Spectral Efficiency of Mixed-ADC Massive MIMO

Hessam Pirzadeh, *Student Member, IEEE*, and A. Lee Swindlehurst, *Fellow, IEEE*

***Abstract*—We study the spectral efficiency (SE) of a mixed-analog-to-digital converter (ADC) massive multiple-input multiple-output (MIMO) system in which K single-antenna users communicate with a base station equipped with M antennas connected to N high-resolution ADCs and $M - N$ 1-b ADCs. This architecture has been proposed as an approach for realizing massive MIMO systems with reasonable power consumption. First, we investigate the effectiveness of mixed-ADC architectures in overcoming the channel estimation error caused by coarse quantization. For the channel estimation phase, we study to what extent one can combat the SE loss by exploiting just $N \ll M$ pairs of high-resolution ADCs. We extend the round-robin training scheme for mixed-ADC systems to include both high-resolution and 1-b quantized observations. Then, we analyze the impact of the resulting channel estimation error in the data detection phase. We consider random high-resolution ADC assignment and also analyze a simple antenna selection scheme to increase the SE. Analytical expressions are derived for the SE for maximum ratio combining and numerical results are presented for zero-forcing detection. Performance comparisons are made against systems with uniform ADC resolution and against mixed-ADC systems without round-robin training to illustrate under what conditions each approach provides the greatest benefit.**

***Index Terms*—Massive multiple-input multiple-output (MIMO), analog-to-digital converter, mixed-ADC, spectral efficiency.**

I. Introduction

THE seminal work of Marzetta introduced massive multiple-input multiple-output (MIMO) as a promising architecture for future wireless systems [2]. In the limit of an infinite number of base station (BS) antennas, it was shown that massive MIMO can substantially increase the network capacity. Another key potential of massive MIMO systems which has also made it interesting from a practical standpoint is its ability of achieving this goal with inexpensive, low-power components [3], [4]. However, preliminary studies on massive MIMO systems have for the most part only analyzed its performance under the assumption of perfect hardware [5], [6]. The impact of hardware imperfections and nonlinearities on massive MIMO systems has recently been investigated in [7]–[12]. Although it is well-known that the dynamic power in massive MIMO systems can be scaled down proportional to $\sqrt{M}$, where M denotes the number of BS antennas, the static power consumption at the BS will increase proportionally to M [8]. Hence, considering hardware-aware design together with power consumption at the BS seems necessary in realizing practical massive MIMO systems.

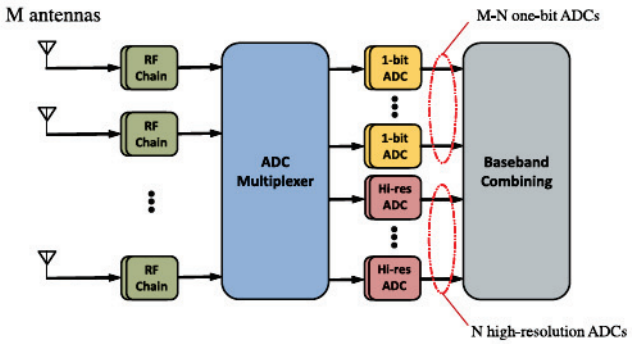

Fig. 1. Mixed-ADC architecture.

Among the various components responsible for power dissipation at the BS, the contribution of analog-to-digital converters (ADCs) is known to be dominant [13]. Consequently, the idea of replacing the high-power high-resolution ADCs with power efficient low-resolution ADCs could be a viable approach to address power consumption concerns at the massive MIMO BSs. The impact of utilizing low-resolution ADCs on the spectral efficiency (SE) and energy consumption of massive MIMO systems has been considered in [14]–[22]. In particular, studies on massive MIMO systems with purely one-bit ADCs show that the high spatial multiplexing gain owing to the use of a large number of antennas is still achievable even with one-bit ADCs [14], [15]. However, many more antennas with one-bit ADCs (at least 2–2.5 times) are required to attain the same performance as in the high-resolution ADCs case.

One of the main causes of SE degradation in purely one-bit massive MIMO systems is the error due to the coarse quantization that occurs during the channel estimation phase. While at low SNR the loss due to one-bit quantization is only about 2 dB, at higher SNRs performance degrades considerably more and leads to an error floor [14]. The SE degradation can be reduced by improving the quality of the channel estimation prior to signal detection. One approach for doing so is to exploit so-called mixed-ADC architectures during the channel estimation phase, in which a combination of low- and high-resolution ADCs are used side-by-side. This architecture is depicted in Fig. 1. Mixed-ADC implementations were introduced in [23],

Manuscript received July 30, 2017; revised November 15, 2017, February 21, 2018, and April 23, 2018; accepted April 25, 2018. Date of publication May 9, 2018; date of current version June 8, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chandra Ramabhadra Murthy. This work was supported in part by the National Science Foundation under Grant ECCS-1547155 and Grant CCF-1703635 and in part by a Hans Fischer Senior Fellowship from the Technische Universität München Institute for Advanced Study. This paper was presented in part at the 7th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, Curacao, Dutch Antilles, December 2017. *(Corresponding author: Hessam Pirzadeh.)*

The authors are with the Center for Pervasive Communications and Computing, University of California—Irvine, Irvine, CA 92697 USA (e-mail: hpirzade@uci.edu; swindle@uci.edu).

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

Digital Object Identifier 10.1109/TSP.2018.2833807

[24] and their performance was studied from an information theoretic perspective via generalized mutual information.

The basic premise behind the mixed-ADC architecture is to achieve the benefits of conventional massive MIMO systems by just exploiting $N \ll M$ pairs of high-resolution ADCs. An SE analysis of mixed-ADC massive MIMO systems with maximum ratio combining (MRC) detection for Rayleigh and Rician fading channels was carried out in [25] and [26], respectively. The SE and energy efficiency of mixed-ADC systems compared with systems composed of one-bit ADCs was studied in [27] for MRC detection, and conditions were derived under which each architecture provided the highest SE for a given power consumption. The advantage of using a mixed-ADC architecture in designing Bayes-optimal detectors for MIMO systems with low-resolution ADCs is reported in [28]. Although the nonlinearity of the quantization process increases the complexity of the optimal detectors, it is shown that adding a small number of high-resolution ADCs to the system allows for less complex detectors with only a slight performance degradation. Moreover, the benefit of using mixed-ADC architectures in massive MIMO relay systems and cloud-RAN deployments is elaborated in [29], [30].

Most existing work in the mixed-ADC massive MIMO literature has assumed either perfect channel state information (CSI) or imperfect CSI with "round-robin" training. In the round-robin training approach [23], [24], [26], the training data is repeated several times and the high-resolution ADCs are switched among the RF chains so that every antenna can have a "clean" snapshot of the pilots for channel estimation. This obviously requires a larger portion of the coherence interval to be devoted to training rather than data transmission. More precisely, for M antennas and N pairs of high-resolution ADCs, M/N pilot signals are required in the single-user scenario to estimate all M channel coefficients with high-resolution ADCs. This issue is pointed out in [23] for the single user scenario and its impact is taken into account. This training overhead will be exacerbated in the multiuser scenario where orthogonal pilot sequences should be assigned to the users. In this case, the training period becomes $(M/N)\eta$, where η represents the length of the pilot sequences (at least as large as the number of user terminals), which could be prohibitively large and may leave little room for data transmission. Hence, it is crucial to account for this fact in any SE analysis of mixed-ADC massive MIMO systems.

In this paper, we examine the channel estimation performance and the resulting uplink SE of mixed-ADC architectures with and without round-robin training, and compare them with implementations that employ uniform ADC quantization across all antennas. The main goals are to determine when, if at all, the benefits of using the round-robin approach with ADC/antenna switching outweigh the cost of increasing the training overhead, and furthermore to examine the question of whether or not one should employ a mixed-ADC architecture in the first place. The contributions of the paper can be summarized as follows.

- We first present an extension of the round-robin training approach that incorporates both high-resolution and one-bit measurements for the channel estimation. The round-robin training proposed in [23], [24], [26] based the channel estimate on only high-resolution observations, assuming that no data was collected from antennas during intervals when they were not connected to the high-resolution ADCs. In contrast, our extension assumes that these antennas collect one-bit observations and combine this data with the high-resolution samples to improve the channel estimation performance.
- We use the Bussgang decompositon [31] to develop a linear minimum mean-squared error (LMMSE) channel estimator based on the combined round-robin measurements and we derive a closed-form expression for the resulting mean-squared error (MSE). We further illustrate the importance of using the Bussgang approach rather than the simpler additive quantization noise model in obtaining the most accurate characterization of the channel estimation performance for round-robin training. The analysis illustrates that the addition of the one-bit observations considerably improves performance at low SNR.
- We perform a spectral efficiency analysis of the mixed-ADC implementation for the MRC and ZF receivers, and obtain expressions for a lower bound on the SE that takes into account the channel estimation error and the loss of efficiency due to the round-robin training. We compare the resulting SE with that achieved by mixed-ADC implementations that do not switch ADCs among the RF chains, and hence do not use round-robin training. We also compare against the SE for architectures that do not mix the ADC resolution across the array, but instead use uniform resolution with a fixed number of comparators for different array sizes. We show that, depending on the SNR, coherence interval, number of high-resolution ADCs, and the choice of the linear receiver, there are situations where each of the considered approaches shows superior performance. In particular, using uniform low-resolution ADCs is better than a mixed-ADC approach for an interference limited system. On the other hand, a mixed-ADC system, even one with round-robin training, is superior at higher SNRs when zero-forcing is used to reduce the interference.
- We analyze the possible SE improvement that can be achieved by using an antenna selection algorithm that connects the high-resolution ADCs to the subset of antennas with the highest channel gain. We analytically derive the SE performance of the antenna selection algorithm for MRC and numerically study its performance for ZF detection, comparing against the simpler approach of assigning the high-resolution ADCs to an arbitrary fixed subset of the RF chains.

In addition to the above contributions, we also discuss some of the issues related to implementing an ADC switch or multiplexer in hardware that allows different ADCs to be assigned to different antennas. We restrict our analysis and numerical examples to a single-carrier flat-fading scenario, although our methodology can be used in a straightforward way to extend the results to frequency-selective fading or multiple-carrier signals (e.g., see our prior work in Section III.B of [14] for the SE analysis of an all-one-bit ADC system for OFDM and frequency selectivity). The reasons for focusing on the single-carrier flat-fading case

are as follows: (1) the mixed-ADC assumption already makes the resulting analytical expressions quite complicated even for the simple flat-fading case, and it would be more difficult to gain insight into the problem if the expressions were further complicated; (2) the original round-robin training idea was proposed in [23] for the single-carrier flat-fading case, and thus we analyze it under the same assumptions; (3) the main conclusions of the paper are based on relative algorithm comparisons for the same set of assumptions, and we expect our general conclusions to remain unchanged if frequency rather than flat fading were considered; and (4) the flat fading case is still of interest in some applications, for example in a micro-cell setting with typical path-length differences of 50–100 m, the coherence bandwidth is between 3–6 MHz, which is not insignificant.

Further assumptions regarding the system model are outlined in the next section. Section III discusses channel estimation using round-robin training, and derives the LMMSE channel estimator that incorporates both the high-resolution and one-bit observations. A discussion of hardware and other practical considerations associated with using a mixed-ADC system with ADC/antenna switching is presented in Section IV. Section V then presents the analysis of the spectral efficiency for MRC and ZF receivers based on the imperfect channel state estimates, including an analytical performance characterization of antenna selection and architectures with uniform ADC resolution across the array. A number of numerical studies are then presented in Section VI to illustrate the relative performance of the algorithms considered.

