One-Bit Massive MIMO: Channel Estimation and High-Order Modulations

Sven Jacobsson[*†], Giuseppe Durisi[*], Mikael Coldrey[†], Ulf Gustavsson[†], Christoph Studer[‡]

[*]Chalmers University of Technology, Gothenburg, Sweden
[†]Ericsson Research, Sweden
[‡]Cornell University, Ithaca, NY, USA

Abstract—We investigate the information-theoretic throughout achievable on a fading communication link when the receiver is equipped with one-bit analog-to-digital converters (ADCs). The analysis is conducted for the setting where neither the transmitter nor the receiver have *a priori* information on the realization of the fading channels. This means that channel-state information needs to be acquired at the receiver on the basis of the one-bit quantized channel outputs. We show that least-squares (LS) channel estimation combined with joint pilot and data processing is capacity achieving in the single-user, single-receive-antenna case.

We also investigate the achievable uplink throughput in a massive multiple-input multiple-output system where each element of the antenna array at the receiver base-station feeds a one-bit ADC. We show that LS channel estimation and maximum-ratio combining are sufficient to support both multiuser operation and the use of high-order constellations. This holds in spite of the severe nonlinearity introduced by the one-bit ADCs.

I. INTRODUCTION

Digital signal processing (DSP) is an integral part of all modern communication systems. In order to process data digitally, the analog baseband signal has to be mapped to the digital domain. This requires conversion both in time (sampling) and amplitude (quantization). The circuit that performs this last operation, known as analog-to-digital converter (ADC), is a necessary component in every system that includes DSP. An ADC with frequency f_s and resolution of n bits maps the continuous-amplitude samples into a set of 2^n quantization levels, by operating $f_s 2^n$ conversion steps per second. A crucial problem with modern ADCs is that the power dissipated per conversion step (a.k.a. Walden's figure of merit [1], [2]) increases dramatically for sampling rates higher than about $100\,\mathrm{MHz}$ [3]. This implies that, for wideband communication systems, the resolution of the ADCs must be kept low to maintain a power budget that is within acceptable levels.

The one-bit resolution case, where the in-phase and the quadrature components of the continuous-valued received samples are quantized separately using one-bit ADCs (zero-threshold comparators), is particularly attractive, because of the resulting low hardware complexity. Indeed, in such a one-bit ADC architecture, there is no need for an automatic gain controller. Communication systems employing one-bit ADCs have been previously analyzed in the context of low-power ultra-wideband systems [4]–[6], and,

more recently, in the context of millimeter-wave communication systems [7], [8], and massive (or large-scale) multiple-input multiple-output (MIMO) systems [9]. In ultra-wideband and in millimeter-wave systems, the motivation for using one-bit ADCs is the large bandwidth of the transmitted signal. In massive MIMO systems, an additional reason is the massification of the number of radio-frequency chains at the base-station (BS), which makes the use of low-cost solutions—such as one-bit ADCs—attractive [9].

Previous Results: A receiver employing one-bit ADCs needs to cope with the severe nonlinearity introduced by such quantizers. In their presence, the signaling schemes and the receiver algorithms employed for the case of high-resolution quantizers become suboptimal. The impact of the one-bit ADC nonlinearity on the performance of communication systems has been analyzed to some extent in the literature. In [10], it is proven that 2-PAM is capacity achieving over a real-valued nonfading single-input single-output (SISO) Gaussian channel. For complex-valued Gaussian channels, QPSK turns out to be optimal. For the general MIMO case, QPSK is not capacity achieving, and the capacity-achieving distribution is unknown.

These results hold under the assumption that the one-bit ADC is a zero-threshold comparator. It turns out that in the low-SNR regime, a zero-threshold comparator is not optimal [11]. The optimal strategy involves the use of *flash signaling* [12, Def. 2] and requires an optimization over the threshold value. Unfortunately, the power gain obtainable using this optimal strategy manifests itself only at extremely low values of spectral efficiency. In the remainder of the paper, we therefore exclusively focus on the zero-threshold architecture.

Moving to Rayleigh-fading channels, QPSK is capacity achieving (again for the SISO case) under the assumption that the receiver has somehow access to perfect channel state information (CSI) [13]. The assumption that perfect CSI is available is, however, not realistic in the one-bit quantized case, since the nonlinear distortion caused by the one-bit quantizers makes it challenging to estimate the fading process perfectly. For the more practically relevant case when the channel is not known *a priori* to the receiver, but must be learnt (for example, via pilot symbols), QPSK is optimal when the SNR exceeds a certain threshold that depends on the coherence time of the fading process [14]. For SNR values that are below this threshold, on-off QPSK is capacity achieving [14].

