

 Open access • Proceedings Article • DOI:10.1109/ICASSP.2017.7953405

Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink — [Source link](#)

A.L. Swindlehurst, Amodh Kant Saxena, Amine Mezghani, Inbar Fijalkow

Institutions: University of California, Irvine, Cergy-Pontoise University

Published on: 05 Mar 2017 - International Conference on Acoustics, Speech, and Signal Processing

Topics: Precoding, Quantization (signal processing), MIMO, Phase-shift keying and Telecommunications link

Related papers:

- [Quantized Precoding for Massive MU-MIMO](#)
- [Transmit processing with low resolution D/A-converters](#)
- [Analysis of One-Bit Quantized Precoding for the Multiuser Massive MIMO Downlink](#)
- [Minimum BER precoding in 1-Bit massive MIMO systems](#)
- [1-bit Massive MU-MIMO Precoding in VLSI](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/minimum-probability-of-error-perturbation-precoding-for-the-3eunrelfru>



HAL
open science

Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink

A Swindlehurst, A Saxena, A Mezghani, I Fijalkow

► **To cite this version:**

A Swindlehurst, A Saxena, A Mezghani, I Fijalkow. Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink. IEEE ICASSP, Mar 2017, New Orleans, United States. pp.6483 - 6487, 10.1109/ICASSP.2017.7953405 . hal-01549575

HAL Id: hal-01549575

<https://hal.archives-ouvertes.fr/hal-01549575>

Submitted on 28 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MINIMUM PROBABILITY-OF-ERROR PERTURBATION PRECODING FOR THE ONE-BIT MASSIVE MIMO DOWNLINK

A. Swindlehurst, A. Saxena, A. Mezghani*

I. Fijalkow

Center for Pervasive Communications and Computing
University of California Irvine
Irvine, CA 92697, USA
{swindle,aksaxena,amezghan}@uci.edu

ETIS, UMR 8051 / ENSEA
Université Cergy-Pontoise, CNRS
F-95000 Cergy, France
inbar.fijalkow@ensea.fr

ABSTRACT

Linear precoders have been shown to perform reasonably well at low SNR when the basestation of a MIMO downlink employs one-bit digital-to-analog converters to quantize the precoder outputs. However, at medium-to-high SNRs, an error floor is encountered due to the coarse quantization. This paper examines methods for slightly perturbing the transmitted signal prior to quantization in an effort to improve downlink performance at higher SNRs. The perturbation is performed with the goal of minimizing the worst-case probability of error among the user terminals, and assumes that the symbols to be transmitted are drawn from a finite alphabet constellation. Two different types of perturbations are studied, and it is found via simulation that the methods can provide dramatic gains in downlink performance.

Index Terms— one-bit quantization, massive MIMO, linear precoding, perturbation precoding, minimum probability of error precoding

1. INTRODUCTION

The use of one-bit quantization has been proposed to reduce the cost, complexity and energy consumption associated with massive MIMO systems. Rather than reducing the number of RF chains and decreasing beamforming flexibility as in hybrid digital-analog approaches [1]-[6], one-bit quantization maintains one RF chain per antenna and reduces the cost per RF chain by replacing each high-precision analog-to-digital converter (ADC) or digital-to-analog converter (DAC) with a simple one-bit implementation. The savings are considerable since energy consumption and cost associated with the ADCs and DACs grow exponentially with the number of quantization bits, and other RF components can be simplified as well. This is especially important for the downlink, where a one-bit

DAC obviates the need for linear amplifiers operating with a high back-off and hence low energy efficiency.

Most research on one-bit quantization for MIMO systems has focused on the uplink (e.g., see [7]-[12]). The downlink case, with one-bit DACs implemented on each RF chain, has received considerably less attention [13, 14, 15, 16]. The impact of one-bit quantization on the performance of linear precoding was studied in [15], and it was shown that the signal-to-quantization-and-interference ratio at the receivers grows approximately linearly with the ratio of the number of antennas to the number of users. Using one-bit DACs at the transmitter is equivalent to restricting the transmitted signal from each antenna to be QPSK. This observation was exploited in [16] to directly design the transmitted signal for each possible combination of symbols desired at the user terminals. Rather than using a linear precoder, a nonlinear minimum probability-of-error criterion was used to directly determine the QPSK symbols for each antenna.