Notation: We use boldface letters to denote vectors, and capitals to denote matrices. The symbols $(.)^*$, $(.)^T$, and $(.)^H$ represent conjugate, transpose, and conjugate transpose, respectively. A circularly-symmetric complex Gaussian (CSCG) random vector with zero mean and covariance matrix $\mathbf{R}$ is denoted $\boldsymbol{v} \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$. The symbol $\|.\|$ represents the Euclidean norm. The $K \times K$ identity matrix is denoted by $\boldsymbol{I}_K$ and the expectation operator by $\mathbb{E}\{.\}$. We use $\mathbf{1}_N$ to denote the $N \times 1$ vector of all ones, and $\mathrm{diag}\{\mathbf{C}\}$ the diagonal matrix formed from the diagonal elements of the square matrix $\mathbf{C}$. For a complex value, $c = c_R + jc_I$, we define $\arcsin(c) \triangleq \arcsin(c_R) + j\arcsin(c_I)$.

II. System Model

Consider the uplink of a single-cell multi-user MIMO system consisting of K single-antenna users that send their signals simultaneously to a BS equipped with M antennas. Assuming a single-carrier frequency flat channel and symbol-rate sampling, the $M \times 1$ signal received at the BS from the K users is given by

$$\boldsymbol{r} = \sum_{k=1}^{K} \sqrt{p_k}\boldsymbol{g}_k s_k + \boldsymbol{n}, \tag{1}$$

where p_k represents the average transmission power from the kth user, $\boldsymbol{g}_k = \sqrt{\beta_k}\boldsymbol{h}_k$ is the channel vector between the kth user and the BS where β_k models geometric attenuation and shadow fading, and $\boldsymbol{h}_k \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_M)$ represents the fast fading and is assumed to be independent of other users' channel vectors. The symbol transmitted by the kth user is denoted by s_k where $\mathbb{E}\{|s_k|^2\} = 1$ and is drawn from a CSCG codebook independent of the other users. Finally, $\boldsymbol{n} \sim \mathcal{CN}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_M)$ denotes additive CSCG receiver noise at the BS. The assumption of symbol-rate sampling means that the matched filter at the receiver must be implemented in the analog domain. Better performance (e.g., higher rates) could be achieved by oversampling the ADCs, particularly those with one-bit resolution.

We consider a block-fading model with coherence bandwidth W_c and coherence time T_c. In this model, each channel remains constant in a coherence interval of length $T = T_c W_c$ symbols and changes independently between different intervals. Note that T is a fixed system parameter chosen as the minimum coherence duration of all users. At the beginning of each coherence interval, the users send their η-tuple mutually orthogonal pilot sequences ($K \leq \eta \leq T$) to the BS for channel estimation. Denoting the length of the training phase as η_{eff}, the remaining $T - \eta_{\text{eff}}$ symbols are dedicated to uplink data transmission.

III. Training Phase

In this section, we investigate the linear minimum mean squared error (LMMSE) channel estimator for different ADC architectures at the BS. In all scenarios, the pilot sequences are drawn from an $\eta \times K$ matrix $\boldsymbol{\Phi}$, where the kth column of $\boldsymbol{\Phi}$, $\boldsymbol{\phi}_k$, is the kth user's pilot sequence and $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{I}_K$. Therefore, the $M \times \eta$ received signal at the BS before quantization becomes

$$\boldsymbol{X} = \sum_{k=1}^{K} \sqrt{\eta p_k}\boldsymbol{g}_k \boldsymbol{\phi}_k^T + \boldsymbol{N}, \tag{2}$$

where $\boldsymbol{N}$ is an $M \times \eta$ matrix with i.i.d. $\mathcal{CN}(0, \sigma_n^2)$ elements. Since the rows of $\boldsymbol{X}$ are mutually independent due to the assumption of spatially uncorrelated Gaussian channels and noise, we can analyze them separately. As a result, we will focus on the mth row of $\boldsymbol{X}$ which is

$$\boldsymbol{x}_m^T = \sum_{k=1}^{K} \sqrt{\eta p_k} g_{mk} \boldsymbol{\phi}_k^T + \boldsymbol{n}_m^T, \tag{3}$$

where g_{mk} is the mth element of the kth user channel vector, $\boldsymbol{g}_k$, and $\boldsymbol{n}_m^T$ is the mth row of $\boldsymbol{N}$. Since the analysis is not dependent on m, hereafter we drop this subscript and denote the received signal at the mth antenna by $\boldsymbol{x}$.

A. Estimation Using One-Bit Quantized Observations

In this subsection, to have a benchmark for comparison purposes, we consider the case in which all antennas at the BS are connected to one-bit ADCs. The received signal $\boldsymbol{x}^T$ after quantization by one-bit ADCs can be written as

$$\boldsymbol{y}_t^T = \mathcal{Q}\left(\boldsymbol{x}^T\right), \tag{4}$$

where the element-wise one-bit quantization operation $\mathcal{Q}(\cdot)$ replaces each input entry with the quantized value $\frac{1}{\sqrt{2}}(\pm 1 \pm j)$, depending on the sign of the real and imaginary parts. According to the Bussgang decomposition [31], the following linear

representation of the quantization can be employed [14]:

$$\mathcal{Q}\left(x^T\right) = \sqrt{\frac{2}{\pi}} x^T \mathbf{D}_{\boldsymbol{x}}^{-\frac{1}{2}} + q_t^T, \tag{5}$$

where $\mathbf{D}_{\boldsymbol{x}} = \mathrm{diag}\{\mathbf{C}_{\boldsymbol{x}}\}$ and $\mathbf{C}_{\boldsymbol{x}}$ denotes autocorrelation matrix of x, which can be calculated as

$$\mathbf{C}_{\boldsymbol{x}} = \sum_{k=1}^{K} \eta p_k \beta_k \phi_k^* \phi_k^T + \sigma_n^2 \mathbf{I}_\eta. \tag{6}$$

In addition, q_t represents quantization noise which is uncorrelated with x and its autocorrelation matrix can be derived based on the *arcsine law* as [32]

$$\mathbf{C}_{\boldsymbol{q}_t} = \frac{2}{\pi} \arcsin\left\{\mathbf{D}_{\boldsymbol{x}}^{-\frac{1}{2}} \mathbf{C}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{x}}^{-\frac{1}{2}}\right\} - \frac{2}{\pi} \mathbf{D}_{\boldsymbol{x}}^{-\frac{1}{2}} \mathbf{C}_{\boldsymbol{x}} \mathbf{D}_{\boldsymbol{x}}^{-\frac{1}{2}}. \tag{7}$$

Much of the existing work on massive MIMO systems with low-resolution ADCs employs the simple additive quantization noise model (AQNM) for their analysis [20]–[22], [25]–[30], [39] which is valid only for low SNRs and does not capture the correlation among the elements of q_t, which turns out to be of crucial importance in our analysis. Hence, we consider the Bussgang decomposition instead and will show its effect on the system performance analysis. Stacking the rows of (5) into a matrix, the one-bit quantized observation at the BS becomes

$$\mathbf{Y} = \sqrt{\frac{2}{\pi}} \mathbf{X} \mathbf{D}_{\boldsymbol{x}}^{-\frac{1}{2}} + \mathbf{Q}, \tag{8}$$

where $\mathbf{Q}$ is an $M \times \eta$ matrix whose mth row is q_t^T. The LMMSE estimate of the channel $G = [g_1, \ldots, g_K]$ based on just one-bit quantized observations (8) is given in the following theorem.

Theorem 1: The LMMSE estimate of the k-th user channel, g_k, given the one-bit quantized observations $\mathbf{Y}$ is [14]

$$\hat{g}_k = \frac{\beta_k}{\beta_k + \sigma_{w_k}^2} \sqrt{\frac{1}{\eta p_k}} \mathbf{Y} \bar{\phi}_k^*, \tag{9}$$

where

$$\bar{\phi}_k \triangleq \sqrt{\frac{\pi}{2}} \mathbf{D}_{\boldsymbol{x}}^{\frac{1}{2}} \phi_k \tag{10}$$

$$\sigma_{w_k}^2 = \frac{1}{\eta p_k} \left(\sigma_n^2 + \bar{\phi}_k^T \mathbf{C}_{\boldsymbol{q}_t} \bar{\phi}_k^*\right). \tag{11}$$

Define the channel estimation error $\varepsilon \triangleq \hat{g}_k - g_k$. Then we have

$$\sigma_{\hat{g}_k}^2 = \frac{\beta_k^2}{\beta_k + \sigma_{w_k}^2} \text{ and } \sigma_{\varepsilon_k}^2 = \frac{\sigma_{w_k}^2 \beta_k}{\beta_k + \sigma_{w_k}^2}, \tag{12}$$

where $\sigma_{\hat{g}_k}^2$ and $\sigma_{\varepsilon_k}^2$ are the variances of the independent zero-mean elements of $\hat{g}_k$ and ε, respectively.

From Theorem 1, it is apparent that in the channel estimation analysis of massive MIMO systems with one-bit ADCs, the estimation error is directly affected not only by the inner product of the pilot sequences, but also by their outer product as well [14]. To get insight into the impact of the one-bit quantization on the channel estimation, in the next corollary we adopt the statistics-aware power control policy proposed in [37]. Apart from its practical advantages, this policy is especially suitable specially for one-bit ADCs since it avoids near-far blockage and hence strong interference. Moreover, this power control approach also leads to simple expressions and provides analytical convenience for our derivation in Section VI. Although not the focus of this paper, we note that in general a massive MIMO system employing a mixed-ADC architecture will be more resilient than an all one-bit implementation to the near-far effect and jamming. This is an interesting topic for further study.

Corollary 1: For the case in which power control is performed, i.e., $p_k = \frac{p}{\beta_k}$ for some fixed value p and for $k \in \mathcal{K} = \{1, \ldots, K\}$, the number of users is equal to the length of pilot sequences, i.e., $\eta = K$, and the pilot matrix satisfies $\Phi\Phi^H = \mathbf{I}_K$, we have

$$\mathbf{C}_{\boldsymbol{x}} = \left(Kp + \sigma_n^2\right) \mathbf{I}_K = \mathbf{D}_{\boldsymbol{x}} \tag{13}$$

$$\mathbf{C}_{\boldsymbol{q}_t} = \left(1 - \frac{2}{\pi}\right) \mathbf{I}_K, \tag{14}$$

which yields

$$\sigma_{\hat{g}_k}^2 = \frac{2}{\pi} \frac{\beta_k}{1 + \frac{\sigma_n^2}{Kp}} \tag{15}$$

$$\sigma_{\varepsilon_k}^2 = \frac{\left(\frac{Kp}{\sigma_n^2}\left(1 - \frac{2}{\pi}\right) + 1\right) \beta_k}{1 + \frac{Kp}{\sigma_n^2}}. \tag{16}$$

Corollary 1 states conditions under which $\mathbf{C}_{\boldsymbol{q}_t}$ is diagonal. In addition, it is evident that the channel estimation suffers from an error floor at high SNRs.

B. Channel Estimation With Few Full Resolution ADCs

Channel estimation with coarse observations suffers from large errors especially in the high SNR regime. On the other hand, while estimating all channels using high-resolution ADCs is desirable, the resulting power consumption burden makes this approach practically infeasible. This motivates the use of a mixed-ADC architecture for channel estimation to eliminate the large estimation error caused by one-bit quantization while keeping the power consumption penalty at an acceptable level. In the approach described in [23], [24], [26], $N \ll M$ pairs of high-resolution ADCs are deployed and switched between different antennas during different transmission intervals in an approach referred to as "round-robin" training. In this approach, the M BS antennas are grouped into M/N sets.[1] In the first training sub-interval, users send their mutually orthogonal pilots to the BS while the N high-resolution ADC pairs are connected to the first set of N antennas. After receiving the pilot symbols from all users in the η-symbol-length training sub-interval, the high-resolution ADCs are switched to the next set of antennas and so on. In this manner, after $(M/N)\eta$ pilot transmissions (M/N sub-intervals), we can estimate each channel based on observations with only high-resolution ADCs. This round-robin channel estimation protocol is illustrated in Fig. 2 for a mixed-ADC system with $M/N = 5$.

Stacking all $N \times \eta$ full-resolution observations into an $M \times \eta$ matrix, $\mathbf{X}$, the LMMSE estimate of the k-th user channel, g_k,

[1] We assume M/N is an integer throughout the paper.