In [7], crude high-SNR bounds are obtained for the capacity

of one-bit-quantized MIMO fading channels, under the ideal assumption that perfect CSI is available to both the transmitter and the receiver. Risi *et al.* [9] recently provided a lower bound on the throughput achievable on massive MIMO uplink channels, when the BS employs one-bit ADCs. The bound suggests that, in some scenarios, massive MIMO may be robust against the coarse output quantization resulting from the use of one-bit ADCs. However, as the lower bound obtained in [9] is based on a suboptimal input distribution, i.e., QPSK, and a suboptimal detection algorithm, i.e., least-squares (LS) channel estimation followed by maximal-ratio combining (MRC) or zero-forcing, its tightness is unclear.

All the results reviewed so far hold under the assumption of Nyquist-rate sampling at the receiver. It is worth pointing out that Nyquist-rate sampling is not necessarily optimal in the presence of quantization at the receiver [15], [16]. Indeed, higher information rates can be achieved by oversampling the received signal. For example, for the complex AWGN case, high-order constellations such as 16-QAM can be supported in the SISO case if one allows for oversampling at the receiver [17].

Contributions: Focusing on Nyquist-rate sampling, and on the scenario where neither the transmitter, nor the receiver have *a priori* CSI, we investigate the rates achievable over Rayleigh block-fading MIMO channels when the receiver is equipped with one-bit ADCs. Our contribution is twofold:

- For the SISO case, we prove that LS channel estimation performed jointly on pilot and data symbols is capacity achieving. In the infinite precision (no quantization) case, the benefit of *joint pilot-data* (JPD) processing has been illustrated, e.g., in [18]–[20], where it is shown that joint processing yields a smaller gap to capacity compared to separate pilot/data processing. Our result shows that, in the one-bit ADC case, the gap to capacity is actually zero. Moreover, LS estimation, although inferior to the optimal maximum a posteriori probability estimator (see [21], [22]), suffices to achieve capacity when combined with JPD processing.
- We also consider the uplink of a massive MIMO system where single-antenna users communicate with a BS equipped with a large antenna array whose elements feed one-bit ADCs. Generalizing the analysis presented in [9], we show that MRC combined with LS channel estimation at the BS is sufficient to support both multi-user operations and the use of high-order constellation such as 16-QAM. Furthermore, the rates achievable with 16-QAM turn out to exceed the ones reported in [9] for QPSK, for SNR values as low as -15 dB, and for antenna arrays of 100 elements or more. Our result suggests that temporal oversampling, as proposed in [17], can be replaced by spatial oversampling through the use of a massive antenna array at the BS.

II. SYSTEM MODEL

We consider a single-cell uplink system as depicted Fig. 1, where K single-antenna users are served by a BS that is equipped with an array of $N \gg K$ antennas. We model the subchannels between each transmit-receive antenna pair as a Rayleigh block-fading channel (see e.g., [23]), i.e., a channel that stays constant
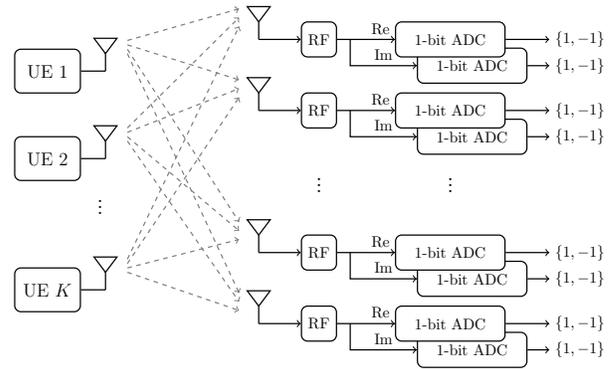

Fig. 1. One-bit massive MIMO uplink system model.

for T channel uses, and evolves independently across blocks of T channel uses. We shall refer to T as the channel coherence time (measured in channel uses). We further assume that the subchannels are mutually independent. The discrete-time complex baseband received signal over all antennas within an arbitrary coherence block and before quantization, is modelled as

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

Here, $\mathbf{X} \in \mathbb{C}^{T \times K}$ denotes the channel input, $\mathbf{H} \in \mathbb{C}^{K \times N}$ is the channel matrix connecting the K users to the N BS antennas. The entries of $\mathbf{H}$ are independent and $\mathcal{CN}(0,1)$ distributed. Furthermore, the matrix $\mathbf{W} \in \mathbb{C}^{T \times N}$, whose entries are independent and $\mathcal{CN}(0,1)$ distributed, stands for the AWGN.