A key observation in much of the work for both the uplink and downlink is that at low signal-to-noise ratio (SNR), the loss in performance due to one-bit quantization is relatively small, only about 2dB (a factor of $2/\pi$) in most cases. This is encouraging since the low SNR regime is a likely operating point for massive MIMO systems that take advantage of array or beamforming gain to make up for low signal power. However, it is still of interest to study methods that minimize the impact of one-bit quantization for medium and higher SNRs.

In this paper we focus on making linear precoding methods more effective in conjunction with one-bit quantization by perturbing the linearly precoded signal to favorably impact the probability of decoding error at the user terminals. We assume that the basestation is transmitting symbols drawn from a finite alphabet, and hence that each user needs only to make a correct detection of its symbol rather than receive something “close” to the symbol itself. With channel state information available at the transmitter, we can determine in advance the effect of the one-bit quantization at each user, and thus we choose an appropriate perturbation that increases the minimum distance to a decision boundary over all the users.

*This work was supported by the National Science Foundation under ECCS-1547155, and by the Technische Universität München Institute for Advanced Study, funded by the German Excellence Initiative and the European Union Seventh Framework Programme under Grant No. 291763, and by the European Union under the Marie Curie COFUND Program.

Instead of searching for an optimal perturbation which would be prohibitively expensive, we show via simulation that making a relatively small and fixed number of perturbations to improve the performance of the worst-case user can lead to a significant improvement in overall system performance.

The next section introduces the data model assumed in the paper and the minimum symbol error probability criterion that we will use to update the perturbation for the transmitted symbol. Section 3 presents the perturbation methods considered in this work and how they are computed, and then simulation results are given in Section 4 to illustrate the performance gains compared with a transmitter with full-precision quantization and one with no perturbation.

2. QUANTIZED LINEAR PRECODING

2.1. Data Model and Assumptions

To illustrate the basic idea of our approach, we assume a flat fading downlink channel involving an M -antenna basestation (BS) and K single antenna user terminals. We will focus on a massive MIMO scenario in which $M \gg K$, although this is not strictly necessary for the proposed methods. We make no assumptions regarding the $K \times M$ channel matrix \mathbf{H} other than it is generically full rank. We let the M -element vector \mathbf{x} represent the transmitted signal at the BS, so that the downlink data model can be represented as

$$\hat{\mathbf{r}} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where the k -th element of $\hat{\mathbf{r}}$, denoted by \hat{r}_k , represents the signal received by user k , and \mathbf{n} is a vector whose K elements are independent identically distributed (i.i.d.) Gaussian noise. We will denote $\mathbf{r} = \mathbf{H}\mathbf{x}$ as the noise-free version of $\hat{\mathbf{r}}$, and we will let H_{km} denote the element of \mathbf{H} in row k and column m .

In [15], it is assumed that the vector \mathbf{x} is the one-bit quantized output of a linear precoder:

$$\mathbf{x} = \mathbb{Q}(\mathbf{P}\mathbf{s}), \quad (2)$$

where \mathbf{s} is the desired K -vector of symbols to be decoded at the users, \mathbf{P} is an $M \times K$ precoder, and the one-bit quantizer $\mathbb{Q}(\cdot)$ operates element-wise on its argument:

$$[\mathbb{Q}(\mathbf{z})]_m = \frac{1}{\sqrt{2}} (\text{sign}(\text{Re}\{z_m\}) + j\text{sign}(\text{Im}\{z_m\})),$$

where $\text{sign}(\cdot)$ returns the sign of the argument, and $\text{Re}\{\cdot\}$ and $\text{Im}\{\cdot\}$ represent the real and imaginary parts of their arguments, respectively. As explained below, in our approach we will perturb the signal in (2) in order to improve the probability of correctly detecting \mathbf{s} at the user terminals. We will focus on perturbed versions of the maximal ratio transmission (MRT) scheme in which $\mathbf{P} = \mathbf{H}^H$, and the zero-forcing (ZF) precoder where $\mathbf{P} = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H)^{-1}$. In order to focus on

the impact of the one-bit quantization, we assume in this paper that the channel matrix \mathbf{H} is known at the transmitter. In practice the channel will be subject to estimation errors that will lead to degraded detection performance.