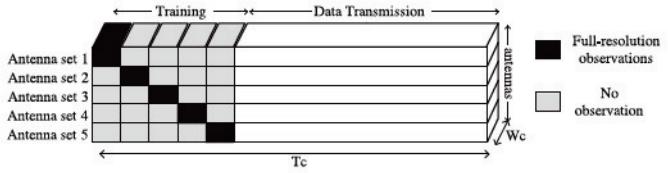

Fig. 2. Transmission protocol for estimation using full-resolution observations.

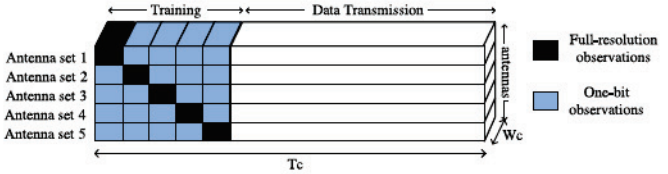

Fig. 3. Transmission protocol for estimation using full-resolution/one-bit observations.

is [5]

$$\hat{g}_k = \frac{1}{1+\frac{\sigma_n^2}{\eta p_k \beta_k}} \frac{1}{\sqrt{\eta p_k}} \mathbf{X}\phi_k^*, \tag{17}$$

and the resulting variances of the channel estimate and the error are given respectively by

$$\sigma_{\hat{g}_k}^2 = \frac{\beta_k}{1+\frac{\sigma_n^2}{\eta p_k \beta_k}} \text{ and } \sigma_{\varepsilon_k}^2 = \frac{\beta_k}{1+\frac{\eta p_k \beta_k}{\sigma_n^2}}. \tag{18}$$

Equation (18) states that by employing only N pairs of high-resolution ADCs and by expending a larger portion of the coherence interval for channel estimation, the channel can be estimated with the same precision as that achieved by conventional high-resolution ADC massive MIMO systems. However, this comes at the high cost of repeating the training data M/N times, which can significantly reduce the time available for data transmission. Indeed, we will see later that in some cases, a mixed-ADC implementation with round-robin training achieves a lower SE than a system with all one-bit ADCs because of the long training interval (even with the improvements we propose below for the round-robin method). However, we will also see that there are other situations for which the mixed-ADC round-robin method provides a large gain in SE. The primary goal of this paper is to elucidate under what conditions these and other competing approaches provide the best performance.

Before analyzing the tradeoff between the gain (lower channel estimation error) and cost (longer training period) of the round-robin approach, in the next subsection we propose channel estimation based on the use of both full-resolution and one-bit data received by the BS in order to further improve the performance of the mixed-ADC architecture with round-robin channel estimation. To our knowledge, this approach has not been considered in prior work on mixed-ADC massive MIMO.

C. Estimation Using Joint Full-Resolution/One-Bit Observations

While channel estimation performance based on coarsely quantized observations suffers from large errors in the high SNR regime, it provides reasonable performance for low SNRs. Hence, in this subsection we consider joint channel estimation based on observations from both high-resolution and one-bit ADCs to further improve the channel estimation accuracy. Unlike the previous subsection in which the one-bit ADCs were not employed, here we incorporate their coarse observations into the channel estimation procedure. The protocol for this method is illustrated in Fig. 3 for a mixed-ADC system with $M/N = 5$. It can be seen that, in addition to one set of full-resolution observations for each antenna, there are $(M/N) - 1$ sets of one-bit observations which are also taken into account for channel estimation. The next theorem characterizes the performance of this approach.

Theorem 2: Stacking all $N \times \eta$ full-resolution observations into an $M \times \eta$ matrix, $\mathbf{X}$, and all $(M/N) - 1$ $N \times \eta$ one-bit quantized observations into $M \times \eta$ matrices, $\mathbf{Y}_t$, $t \in \mathcal{T} = \{1, ..., M/N - 1\}$, the LMMSE estimate of the k-th user channel, g_k, is

$$\hat{g}_k = \sqrt{\frac{1}{\eta p_k}} \left(w_{\infty_k} \mathbf{X}\phi_k^* + w_{1_k} \sum_{t=1}^{\frac{M}{N}-1} \mathbf{Y}_t \bar{\phi}_k^* \right), \tag{19}$$

where

$$w_{\infty_k} = \frac{\frac{\eta p_k}{\sigma_n^2}}{\frac{1}{\beta_k} + \frac{\eta p_k}{\sigma_n^2} + \varsigma_k(p_k)} \tag{20}$$

$$w_{1_k} = \frac{\left(\frac{M}{N} - 1\right)^{-1} \varsigma_k(p_k)}{\frac{1}{\beta_k} + \frac{\eta p_k}{\sigma_n^2} + \varsigma_k(p_k)} \tag{21}$$

$$\varsigma_k(p_k) = \frac{\left(\frac{M}{N} - 1\right)}{\sigma_{w_k}^2 + \left(\frac{M}{N} - 2\right) \varrho_k} \tag{22}$$

$$\sigma_{w_k}^2 = \frac{1}{\eta p_k} \left(\sigma_n^2 + \bar{\phi}_k^T \mathbf{C}_{\boldsymbol{q}_t} \bar{\phi}_k^* \right) \tag{23}$$

$$\varrho_k = \frac{1}{\eta p_k} \bar{\phi}_k^T \bar{\mathbf{C}}_{\boldsymbol{q}_t} \bar{\phi}_k^* \tag{24}$$

$$\bar{\mathbf{C}}_{\boldsymbol{q}_t} = \frac{2}{\pi} \arcsin\left\{ \bar{\mathbf{D}}_{\boldsymbol{x}}^{-\frac{1}{2}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{D}}_{\boldsymbol{x}}^{-\frac{1}{2}} \right\} - \frac{2}{\pi} \bar{\mathbf{D}}_{\boldsymbol{x}}^{-\frac{1}{2}} \bar{\mathbf{C}}_{\boldsymbol{x}} \bar{\mathbf{D}}_{\boldsymbol{x}}^{-\frac{1}{2}} \tag{25}$$

$$\bar{\mathbf{C}}_{\boldsymbol{x}} = \sum_{k=1}^{K} \eta p_k \beta_k \phi_k^* \phi_k^T \tag{26}$$

$$\bar{\mathbf{D}}_{\boldsymbol{x}} = \text{diag}\{\bar{\mathbf{C}}_{\boldsymbol{x}}\}. \tag{27}$$

This approach yields the following variances for the channel estimate and the estimation error, respectively:

$$\sigma_{\hat{g}_k}^2 = \frac{\frac{\eta p_k}{\sigma_n^2} + \varsigma_k(p_k)}{\frac{1}{\beta_k} + \frac{\eta p_k}{\sigma_n^2} + \varsigma_k(p_k)} \beta_k \tag{28}$$

$$\sigma_{\varepsilon_k}^2 = \frac{1}{\frac{1}{\beta_k} + \frac{\eta p_k}{\sigma_n^2} + \varsigma_k(p_k)}. \tag{29}$$

Proof: See Appendix A. ■

Theorem 2 demonstrates the optimal approach for combining the observations from high-resolution and one-bit ADCs. In addition, this highlights the importance of considering the correlation among the one-bit observations in the analysis of mixed-

ADC channel estimation, something that could not be addressed by the widely-used AQNM approach. More precisely, it can be seen that the impact of joint high-resolution/one-bit channel estimation is manifested in the variance of the channel estimation error by the term $\varsigma_k(p_k)$. To see this, assume that the correlation among one-bit observations in different training sub-intervals is ignored (as would be the case with the AQNM approach). As shown in the appendix, this is equivalent to setting $\varrho_k = 0$ in (24). Under this assumption, $\varsigma_k(p_k)$ becomes

$$\varsigma_{k_0}(p_k) = \frac{\left(\frac{M}{N} - 1\right)}{\sigma_{w_k}^2} > \varsigma_k(p_k), \tag{30}$$

and thus, $\sigma_{\varepsilon_k}^2 > \sigma_{\varepsilon_{k_0}}^2$ where $\sigma_{\varepsilon_{k_0}}^2$ denotes the estimation error for $\varrho_k = 0$. Consequently, the AQNM model yields an overly optimistic assessment of the channel estimation error compared with the more accurate Bussgang analysis. We will see below that the impact of the AQNM approximation is significant for mixed-ADC channel estimation.

The next corollary provides insight into the impact of the system parameters on the joint high-resolution/one-bit LMMSE estimation.

Corollary 2: For the case in which power control is performed, i.e., $p_k = \frac{p}{\beta_k}$ for $k \in \mathcal{K} = \{1, \ldots, K\}$, the number of users is equal to the length of pilot sequences, i.e., $\eta = K$, and the pilot matrix satisfies $\mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I}_K$, we have

$$\bar{\mathbf{C}}_{\boldsymbol{x}} = Kp\mathbf{I}_K = \bar{\mathbf{D}}_{\boldsymbol{x}}, \tag{31}$$

and

$$\bar{\mathbf{C}}_{\boldsymbol{q}_t} = \left(1 - \frac{2}{\pi}\right)\mathbf{I}_K, \tag{32}$$

which yields

$$\sigma_{\hat{g}_k}^2 = \frac{\frac{Kp}{\sigma_n^2} + \varsigma(p)}{1 + \frac{Kp}{\sigma_n^2} + \varsigma(p)}\beta_k \text{ and } \sigma_{\varepsilon_k}^2 = \frac{1}{1 + \frac{Kp}{\sigma_n^2} + \varsigma(p)}\beta_k, \tag{33}$$

where

$$\varsigma(p) = \frac{\left(\frac{M}{N} - 1\right)}{\frac{\pi}{2}\frac{\sigma_n^2}{Kp} + \left(\frac{\pi}{2} - 1\right)\left(\frac{M}{N} - 1\right)}. \tag{34}$$

In addition,

$$w_\infty = \frac{\frac{Kp}{\sigma_n^2}}{1 + \frac{Kp}{\sigma_n^2} + \varsigma(p)} \text{ and } w_1 = \frac{\left(\frac{M}{N} - 1\right)^{-1}\varsigma(p)}{1 + \frac{Kp}{\sigma_n^2} + \varsigma(p)}, \tag{35}$$

where w_∞ and w_1 denote the weights of the high-resolution and one-bit observations in the LMMSE estimation, respectively.

Corallary 2 states that in contrast to Theorem 1 where the correlation among one-bit observations within each training sub-interval can be eliminated by carefully selecting the system parameters as in Corollary 1, we cannot overcome the correlation among one-bit observations from different training sub-intervals. This phenomenon makes the addition of the one-bit observations less useful especially in the high SNR regime. For instance, in the asymptotic case, as the SNR $= \frac{p}{\sigma_n^2}$ goes to infinity, we have

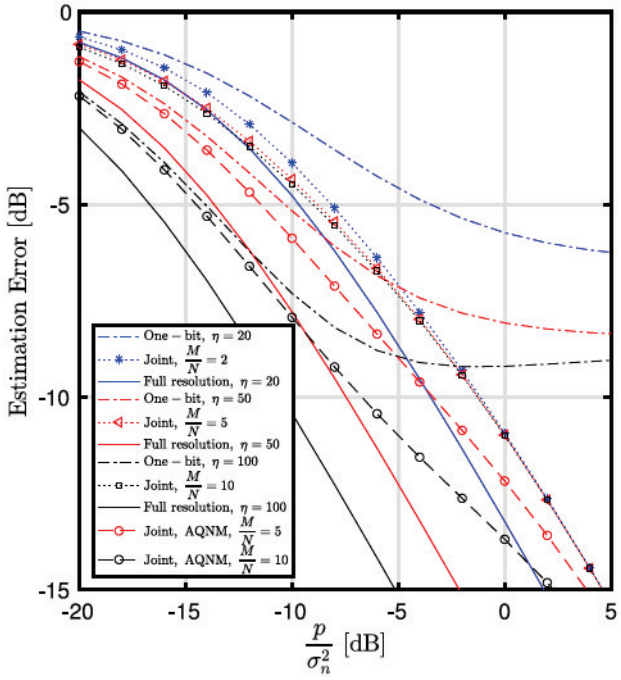

Fig. 4. Channel estimation error $\sigma_{\varepsilon_k}^2/\beta_k$ versus p/σ_n^2.

$$\varsigma \longrightarrow \frac{1}{\frac{\pi}{2} - 1}, \tag{36}$$

$$w_\infty \longrightarrow 1,\ w_1 \longrightarrow 0. \tag{37}$$

It is apparent from (36) that in the asymptotic regime ς tends to a finite value and also is independent of M/N. Moreover, (37) implies that the optimal approach for high SNRs is to estimate the channel based solely on the high-resolution observations.