The real and the imaginary components of the received signal at each antenna are quantized separately using a one-bit ADC. Let $\mathcal{R} = \{1 + j, -1 + j, -1 - j, 1 - j\}$ be the set of possible quantization outcomes. It will be convenient to describe the joint operation of all $2N$ one-bit ADCs at the BS through the function $\mathcal{Q}(\cdot) : \mathbb{C}^{T \times N} \to \mathcal{R}^{T \times N}$ that maps the output matrix $\mathbf{Y}$ with entries $\{y_{t,n}\}$ into the quantized output matrix $\mathbf{R}$ with entries $\{r_{t,n}\}$ according to $\Re\{r_{t,n}\} = \text{sign}\{\Re\{y_{t,n}\}\}$ and $\Im\{r_{t,n}\} = \text{sign}\{\Im\{y_{t,n}\}\}$. Using this convention, we can write the quantized output matrix as

$$\mathbf{R} = \mathcal{Q}(\mathbf{Y}) = \mathcal{Q}(\mathbf{X}\mathbf{H} + \mathbf{W}). \tag{2}$$

We shall consider the scenario where neither the users nor the BS are aware of the realizations of the channel matrix $\mathbf{H}$ (no *a priori* CSI), and where coding is performed over many coherence blocks. For this scenario, the channel sum-rate capacity C is

$$C(\rho) = \sup I(\mathbf{X}; \mathbf{R}) \tag{3}$$

where the supremum is computed over all input probability distributions on $\mathbf{X}$ for which the columns of $\mathbf{X}$ are independent, and the following average-power constraint holds:

$$\mathbb{E}\left[\text{tr}\{\mathbf{X}\mathbf{X}^H\}\right] \leq KT\rho. \tag{4}$$

Since the noise variance is normalized to one, we can think of ρ as the SNR. The sum-rate capacity (3) is, in general, not known in closed form, even in the infinite-precision case, for which tight bounds have been recently reported in [24].

III. SISO CASE

We focus in this section on the scenario where there is only a single active user in the system and where the BS has a single receive antenna. For this case, the input-output relation (2) reduces to

$$\mathbf{r} = \mathcal{Q}(\mathbf{y}) = \mathcal{Q}(\mathbf{x}h + \mathbf{w}). \tag{5}$$

Here, $\mathbf{x} \in \mathbb{C}^T$ and $\mathbf{w} \in \mathbb{C}^T$ are the input and noise vectors, respectively, and h denotes the fading channel, which remains constant over the coherence block. For this case, the capacity (3) is known [14, Th. 1] and given by

$$C(\rho) = \begin{cases} \dfrac{\rho}{\rho_c} R_{\mathrm{QPSK}}(\rho_c), & \rho \le \rho_c \\ R_{\mathrm{QPSK}}(\rho), & \rho > \rho_c. \end{cases} \tag{6}$$

Here, R_{QPSK} denotes the rate achievable with QPSK,

$$R_{\mathrm{QPSK}}(\rho) = 2 + \frac{2}{T} \sum_{k=0}^{T} \binom{T}{k} \beta(k, T-k) \log_2 \beta(k, T-k) \tag{7}$$

where

$$\beta(a,b) = \mathbb{E}_g \Big[\Phi \left(-g\sqrt{\rho} \right)^a \Phi \left(g\sqrt{\rho} \right)^b \Big] \tag{8}$$

with $g \sim \mathcal{N}(0,1)$ and $\Phi(x)$ denoting the cumulative distribution function of a standard normal random variable. Furthermore, the SNR threshold ρ_c in (6) is the solution of the following optimization problem:

$$\rho_c = \arg\max_{\rho \ge 0} \frac{1}{\rho} R_{\mathrm{QPSK}}(\rho). \tag{9}$$

A common approach to transmitting information over fading channels whose realizations are not known *a priori* to the receiver is to reserve a certain number of channel uses at the beginning of each coherence block for the transmission of pilot symbols [25]. These pilot symbols are then used at the receiver to estimate the fading channel. Assume that $0 \le P \le T$ pilots are used and let $\mathbf{x}^{(p)}$ be the P-dimensional vector containing these pilot symbols. Similarly, let $\mathbf{r}^{(p)}$ be the corresponding one-bit quantized channel output. The pair $(\mathbf{x}^{(p)}, \mathbf{r}^{(p)})$ can be used at the receiver to estimate the channel h. As in [9], we shall focus on LS estimation because of its low complexity. When the pilots are QPSK symbols, the LS estimate $\hat{h}$ of h is

$$\hat{h} = \frac{1}{P\sqrt{\rho}} (\mathbf{x}^{(p)})^H \mathbf{r}^{(p)}. \tag{10}$$

