [53], [54], [55], [56], [57], [58].

A. Noise covariance matrix \( C_{n_p} \)

With (3) the auto-correlation matrix \( C_{n_p} \) in (17) is calculated as

\[ C_{n_p} = \sigma^2_n (I_{N_r} \otimes G G^H). \]

(18)

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

\[ C_{n_p} = \sigma^2_n I_{N_r}. \]

(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 \( G G^H \).

B. Standard LS Channel Estimator

The work in [59] 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

\[ \hat{h}_{\text{Standard LS}} = \arg \min_{h'} ||y_{Q_p} - \Phi_p \hat{h}'||^2 \]

\[ = (\tilde{\Phi}_p^H \tilde{\Phi}_p)^{-1} \tilde{\Phi}_p^H y_{Q_p}. \]

(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 \( h' \) scales with the amplitude associated with the quantizer, which then corresponds to a biased estimation.

\(^1\)In this paper, we assume the quantization noise \( n_q \) is Gaussian distributed with zero mean and covariance \( C_{n_q} \).

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:

\[ \hat{h}_{\text{LRA-LS}} = \arg \min_{h'} ||y_{Q_p} - \tilde{\Phi}_p h'||^2 \]

\[ = (\tilde{\Phi}_p^H \tilde{\Phi}_p)^{-1} \tilde{\Phi}_p^H y_{Q_p}. \]

(21)

Compared to the standard LS channel estimator, the proposed estimator has taken \( R_{h'} \) into consideration in order to obtain the linear operator \( A_p \).

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

\[ W_{\text{LMMSE}} = \arg \min_{W} E\{||h' - W y_{Q_p}||^2\} \]

\[ = R_{h'} \tilde{\Phi}_p^H C_{y_{Q_p}}^{-1}. \]

(22)

where

\[ C_{y_{Q_p}} = \frac{2}{\pi} \sin^{-1}(K C_{y_p} K) + j \sin^{-1}(K C_{y_p}^I K) \].

(23)

The resulting LRA-LMMSE channel estimator is then

\[ \hat{h}_{\text{LRA-LMMSE}} = R_{h'} \tilde{\Phi}_p^H C_{y_{Q_p}}^{-1} y_{Q_p}. \]

(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

\[ y_{Q_p} = \Phi_p^H h_{mr} + n_{Q_p}. \]

(25)

where \( y_{Q_p} = [y_{Q_p}(1), y_{Q_p}(2), ..., y_{Q_p}(M \tau)]^T \) and \( h_{mr} \in \mathbb{C}^{N_r \times 1} \) is the \( n_r \)th row of \( H' \). Different from \( \tilde{\Phi}_p \) in (14), \( \Phi_p^H \in \mathbb{C}^{M \tau \times N_t} \) is an equivalent pilot matrix to the \( n_r \)th receive antenna. The sliding window based technique [26] (shown in Fig. 2) 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 $M_{\text{win}}$ oversampled samples and the second contains the next $M_{\text{win}}$ samples until the last window. Note that only one symbol-rate-sampled symbol (or $M$ oversampled samples) is shifted for the subsequent window.

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

Based on (25), the received signal at the $n$th window can be expressed as

$$y_{Q_p}^{\text{nr}}(n) = \tilde{\Phi}_p^{\text{nr}}(n)H^{\text{nr}} + \tilde{n}_p^{\text{nr}}(n),$$

(26)

where $y_{Q_p}^{\text{nr}}(n) = [y_{Q_p}^{\text{nr}}(M(n - 1) + 1), \ldots, y_{Q_p}^{\text{nr}}(Mn + M_{\text{win}})]^T$ and $\tilde{\Phi}_p^{\text{nr}}(n) = A_p^{\text{nr}}(n)\Phi_p^{\text{nr}}(n) \in \mathbb{C}^{M_{\text{win}} \times N_t}$ contains the transmit pilot sequences in the $n$th window.