2.2. Minimum Probability of Error Precoding

We assume that the symbols that the BS desires to send to the users are drawn from a finite alphabet with defined symbol decision boundaries. Each user compares its received signal \hat{r}_k with these decision boundaries and takes the closest element of the alphabet \hat{s}_k as the symbol that was transmitted to it. As such, it is often not necessary that r_k be “close” to s_k in any sense, just that r_k lie somewhere within the decision region associated with the desired symbol s_k . In fact, for some constellations (e.g., PSK), a lower probability of error is obtained when r_k moves farther away from s_k such that its distance to the decision boundaries is larger.

For example, consider the case of QPSK symbols depicted in Figure 1. In this figure, the desired symbol for user k is $1 + j$ in the upper-right quadrant, but any received signal \hat{r}_k that lies in the upper-right quadrant will be decoded as the correct s_k , and the farther the received signal from the decision boundary, the lower the error probability. Received signals outside the correct quadrant have a positive distance to the boundary indicated by the lines with arrows, those inside the quadrant have a negative distance. While this distance is specific to a single user, its value will not only depend on s_k , but also on the channel and on the transmitted signal \mathbf{x} , which in turn will depend on the other symbols \mathbf{s} . Thus, we will let the function $d_k(\mathbf{x}|\mathbf{s}, \mathbf{H})$, more concisely d_k , represent the distance of \hat{r}_k to the nearest decision boundary for user k .

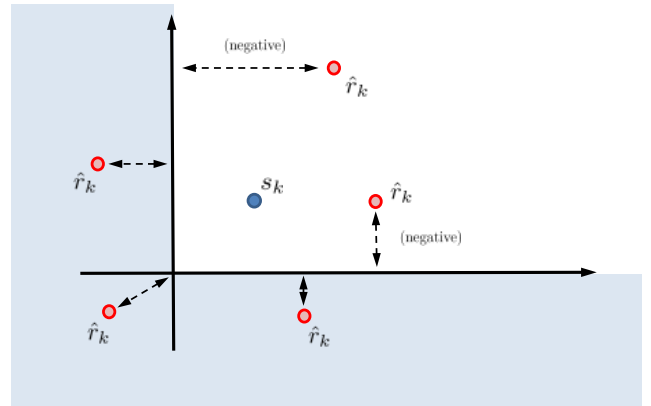


Fig. 1. Distance metric for minimizing probability of decoding error for symbol s_k ; the distance is shown for several possible received signals \hat{r}_k .

As discussed in [16], there are several ways to aggregate these distances to provide a single figure of merit for the downlink. These include minimizing the sum $\sum_k d_k$, prod-

uct $\prod_k d_k$, or (as we focus on in this paper) minimizing the maximum distance over all users:

$$\mathbf{x} = \arg \min_{\mathbf{x}} \max_k d_k(\mathbf{x}|\mathbf{s}, \mathbf{H}) . \quad (3)$$

Solving (3) over all 4^M possible \mathbf{x} is prohibitively expensive, so in this paper we focus on choosing \mathbf{x} to be a perturbed version of a quantized linear precoder. We will consider the two general classes of perturbations outlined below.

Direct Perturbation (DP) – Here we choose $\mathbf{x} = \mathbb{Q}(\mathbf{P}\mathbf{s}) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is an M -element perturbation vector that must be constrained such that the elements of \mathbf{x} are restricted to be $\pm 1 \pm j$. The idea is that by changing the symbols transmitted by a small number of the antennas, received signals near decision boundaries can be made more robust without affecting decisions made by other users.

Symbol Perturbation (SP) – In this case, the perturbation is made to the symbols: $\mathbf{x} = \mathbb{Q}(\mathbf{P}(\mathbf{s} + \boldsymbol{\epsilon}))$, to reflect the idea that it is not necessary to force recovery of the exact symbols \mathbf{s} , just received signals within the correct decision regions. A small perturbation to one of the user symbols can help move the received signal for another user deeper into the decision region.

3. PERTURBATION PRECODING

In this section we present the details of the two approaches mentioned above. Even though these approaches work with a constrained set of possible values for \mathbf{x} , for large M it is still unlikely that an optimization such as (3) can be implemented over all possible perturbations $\boldsymbol{\epsilon}$. Furthermore, the quantization non-linearity makes implementation of gradient-based methods problematic. Instead, we will show that simply considering a small constrained set of $\boldsymbol{\epsilon}$ can yield a significant benefit rather than optimizing over all possible values of $\boldsymbol{\epsilon}$.