Under the assumption that the $T - P$ data symbols are drawn independently from the same input distribution, one obtains then the following lower bound on C:

$$C(\rho) \ge \frac{T-P}{T} I(x; r \mid \hat{h}) \tag{11}$$

$$= 2\frac{T-P}{T} \left[1 - \sum_{\ell=0}^{P} \binom{P}{\ell} \beta(\ell, P-\ell) H_b \left(\frac{\beta(\ell+1, P-\ell)}{\beta(\ell, P-\ell)} \right) \right]. \tag{12}$$

Here, $H_b(\cdot)$ denotes the binary entropy function. The inequality in (11) follows from standard manipulations on the mutual
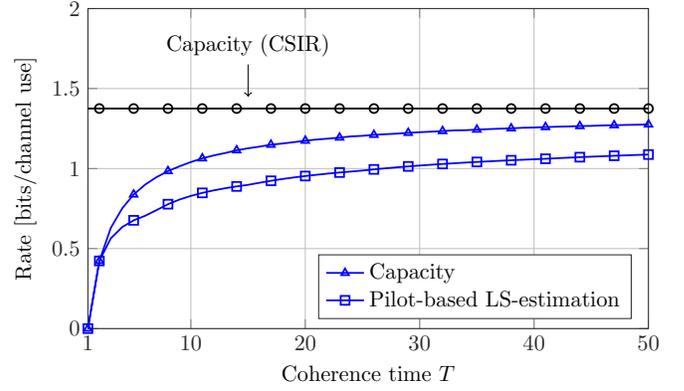

Fig. 2. Comparison of capacity (6) and the pilot-based LS-estimation lower bound (12) for the SISO case when $\rho = 10$ dB.

information (see, e.g., [20]); the equality (12) is proven in Appendix A.

In Fig. 2, we plot the capacity (6) and the pilot-based LS-estimation lower bound (12) for the case $\rho = 10$ dB. Note that, for this ρ value, $C(\rho) = R_{\mathrm{QPSK}}(\rho)$ for all T. The number of pilots in (12) is optimized for each value of T. For reference, we also depict the perfect receiver-CSI capacity [13]. As T grows large, the gap between (6) and the perfect receiver-CSI capacity decreases. The gap between the pilot-based LS-estimation lower bound (12) and capacity (6) is essentially constant over the considered range of T values. One exception is the case $T = 2$, for which there seems to be no gap. Indeed, the following result holds (see [26, Lem. 1]).

Lemma 1: The RHS of (12) coincides with the rate acheivable with QPSK (7) for the case $T = 2$.

It is well known that the pilot-based lower bound (12) can be improved by using also the channel outputs corresponding to the data symbols to improve the channel estimate [18]–[20]. This approach is sometimes referred to as JPD processing. Perhaps surprisingly, for the one-bit quantization case, LS estimation combined with JPD processing turns out to be optimal, as formalized in the following theorem.

Theorem 2: For the channel (5), LS estimation combined with JPD processing achieves the channel capacity (6).

Proof: See Appendix B. ∎

This result implies that if one allows for JPD processing, LS channel estimation is optimal, and there is no need to use more sophisticated channel-estimation techniques such as the one recently proposed in [27].

IV. MASSIVE MIMO CASE

Motivated by the results obtained for the SISO case, we now assess the rates achievable with LS estimation in a multiuser massive MIMO uplink scenario. To limit the receiver complexity, we shall only consider the pilot-based version of the LS estimation algorithm (no JPD processing). Indeed, JPD processing is in general computationally demanding [20] and may be not suitable for massive MIMO. We shall also assume that the receiver

(a) $N = 40$ antennas, $\rho = 0$ dB.   (b) $N = 400$ antennas, $\rho = 0$ dB.

(c) $N = 400$ antennas, $\rho = 20$ dB.   (d) $N = 400$ antennas, $\rho = 20$ dB, fully correlated channel coefficients.
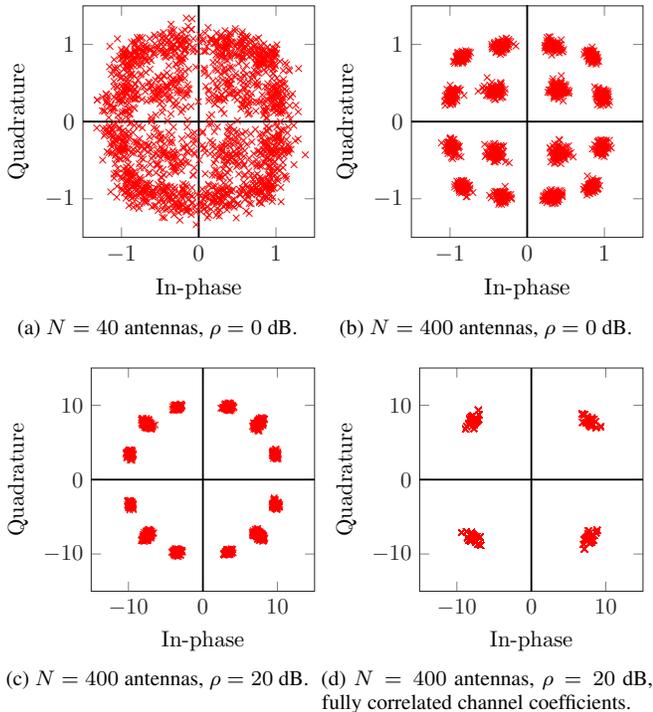
Fig. 3. Single-user MRC outputs with LS channel estimation as a function of the number of receive antennas and the SNR; 16-QAM inputs.