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

$$\hat{h}^{\text{nr}}_{\text{LRA-LMS}}(n) = \arg\min_{h^{\text{nr}}(n)} \sum_{n=1}^{T-l_{\text{win}}+1} ||y_{Q_p}^{\text{nr}}(n) - \tilde{\Phi}_p^{\text{nr}}(n)\hat{h}^{\text{nr}}(n)||^2,$$

(27)

where $\hat{h}^{\text{nr}}(n)$ is the instantaneous estimate of $h^{\text{nr}}$ in the $n$th window.

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

$$\frac{\partial}{\partial h^{\text{nr}}(n)^H} \sum_{n=1}^{T-l_{\text{win}}+1} ||y_{Q_p}^{\text{nr}}(n) - \tilde{\Phi}_p^{\text{nr}}(n)\hat{h}^{\text{nr}}(n)||^2$$

$$= \sum_{n=1}^{T-l_{\text{win}}+1} -\tilde{\Phi}_p^{\text{nr}}(n)^H(y_{Q_p}^{\text{nr}}(n) - \tilde{\Phi}_p^{\text{nr}}(n)\hat{h}^{\text{nr}}(n))$$

$$= \sum_{n=1}^{T-l_{\text{win}}+1} -\tilde{\Phi}_p^{\text{nr}}(n)^He^{\text{nr}}(n).$$

The recursion of the proposed LRA-LMS algorithm is

$$h^{\text{nr}}(n + 1) = h^{\text{nr}}(n) + \mu\tilde{\Phi}_p^{\text{nr}}(n)^H e^{\text{nr}}(n),$$

(29)

where the constant step size $\mu$ fulfills

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

(30)

$\gamma_{\text{max}}$ is the largest eigenvalue of $C_{\Phi_p^{\text{nr}}}(n)$, which is $E\{\tilde{\Phi}_p^{\text{nr}}(n)(\tilde{\Phi}_p^{\text{nr}}(n))^H\}$.

Proof: See Appendix B.

The proposed adaptive channel estimator is summarized in Algorithm 1, where $x_p(n) \in \mathbb{C}^{\text{len}N_t \times 1}$ contains the pilot symbols in the $n$th window. Both $e_{\text{in}}^n \in \mathbb{R}^{l_{\text{win}} \times 1}$ and $e_{\text{out}}^n \in \mathbb{R}^{N_r \times 1}$ represent all-zero column vectors except that the $n$th elements are ones. Fig. 3 shows the convergence performance of the proposed LRA-LMS channel estimator for each receive antenna. The proposed estimator achieves its steady state after $\tau = 40$.

**Algorithm 1 Proposed LRA-LMS Channel Estimator**

1: Parameters:
   $\mu$: forgetting factor
2: Initialization:
   $h_{\text{LRA-LMS}}^{\text{nr}}(1) = 0_{N_t \times 1}$
3: Iteration:
   for $n_r = 1 : N_r$ do
   5:   for $n = 1 : T - l_{\text{win}} + 1$ do
   6:     $\tilde{\Phi}_p^{\text{nr}}(n) = [x_p(n) \otimes Z(I_{\text{win}} \otimes u)]$
   7: $\frac{1}{2}\text{diag}(C_{\Phi_p^{\text{nr}}}(n))^{-1/2};$
   8: $A_p^{\text{nr}}(n) = \sqrt{\frac{1}{2N_t}} C_{\Phi_p^{\text{nr}}}(n) + \sigma_n^2 G G^H$;
   9: $\Phi_p^{\text{nr}}(n) = A_p^{\text{nr}}(n)\Phi_p^{\text{nr}}(n);$
   10: $e^{\text{nr}}(n) = y_{Q_p}^{\text{nr}}(n) - \Phi_p^{\text{nr}}(n)h^{\text{nr}}(n);$ 
   11: $h^{\text{nr}}(n + 1) = h^{\text{nr}}(n) + \mu\tilde{\Phi}_p^{\text{nr}}(n)^He^{\text{nr}}(n);$ 
   13: end for

![Convergence of the LRA-LMS channel estimator with $N_t = 8$ and $N_r = 64$ at SNR = 20dB.](image)

**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 $R_{h'}$ is an identity matrix. Table I shows the total required complex additions/subtractions and multiplications/divisions for obtaining the channel estimate $h'$. More intuitively, Fig. 4 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...
TABLE I: Computational complexity of different channel estimators

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

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.