3.1. Direct Perturbation Precoding

The DP approach takes an initial quantized transmit vector \mathbf{x} and strategically perturbs its elements to improve the performance of the worst-case user. The algorithm operates at each step by finding the user with the maximum distance d_k , and then finding the element of \mathbf{x} that makes the largest contribution to d_k . Since

$$r_k = \sum_{m=1}^M H_{km} x_m ,$$

the “worst” element of \mathbf{x} from the point of view of user k is the one for which $H_{km} x_m$ is the farthest from the decision boundary, or in other words the one for which $d_k(x_m|s_k, H_{km})$ is largest. This element of \mathbf{x} is thus adjusted to one of the other three possible QPSK symbols if that leads to a reduction in the distance for that user; otherwise it

remains unchanged. This process is repeated a fixed number of times, with only one perturbation allowed per element of \mathbf{x} . Mathematically, the implementation of the algorithm can be described as follows:

1. Initialize $p = 0$, number of perturbations N_p , the index set $\mathcal{M} = \emptyset$ and find an initial $\mathbf{x}^{(0)}$ using (2) for some linear precoder \mathbf{P} .
2. Find $k^{(p)} = \arg \max_k d_k(\mathbf{x}^{(p)}|\mathbf{s}, \mathbf{H})$
3. Find $m^{(p)} = \arg \max_{m \notin \mathcal{M}} d_{k^{(p)}}(x_m|s_k, H_{km})$
4. Add $m^{(p)}$ to \mathcal{M}
5. Set element $m^{(p)}$ of $\mathbf{x}^{(p)}$ equal to the value $\pm 1 \pm j$ that minimizes $\max_k d_k(\mathbf{x}^{(p)}|\mathbf{s}, \mathbf{H})$
6. $p \rightarrow p + 1$
7. if $p = N_p$, set $\mathbf{x} = \mathbf{x}^{(p)}$ and end; otherwise return to step 2.

Since it is relatively simple to update $d_k(\mathbf{x}^{(p)}|\mathbf{s}, \mathbf{H})$ when only one element of $\mathbf{x}^{(p)}$ has changed, for large M most of the computational burden comes in step 3, where M separate values have to be checked for the worst-case contribution to $r_{k^{(p)}}$. The complexity of the algorithm can be controlled by choosing the maximum number of potential perturbations N_p that are allowed.

3.2. Symbol Perturbation Precoding

Instead of directly perturbing the elements of \mathbf{x} , the SP approach perturbs the values of the K symbols \mathbf{s} that are used in the linear precoder. Also, rather than perturbing the symbols one at a time as we did with \mathbf{x} , we choose a set of random perturbations $\{\boldsymbol{\epsilon}_p, p = 1, \dots, N_p\}$, and then select the $\boldsymbol{\epsilon}_p$ that results in the smallest value for d_k . There are a number of possible perturbations that can be used; we have found that the following works well:

$$\boldsymbol{\epsilon}_k = c(u_k + jv_k) , \quad (4)$$

where c is a constant gain and u_k and v_k are uniform random variables on $[0, 1]$. The details of the algorithm are given in the following steps:

1. Initialize $p = 1$, the “size” c of the perturbation, the number of perturbations N_p , and find an initial $\mathbf{x}^{(0)}$ using (2) for some linear precoder \mathbf{P} . Calculate $\delta(0) = \max_k d_k(\mathbf{x}^{(0)}|\mathbf{s}, \mathbf{H})$.
2. Generate $\boldsymbol{\epsilon}^{(p)}$ according to (4).
3. Calculate $\mathbf{x}^{(p)} = \mathbb{Q}(\mathbf{P}(\mathbf{s} + \boldsymbol{\epsilon}^{(p)}))$
4. Calculate $\delta(p) = \max_k d_k(\mathbf{x}^{(p)}|\mathbf{s}, \mathbf{H})$
5. $p \rightarrow p + 1$
6. if $p < N_p$, return to step 2.
7. Find $p^* = \arg \min_p \delta(p)$ and set $\mathbf{x} = \mathbf{x}^{(p^*)}$.