The error for the three channel estimation approaches in Eqs. (12), (18), and (29) is depicted in Fig. 4 for a case with $M = 100$ antennas, $K = 10$ users, and various numbers of high-resolution ADCs, N and training lengths η. The label "Joint" refers to round-robin channel estimation that includes the one-bit observations as described in the previous section, "Full resolution" indicates the performance achieved using a full array of high-resolution ADCs, and "One-bit" refers to the performance of an all-one-bit architecture. We also plot the performance predicted for the Joint approach based on the AQNM analysis, which ignores the correlation among the one-bit observations. We see that the AQNM-based analysis yields an overly optimistic prediction for the channel estimation error. In particular, unlike AQNM, the more accurate Bussgang analysis shows that channel estimation with all an one-bit BS actually outperforms the Joint method for low SNRs, a critical observation in analyzing whether or not a mixed-ADC implementation makes sense. However, we see that the mixed-ADC architecture eventually overcomes the error floor of the all one-bit system for high SNRs and in such cases can reduce the estimation error dramatically. Fig. 4 focuses on channel estimation performance, but does not reflect the full impact of the round-robin training on the overall system spectral efficiency, since reducing N increases

the amount of training required by the round-robin method. This will be taken into account when we analyze the SE in Section V.

IV. Practical Considerations

The improvement in channel estimation performance provided by the round-robin training clearly comes at the expense of a significantly increased training overhead. For example, consider a simple worst-case example with a 400 Hz Doppler spread in a narrowband channel of 400 kHz bandwidth; in this case, the coherence time is roughly 1000 symbols. For higher bandwidths or smaller cells with lower mobility, the coherence time can easily approach 10,000 symbols or more. A mixed-ADC array of 128 antennas with 16 high-resolution ADCs would require repeating the pilots 8 times, which for 20 users would amount to 160 symbols, or 16% of the coherence time when $T = 1000$ symbols. This is a relatively high price to pay, and as we will see later, in many instances the resulting loss in SE cannot be offset by the improved channel estimate. However, we will also see that on the other hand, there are other situations where the opposite is true, where the round-robin method leads to significant gains in SE even taking the training overhead into account.

Besides the extra training overhead, the round-robin method has the disadvantage of requiring extra RF switching or multiplexing hardware prior to the ADCs, as shown in Fig. 1. It is unlikely that a single large $M \times M$ multiplexer would be used for this purpose, since complete flexibility in assigning a given high-resolution ADC to any possible antenna is not needed. A more likely architecture would employ a bank of smaller multiplexers that allows one high-resolution ADC to be switched among a smaller subarray of antennas, ensuring that each RF chain has access to high-resolution training data during one of the round-robin intervals. Such an approach is similar to the simplified "subarray switching" schemes proposed for antenna selection in massive MIMO [33]–[35]. In an interesting earlier example, a large 108×108 multiplexer chipset for a local area network application was developed in [36], composed of several 36×36 differential crosspoint ASIC switches that consume less than 100 mW each, with a bandwidth of 140 MHz and a 0 dB insertion loss.

In the 20 years since [36], RF switch technology has advanced considerably. For the example discussed above involving a 128-element array with 16 high-resolution ADCs and 112 one-bit ADCs, the multiplexing could be achieved using 16 8×8 analog switches arranged in parallel. Consider the Analog Devices ADV3228 8×8 crosspoint switch as an example of an off-the-shelf component for such an architecture.[2] The ADV3228 has a 750 MHz bandwidth, a switching time of 15 ns, and a power consumption of 500 mW, which is similar to that of an 8-bit ADC (for example, see Texas Instruments' ADC08B200 8-bit 200 MS/s ADC[3]). Since the switches can be implemented at a lower intermediate frequency prior to the I-Q demodulation, only one per subarray is required, and thus the total power consumption of the switches would be less than half that of the ADCs.

Note that for the vast majority of the coherence time, the switch is idle. To accommodate the round-robin training, the switches only need to be operated $\frac{M}{N} - 1$ times, once for every repetition of the training data. This reduces the actual power consumption to below the specification, and further reduces the impact of the additional training. Short guard intervals would need to be inserted between the training intervals to account for the switching transients, but these will typically not impact the SE. For the example discussed above with 128 antennas and 8 switches, 7 switching events are required for a total switching time of 105 ns, which is insignificant compared to the coherence time of 2.5 ms at a 400 Hz Doppler.

The insertion loss of the analog switches would also have to be taken into account in an actual implementation, since this will directly reduce the overall SNR of the received signals. Harmonic interference due to nonlinearities in the switch are likely not an issue; for example, the specifications for a Texas Instruments switch (LMH6583) similar to the ADV3228 indicate that the power of the second and third harmonic distortions were -76 dBc. Furthermore, it has been shown that the use of signal combining with a massive antenna array provides significant robustness to such nonlinearities and other hardware imperfections [7]–[12].

V. Spectral Efficiency

Although channel estimation with a mixed-ADC architecture using round-robin training can substantially improve the channel estimation accuracy, it requires a longer training interval and, therefore, leaves less room for data transmission in each coherence interval. More precisely, $(M/N)\eta$ symbol transmissions are required for round-robin channel estimation which could be large when the number of high-resolution ADCs, N, is small.[4] Despite losing a portion of the coherence interval for channel estimation due to the mixed-ADC architecture, the improvement in the signal-to-quantization-interference-and-noise ratio (SQINR) can be significant owing to more accurate channel estimation, and thus a higher rate would be expected during this shorter data transmission period. In this section, we study this system performance trade-off in terms of spectral efficiency for the three mentioned channel estimation approaches.

In the data transmission phase, all users simultaneously send their data symbols to the BS. To begin, assume the antennas are ordered so that the last N antennas are connected to high-resolution ADCs in this phase. A more thoughtful assignment of the high-resolution ADCs will be considered below. From equation (1), and based on the Bussgang decomposition, the

[2]See http://www.analog.com/en/products/switches-multiplexers/buffered-analog-crosspoint-switches/adv3228.html#product-overview for product details.

[3]http://www.ti.com/product/ADC08B200/technicaldocuments

[4]Note that in designing a mixed-ADC system with round-robin channel training, one should consider the ratio M/N in scaling the system instead of just increasing the number of antennas M. In particular, increasing the number of BS antennas requires increasing of the high-resolution ADCs, N, as well.

received signal at the BS after one-bit quantization is

$$\boldsymbol{y}_d = \begin{bmatrix} \sqrt{\frac{2}{\pi}}\bar{\mathbf{D}}^{-\frac{1}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix} \boldsymbol{r} + \underbrace{\begin{bmatrix} \bar{\boldsymbol{q}}_d \\ \mathbf{0} \end{bmatrix}}_{\boldsymbol{q}_d} \tag{38}$$

$$\bar{\mathbf{D}} = \operatorname{diag}\{\mathbf{C}_{\boldsymbol{r}}\} \tag{39}$$

$$\mathbf{C}_{\boldsymbol{r}} = \sum_{k=1}^{K} p_k \bar{\boldsymbol{g}}_k \bar{\boldsymbol{g}}_k^H + \sigma_n^2 \mathbf{I}_{M-N}, \tag{40}$$

where $\bar{\boldsymbol{g}}_k$ denotes the $M-N$ elements of $\boldsymbol{g}_k$ corresponding to the $M-N$ one-bit ADCs and $\bar{\boldsymbol{q}}_d$ is the $(M-N)\times 1$ quantization noise in the data transmission phase. It is apparent that the covariance matrix in (40) is not diagonal which makes analytical tractability difficult. However, by adopting statistics-aware power control [37], i.e., $p_k = \frac{p}{\beta_k}$, and assuming that the number of users is relatively large (typical for massive MIMO systems), channel hardening occurs [14], and (40) can be approximated as

$$\mathbf{C}_{\boldsymbol{r}} \cong \left(Kp + \sigma_n^2\right) \mathbf{I}_{M-N} = \bar{\mathbf{D}}. \tag{41}$$

As a result, according to the arcsine law (see (7)), the covariance matrix of the quantization noise in the data transmission phase becomes $\mathbf{C}_{\bar{\boldsymbol{q}}_d} \cong (1-2/\pi)\mathbf{I}_{M-N}$ and (38) simplifies to

$$\boldsymbol{y}_d \cong \mathbf{A}\left(\sum_{k=1}^{K} \sqrt{p}\boldsymbol{h}_k s_k + \boldsymbol{n}\right) + \boldsymbol{q}_d \tag{42}$$

$$\mathbf{A} = \begin{bmatrix} \alpha \mathbf{I}_{M-N} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_N \end{bmatrix},$$

where $\alpha \triangleq \sqrt{\frac{2}{\pi}\frac{1}{(Kp+\sigma_n^2)}}$.

For data detection, the BS selects a linear receiver $\boldsymbol{W} \in \mathbb{C}^{M\times K}$ as a function of the channel estimate. Note that the quantization model considered in (4) and (5) does not preserve the power of the input of the quantizer since the power of the output is forced to be 1. Thus we premultiply the received signal as follows to offset this effect:

$$\hat{\boldsymbol{y}}_d = \mathbf{A}^{-1}\boldsymbol{y}_d. \tag{43}$$

By employing the linear detector $\mathbf{W}$, the resulting signal at the BS is

$$\hat{\boldsymbol{s}} = \boldsymbol{W}^H \hat{\boldsymbol{y}}_d. \tag{44}$$

Thus, the kth element of $\hat{\boldsymbol{s}}$ is

$$\hat{s}_k = \sqrt{p}\boldsymbol{w}_k^H \boldsymbol{h}_k s_k + \sqrt{p}\sum_{i=1, i\neq k}^{K} \boldsymbol{w}_k^H \boldsymbol{h}_i s_i + \boldsymbol{w}_k^H \boldsymbol{n} + \boldsymbol{w}_k^H \mathbf{A}^{-1}\boldsymbol{q}_d, \tag{45}$$

where $\boldsymbol{w}_k$ is the kth column of $\boldsymbol{W}$. We assume the BS treats $\boldsymbol{w}_k^H \boldsymbol{h}_k$ as the gain of the desired signal and the other terms of (45) as Gaussian noise when decoding the signal.[5] Consequently, we can use the classical bounding technique of [37] to derive an approximation for the ergodic achievable SE at the kth user as

$$\mathcal{S}_k = \mathcal{R}\left(\mathrm{SQINR}_k\right), \tag{46}$$

where the effective SQINR_k is defined by (47) shown at the bottom of this page, and $\mathcal{R}(\theta) \triangleq (1-\eta_{\text{eff}}/T)\log_2(1+\theta)$ where η_{eff} represents the training duration which is η and $(M/N)\eta$ for the pure one-bit and mixed-ADC architectures, respectively.

A. MRC Detection

1) Random Mixed-ADC Detection: In this subsection, we consider the case in which the high-resolution ADCs are connected to an arbitrary set of N antennas. Denoting the estimate of the channel by $\hat{\mathbf{H}} = [\hat{\boldsymbol{h}}_1, \ldots, \hat{\boldsymbol{h}}_K]$, setting $\mathbf{W} = \hat{\mathbf{H}}$, and following the same reasoning as in [14], the SE of the mixed-ADC architecture with MRC detection can be derived as

$$\mathcal{S}_k^{\mathrm{MRC}} = \mathcal{R}\left(\frac{pM\sigma_{\hat{h}}^2}{pK + \sigma_n^2 + \frac{\left(1-\frac{2}{\pi}\right)}{\alpha^2}\left(1-\frac{N}{M}\right)}\right), \tag{48}$$

where the channel estimate variance $\sigma_{\hat{h}}^2 = \sigma_{\hat{g}_k}^2/\beta_k$ depends on the estimation approach as denoted in (12), (18), and (28).