Fig. 4: Computational complexity comparison between different channel estimators in an oversampled system $M = 3$ with $\tau = 20$, $t_{\text{win}} = 3$ and $N_t = 8$.

G. Estimation of $R_{h'}$

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

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

(31)

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

$$y_Q(n) = Q[(Hx(n) + n(n))]$$

$$= Q[(x_p^T(n) \otimes I_{N_r} \otimes Z'u)h' + n(n)],$$

(32)

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

$$\hat{h}'(n) = (x_p^T(n) \otimes I_{N_r} \otimes Z'u)^+y_Q(n),$$

(33)

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

IV. CRÅÎ®-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 $R_{h'}$ is known at the receiver, namely Bayesian CRB with known $R_{h'}$ and general CRB with estimated $R_{h'}$.

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

$$\begin{bmatrix} y_R' \\ y_I' \end{bmatrix} = \begin{bmatrix} \Phi_R & -\Phi_I \\ \Phi_I & \Phi_R \end{bmatrix} \begin{bmatrix} h_R' \\ h_I' \end{bmatrix} + \begin{bmatrix} n_R \\ n_I \end{bmatrix}. \quad (34)$$

Let $\tilde{h}' = [h_R'; h_I']$ be the unknown parameter vector, since the real and imaginary parts are independent, the Bayesian information matrix (BIM) [60] for the quantized signal is defined as

$$J_{y_Q}(\tilde{h}') = J_{y_Q^R}(\tilde{h}') + J_{y_Q^I}(\tilde{h}'),$$

(35)

where

$$[J_{y_Q^R}(\tilde{h}')]_{ij} \triangleq E_{y_Q^R}(\tilde{h}') \left\{ \frac{\partial \ln p(y_Q^R(y', \tilde{h}'))}{\partial [h']_i} \frac{\partial \ln p(y_Q^R(y', \tilde{h}'))}{\partial [h']_j} \right\}, \quad (36)$$

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

$$[J_{y_Q^R}(\tilde{h}')]_{ij} = [J_{y_Q^R}(\tilde{h}')]_{ij} + [J_{y_Q^R}(\tilde{h}')]_{ij},$$

(37)

where

$$[J_{y_Q^R}(\tilde{h}')]_{ij} \triangleq E_{y_Q^R}(\tilde{h}') \left\{ \frac{\partial \ln p(y_Q^R(y', \tilde{h}'))}{\partial [h']_i} \frac{\partial \ln p(y_Q^R(y', \tilde{h}'))}{\partial [h']_j} \right\}, \quad (38)$$

$$[J_{y_Q^R}(\tilde{h}')]_{ij} \triangleq E_{y_Q^R}(\tilde{h}') \left\{ \frac{\partial \ln p(y_Q^R(y', \tilde{h}'))}{\partial [h']_i} \frac{\partial \ln p(y_Q^R(y', \tilde{h}'))}{\partial [h']_j} \right\}. \quad (39)$$

To transform the real-valued $J_{y_Q}(\tilde{h}')$ back to the complex domain $J_{y_Q}(\tilde{h}')$, $J_{y_Q}(\tilde{h}')$ is defined with the following structure:

$$J_{y_Q}(\tilde{h}') = \begin{bmatrix} J_{y_Q^R}^R(\tilde{h}') & J_{y_Q^R}^I(\tilde{h}') \\ J_{y_Q^I}^R(\tilde{h}') & J_{y_Q^I}^I(\tilde{h}') \end{bmatrix}.$$
and apply the chain rule to get:
\[ \mathbf{J}_{\mathbf{y}}(\mathbf{h}') = \frac{1}{4} (\mathbf{J}_{\mathbf{y}}^{RR}(\mathbf{h}') + \mathbf{J}_{\mathbf{y}}^{JJ}(\mathbf{h}')) + \frac{j}{4} (\mathbf{J}_{\mathbf{y}}^{RI}(\mathbf{h}') - \mathbf{J}_{\mathbf{y}}^{JR}(\mathbf{h}')), \]
where \( \mathbf{J}_{\mathbf{y}}^{RR}(\mathbf{h}') \), \( \mathbf{J}_{\mathbf{y}}^{RI}(\mathbf{h}') \), \( \mathbf{J}_{\mathbf{y}}^{JR}(\mathbf{h}') \) and \( \mathbf{J}_{\mathbf{y}}^{JJ}(\mathbf{h}') \) have the same dimensions \( N_c \times N_c \). The variance of the estimator
\( \mathbf{h}'(\mathbf{y}) \) is lower bounded by
\[ \text{Var} \{ \mathbf{h}'(\mathbf{y}) \} \geq \mathbf{J}_{\mathbf{y}}^{-1}(\mathbf{h}'). \]