employs MRC to separate the information streams associated with the different users.

We assume that the users coordinate the pilot-transmission phase. Specifically, they transmit their pilot sequences in a round-robin fashion. The channel estimates are then used to obtain the MRC coefficients. Differently from [9], where a similar setup is considered, we focus on high-order modulations and not only on QPSK. Indeed, although QPSK is optimal in the SISO case, the use of multiple antennas at the receiver opens up the possibility to use high-order modulation formats. This is demonstrated in Fig. 3 where we plot the MRC receiver output corresponding to 16-QAM data symbols for the case when a single user, alone in the cell, transmits also $P = 20$ pilots to let the BS acquire LS channel estimates. As the size of the receiver antenna array increases, the 16-QAM constellation becomes progressively distinguishable (see Fig. 3b), provided that ρ is not too high.

Additive noise is one of the factors that enables the detection of the 16-QAM constellation; the other is the independent phase of the fading coefficients associated with each receive antenna. Recall in fact that, due to the one-bit quantizer, the quantized output at each receive antenna belongs to the set $\mathcal{R}$ of cardinality 4. These 4 possible channel outputs are then averaged by the MRC filter to produce a channel output (a scalar) that belongs to an alphabet with much higher cardinality. The cardinality depends on the number of pilots and receive antennas. The key observation is that the inner points of the 16-QAM constellation are more likely to be erroneously detected. This results in a smaller averaged value after MRC than for the outer constellation
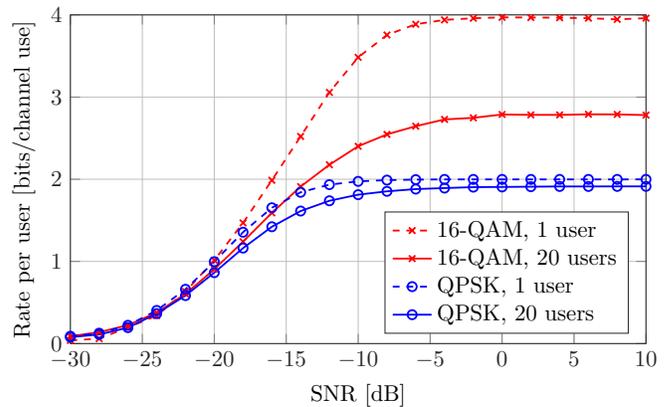

Fig. 4. Per-user achievable rate with LS estimation and MRC as a function of ρ; $T = 1000$, $N = 400$; the number of pilots P is optimized for each value of ρ.

points.

To highlight the importance of the additive noise, we consider in Fig. 3c the case $\rho = 20$ dB. Since the additive noise is negligible, all 16-QAM constellation points are detected correctly with high probability. As a result, the output of the MRC filter lies approximately on a circle, which suggests that the amplitude of the transmitted signal cannot be used to convey information. When the noise is negligible and all fading coefficients are fully correlated, the constellation collapses to a noisy QPSK diagram (Fig. 3d). In this case, high-order modulations are not supported by the channel.

The achievable rate $R_{\mathrm{MRC}}^{(k)}$ for user $k = 1, \ldots, K$ with LS estimation and MRC is

$$R_{\mathrm{MRC}}^{(k)}(\rho) = \frac{T - P}{T} I(x^{(k)}; \tilde{x}^{(k)} \,|\, \hat{\mathbf{H}}). \tag{13}$$