From (48), it can be observed that the gain of exploiting the mixed-ADC architecture is manifested in the SE expressions by two factors, channel estimation improvement by a factor of $\sigma_{\hat{h}}^2$, and quantization noise reduction by a factor of $1-N/M$.

2) Mixed-ADC Detection With Antenna Selection: Having an accurate channel estimate can help us to employ the N high-resolution ADCs in an intelligent manner to further improve the performance of the mixed-ADC architecture. A careful look at the SQINR expression in (47) reveals that the effect of one-bit quantization on the SE is manifested by the last term of the denominator. Hence, one can maximize the SE by minimizing this term through smart use of the N high-resolution ADCs. We refer to this approach as *Mixed-ADC with Antenna Selection*. We consider an antenna selection scheme suggested by the SQINR expression in (47). In this approach, the N high-resolution ADCs are connected to the antennas corresponding to rows of $\hat{\mathbf{H}}$ with the largest energy, i.e. $\sum_{k=1}^{K}|\hat{h}_{mk}|^2$. Besides numerical evaluation in Section VI, in Theorem 3 we derive a bound for the SE achieved by MRC detection with antenna selection.

[5]Note that in general, the quantization noise is not Gaussian. However, to derive a lower bound for the SE, we assume it is Gaussian with covariance $\mathbf{C}_{\boldsymbol{q}_d}$.

$$\mathrm{SQINR}_k = \frac{p\left|\mathbb{E}\left\{\boldsymbol{w}_k^H \boldsymbol{h}_k\right\}\right|^2}{p\sum_{i=1}^{K}\mathbb{E}\left\{\left|\boldsymbol{w}_k^H \boldsymbol{h}_i\right|^2\right\} - p\left|\mathbb{E}\left\{\boldsymbol{w}_k^H \boldsymbol{h}_k\right\}\right|^2 + \sigma_n^2\mathbb{E}\left\{\|\boldsymbol{w}_k\|^2\right\} + \alpha^{-2}\mathbb{E}\left\{\boldsymbol{w}_k^H \mathbf{C}_{\boldsymbol{q}_d}\boldsymbol{w}_k\right\}} \tag{47}$$

Theorem 3: The spectral efficiency of the mixed-ADC system with antenna selection and an MRC receiver is lower bounded by

$$\bar{\mathcal{S}}_k^{\mathrm{MRC}} = \mathcal{R}\left(\frac{pM\sigma_{\hat{h}}^2}{pK + \sigma_n^2 + \frac{\left(1-\frac{2}{\pi}\right)}{MK\alpha^2}\left(\sum_{m=1}^{M-N}\chi_m\right)}\right), \quad (49)$$

where χ_m is defined at the top of the next page, and $\mathcal{F}_A$ denotes the Lauricella function of type A [45], (50) shown at the bottom of this page.

Proof: See Appendix B. ∎

The lower bound (49) explicitly reflects the benefit of antenna selection in the data transmission phase. By comparing (49) with (48), it is evident that antenna selection has improved the SE by replacing $1 - N/M$ by $\frac{\sum_{m=1}^{M-N}\chi_m}{MK}$. In Section VI we illustrate how antenna selection improves SE for different SNRs. Note that Theorem 3 assumes the ability to make an arbitrary assignment of the high-resolution ADCs to different RF chains, which may not be possible if the ADC multiplexing is implemented by a bank of subarray switches. In the numerical results presented later, we show that this does not lead to a significant degradation in performance.

B. ZF Detection

In this section, we study the SE of the mixed-ADC architecture with ZF detection. To design a mixed-ADC adapted ZF detector, we re-write the last two terms of the denominator of (47) as follows:

$$\boldsymbol{w}_k^H\left(\sigma_n^2\mathbf{I}_M + \alpha^{-2}\mathbf{C}_{\boldsymbol{q}_d}\right)\boldsymbol{w}_k = \left[\mathbf{W}^H\mathbf{C}_{n_{\mathrm{eff}}}\mathbf{W}\right]_{kk}, \quad (51)$$

where $\mathbf{C}_{n_{\mathrm{eff}}} = \sigma_n^2\mathbf{I}_M + \alpha^{-2}\mathbf{C}_{\boldsymbol{q}_d}$. Accordingly, the ZF detector for the mixed-ADC architecture can be written as

$$\mathbf{W} = \mathbf{C}_{n_{\mathrm{eff}}}^{-1}\hat{\mathbf{H}}\left(\hat{\mathbf{H}}^H\mathbf{C}_{n_{\mathrm{eff}}}^{-1}\hat{\mathbf{H}}\right)^{-1}. \quad (52)$$

Plugging (52) into (47) yields (53) shown at the bottom of this page. Similar to the MRC case, the SQINR in (53) suggests the same antenna selection approach for ZF detection. In general, calculating the expected values in (53) is not tractable neither for arbitrary-antenna mixed-ADC detection nor mixed-ADC with antenna selection. Hence, we numerically evaluate the performance of mixed-ADC with ZF detection in the next section.

C. Massive MIMO With Uniform ADC Resolution

Contrary to the mixed-ADC architecture where the ADC comparators are concentrated in a few antennas, uniformly spreading the comparators over the array is an alternative approach [19]–[21], [41], [44]. In this subsection, we provide the SE expressions for such systems. These expressions will be used in the next section for performance comparisons with the mixed-ADC architecture.

The SE for the case of all one-bit ADCs was derived in [14] using the Bussgang decomposition. For ADC resolutions of 2 bits or higher, the AQNM model is sufficiently accurate. Using AQNM and following the same reasoning as in [21], [41], [44], the SE of a massive MIMO system with uniform resolution ADCs can be derived as

$$\tilde{\mathcal{S}}_k^{\mathrm{MRC}} = \mathcal{R}\left(\frac{pM\tilde{\sigma}_{\hat{h}}^2}{pK + \sigma_n^2 + \frac{(1-\alpha_0)}{\alpha_0^2}\left(p\left(\tilde{\sigma}_{\hat{h}}^2 + K\right) + \sigma_n^2\right)}\right) \quad (54)$$

$$\tilde{\mathcal{S}}_k^{\mathrm{ZF}} = \mathcal{R}\left(\frac{p\left(M-K\right)\tilde{\sigma}_{\hat{h}}^2}{pK\left(1-\tilde{\sigma}_{\hat{h}}^2\right) + \sigma_n^2 + \frac{(M-K)\tilde{\sigma}_{\hat{h}}^2}{\alpha^2}\mathbb{E}\left\{\boldsymbol{w}_k^H\mathbf{C}_0\boldsymbol{w}_k\right\}}\right), \quad (55)$$

for MRC and ZF detection, respectively. In (54) and (55),

$$\tilde{\sigma}_{\hat{h}}^2 = \frac{\alpha_0^2\eta p}{\alpha_0^2\eta p + \alpha_0^2\sigma_n^2 + \alpha_0\left(1-\alpha_0\right)\left(pK+\sigma_n^2\right)}, \quad (56)$$

α_0 is a scalar depending on the ADC resolution and can be found in Table I of [21], $\boldsymbol{w}_k$ is the kth column of $\mathbf{W} = \hat{\mathbf{H}}\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}$, and $\mathbf{C}_0$ denotes the covariance matrix of the quantization noise based on the AQNM model [21]. The detailed calculation of $\mathbb{E}\left\{\boldsymbol{w}_k^H\mathbf{C}_0\boldsymbol{w}_k\right\}$ in (55) is provided in [44] which we do not include here for the sake of brevity.

VI. Numerical Results

By substituting from (12), (18), and (28) into (48), (49), and (53), we can evaluate the performance of mixed-ADC architectures for different system settings. For all of the following experiments, we consider a system with $M = 100$ antennas at the BS, and $K = 10$ users. Also, we assume the power control approach of [37] is used, so that $p_k\beta_k = p$ for all k. We also assume that

$$\chi_m = \frac{M!}{(m-1)!\,(M-m)!}\sum_{\ell=0}^{M-m}(-1)^{-\ell}\binom{M-m}{\ell}\left(\Gamma(K)\right)^{-m-\ell}K^{1-m-\ell}\Gamma\left(1+K\left(m+\ell\right)\right)$$
$$\times\,\mathcal{F}_A^{(m+\ell-1)}\left(1+K\left(m+\ell\right);K,\ldots,K;K+1,\ldots,K+1;-1,\ldots,-1\right) \quad (50)$$

$$\mathrm{SQINR}_k^{\mathrm{ZF}} = \frac{p}{pK\left(1-\sigma_{\hat{h}}^2\right)\mathbb{E}\left\{\left[\left(\hat{\mathbf{H}}^H\mathbf{C}_{n_{\mathrm{eff}}}^{-1}\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^H\mathbf{C}_{n_{\mathrm{eff}}}^{-2}\hat{\mathbf{H}}\left(\hat{\mathbf{H}}^H\mathbf{C}_{n_{\mathrm{eff}}}^{-1}\hat{\mathbf{H}}\right)^{-1}\right]_{kk}\right\} + \mathbb{E}\left\{\left[\left(\hat{\mathbf{H}}^H\mathbf{C}_{n_{\mathrm{eff}}}^{-1}\hat{\mathbf{H}}\right)^{-1}\right]_{kk}\right\}} \quad (53)$$

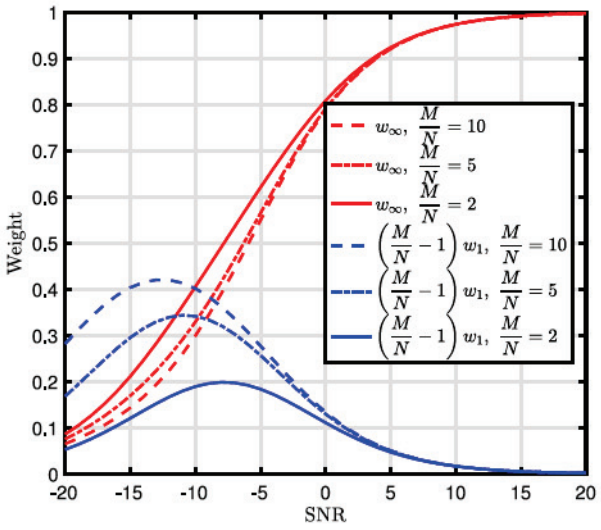

Fig. 5. Weights used in the LMMSE channel estimator for high-resolution and one-bit observations.

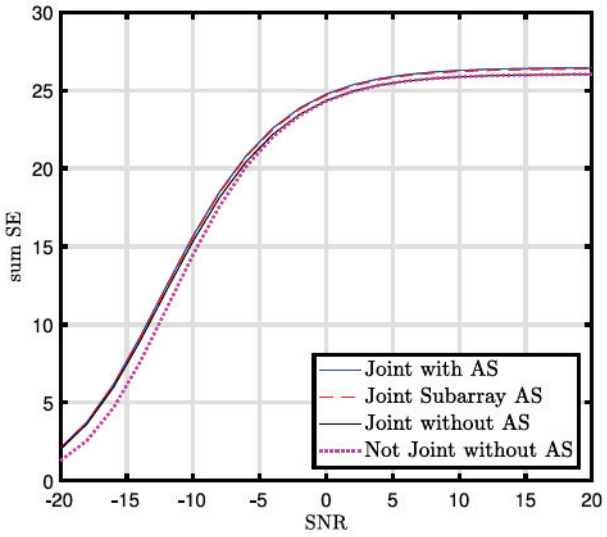

Fig. 6. Sum SE for MRC detection versus SNR for $M = 100$, $N = 20$, and $T = 400$.