While this method eliminates the search over M elements for each p , it requires the computation of $\mathbf{P}(\mathbf{s} + \boldsymbol{\epsilon}^{(p)})$ and $\mathbf{H}\mathbf{x}^{(p)}$ for each perturbation. Again, the computational load can be controlled by the choice of N_p .

4. NUMERICAL RESULTS

For the numerical examples presented here, we assume a basestation with $M = 128$ antennas serving $K = 8$ single-antenna users. The channel \mathbf{H} is composed of circular complex independent identically distributed (c.c.i.i.d) Gaussian random variables with zero mean and variance ρ . The additive noise at the receivers is also c.c.i.i.d Gaussian with zero mean and unit variance, and thus ρ represents the per-antenna SNR. The elements of the signal vector \mathbf{s} are independent QPSK symbols. Our performance metric is the symbol error rate (SER) averaged over the $K = 8$ users.

Figure 2 shows the average SER achieved by several algorithms based on the MRT precoder. The curve labeled MRT refers to the MRT precoder with a full-precision DAC at the transmitter, Q-MRT refers to the quantized MRT precoder with no perturbation, while Q-MRT SP and Q-MRT DP refer to the direct- and symbol-perturbed versions of Q-MRT. Figure 3 shows the same set of results, but for algorithms based on the ZF precoder. The DP approach was implemented with 25 perturbations to the elements of the original \mathbf{x} achieved by either Q-MRT or Q-ZF, and the SP approach used 25 different random perturbation vectors with $c = 0.25$ and at each trial chose the one that resulted in the best performance for the worst-case user.

We see that the perturbation precoders provide a gain in SER when the SNR is 0dB or higher, and the gain becomes more and more significant as the SNR increases. The SP version of Q-MRT eventually achieves the same performance as the unquantized MRT, and the DP version dramatically outperforms the unquantized MRT at high SNR. Both the SP and DP approaches perform similarly for ZF, and come within 3-4dB of the unquantized ZF precoder. Figure 4 shows how performance improves with the number of perturbations considered for the Q-MRT approach with DP, indicating that even better performance may be obtained if $N_p > 25$ is possible.

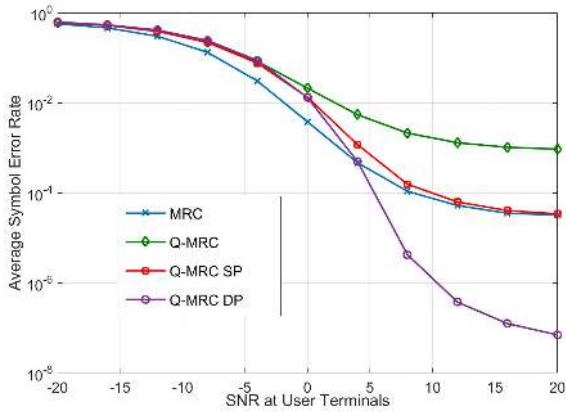


Fig. 2. Average SER performance for MRT algorithms.

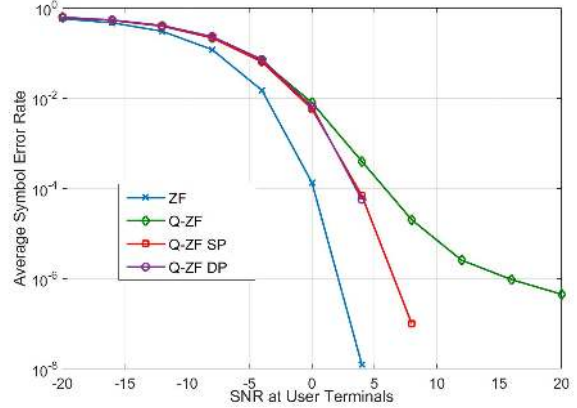


Fig. 3. Average SER performance for ZF algorithms.

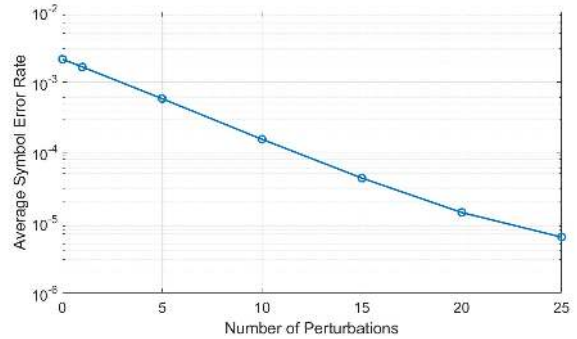


Fig. 4. Performance improvement with N_p for Q-MRT DP.