The mutual information between the channel input $x^{(k)}$ and the MRC receiver output $\tilde{x}^{(k)}$ can be computed by mapping $\tilde{x}^{(k)}$ to points over a regular grid in the complex plane as described in [9]. With this technique, one obtains a lower bound on $R_{\mathrm{MRC}}^{(k)}(\rho)$ [28, p. 3503] that becomes increasingly tight as the grid spacing is driven to zero. The conditional probability mass functions needed for the evaluation of the mutual information are computed using Monte-Carlo techniques.[1] Since all the users in the system are assumed to be statistically equivalent, we have that

$$C(\rho) \geq K R_{\mathrm{MRC}}^{(1)}(\rho). \tag{14}$$

In Fig. 4 we compare the rates achievable with QPSK and 16-QAM as a function of ρ. The number of receive antennas is $N = 400$, the coherence time is $T = 1000$, and we consider both the case when the number of users K is 1 and 20. The number of transmitted pilots is optimized for every ρ value. We see that 16-QAM outperforms QPSK already at SNR values as low as -15 dB. Furthermore, the full 16-QAM rate of 4 bits per channel use can be achieved in the single-user case for large SNR values. Note that if ρ is further increased, the 16-QAM rate starts decreasing, because the constellation collapses

[1]The numerical routines that are used to evaluate (13) can be downloaded at https://github.com/infotheorychalmers/one-bit_massive_MIMO.
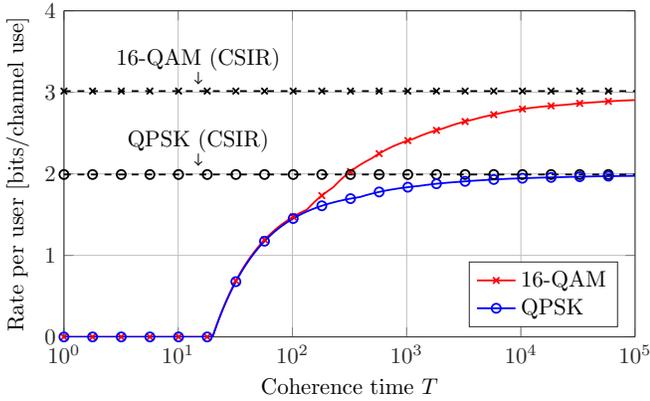
Fig. 5. Per-user achievable rate with LS estimation and MRC per user as a function of T; $N = 400$, $\rho = -10$ dB, $K = 20$; the number of pilots P is optimized for each value of T.
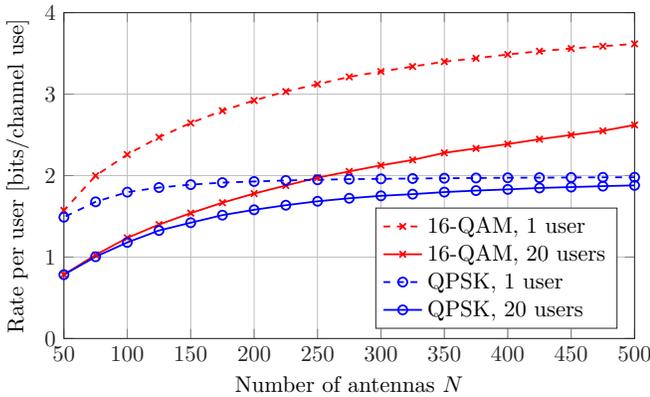

Fig. 6. Per-user achievable rates as a function of the number of antennas; $T = 1000$, $\rho = -10$ dB; the number of pilots P is optimized for each value of N.

to a circle (cf. Fig. 3c). Note also that, when QPSK is used, the difference in achievable rate between the case $K = 1$ and $K = 20$ is marginal—an observation already reported in [9]. On the contrary, the difference in achievable rates between single- and multi-user case is more pronounced when 16-QAM is used. This suggests that, with 16-QAM, the system becomes interference limited, and that the one-bit quantizers partly destroy the orthogonality between the fading channels associated with different users.

In Fig. 5, we plot the per-user achievable rates as a function of T for $\rho = -10$ dB, $N = 400$, and $K = 20$. The number of pilot symbols is again optimized for each value of T. We also depict the achievable rates for the perfect receiver-CSI case. Similarly to the SISO case, as T increases, the per-user achievable rates approach the perfect-receive CSI rate. However, this convergence occurs at a much slower pace than for the infinite-precision case (cf. [29], [24]). This suggests that the one-bit ADC architecture may be unsuitable for high-mobility scenarios. Note also that the achievable rate is zero when $T \leq 20$. In fact, when orthogonal pilot sequences are transmitted, at least 20 pilot symbols are required when $K = 20$.

Finally, in Fig. 6 we plot the per-user achievable rates as

a function of the number of antennas. Here, $\rho = -10$ dB, and $T = 1000$. As in the previous cases, the number of pilot symbols is optimized for each value of N. We note that 16-QAM outperforms QPSK also when the number of receive antennas is much smaller than 400. We note also that, when QPSK is used, the achievable rate saturate rapidly as the number of receive antennas is increased.