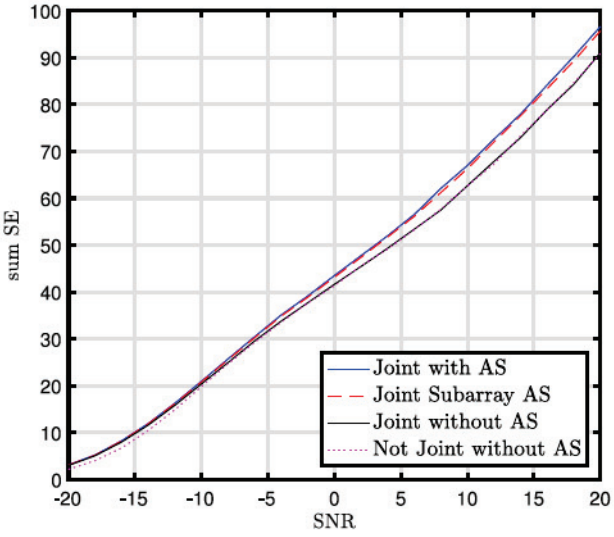

Fig. 7. Sum SE for ZF detection versus SNR for $M = 100$, $N = 20$, and $T = 400$.

an optimal resource allocation has been performed [41], [42] such that the training length, η_{eff}, transmission power during the training phase, p_t, and data transmission phase, p_d are optimized under a power constraint $\eta_{\text{eff}} p_t + (T - \eta_{\text{eff}}) p_d = P_{\text{ave}} T$. In the following figures, the SNR is defined as $\text{SNR} \triangleq P_{\text{ave}}/\sigma_n^2$.

Figure 5 illustrates the optimal weights for combining high-resolution and one-bit observations for the joint high-resolution/one-bit LMMSE channel estimation. Interestingly, it can be seen that when M/N is large, the one-bit observations are emphasized in the low SNR regime relative to the high-resolution observations. In addition, in contrast to the weights for the high-resolution observations, which rise monotonically with increasing SNR, the weight for the one-bit observations grows at first and then decreases to zero.

To study the performance improvement due to joint channel estimation and antenna selection in mixed-ADC massive MIMO, the sum SE for the MRC and ZF detectors for a system with coherence interval $T = 400$ symbols and $N = 20$ high-resolution ADCs is depicted in Fig. 6 and Fig. 7, respectively. In these and subsequent figures, "Joint with AS" indicates that the channel estimation was performed with both one-bit and high-resolution ADCs and that antenna selection (AS) was used for data detection, "Joint without AS" represents the same case without antenna selection, "Joint Subarray AS" means that the antenna selection only occurred within each M/N-element subarray (one high-resolution ADC assigned to the strongest channel within each subarray), and "Not Joint without AS" represents the case in which channel is estimated based on only high-resolution observations and no antenna selection is employed. For both MRC and ZF, it can be seen that antenna selection slightly improves the SE for high SNRs, where the channel estimation is most accurate. At low SNR, we see that joint channel estimation provides a gain from the use of one-bit ADCs, which provide useful information at these SNRs. We also see that the constrained AS required when the switching is only performed within subarrays provides nearly identical performance to the case where arbitrary AS is allowed.

Note that the main reason for the small gain for antenna selection is due to the fact that, with multiple users, selecting a given antenna does not benefit all users simultaneously, and the strong users responsible for a given antenna being selected will in general be different for different antennas. Thus, the improvement due to increased signal-to-noise ratio for some users is somewhat offset by the fact that other users may experience a lower SNR on those same antennas. We would see a much larger benefit for antenna selection if only a single user were present.

Figures 8 and 9 provide a comparison among a mixed-ADC massive MIMO system with joint channel estimation and antenna selection, an all-one-bit architecture ("One-bit"), and a mixed-ADC without round-robin training for which the high-resolution ADCs are connected to a fixed set of antennas without ADC switching or antenna selection ("Non-round-robin")

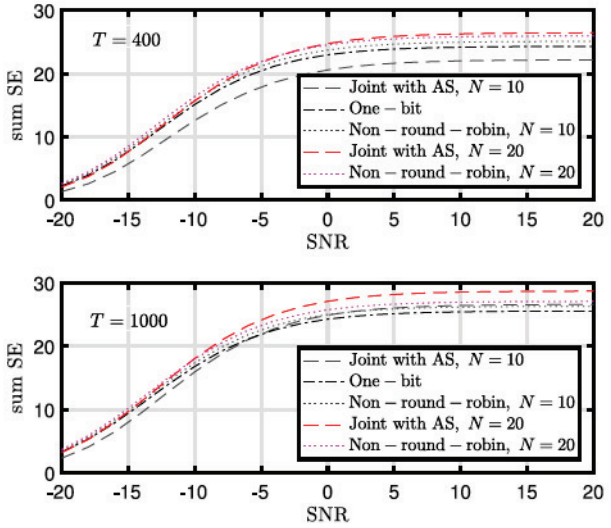

Fig. 8. Sum SE for MRC detection versus SNR for $M = 100$, $N = 20$, 10, and $T = 400$, 1000.

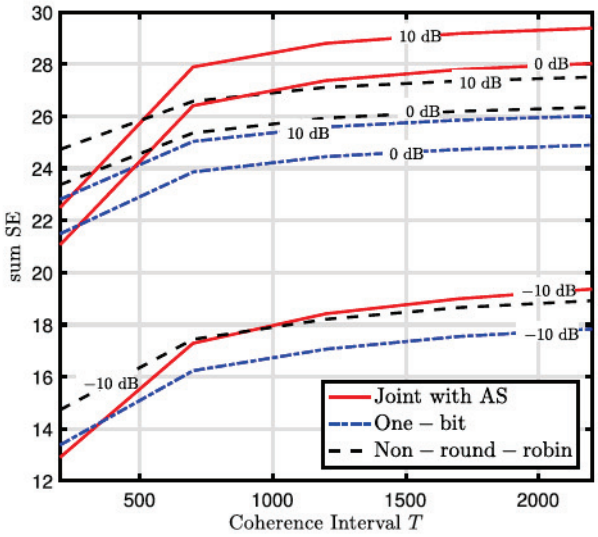

Fig. 10. Sum SE for MRC detection versus T for $M = 100$, $N = 20$, and $SNR = -10, 0, 10$ dB.

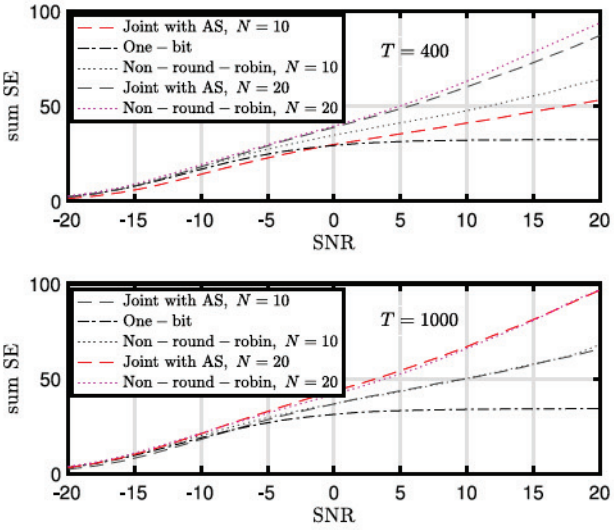

Fig. 9. Sum SE for ZF detection versus SNR for $M = 100$, $N = 20$, 10, and $T = 400$, 1000.

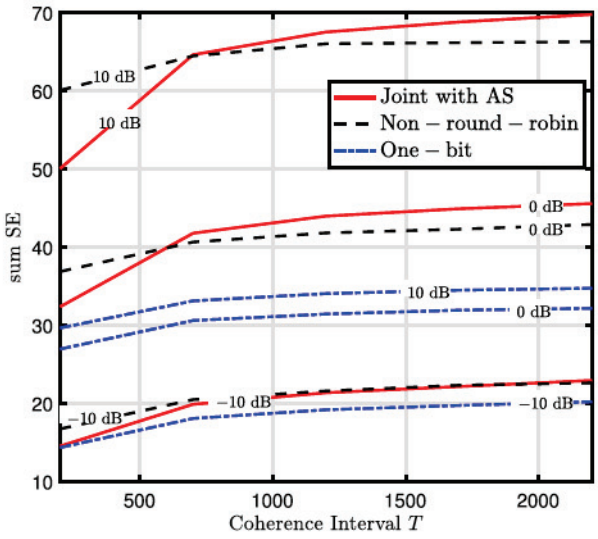

Fig. 11. Sum SE for ZF detection versus T for $M = 100$, $N = 20$, and $SNR = -10, 0, 10$ dB.

[27]. Since mixed-ADC channel estimation improves the channel estimation accuracy by expending a larger portion of the coherence interval for training, its benefit is directly related to the length of the coherence interval. For MRC detection, when $T = 400$, the mixed-ADC architecture performs better than the all-one-bit architecture for $N = 20$, but when $N = 10$ the all-one-bit architecture is better due to the larger training overhead incurred when N is smaller. However, for $T = 1000$, mixed-ADC outperforms the all-one-bit architecture at high SNRs for both $N = 10$, 20, while the all-one-bit case is still better for $N = 10$ at low SNRs. Round-robin training provides better SE performance at high SNR when $N = 20$ compared to the case without antenna switching, especially for the larger coherence interval. However, for other cases, the round-robin training overhead significantly reduces the SE, especially for $N = 10$ and the shorter coherence interval.

For ZF detection, we see that the mixed-ADC architectures can provide very large gains in SE compared to the one-bit case at high SNRs, regardless of T. For low SNRs, there is little to no improvement. These cases still do not show a significant benefit for round-robin training compared with a fixed ADC assignment; only when $N = 20$ and $T = 1000$ do we see a slight improvement.

For $N = 20$, Figs. 10 and 11 show how the coherence interval T impacts the effectiveness of the mixed-ADC architecture for MRC and ZF detectors, respectively. For mixed-ADC MRC detection, it is apparent that the best choice among the three architectures (all one-bit, mixed-ADC with and without round-robin training) depends on the SNR operating point and the length of the coherence interval. The advantage of round-robin training becomes apparent for long coherence intervals, where the increased training length has a smaller impact. The gain

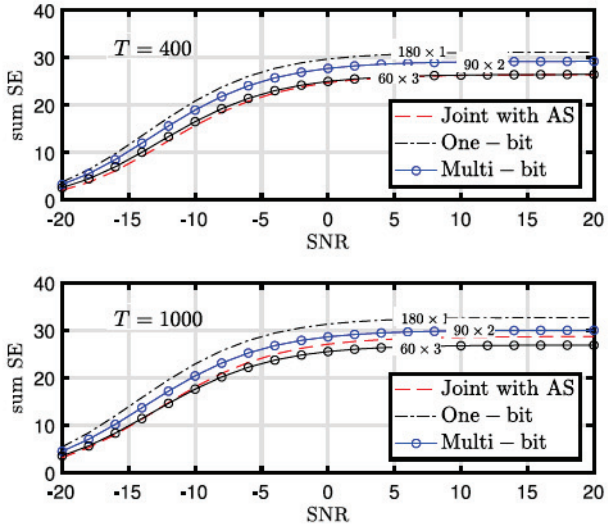

Fig. 12. Sum SE for MRC detection versus SNR for 180 comparators and $T = 400, 1000$.

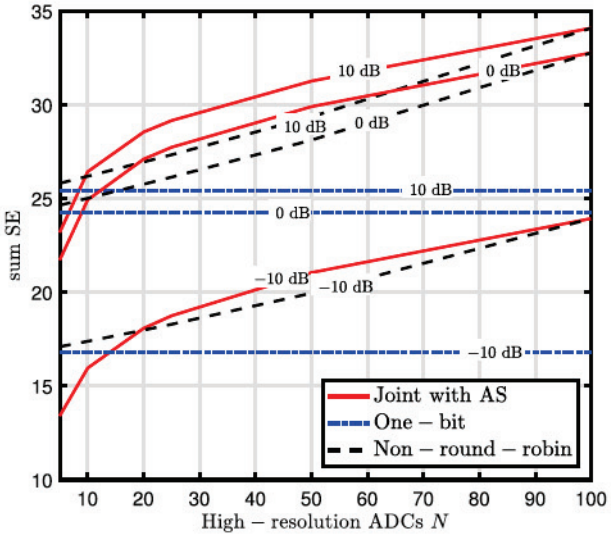

Fig. 14. Sum SE for MRC detection versus N for $SNR = -10, 0, 10$ dB and $T = 1000$.