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 \( \mathbf{C}_n = \sigma^2 \mathbf{I}_{N_c} \). With the independence of the real and imaginary parts, the log-likelihood function can be expressed as
\[ \ln p(\mathbf{y} | \mathbf{h}') = \sum_{k=1}^{N_c} \ln p([\mathbf{y}]_k | \mathbf{h}') + \ln p([\mathbf{y}_R]_k | \mathbf{h}') \]
with
\[ p([\mathbf{y}]_k | \mathbf{h}') = \pm \frac{1}{\sqrt{2}} | \mathbf{h}' \rangle \mathbf{Q} \left( \frac{\left[ \Phi_R^R \mathbf{h}' - \Phi_I^I \mathbf{h}' \right]}{\sigma_n/\sqrt{2}} \right) \]
and
\[ p([\mathbf{y}]_k | \mathbf{h}') = \pm \frac{1}{\sqrt{2}} | \mathbf{h}' \rangle \mathbf{Q} \left( \frac{\left[ \Phi_I^I \mathbf{h}' + \Phi_R^R \mathbf{h}' \right]}{\sigma_n/\sqrt{2}} \right) \]
where \( Q(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp(-u^2) du \). Inserting (43) into (38), we obtain
\[ \mathbf{J}_{\mathbf{y}}(\mathbf{h}')_{ij} = - \frac{\partial^2 \ln p(\mathbf{y} | \mathbf{h}')}{\partial [\mathbf{h}'][j] \partial [\mathbf{h}'][j]} \]
\[ = \mathbf{J}_{\mathbf{y}}^{RR}(\mathbf{h}')_{ij} + \mathbf{J}_{\mathbf{y}}^{JJ}(\mathbf{h}')_{ij}. \]

With the derivative of the \( Q(x) \) function, the real part in (38) \( \mathbf{J}_{\mathbf{y}}(\mathbf{h}')_{ij} \) is given by
\[ \mathbf{J}_{\mathbf{y}}^{RR}(\mathbf{h}')_{ij} = \sum_{k=1}^{N_c} \ln p([\mathbf{y}]_k | \mathbf{h}') \]
\[ = \frac{1}{\sigma_n^2} \mathbf{C}_n^{-1} \mathbf{C}_n \mathbf{C}_n^{-1} \mathbf{C}_n \]
\[ \times \sum_{k=1}^{N_c} \exp(-\left[ \frac{\left[ \Phi_R^R \mathbf{h}' - \Phi_I^I \mathbf{h}' \right]}{\sigma_n/\sqrt{2}} \right]^2) \]
\[ \times \left( \frac{\left[ \Phi_R^R \mathbf{h}' - \Phi_I^I \mathbf{h}' \right]}{\sigma_n/\sqrt{2}} \right) \left( \frac{\left[ \Phi_I^I \mathbf{h}' + \Phi_R^R \mathbf{h}' \right]}{\sigma_n/\sqrt{2}} \right) \]
\[ \mathbf{J}_{\mathbf{y}}^{RR}(\mathbf{h}')_{ij} = - \frac{1}{2} \mathbf{C}_n^{-1}. \]

The derivation for the imaginary part \( \mathbf{J}_{\mathbf{y}}^{JJ}(\mathbf{h}')_{ij} \) is analogous.