V. CONCLUSIONS

We have analyzed the performance of a one-bit quantized receiver architecture operating over a Rayleigh block-fading channel whose realizations are not known *a priori* to transmitter and receiver. We have demonstrated that, for the SISO case, a signaling scheme based on LS estimation and JPD processing is capacity achieving. For the one-bit massive MIMO case, we have shown that, in contrast to the SISO case, high-order constellations can be used to convey information at higher rates than with QPSK. This holds in spite of the nonlinearity introduced by the one-bit quantizer and in spite of the multiuser interference. Similar results hold for the case when zero-forcing instead of MRC is used (see [26]). Note also that constellations that are optimized for the nonlinearity introduced by the one-bit quantizers may yield higher achievable rates than the 16-QAM constellation analyzed in this paper. Extension of our analysis to the case when the ADCs have 2 or 3 bits of resolution is also of interest.

APPENDIX A
PROOF OF (12)

Both the pilot and the data symbols are assumed to belong to a QPSK constellation. By symmetry, the rates achievable on the SISO channel (5) with QPSK inputs are twice as high as the rates achievable with BPSK. Hence, in the remainder of the proof, we shall consider a real-valued version of the channel input-output relation (5), where h and $\mathbf{w}$ are real Gaussian, and the input vector $\mathbf{x}$ consists of BPSK symbols. Let ℓ denote the number of sign mismatches between the BPSK vectors $\mathbf{x}^{(p)}$ and the quantized received vector $\mathbf{r}^{(p)}$. Since for the real case there exists a one-to-one relation between the LS estimate $\hat{h}$ in (10) and ℓ, we conclude that $I(x; r \mid \hat{h}) = I(x; r \mid \ell)$. To evaluate this mutual information, we need the conditional probability mass function $p_{r|x,\ell}$, which can be expressed as follows

$$p_{r|x,\ell}(r|x,\ell) = \frac{\mathbb{E}_h\big[p_{\ell\,|\,h}(\ell\,|\,h)\Phi(rhx)\big]}{p_\ell(\ell)}. \tag{15}$$

Using a similar approach as the one detailed in [14], one can show that

$$p_{\ell\,|\,h}(\ell\,|\,h) = \binom{P}{\ell}\Phi(-h\sqrt{\rho})^\ell\Phi(h\sqrt{\rho})^{P-\ell} \tag{16}$$

and

$$p_l(\ell) = \binom{P}{\ell}\beta(\ell, P - \ell). \tag{17}$$

Substituting (16) and (17) into (15), and then using the definition of mutual information, and that

$$\beta(\ell + 1, P - l) + \beta(\ell, P - \ell + 1) = \beta(\ell, P - \ell) \tag{18}$$

one obtains (12) (see [26, App. B] for details).

APPENDIX B
PROOF OF THEOREM 2

Our JPD processing lower bound is based on the following scheme. The first transmit symbol in each coherence block is a pilot symbol. To decode the nth symbol in the block, we rely on the LS channel estimate obtained on the basis of the past $n-1$ symbols (1 pilot symbol and $n-2$ data symbols). This scheme yields the following achievable rate:

$$R_{\mathrm{JPD}}^{(T)}(\rho) = \frac{1}{T}\sum_{n=2}^{T} I\big(x_n; r_n \,|\, \hat{h}(n-1)\big). \tag{19}$$

Here, we have indicated the channel estimate by $\hat{h}(n-1)$ to clarify the number of input symbols that are used to estimate the channel. As in Appendix A, it is sufficient to focus on a real-valued version of the channel input-output relation (5), where h and $\mathbf{w}$ are real Gaussian, and the input vector $\mathbf{x}$ consists of BPSK symbols.

To establish Theorem 2, it is then sufficient to show that (19) coincides with

$$R_{\mathrm{BPSK}}^{(T)}(\rho) = 1 + \frac{1}{T}\sum_{k=0}^{T}\binom{T}{k}\beta(k, T-k)\log_2\beta(k, T-k). \tag{20}$$

The final result (6) is then established by replacing BPSK with on-off BPSK. See [14] for the details. Our proof that (19) coincides with (20) is by induction. We start by noting that when $T = 2$, (19) coincides with the RHS of (12). Equality between (19) and (20) for this case then follows from Lemma 1.