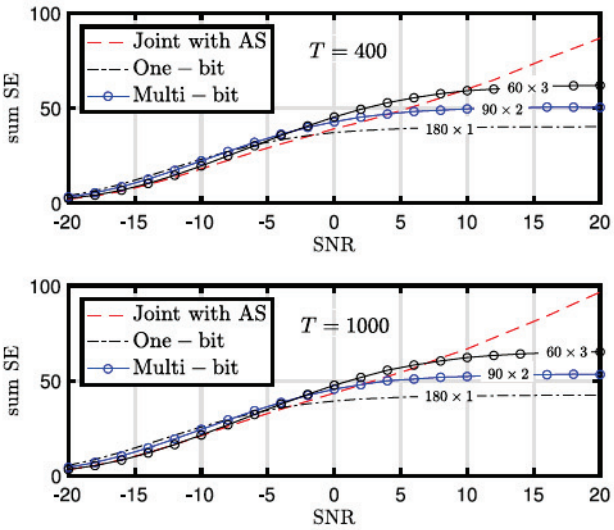

Fig. 13. Sum SE for ZF detection versus SNR for 180 comparators and $T = 400, 1000$.

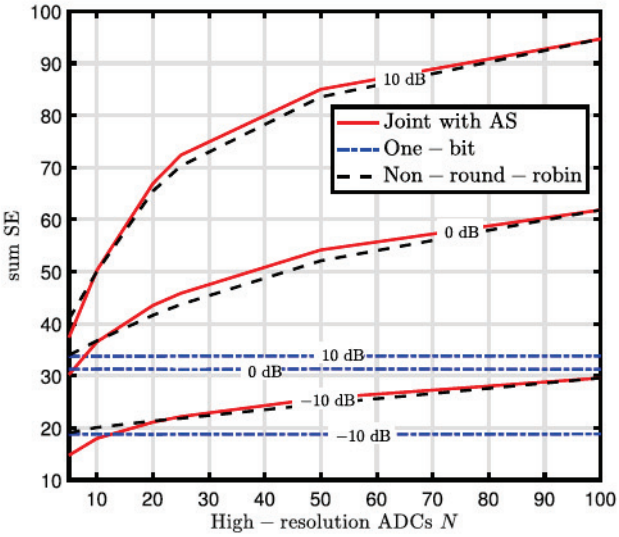

Fig. 15. Sum SE for ZF detection versus N for $SNR = -10, 0, 10$ dB and $T = 1000$.

for round-robin training is greatest at higher SNRs. For shorter coherence intervals, mixed ADC with fixed antenna/ADC assignments provides the best SE, with the largest gains again coming at higher SNRs. For this value of N, the all-one-bit system generally has the lowest SE, although the difference is not large for MRC.

The next example investigates the impact of distributing the resolution (i.e., the comparators of the ADCs) across the array with different numbers of antennas. If we assume that the "high-resolution" ADCs consist of 5 bits [43], a mixed-ADC architecture with $N = 20$ high-resolution and $M - N = 80$ one-bit ADCs will have 180 total comparators. Figs. 12 and 13 illustrate the SE achieved by distributing the 180 comparators across arrays of different length for MRC and ZF detection, respectively. In these figures, "Joint with AS" and "Non-round-robin" refer to mixed-ADC architectures with $N = 20$ 5-bit ADCs and $M - N = 80$ one-bit ADCs, "One-bit" corresponds to $M = 180$ antennas with one-bit ADCs, and "Multi-bit" indicates a system with either $M = 90$ 2-bit ADCs or $M = 60$ 3-bit ADCs. As we see in the figures, it can be inferred that for MRC detection, which is interference limited, it is better to have a larger number of antennas with lower-resolution ADCs instead of equipping the BS with fewer antennas and high resolution ADCs. This is consistent with the results of [30], [39], and is due to the fact that a larger number of antennas helps the system to more effectively cancel the interference. On the other hand, for ZF detection which is noise limited, the use of high-resolution ADCs avoids additional quantization noise imposed by the low-resolution ADCs, and is more beneficial than having a larger number of antennas with low-resolution ADCs at high SNR.

Finally, Figs. 14 and 15 show the impact of the number of high-resolution ADCs in a mixed-ADC system with

$M = 100$ antennas, $K = 10$ users, and various numbers N of high-resolution ADCs, where $N = 100$ denotes the all-high-resolution system. It is apparent that with a large enough coherence interval and a sufficient number of high-resolution ADCs, the mixed-ADC implementation with joint round-robin channel estimation and antenna selection outperforms the all-one-bit architecture and mixed-ADC without round-robin training. The gains are greatest when ZF detection is used and the SNR is high, but such gains must be weighed against the increased power consumption and hardware complexity.

VII. Conclusion

We studied the spectral efficiency of mixed-ADC massive MIMO systems with either MRC or ZF detection. We showed that properly accounting for the impact of the quantized receivers using the Bussgang decomposition is important for obtaining an accurate analysis of the SE. We introduced a joint channel estimation approach to leverage both high-resolution ADCs and one-bit ADCs and our analytical and numerical results confirmed the benefit of joint channel estimation for low SNRs.

Mixed-ADC detection with MRC and ZF detectors and antenna selection were also studied. Analytical expressions were derived for MRC detection and a numerical performance analysis was performed for ZF detection. It was shown that antenna selection provides a slight advantage for high SNRs while this advantage tends to disappear for low SNRs.

We showed that the SNR, the number of high-resolution ADCs and the length of the coherence interval play a pivotal role in determining the performance of mixed-ADC systems. We showed that, in general, mixed-ADC architectures will have the greatest benefit compared to implementations with all low-resolution ADCs when ZF detection is used and the SNR is relatively high. In such cases, the gain of the mixed-ADC approach can be substantial. Gains are also possible for MRC, but they are not as significant, and require larger numbers of high-resolution ADCs to see a benefit compared with the ZF case. The more complicated mixed-ADC approach based on ADC switching and round-robin training can achieve the best performance in some cases, particularly when the coherence interval is long and more high-resolution ADCs are available to reduce the number of training interval repetitions. Otherwise, a mixed-ADC implementation without ADC switching and extra training is preferred.

Appendix

A. Proof of Theorem 2

From (2), the observations from the high-resolution ADCs can be written as

$$\boldsymbol{v}(0) = \sqrt{\frac{1}{\eta p_k}}\mathbf{X}\boldsymbol{\phi}_{\boldsymbol{k}}^{*} = \boldsymbol{g}_k + \tilde{\boldsymbol{n}}(0), \tag{57}$$

where $\tilde{\boldsymbol{n}}(0) \sim \mathcal{CN}(\mathbf{0}, \frac{\sigma_n^2}{\eta p_k}\boldsymbol{I}_M)$. In addition, from (8), the observations from the one-bit ADCs become

$$\boldsymbol{v}(t) = \sqrt{\frac{1}{\eta p_k}}\mathbf{Y}_t\bar{\boldsymbol{\phi}}_{\boldsymbol{k}}^{*} = \boldsymbol{g}_k + \tilde{\boldsymbol{n}}(t) + \tilde{\boldsymbol{q}}(t),\ t \in \mathcal{T}, \tag{58}$$

where $\tilde{\boldsymbol{n}}(t) \sim \mathcal{CN}(\mathbf{0}, \frac{\sigma_n^2}{\eta p_k}\boldsymbol{I}_M)$ is independent of $\tilde{\boldsymbol{n}}(t')$ for $t \neq t'$, and $\tilde{\boldsymbol{q}}(t) = \sqrt{\frac{1}{\eta p_k}}\mathbf{Q}(t)\bar{\boldsymbol{\phi}}_{\boldsymbol{k}}^{*}$. Since the elements of $\boldsymbol{v}(t)$ are independent, we can estimate the mth channel g_{mk} separately. Therefore, stacking all the observations in a vector, we can write

$$\underbrace{\begin{bmatrix} v_m(0) \\ \vdots \\ v_m(t) \\ \vdots \\ v_m(\frac{M}{N}-1) \end{bmatrix}}_{\mathbf{v}} = \underbrace{\begin{bmatrix} 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\mathbf{1}_{\frac{M}{N}}} g_{mk} + \underbrace{\begin{bmatrix} \tilde{n}_m(0) \\ \vdots \\ \tilde{n}_m(t) + \tilde{q}_m(t) \\ \vdots \\ \tilde{n}_m(\frac{M}{N}-1) + \tilde{q}_m(\frac{M}{N}-1) \end{bmatrix}}_{\mathbf{u}}. \tag{59}$$

As a result, the LMMSE estimation of the mth channel coefficient for the kth user is [40]

$$\hat{g}_{mk} = \left(\frac{1}{\beta_k} + \mathbf{1}_{\frac{M}{N}}^T\mathbf{C}_{\mathbf{u}}^{-1}\mathbf{1}_{\frac{M}{N}}\right)^{-1}\mathbf{1}_{\frac{M}{N}}^T\mathbf{C}_{\mathbf{u}}^{-1}\mathbf{v}. \tag{60}$$

In Eq. (60), $\mathbf{C}_{\mathbf{u}}$ denotes the covariance matrix of $\mathbf{u}$ which is a block diagonal matrix of the form

$$\mathbf{C}_{\mathbf{u}} = \begin{bmatrix} \frac{\sigma_n^2}{\eta p_k} & 0 & \dots & 0 \\ 0 & \sigma_{w_k}^2 & \dots & \varrho_k \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \varrho_k & \dots & \sigma_{w_k}^2 \end{bmatrix} = \begin{bmatrix} \frac{\sigma_n^2}{\eta p_k} & \mathbf{0} \\ \mathbf{0} & \mathbf{S} \end{bmatrix}, \tag{61}$$

where

$$\varrho_k = \mathbb{E}\{(\tilde{n}_m(t) + \tilde{q}_m(t))(\tilde{n}_m(t') + \tilde{q}_m(t'))^{*}\},\ t \neq t', \tag{62}$$

can be easily calculated with the aid of the Bussgang decomposition and the arcsine law as in (24). Substituting (61) into (60), we have

$$\hat{g}_{mk} = \left(\frac{1}{\beta_k} + \frac{\eta p_k}{\sigma_n^2} + \mathbf{1}_{\frac{M}{N}-1}^T\mathbf{S}^{-1}\mathbf{1}_{\frac{M}{N}-1}\right)^{-1} \times \left[\frac{\sigma_n^2}{\eta p_k}\ \mathbf{1}_{\frac{M}{N}-1}^T\mathbf{S}^{-1}\right]\mathbf{v}. \tag{63}$$

To calculate the inverse of the matrix $\mathbf{S}$, we re-write it as

$$\mathbf{S} = \left(\sigma_{w_k}^2 - \varrho_k\right)\mathbf{I}_{\frac{M}{N}-1} + \varrho_k\mathbf{1}_{\frac{M}{N}-1}\mathbf{1}_{\frac{M}{N}-1}^T, \tag{64}$$

and use Woodbury's matrix identity:

$$\mathbf{S}^{-1} = \frac{1}{\sigma_{w_k}^2 - \varrho_k}\mathbf{I}_{\frac{M}{N}-1} - \frac{1}{\left(\sigma_{w_k}^2 - \varrho_k\right)^2}\left(\frac{1}{\varrho_k} + \frac{(\frac{M}{N}-1)}{\sigma_{w_k}^2 - \varrho_k}\right)^{-1}\mathbf{1}_{\frac{M}{N}-1}\mathbf{1}_{\frac{M}{N}-1}^T, \tag{65}$$

which yields

$$\mathbf{1}_{\frac{M}{N}-1}^T\mathbf{S}^{-1} = \frac{1}{\sigma_{w_k}^2 + \left(\frac{M}{N}-2\right)\varrho_k}\mathbf{1}_{\frac{M}{N}-1}^T, \tag{66}$$

$$\mathbf{1}_{\frac{M}{N}-1}^T\mathbf{S}^{-1}\mathbf{1}_{\frac{M}{N}-1} = \frac{\left(\frac{M}{N}-1\right)}{\sigma_{w_k}^2 + \left(\frac{M}{N}-2\right)\varrho_k}. \tag{67}$$

Substituting (66) and (67) into (63) completes the proof.

