

## SPARSE LINEAR PRECODERS FOR MITIGATING NONLINEARITIES IN MASSIVE MIMO

Amine Mezghani<sup>1</sup>, Daniel Plabst<sup>4</sup>, Lee A. Swindlehurst<sup>2</sup>, Inbar Fijalkow<sup>3</sup>, Josef A. Nossek<sup>4</sup><sup>1</sup>University of Manitoba, MB, Canada; <sup>3</sup>ETIS, CY Cergy Paris Univ, ENSEA, CNRS, Paris, France  
<sup>2</sup>University of California, Irvine, CA, USA; <sup>4</sup>Technische Universität München, Germany

## ABSTRACT

Dealing with nonlinear effects of the radio-frequency (RF) chain is a key issue in the realization of very large-scale multi-antenna (MIMO) systems. Achieving the remarkable gains possible with massive MIMO requires that the signal processing algorithms systematically take into account these effects. Here, we present a computationally-efficient linear precoding method satisfying the requirements for low peak-to-average power ratio (PAPR) and low-resolution D/A-converters (DACs). The method is based on a sparse regularization of the precoding matrix and offers advantages in terms of precoded signal PAPR as well as processing complexity. Through simulation, we find that the method substantially improves conventional linear precoders.

## 1. INTRODUCTION

The use of large-scale multiple antenna systems, known as Massive MIMO, has emerged as a leading concept for advancing wireless communications infrastructure. The design of such systems, however, comes with inherent challenges in terms of analog hardware and computational complexity. The rapidly growing pace of scale in antenna systems for communications necessitates methods to reduce the RF hardware complexity and power consumption, such as reducing the resolution of the data converters and operating the power-amplifiers with constant-envelope signals. In particular, the power amplifier (PA) typically accounts for a substantial amount of the power consumption in a base station (BS) [1]. Therefore, it is desirable in terms of power efficiency to run the PA in the saturation region where most of the available PA power is radiated and less power is lost as heat.

Operating PAs in the saturation region and using low-resolution D/A-converters (DACs), however, implies high distortions and nonlinearities that are introduced to the signals, calling for new processing techniques to mitigate their effects. In [2], the impact of power amplifier nonlinearity has been analyzed and it was shown that both single-carrier (SC) and OFDM transmission are affected similarly if conventional linear precoding is used since the peak-to-average power ratio (PAPR) increases with multi-user processing [3]. Standard and improved linear zero-forcing and minimum-mean-squared error (MMSE) precoder designs for (massive)

MIMO with low-resolution DACs and their performance have been studied in [4, 5, 6, 7, 8, 9], showing satisfactory performance for small loading factors and well-conditioned channels such as i.i.d. channels. Achievable rates for multi-user MISO systems with low-resolution D/A converters were considered in [10]. Nonlinear *Tomlinson-Harashima Precoding* has been considered in [11] for low-resolution DACs, and achieves better performance than pure linear methods. Other works such as [12, 13, 14, 15, 16, 17, 18] show that nonlinear symbol-by-symbol precoding schemes outperform linear precoders under hardware impairments at the cost of an increased computational complexity.

Motivated by the simplicity of linear methods, we reconsider the problem of optimizing linear precoding techniques and present a new approach where SC transmission can still preserve its advantage in terms of PAPR compared to OFDM despite the multi-user processing. The new precoders are used in particular for mitigating the effects of constant envelope transmission as well as 1-bit DACs. To this end, a new cost functions with  $\ell_1$  regularization is investigated. The new class of linear precoders based on  $\ell_1$  matrix norm optimization significantly outperforms conventional designs for QPSK as well as higher-order modulations and allows for reduced computational complexity.