We now assume that JPD processing achieves (20) for a given coherence time T. We need to prove that the same holds when the coherence time is $T+1$. Note that

$$R_{\mathrm{JPD}}^{(T+1)}(\rho) = \frac{1}{T+1}\Big[TR_{\mathrm{JPD}}^{(T)}(\rho) + I\big(x_{T+1}; r_{T+1}\,|\,\hat{h}(T)\big)\Big]. \tag{21}$$

By the induction hypothesis, we can replace $R_{\mathrm{JPD}}^{(T)}(\rho)$ by $R_{\mathrm{BPSK}}^{(T)}(\rho)$. Furthermore, we can replace the mutual information on the RHS of (21) with the RHS of (12), evaluated for $P = T$. The desired result then follows by using (18), the following binomial equality

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \tag{22}$$

and by performing simple algebraic manipulations.

REFERENCES

[1] R. H. Walden, "Analog-to-digital converter technology comparison," in *IEEE GaAs IC Symposium Technical Digest*, Philadelphia, PA, Oct. 1994, pp. 217–219.

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

[3] B. Murmann, "ADC Performance Survey 1997-2014." [Online]. Available: http://web.stanford.edu/~murmann/adcsurvey.html

[4] S. Hoyos, B. Sadler, and G. Arce, "Monobit digital receivers for ultraw-ideband communications," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1337–1344, Jul. 2005.

[5] A. Mezghani and J. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Nice, France, Jun. 2007, pp. 1286–1289.

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

[7] J. Mo and R. W. Heath Jr, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," Oct. 2014. [Online]. Available: http://arxiv.org/abs/1410.7353

[8] A. Ulusoy et al., "A 60 GHz multi-Gb/s system demonstrator utilizing analog synchronization and 1-bit data conversion," in *IEEE Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)*, Austin, TX, Jan. 2013, pp. 99–101.

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

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

[11] T. Koch and A. Lapidoth, "At low SNR, asymmetric quantizers are better," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5421–5445, Sep. 2013.

[12] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[13] S. Krone and G. Fettweis, "Fading channels with 1-bit output quantization: Optimal modulation, ergodic capacity and outage probability," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Dublin, Ireland, Aug. 2010.

[14] A. Mezghani and J. Nossek, "Analysis of Rayleigh-fading channels with 1-bit quantized output," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Toronto, ON, Canada, Jul. 2008, pp. 260–264.

[15] S. Shamai (Shitz), "Information rates by oversampling the sign of a bandlimited process," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1230–1236, Jun. 1994.

[16] T. Koch and A. Lapidoth, "Increased capacity per unit-cost by oversampling," in *IEEE 26th Convention of Electrical and Electronics Engineers in Israel (IEEEI),*, Eliat, Israel, Nov. 2010, pp. 684 –688.

[17] S. Krone and G. Fettweis, "Capacity of communications channels with 1-bit quantization and oversampling at the receiver," in *IEEE Sarnoff Symp. (SARNOFF)*, Newark, NJ, May 2012.

[18] M. Coldrey and P. Bohlin, "Training-based mimo systems: Part ii–improvements using detected symbol information," *IEEE Trans. Signal Process.*, vol. 56, no. 1, pp. 296–303, Jan. 2008.

[19] N. Jindal, A. Lozano, and T. Marzetta, "What is the value of joint processing of pilots and data in block-fading channels?" in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun. 2009, pp. 2189–2193.

[20] M. Dorpinghaus, A. Ispas, and H. Meyr, "On the gain of joint processing of pilot and data symbols in stationary rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2963–2982, May 2012.

[21] A. Zymnis, S. Boyd, and E. Candés, "Compressed sensing with quantized measurements," *IEEE Signal Process. Lett.*, vol. 17, no. 2, pp. 149–152, Feb. 2010.

[22] C. Studer and G. Durisi, "Quantized massive MIMO-OFDM uplink," 2015, in preparation.

[23] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.

[24] R. Devassy, G. Durisi, J. Östman, W. Yang, T. Eftimov, and Z. Utkovski, "Finite-SNR bounds on the sum-rate capacity of Rayleigh block-fading multiple-access channels with no a priori CSI," Dec. 2014. [Online]. Available: http://arxiv.org/abs/1501.01957

[25] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[26] S. Jacobsson, "Throughput analysis of massive MIMO uplink with one-bit ADCs," Master's thesis, Chalmers University of Technology, Gothenburg, Sweden, Jan. 2015.

[27] C.-K. Wen, S. Jin, K.-K. Wong, C.-J. Wang, and G. Wu, "Joint channel-and-data estimation for large-MIMO systems with low-precision ADCs," Jan. 2015. [Online]. Available: http://arxiv.org/abs/1501.05580

[28] D. Arnold, H.-A. Loeliger, P. Vontobel, A. Kavcic, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

[29] F. Rusek, A. Lozano, and N. Jindal, "Mutual information of IID complex Gaussian signals on block Rayleigh-faded channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 331–340, Jan. 2012.