By assuming that \( \mathbf{h}' \) is Gaussian distributed with zero mean and covariance matrix \( \mathbf{C}_h = \frac{1}{2} \mathbf{I}_2 \otimes \mathbf{C}_n \), \( \mathbf{n} \) is \( \mathbf{n} \) yields
\[ \ln p(\mathbf{h}') = -\frac{1}{2} N_c N_t \ln(2\pi) - \frac{1}{2} \mathbf{h}'^T \mathbf{C}_h^{-1} \mathbf{h}'. \]

Substituting (48) into (39), we obtain
\[ \mathbf{J}_{\mathbf{y}}(\mathbf{h}') = 2 \mathbf{J}_{\mathbf{y}}^{RR}(\mathbf{h}') = 2 \mathbf{C}_h^{-1}. \]

Finally, the resulting BIM is the summation of (46) and (49) as described by
\[ \mathbf{J}_{\mathbf{y}}(\mathbf{h}') = \mathbf{J}_{\mathbf{y}}^{D}(\mathbf{h}') + \mathbf{J}_{\mathbf{y}}^{P}(\mathbf{h}'). \]
Lemma 1. The proposed LRA channel estimators with combination of estimated $\hat{R}_h$ are biased channel estimators.

Proof: See Appendix C.

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

$$\text{Cov}(\hat{h}_{\text{bias}}') \geq \frac{\partial E^R[\hat{h}_{\text{bias}}]}{\partial h'} \left( J_{y_Q}^{(R)}(h')^{-1} \frac{\partial E^R[\hat{h}_{\text{bias}}]}{\partial h'} \right)^T $$

(57)

$$\text{Cov}(\hat{h}_{\text{bias}}') \geq \frac{\partial E^I[\hat{h}_{\text{bias}}]}{\partial h'} \left( J_{y_Q}^{(I)}(h')^{-1} \frac{\partial E^I[\hat{h}_{\text{bias}}]}{\partial h'} \right)^T $$

(58)

where $J_{y_Q}^{(R)}(h')$ and $J_{y_Q}^{(I)}(h')$ are defined by

$$[J_{y_Q}^{(R)}(h')]_{ij} \triangleq E \left\{ \frac{\partial \ln p(y_Q | h')}{\partial [h']_i} \frac{\partial \ln p(y_Q | h')}{\partial [h']_j} \right\} $$

(59)

$$[J_{y_Q}^{(I)}(h')]_{ij} \triangleq E \left\{ \frac{\partial \ln p(y_Q | h')}{\partial [h']_i} \frac{\partial \ln p(y_Q | h')}{\partial [h']_j} \right\} $$

(60)

which are the upper left and lower right part of the $J_{y_Q}^{(D)}(h')$ (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 $10 \log_10(\frac{2}{\sigma^2})$. 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 $l_{\text{win}}$ is chosen as three to ensure low computational complexity. The step size $\mu$ is optimized according to the oversampling factor and SNR. In the simulation, $\mu$ 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. We remark that more sophisticated detectors could also be considered [61], [62], [63], [64], [65], [66], [61], [67], [45], [68], [69], [70], [71], [72], [73], [74], [75], [76], [77] along with precoders [78], [79], [80], [81], [82], [83].

The performance of the channel estimators is evaluated based on the channel model simulated in [84]. The channel for user $n_t$ is assumed Rayleigh distributed

$$h'_{n_t} = R_{r,n_t} h'_{w,n_t} $$

(61)

where $R_{r,n_t}$ denotes the receive correlation matrix with the following form

$$R_{r,n_t} = \begin{bmatrix}
1 & \rho_n t^{(N_t-1)} & \cdots & \rho_n^{(N_t-2)} \\
\rho_n^{*} & 1 & \cdots & \rho_n^{* (N_t-2)} \\
\vdots & \vdots & \ddots & \vdots \\
\rho_n^{*(N_t-1)} & \rho_n^{*(N_t-2)} & \cdots & 1
\end{bmatrix} $$

(62)

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

$$H' = [h'_{1}, h'_{2}, \cdots, h'_{n_t}] $$

(63)

and $R_{h'}$ is calculated as

$$R_{h'} = \begin{bmatrix}
R_{r,1} & 0 & \cdots & 0 \\
0 & R_{r,2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & R_{r,n_t}
\end{bmatrix} $$