B. Proof of Theorem 3

Denote the energy of the mth row, $m \in \mathcal{M} = \{1, \ldots, M\}$, of $\hat{\mathbf{H}}$ by $\mathcal{E}_m$, i.e.,

$$\mathcal{E}_m \triangleq \sum_{k=1}^{K} \left| \hat{h}_{mk} \right|^2. \tag{68}$$

To do antenna selection, we must connect the N high-resolution ADCs to the antennas corresponding to the largest $\mathcal{E}_m$. Suppose that the indices of the N antennas to which the high-resolution ADCs are connected are contained in the set $\mathcal{N}$. Hence, we have

$$\sum_{k=1}^{K} \mathbb{E}\left\{ \hat{\boldsymbol{h}}_k^H \mathbf{C}_{\boldsymbol{q}_d} \hat{\boldsymbol{h}}_k \right\} = K\mathbb{E}\left\{ \hat{\boldsymbol{h}}_k^H \mathbf{C}_{\boldsymbol{q}_d} \hat{\boldsymbol{h}}_k \right\} = \left(1 - \frac{2}{\pi}\right) \sum_{\mathcal{M}\setminus\mathcal{N}} \mathbb{E}\{\mathcal{E}_m\}. \tag{69}$$

Equation (69) provides a criterion for connecting the N high-resolution ADCs in the data transmission phase. In fact, it states that, for the MRC receiver, the expected value in (69) will be minimized if the high-resolution ADCs are connected to the antennas corresponding to the largest $\mathcal{E}_m$. Denote $\mathcal{E}_{(m)}$ as the mth smallest value of $\mathcal{E}_m$, i.e.,

$$\mathcal{E}_{(1)} \leq \mathcal{E}_{(2)} \leq \cdots \leq \mathcal{E}_{(M)}.$$

Hence, $\mathcal{E}_{(m)}$ is the mth order statistic, and assuming that the $\mathcal{E}_{(m)}$ are statistically independent and identically distributed, we have [46]

$$\mathbb{E}\{\mathcal{E}_{(m)}\} = M\binom{M-1}{m-1} \int_{-\infty}^{\infty} x \left[F(x)\right]^{m-1} \left[1 - F(x)\right]^{M-m} dF(x), \tag{70}$$

where x is the realization of $\mathcal{E}_{(m)}$ and $F(x)$ is the cumulative distribution function of $\mathcal{E}_m$. For the case that we have considered, where the channel coefficients are i.i.d. Rayleigh distributed, the $\mathcal{E}_m$ are independent Gamma random variables with

$$F(x) = \gamma\left(\frac{x}{\sigma_{\hat{h}}^2}, K\right), \tag{71}$$

where $\gamma(.,.)$ denotes the incomplete Gamma function. From [47], the integral (70) can be calculated in closed form for Gamma random variables as

$$\mathbb{E}\{\mathcal{E}_{(m)}\} = \sigma_{\hat{h}}^2 \chi_m. \tag{72}$$

This is in contrast to the unordered case where $\mathbb{E}\{\mathcal{E}_m\} = K\sigma_{\hat{h}}^2$. As a result

$$\min\left\{ \mathbb{E}\left\{ \hat{\boldsymbol{h}}_k^H \mathbf{C}_{\boldsymbol{q}_d} \hat{\boldsymbol{h}}_k \right\} \right\} = \left(1 - \frac{2}{\pi}\right) \frac{\sigma_{\hat{h}}^2}{K} \sum_{m=1}^{M-N} \chi_m. \tag{73}$$

The remaining terms in (47) can be calculated similar to the case where the high-resolution ADCs are connected to arbitrary antennas. Plugging these terms and (73) into (47) and some algebraic manipulation results in (49).

REFERENCES

[1] H. Pirzadeh and A. Swindlehurst, "Analysis of MRC for mixed-ADC massive MIMO," in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adapt. Process.*, 2017, pp. 1–5.

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

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

[4] 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.

[5] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[6] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 172–179, Feb. 2013.

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

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

[9] C. Mollén, E. Larsson, and T. Eriksson, "Waveforms for the massive MIMO downlink: Amplifier efficiency, distortion, and performance," *IEEE Trans. Commun.*, vol. 64, no. 12, pp. 5050–5063, Dec. 2016.

[10] C. Mollén, U. Gustavsson, T. Eriksson, and E. Larsson, "Spatial characteristics of distortion radiated from antenna arrays with transceiver nonlinearities," arXiv:1711.02439, 2017.

[11] C. Mollén, E. G. Larsson, U. Gustavsson, T. Eriksson, and R. W. Heath, "Out-of-band radiation from large antenna arrays," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 196–203, Apr. 2018.

[12] C. Mollén, U. Gustavsson, T. Eriksson, and E. Larsson, "Impact of spatial filtering on distortion from low-noise amplifiers in massive MIMO base stations," arXiv:1712.09612, 2017.

[13] Q. Bai and J. A. Nossek, "Energy efficiency maximization for 5G multi-antenna receivers," *Trans. Emerg. Telecommun. Technol.*, vol. 26, no. 1, pp. 3–14, 2015.

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

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

[16] 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.

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

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

[19] M. Sarajlić, L. Liu, and O. Edfors, "When are low resolution ADCs energy efficient in massive MIMO?" *IEEE Access*, vol. 5, pp. 14837–14853, 2017.

[20] D. Verenzuela, E. Björnson, and M. Matthaiou, "Hardware design and optimal ADC resolution for uplink massive MIMO systems," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.

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

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

[23] N. Liang and W. Zhang, "Mixed-ADC massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 983–997, Apr. 2016.
[24] N. Liang and W. Zhang, "Mixed-ADC massive MIMO uplink in frequency-selective channels," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4652–4666, Nov. 2016.
[25] W. Tan, S. Jin, C. Wen, and Y. Jing, "Spectral efficiency of mixed-ADC receivers for massive MIMO systems," *IEEE Access*, vol. 4, pp. 7841–7846, 2016.
[26] J. Zhang, L. Dai, Z. He, S. Jin, and X. Li, "Performance analysis of mixed-ADC massive MIMO systems over Rician fading channels," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1327–1338, Jun. 2017.
[27] H. Pirzadeh and A. Swindlehurst, "Spectral efficiency under energy constraint for mixed-ADC MRC massive MIMO," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1847–1851, Oct. 2017.
[28] T. C. Zhang, C. K. Wen, S. Jin, and T. Jiang, "Mixed-ADC massive MIMO detectors: Performance analysis and design optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7738–7752, Nov. 2016.
[29] J. Liu, J. Xu, W. Xu, S. Jin, and X. Dong, "Multiuser massive MIMO relaying with mixed-ADC receiver," *IEEE Signal Process. Lett.*, vol. 24, no. 1, pp. 76–80, Dec. 2016.
[30] J. Park, S. Park, A. Yazdan, and R. W. Heath,"Optimization of mixed-ADC multi-antenna systems for cloud-RAN deployments," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3962–3975, Sep. 2017.
[31] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," Res. Lab. Electron., Massachusetts Inst. Technol., Cambridge, MA, USA, Tech. Rep. 216, 1952.
[32] 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.
[33] A. Garcia-Rodriguez, C. Masouros, and P. Rulikowski, "Reduced switching connectivity for large scale antenna selection," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 2250–2263, May 2017.
[34] Y. Gao, H. Vinck, and T. Kaiser, "Massive MIMO antenna selection: Switching architectures, capacity bounds, and optimal antenna selection algorithms," *IEEE Trans. Signal Process.*, vol. 66, no. 5, pp. 1346–1360, Mar. 2018.
[35] X. Gao, O. Edfors, F. Tufvesson, and E. Larsson, "Multi-switch for antenna selection in massive MIMO," in *Proc. IEEE Global Commun. Conf.*, San Diego, CA, USA, 2015, pp. 1–6.
[36] A. Le Fevre and R. Flett, "A 100 Mb/s multi-LAN crosspoint chip set for cable management," *IEEE J. Solid-State Circuits*, vol. 32, no. 7, pp. 1115–1121, Jul. 1997.
[37] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, Feb. 2016.
[38] 2006. [Online]. Available: http://www.analog.com/media/en/newsmarketing-collateral/product-selection-guide/HighSpeedSwitches.pdf
[39] H. Pirzadeh and A. Swindlehurst, "On the optimality of mixed-ADC massive MIMO with MRC detection," in *Proc. Int. ITG Workshop Smart Antennas*, 2017, pp. 1–6.
[40] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[41] L. Fan, D. Qiao, S. Jin, C. K. Wen, and M. Matthaiou, "Optimal pilot length for uplink massive MIMO systems with low-resolution ADC," in *Proc. IEEE Sensor Array Multichannel Signal Process. Workshop*, Rio de Janeiro, Brazil, Jul. 2016, pp. 1–5.
[42] H. Q. Ngo, M. Matthaiou, and E. G. Larsson, "Massive MIMO with optimal power and training duration allocation," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 605–608, Dec. 2014.
[43] K. Roth, H. Pirzadeh, A. L. Swindlehurst, and J. A. Nossek, "A comparison of hybrid beamforming and digital beamforming with low-resolution ADCs for multiple users and imperfect CSI," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 3, pp. 484–498, 2018.
[44] D. Qiao, W. Tan, Y. Zhao, C.-K. Wen, and S. Jin, "Spectral efficiency for massive MIMO zero-forcing receiver with low-resolution ADC," in *Proc. IEEE Wireless Commun. Signal Process.*, Yangzhou, China, Oct. 2016, pp. 1–6.
[45] Q. Shi and Y. Karasawa, "Some applications of Lauricella hypergeometric function F_A in performance analysis of wireless communications," *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 581–584, pp. 1–6, May 2012.
[46] H. A. David, *Order Statistics*, 2nd ed. New York, NY, USA: Wiley, 1981.
[47] S. Nadarajah and M. Pal, "Explicit expressions for moments of gamma order statistics," *Bull. Brazilian Math. Soc., New Ser.*, vol. 39, no. 1, pp. 45–60, Mar. 2008.

Hessam Pirzadeh (S'15) received the B.Sc. degree in engineering science from the University of Tehran, Tehran, Iran, in 2013, and the M.Sc. degree in electrical engineering from Iran University of Science and Technology, Tehran, Iran, in 2016. Since 2016, he has been working toward the Ph.D. degree with the Center for Pervasive Communications and Computing, University of California—Irvine, Irvine, CA, USA. His research interests include signal processing and wireless communications with an emphasis on massive MIMO systems.

A. Lee Swindlehurst received the B.S. and M.S. degrees from Brigham Young University (BYU), Provo, UT, USA, and the Ph.D. degree from Stanford University, Stanford, CA, USA, in 1985, 1986, and 1991, respectively, all in electrical engineering. From 1990 to 2007, he was with the Department of Electrical and Computer Engineering, BYU, where he was a Department Chair from 2003 to 2006. During 1996–1997, he held a joint appointment as a Visiting Scholar with Uppsala University, Uppsala, Sweden, and the Royal Institute of Technology, Stockholm, Sweden. From 2006 to 2007, he was on leave working as Vice-President of Research for ArrayComm LLC, San Jose, CA, USA. Since 2007, he has been a Professor with the Electrical Engineering and Computer Science Department, University of California—Irvine, Irvine, CA, USA, where he was an Associate Dean for research and graduate studies with the Samueli School of Engineering from 2013 to 2016. During 2014–2017, he was also a Hans Fischer Senior Fellow with the Institute for Advanced Studies, Technical University of Munich. His current research focuses on array signal processing for radar, wireless communications, and biomedical applications, and he has more than 300 publications in these areas. He was the inaugural Editor-in-Chief for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He was the recipient of the 2000 IEEE W. R. G. Baker Prize Paper Award, the 2006 IEEE Communications Society Stephen O. Rice Prize in the Field of Communication Theory, the 2006 and 2010 IEEE Signal Processing Society's Best Paper Awards, and the 2017 IEEE Signal Processing Society Donald G. Fink Overview Paper Award.