## 2. SYSTEM MODEL

Consider a downlink system with nonlinear transmitters:

$$\hat{s} = \mathbf{H} \cdot \mathcal{Q}(\mathbf{P} \cdot \mathbf{s}) + \boldsymbol{\eta}, \quad (1)$$

where  $\mathbf{H} \in \mathbb{C}^{K \times N}$  is the channel matrix with  $K$  single-antenna users and  $N$  transmit antenna, while  $\mathbf{P} \in \mathbb{C}^{N \times K}$  is the linear precoder and  $\mathcal{Q}(\cdot)$  represent the transmit nonlinearity. The transmit symbol vector  $\mathbf{s}$  is assumed to be either an OFDM or a single-carrier signal and the noise vector  $\boldsymbol{\eta}$  is i.i.d. circular symmetric complex Gaussian with variance  $\sigma_{\eta}^2$ .

One possible design criterion for  $\mathbf{P}$  is that it be zero-forcing, i.e.,  $\mathbf{H} \cdot \mathbf{P} = \mathbf{I}$  if quantization is ignored. However since usually  $N > K$ , the solution is not unique. Thus the question that arises is what is an appropriate solution for mitigating nonlinearities such as clipping at the power amplifier or one-bit quantization at the DAC. We will focus on two

types of nonlinearities: (i) One-bit DAC quantization:

$$\mathcal{Q}(\mathbf{y}) = \frac{1}{\sqrt{2}}\text{sign}(\text{Re}\{\mathbf{y}\}) + \frac{j}{\sqrt{2}}\text{sign}(\text{Im}\{\mathbf{y}\}). \quad (2)$$

and (ii) constant envelope transmission

$$\mathcal{Q}(\mathbf{y}) = \exp(j\angle \mathbf{y}), \quad (3)$$

where the nonlinear operation is applied element-wise. In both cases, we obtain constant instantaneous power per antenna equal to one. We study in the next section the impact of OFDM or single-carrier processing on the PAPR and how standard linear multiuser precoding techniques affect those signals.

### 3. PAPR COMPARISON BETWEEN SINGLE-CARRIER AND OFDM SYSTEMS

In this section, we show that for the same power spectral density and the same performance under AWGN (with perfect orthogonal transmission), the worst-case peak power in the single-carrier (SC) systems grows only logarithmically with the spectral rolloff (steepness of the spectrum in the transmission band), whereas the worst-case peak power in OFDM grows linearly with the rolloff (i.e., proportional to the number of sub-carriers).

#### 3.1. OFDM vs. block-wise single-carrier transmission

We compare OFDM and block-wise single-carrier modulation (also known as SC-FDMA in LTE) assuming an over-sampling (interpolation) factor of two. In OFDM, a block of  $M$  data symbols (consisting in general of QAM symbols) is mapped to the lower frequency points of a  $2M$ -point IFFT with zero-padding. In the block-wise single-carrier scheme, an additional  $M$ -point FFT is used to precode the data, resulting in a kind of data mapping in the time domain. The single-carrier processing at the transmitter can also be regarded as a cyclic convolution of data blocks with the impulse response of a discrete rectangular filter with a window size of  $M$  out of  $2M$  frequency points,

$$P_k = \begin{cases} \frac{1}{\sqrt{M}} & \text{for } k < \frac{M}{4} \text{ or } 2M - k < \frac{M}{4} - 1 \\ \frac{1}{2\sqrt{M}} & \text{for } k = \frac{M}{4} \text{ or } 2M - k = \frac{M}{4} - 1, \end{cases} \quad (4)$$

and zero elsewhere, with  $0 \leq k < 2M$ . The impulse response of the cyclic pulse shaping filter is given by applying an IFFT

$$p[n] = \begin{cases} 1 & \text{for } n = 0 \\ \frac{\sin(\pi \frac{n}{2})}{M \sin(\pi \frac{n}{2M})} & \text{otherwise} \end{cases}, \quad 0 \leq n < 2M. \quad (5)$$

Both systems have the same spectral density since the additional  $M$ -point FFT in SC-FDMA is a unitary transformation that does not change the second-order statistics. The resulting steepness of the spectral confinement is proportional

to  $M$ . Also, in the AWGN case, both systems provide perfect orthogonality between data symbols and thus achieve the same BER performance. However, we show that these systems have significantly different behavior in terms of PAPR.