5. CONCLUSIONS

This paper has presented methods for improving the performance of a massive MIMO downlink when the output of standard linear precoders is quantized by one-bit DACs at the basestation transmitter. The methods exploit the fact that the impact of the one-bit DACs can be determined prior to transmission, and they operate by making perturbations to the quantized downlink signal prior to transmission in order to improve the worst-case performance over the user terminals. Taking advantage of the assumption that the transmitted symbols are drawn from a finite alphabet, a performance metric for the perturbations is introduced that attempts to directly minimize the probability of decoding error, rather than for example minimizing the distance to the desired symbols. Simulations demonstrate that these perturbation approaches can achieve a dramatic improvement in performance at medium to high SNRs; in fact, the DP approach allows the quantized MRT precoder to significantly outperform standard MRT employing full precision DACs.

6. REFERENCES

- [1] S. Han, C. I. I, Z. Xu, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, January 2015.
- [2] W. Roh, J. Y. Seol, J. Park, B. Lee, J. Lee, Y. Kim, J. Cho, K. Cheun, and F. Aryanfar, "Millimeter-wave beamforming as an enabling technology for 5G cellular communications: Theoretical feasibility and prototype results," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 106–113, February 2014.
- [3] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, Oct 2014.
- [4] A. Alkhateeb, G. Leus, and R. W. Heath, "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Transactions on Wireless Communications*, vol. 14, no. 11, pp. 6481–6494, Nov 2015.
- [5] X. Gao, L. Dai, S. Han, C. L. I, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for mmwave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, April 2016.
- [6] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, "Mimo precoding and combining solutions for millimeter-wave systems," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 122–131, December 2014.
- [7] N. Liang and W. Zhang, "A mixed-ADC receiver architecture for massive MIMO systems," in *Information Theory Workshop - Fall (ITW), 2015 IEEE*, Oct 2015, pp. 229–233.
- [8] S. Jacobsson, G. Durisi, M. Coldrey, U. Gustavsson, and C. Studer, "One-bit massive MIMO: Channel estimation and high-order modulations," in *2015 IEEE International Conference on Communication Workshop (ICCW)*, June 2015, pp. 1304–1309.
- [9] Yongzhi Li, Cheng Tao, Liu Liu, Gonzalo Seco-Granados, and A. Swindlehurst, "Channel estimation and uplink achievable rates in one-bit massive MIMO systems," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM) (SAM 2016)*, Rio de Janeiro, Brazil, July 2016.
- [10] J. Mo and R. W. Heath, "Capacity analysis of one-bit quantized MIMO systems with transmitter channel state information," *IEEE Transactions on Signal Processing*, vol. 63, no. 20, pp. 5498–5512, Oct 2015.
- [11] A. Mezghani and J. A. Nossek, "On ultra-wideband MIMO systems with 1-bit quantized outputs: Performance analysis and input optimization," in *2007 IEEE International Symposium on Information Theory*, June 2007, pp. 1286–1289.
- [12] Amine Mezghani and Josef A Nossek, "Capacity lower bound of MIMO channels with output quantization and correlated noise," in *IEEE International Symposium on Information Theory Proceedings (ISIT)*, 2012.
- [13] A. Mezghani, R. Ghiat, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *Proc. 16th IEEE International Conference on Electronics, Circuits, and Systems, 2009.*, Dec 2009, pp. 683–686.
- [14] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE precoder for massive MIMO using 1-bit quantization," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 3381–3385.
- [15] A. K. Saxena, I. Fijalkow, and A. Swindlehurst, "On one-bit quantized ZF precoding for the multiuser massive MIMO downlink," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM) (SAM 2016)*, Rio de Janeiro, Brazil, July 2016.
- [16] H. Jedda, J. A. Nossek, and A. Mezghani, "Minimum BER precoding in 1-bit massive MIMO systems," in *2016 IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM) (SAM 2016)*, Rio de Janeiro, Brazil, July 2016.