(64)

A. $R_{h'}$ is known at the receiver

In this subsection, we evaluate the performance of the proposed LRA channel estimators with known $R_{h'}$ at the receiver. Fig. 5a and Fig. 5b 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 $H'$ row by row. This approach can largely reduce the computational cost (shown in Fig. 4). Note that this separation into several rows may overlook the correlation of receive antennas. More specifically, the proposed LRA-LMS treats $R_{r,n_t}$ as an identity matrix. As an amendment, the resulting estimated channel matrix $h'_{\text{LRA-LMS}}$ needs to be multiplied with the square root of the receive correlation matrix $R_{r,n_t}^{\frac{1}{2}}$, which can be derived from $R_{h'}$ 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 ($\leq 5$ dB), whereas at high SNR this performance gap becomes large.

The Bayesian CRBs illustrated in Section IV-A are also depicted in Fig. 5. 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.
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. 6, where the oversampled systems obviously outperform the non-oversampled systems. As described in III-A, Fig. 7 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 $\mathbf{C}_n$ is not a diagonal matrix in oversampled systems. It can be seen that at low SNR ($\leq 10$ dB) the performance of simplified LMMSE [24] is worse than the proposed LRA-LMMSE, although they converge together at high SNR ($\geq 10$ 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. $\mathbf{R}_{h'}$ is unknown at the receiver

Practically, $\mathbf{R}_{h'}$ is not known at the receiver. Fig. 8 shows the MSE performance of the LRA channel estimators by using the proposed adaptive recursion to estimate $\hat{\mathbf{R}}_{h'}$, where $\lambda$ is set to 0.99. It can be seen that the performance remains almost the same as Fig. 5a, which shows that the proposed estimation of $\hat{\mathbf{R}}_{h'}$ 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 $\frac{\partial \mathbb{E}\{\hat{\mathbf{h}}^{N_r}R/I\}}{\partial \mathbf{h}'}$, 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. 9 shows the normalized MSE performance of the LRA-LS channel estimator with estimated $\hat{\mathbf{R}}_{h'}$ in (31) for estimating the first $N_r$ elements of $\mathbf{h}^R$ and its corresponding numerically calculated general CRBs under uncorrelated channels ($|\rho| = 0$). More

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.
These steps are repeated until all the elements in $\partial E$ are obtained.

Specifically, each element of the gradient vector $\frac{\partial E}{\partial h^{R/I}}$ is calculated with the following steps:

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

These steps are repeated until all the elements in $\frac{\partial E}{\partial h^{R/I}}$ are obtained.

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 (a few milliwatts) [85], [11], [7]. As one example, Fig. 11 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 [86]

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

where $P_{\text{BB}}$, $P_{\text{LO}}$, $P_{\text{LNA}}$, $P_{\text{H}}$, $P_{\text{M}}$ and $P_{\text{AGC}}$ denote the power consumption in the baseband processor, local oscillator (LO), low noise amplifier (LNA), $\frac{f}{2}$ 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 $P_{\text{ADC}}$ is calculated as

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

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 autocorrelation 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 (29)

Recall the optimization problem

\[ W_{\text{LMMSE}} = \arg \min_W E\{ ||h' - W y_{\Omega_p}||^2 \}. \]  

(67)

Taking the partial derivative with respect to \( W^H \), we obtain

\[ \frac{\partial E\{ ||h' - W y_{\Omega_p}||^2 \}}{\partial W^H} = -E\{h' y_{\Omega_p}^H\} + WE\{y_{\Omega_p} y_{\Omega_p}^H\}. \]  

(68)

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

\[ W_{\text{LMMSE}} = E\{h'y_{\Omega_p}^H\} E\{y_{\Omega_p} y_{\Omega_p}^H\}^{-1} \]
\[ = (E\{h'h_p^H\} \tilde{\Phi}_p^H + E\{h'n_p^H\})C_{\Omega_p}^{-1}. \]  

(69)

Since \( h' \) is uncorrelated with \( n_p \) and \( n_q \) [11], we have

\[ E\{h'n_p^H\} = E\{h'(A_p n_p + n_q)^H\} = 0. \]  