#### 3.2. Peak-to-average power analysis

For simplicity, in this brief analysis we assume the input to be BPSK. For OFDM, the maximum peak-to-average-power is  $M$  and is achieved when all the inputs are identical. In the single-carrier case the peak power is achieved at the intermediate time instants between two consecutive symbols, when all symbols add coherently. Therefore, we get the peak power

$$\begin{aligned} \left( \sum_{n=0}^{M-1} |p[2n+1]| \right)^2 &= \left( \sum_{n=0}^{M-1} \left| \frac{\sin(\pi \frac{2n+1}{2})}{M \sin(\pi \frac{2n+1}{2M})} \right| \right)^2 \\ &= \left( \sum_{n=0}^{M-1} \frac{1}{M \sin(\pi \frac{2n+1}{2M})} \right)^2 \\ &\leq \left( \int_{\frac{1}{4M}}^{1-\frac{1}{4M}} \frac{1}{\sin(\pi x)} dx \right)^2 + \frac{1}{M \sin(\frac{\pi}{2M})} \\ &\approx \frac{4}{\pi^2} \left( \log \left( \sin \left( \frac{\pi}{8M} \right) \right) \right)^2 \\ &\approx \frac{4}{\pi^2} \left( \log \left( \frac{\pi}{8M} \right) \right)^2, \end{aligned} \quad (6)$$

where the approximations hold for large  $M$ . Clearly, we see that the peak power of the single-carrier system only increases logarithmically with  $M$ , showing its superiority compared to OFDM in terms of PAPR. Standard linear multiuser precoding, however, diminishes this property due to the central limit theorem [3]. Our goal is to preserve this advantage in massive MIMO using sparse precoding.

### 4. SPARSE PRECODING DESIGN

The most common choice for the zero-forcing precoder minimizes the Frobenius norm of  $\mathbf{P}$ , i.e., the transmit power:

$$\min_{\mathbf{P}} \|\mathbf{P}\|_{2,2}^2 \quad \text{s.t. } \mathbf{H} \cdot \mathbf{P} = \mathbf{I}, \quad (7)$$

where the  $\ell_{p,q}$  matrix norm is defined as

$$\|\mathbf{A}\|_{p,q}^q = \sum_j \left( \sum_i |a_{i,j}|^p \right)^{q/p}. \quad (8)$$

The solution to this problem is famously given by the pseudo-inverse of  $\mathbf{H}$

$$\mathbf{P} = \mathbf{H}^H (\mathbf{H} \mathbf{H}^H)^{-1}, \quad (9)$$

which aims at minimizing the transmit power while fulfilling the zero-forcing requirement, or equivalently it maximizes the

signal-to-noise ratio at the user terminals for a given power constraint. However, this solution results in a dense matrix  $\mathbf{P}$  (even when  $\mathbf{H}$  is sparse), leading to the fact that the distribution of  $\mathbf{P} \cdot \mathbf{s}$  converges element-wise to a Gaussian distribution with high kurtosis. This happens even when  $\mathbf{s}$  is QPSK, resulting in high distortion at the DACs or amplifiers and eventually producing an error floor at high SNR. Therefore, an alternative formulation is presented in the next section.

## 5. NEW APPROACH: FORMULATION BASED ON THE $\ell_{1,2}$ NORM

If the input  $\mathbf{s}$  is a QPSK signal and the desired output of the precoder  $\mathbf{P}\mathbf{s}$  should be constant envelope or even QPSK (in the case of one-bit DACs), then we propose to design the precoder to have sparse rows, in order to preserve the statistical properties of  $\mathbf{s}$  as much as possible after precoding. To this end, we use the  $\ell_{1,2}$  matrix norm

$$\min_{\mathbf{P}} \left\| \mathbf{P}^T \right\|_{1,2}^2 = \sum_i \left( \sum_j |p_{i,j}| \right)^2 \quad \text{s.t. } \mathbf{H} \cdot \mathbf{P} = \mathbf{I}. \quad (10)$$

This problem is convex and is solved efficiently. For constant envelope signaling,  $|p_{i,j}|$  is the Euclidean norm. For the one-bit DAC case, as it is applied separately to the inphase and quadrature components, it is appropriate to use a real valued representation of the channel and to redefine the norm as

$$|p_{i,j}| \stackrel{1\text{-bit}}{=} |\text{Re}\{p_{i,j}\}| + |\text{Im}\{p_{i,j}\}|. \quad (11)$$

Another important advantage of sparse precoding is the reduced computation for the matrix-vector multiplication.

### 5.1. Combined $\ell_{1,2}/\ell_{2,2}$ regularization

As we will see in the simulations, the  $\ell_{2,2}$  formulation performs better at low SNR while the  $\ell_{1,2}$  approach performs better at high SNR. This is because the quantization effects dominate the noise only for high SNR values. This suggests the combination of both norms in a way that the  $\ell_{2,2}$  norm is effective at low SNR while the  $\ell_{1,2}$  norm becomes active at high SNR values, similar to the *elastic net* approach of [19]:

$$\min_{\mathbf{P}} \left\| \mathbf{H}\mathbf{P} - \mathbf{I} \right\|_{2,2}^2 + \frac{1}{2} \left( \frac{K \cdot \sigma_\eta^2}{N} \left\| \mathbf{P}^T \right\|_{2,2}^2 + \lambda \left\| \mathbf{P}^T \right\|_{1,2}^2 \right), \quad (12)$$

where  $\lambda$  is a constant corresponding to the value  $K \cdot \sigma_\eta^2 / N$  at exactly the crossing point of both BER curves (c.f. Fig. 2). This combined  $\ell_{1,2}/\ell_{2,2}$  regularized optimization can be solved iteratively using the iterative shrinkage-thresholding algorithm (ISTA) [20] with an appropriate stepsize  $\mu$

$$\begin{aligned} \mathbf{P}^{\ell+1} &= \exp \left( \text{j} \angle (\mathbf{P}^\ell - \mu \Delta^\ell) \right) \circ \\ &\max \left( \left| \mathbf{P}^\ell - \mu \Delta^\ell \right| - \mu \frac{\lambda}{2} \cdot \left( \sum_k |\mathbf{p}_k^\ell| \right) \cdot \mathbf{1}^T, \mathbf{0} \right), \end{aligned} \quad (13)$$

where  $\circ$  represents the Hadamard product and

$$\Delta^\ell = \left( \mathbf{H}^H \mathbf{H} + \frac{1}{2} \frac{K \cdot \sigma_\eta^2}{N} \cdot \mathbf{I} \right) \mathbf{P}^\ell - \mathbf{H}^H. \quad (14)$$

In the iterative formula (13), the absolute values are taken elementwise (note that for 1-bit we define  $|x + yj| = |x| + |y|$ ) and  $\mathbf{p}_k$  are the column vectors of  $\mathbf{P}$ .

### 5.2. Superposition coding for higher-order modulation

Using higher-order modulation such as 16-QAM for better spectral efficiency while the base station only employs 1-bit DACs at each antenna may seem to be contradictory at first glance. However since the number of antennas is much larger than the number of users, such an idea can be achieved in principle with an appropriate transmit strategy. We propose such a method, still based on linear precoding, which is applicable for square QAM constellations. To this end, we apply the concept of superposition coding [7], to reformulate the problem again based on QPSK symbols that superimpose "over-the-air" through the channel  $\mathbf{H}$  to form the desired QAM constellation at the user terminals. For simplicity, let us consider the 16-QAM case, which can be represented as the sum of two QPSK signals: One least significant symbol (LSB) and one most significant symbol (MSB)