(70)

The resulting LRA-LMMSE channel estimator is

\[ \hat{h}'_{\text{LRA-LMMSE}} = R_{h'} \tilde{\Phi}_p^H C_{\Omega_p}^{-1} y_{\Omega_p}. \]  

(71)

APPENDIX B
PROOF OF (29)

Defining \( \epsilon(n) = h''_p(n) - h'_p(n) \) and inserting it into (29), we obtain

\[
\begin{align*}
\epsilon(n+1) &= \epsilon(n) + \mu \hat{\Phi}^H_p(n) H (y_{\Omega_p}^n(n) - \hat{\Phi}^H_p(n) h''_p(n)) \\
&= \epsilon(n) + \mu \hat{\Phi}^H_p(n) H y_{\Omega_p}^n(n) \\
&\quad - \mu \hat{\Phi}^H_p(n) H \hat{\Phi}^H_p(n) (\epsilon(n) + h''_p(n)) \\
&= (I - \mu \hat{\Phi}^H_p(n) H \hat{\Phi}^H_p(n)) \epsilon(n) \\
&\quad + \mu \hat{\Phi}^H_p(n) H (y_{\Omega_p}^n(n) - \hat{\Phi}^H_p(n) h''_p(n)).
\end{align*}
\]  

(72)

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

\[ E\{\epsilon(n+1)\} = (I - \mu E\{\hat{\Phi}^H_p(n) H \hat{\Phi}^H_p(n))\} E\{\epsilon(n)\}. \]  

(73)

With the eigenvalue decomposition \( E\{\hat{\Phi}^H_p(n) H \hat{\Phi}^H_p(n)) = Q \Gamma Q^H \), (73) can be written as

\[ Q^H E\{\epsilon(n+1)\} = Q^H (I - \mu \Gamma) Q E\{\epsilon(n)\}, \]  

(74)

where \( Q \) is a unitary matrix and \( \Gamma \) is a diagonal matrix, whose diagonal entries are the eigenvalues of \( E\{\hat{\Phi}^H_p(n) H \hat{\Phi}^H_p(n)) \). With \( u(n) = Q^H E\{\epsilon(n)\} \), (74) is then

\[ u(n+1) = (I - \mu \Gamma) u(n). \]  

(75)

Decoupling the matrix form into individual elements we get

\[ u_{n_1}(n+1) = (1 - \mu \gamma_{n_1}) u_{n_1}(n) \\
= (1 - \mu \gamma_{n_1})^{t-n_1} u_{n_1}(1), \quad n_1 = 1, \ldots, N_t. \]  

(76)

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

\[ |1 - \mu \gamma_{n_1}| < 1. \]  

(77)

The stability condition is then given by

\[ 0 < \mu < \frac{2}{\gamma_{\text{max}}}, \]  

(78)

where \( \gamma_{\text{max}} \) is the largest eigenvalue of \( E\{\hat{\Phi}^H_p(n) H \hat{\Phi}^H_p(n)) \).

APPENDIX C
PROOF OF LEMMA 1

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

\[ E\{\hat{h}'(n)\} = E\{ (x_p^T(n) \otimes I_N, \otimes Z'u)^* y_{\Omega}^n(n) \}. \]  

(79)

From the Bussgang theorem (32) can be decomposed as

\[ y_{\Omega}^n(n) = Q \left( (x_p^T(n) \otimes I_N, \otimes Z'u)h' + n(n) \right) \\
= A'_p(n) (x_p^T(n) \otimes I_N, \otimes Z'u) h' + n(n) + n_q(n), \]  

(80)

where \( A'_p(n) \) is the linear operator and \( n_q(n) \) is the statistically equivalent quantizer noise. Substituting (80) into (79) and with \( \Phi'(n) = (x^T(n) \otimes I_N, \otimes Z'u) \), we obtain

\[ E\{\hat{h}'(n)\} = E\{(\Phi'(n) + A'_p(n) (\Phi'(n) h' + n(n)) + n_q(n)) \} \\
= E\{(\Phi'(n) + A'_p(n) \Phi'(n) h') \} \\
+ E\{ A'_p(n) n_q(n) \} + E\{ \Phi'(n) n_q(n) \}. \]  