$$\mathbf{s}_{16\text{QAM}} = (\mathbf{I}_K \otimes [1, 2]) \cdot \mathbf{s}_{\text{QPSK}} = \mathbf{\Pi} \cdot \mathbf{s}_{\text{QPSK}}, \quad (15)$$

where  $\mathbf{s}_{\text{QPSK}} \in \{\pm 1 \pm j\}^{2K}$ . Now, instead of designing a precoder matrix  $\mathbf{P}$  to be applied directly to  $\mathbf{s}_{16\text{QAM}}$ , we design a larger precoder  $\mathbf{P} \in \mathbb{C}^{N \times 2K}$  that is applied to  $\mathbf{s}_{\text{QPSK}}$ , and we let the corresponding signals superimpose through the channel and form the desired signal  $\mathbf{s}_{16\text{QAM}}$  at the user terminals. By doing so and additionally encouraging  $\mathbf{P}$  to be sparse, we can produce a binary-like signal at the output of the precoder and mitigate the issue of the 1-bit DACs, while at the receiver side a higher-order modulated signal is generated through the channel.

Unfortunately, with the new constraint  $\mathbf{H}\mathbf{P} = \mathbf{\Pi}$ , the Restricted Isometry Property (RIP) condition is not valid for this problem meaning that the  $\ell_1$  solution might not be sparse as desired. In fact, due to symmetry considerations (same channel for both MSB and LSB bits), the solution will present the same symmetry, i.e., the precoding vector for the most significant bit is just twice the one for the LSB bit. To break this symmetry, we propose to add some small non-convex perturbation of the original problem in order to encourage sparse solutions having different precoding vectors for both bits:

$$\begin{aligned} \min_{\mathbf{P}} & \left\| \mathbf{H}\mathbf{P} - \mathbf{\Pi} \right\|_{2,2}^2 + \\ & \frac{1}{2} \left( \frac{K \cdot \sigma_\eta^2}{N} \left\| \mathbf{P}^T \right\|_{2,2}^2 + \lambda \left( \left\| \mathbf{P}^T \right\|_{1,2}^2 + \text{tr}(\mathbf{P}\mathbf{J}\mathbf{P}^H) \right) \right), \end{aligned} \quad (16)$$

with  $\mathbf{J} = \mathbf{I}_K \otimes (\mathbf{1}\mathbf{1}^T - \mathbf{I}_2)$ . The term  $\text{tr}(\mathbf{P}\mathbf{J}\mathbf{P}^H)$  represents the non-convex perturbation for breaking the symmetry. Again, the iterative shrinkage-thresholding algorithm can be used to solve this optimization (locally).

## 6. SIMULATION RESULTS

We consider first a system with  $N = 30$  antennas,  $K = 5$  users, and QPSK constellations. The channel entries are  $h_{ij} \sim \mathcal{CN}(0, 1)$ . In the simulations, we used a conservative value for the stepsize  $\mu = 0.01$ . Accelerated ISTA versions (such as FISTA) can be also used for significantly faster convergence rate [20]. In Fig. 1, the complementary cumulative distribution function (CCDF) of the output power is shown for the new precoding technique with  $\ell_1$  regularization compared to the conventional ZF-technique for different modulation formats. For comparison, the CCDF without precoding is also shown. With the new precoder, the PAPR is only increased marginally, while SC-FDMA and RRC pulse shaping still preserve their advantage compared to OFDM.

The BER performance is shown in Fig. 2 and Fig. 3 under constant envelope and 1-bit DAC constraints, respectively. In both cases, we observe substantial improvements at high SNR with sparse precoding. As explained earlier, the elastic net approach with the combined  $\ell_1$  and  $\ell_2$  regularization provides the best performance over the entire SNR range as it optimizes the trade-off between beamforming gain and PAPR reduction.