(81)
Since $\Phi(n)$ and $n(n)$ are uncorrelated and $E\{n(n)\} = 0$, we have

\[ E\{\Phi(n)^{+} A'(p) n(n)\} = 0. \]  

(82)

Similarly,

\[ E\{\Phi(n)^{+} n_{p}(n)\} = 0. \]  

(83)

Equation (81) can be further simplified as

\[ E\{\hat{h}(n)\} = E\{\Phi(n)^{+} A'(p) \Phi(n)\} \hat{h}'. \]  

(84)

The matrix $A'(p)$ depends on $R_k$, 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 [60]. With the analysis above, it is concluded that the adaptive estimator $R_k$ 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. Academic Press, 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. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, 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 Commun. Mag., vol. 52, no. 2, pp. 186–195, Feb. 2014.

[5] L. T. N. Landau, M. Dörpinghaus, and G. P. Fettweis, “1-bit quantization and oversampling for large-scale multiple-antenna systems,” in Proc. IEEE Glob. Commun. Conf. (GLOBECOM), Dec. 2014.

[6] 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.

[7] J. Mo and R. W. Heath, “High SNR capacity of millimeter wave MIMO systems with one-bit quantization,” in 2014 48th Asilomar Conference on Signals, Systems and Computers, Nov. 2014, pp. 957–961.

[8] D. Persson, T. Söderström, and M. Haardt, “Identifying low-complexity 1-bit systems using LRA,” IEEE Trans. Signal Process., vol. 63, no. 18, pp. 4699–4711, Sep. 2015.

[9] M. A. John and C. Studer, “On the capacity of 1-bit massive MIMO systems and performance analysis,” IEEE Trans. Wireless Commun., vol. 17, no. 12, pp. 7952–7964, Dec. 2018.

[10] L. T. N. Landau and R. C. Lamare, “One-bit quantization and oversampling for receivers employing 1-bit quantization and oversampling,” in IEEE 19th Int. Workshop on Signal Process. Advances in Wireless Comm. (SPAWC), Jun. 2018.

[11] C. Krone and G. Fettweis, “Capacity of communications channels with 1-bit quantization and oversampling and at the receiver,” in 2012 35th IEEE Sarnoff Symposium, May 2012.

[12] T. Wang, R. C. de Lamare, and P. D. Mitchell, “Low-complexity set-partitioning based channel estimation for massive MIMO systems,” IEEE Trans. Signal Process., vol. 62, no. 9, pp. 2294–2303, Sep. 2014.

[13] J. Sel. Areas Commun., vol. 17, no. 1, pp. 40–60, Jan. 2013.

[14] Z. Shao, L. T. N. Landau, and R. C. Lamare, “Sliding window based linear signal detection using 1-bit quantization and oversampling for large-scale multiple-antenna systems,” in 2018 IEEE Statistical Signal Processing Workshop (SSP), Jun. 2018, pp. 183–187.

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

[16] C. Studer, J. Choi, E. G. Larsson, R. W. Heath, “One-bit ADCs in wideband massive MIMO systems with OFDM transmission,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Mar. 2016, pp. 3386–3390.
[83] L. T. N. Landau and R. C. de Lamare, “Branch-and-bound precoding for multiuser mimo systems with 1-bit quantization,” *IEEE Wireless Communications Letters*, vol. 6, no. 6, pp. 770–773, Dec 2017.

[84] B. Clerckx, G. Kim, and S. Kim, “Correlated fading in broadcast MIMO channels: Curse or blessing?” in *2008 IEEE Global Telecommunications Conference*, Nov 2008.

[85] J. Singh, S. Ponnuru, and U. Madhow, “Multi-Gigabit communication: the ADC bottleneck1,” in *2009 IEEE International Conference on Ultra-Wideband*, Sep. 2009, pp. 22–27.

[86] 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,” *Transactions on Emerging Telecommunications Technologies*, vol. 30, no. 5, p. 3549, 2019.