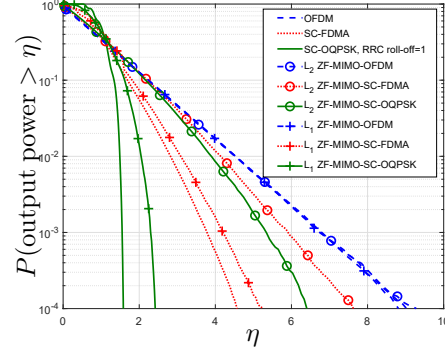
Finally, we consider 16-QAM modulation with superposition coding and 1-bit DACs for  $K = 8$  and  $N = 400$  in Fig. 4. At the receive side, we use the following blind estimation method for the scaling factor for each user prior to detection, as proposed in [21]:

$$f_k = T \cdot \frac{\mathbb{E}[\text{Re}\{s\} + |\text{Im}\{s\}|]}{\sum_{t=1}^T |\text{Re}\{\hat{s}_k[t]\}| + |\text{Im}\{\hat{s}_k[t]\}|}, \quad (17)$$

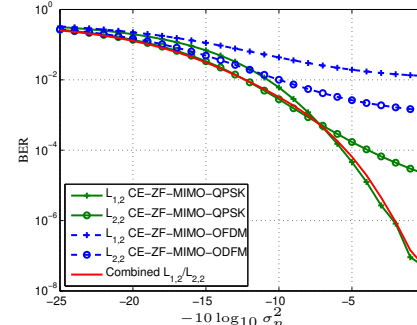
where  $T$  is the length the received sequence. Again, sparse linear precoding combined with superposition coding is clearly advantageous at high SNR. The proposed 16-QAM precoder design also better exploits the higher antenna count than standard ZF at higher SNR values.

## 7. CONCLUSION

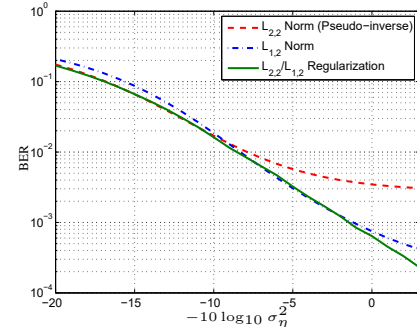
We introduced a novel sparse precoding method to cope with the PA and DAC issues in large-scale MIMO. We found the method to substantially improve the performance of standard linear precoders while reducing the amount of computation required to implement the linear precoder. With the increased demands for antenna element counts, we believe techniques offering reduced analog and digital hardware complexity such as the method presented here will be an essential ingredient in the development of future ultra-massive MIMO systems.



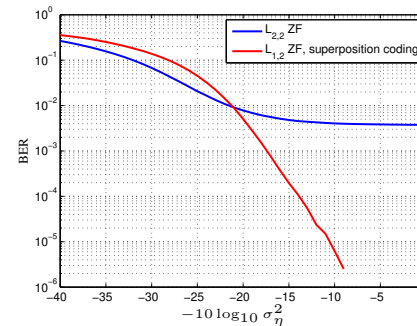
**Fig. 1.** OFDM vs. single-carrier in terms of the CCDF of instantaneous output power for QPSK input and  $M = 128$ ,  $N = 30$  antennas, and  $K = 5$  users.



**Fig. 2.** BER Performance with constant envelope precoding for the downlink with  $K = 5$  users and  $N = 30$  antennas, QPSK.



**Fig. 3.** BER for  $K = 5$  and  $N = 30$ , QPSK, with 1-bit DACs.



**Fig. 4.** BER for  $K = 8$  and  $N = 400$ , 16QAM, 1-bit DACs

## 8. REFERENCES

- [1] O. Arnold, F. Richter, G. Fettweis, and O. Blume, "Power consumption modeling of different base station types in heterogeneous cellular networks," in *2010 Future Network Mobile Summit*, 2010.
- [2] C. Mollén, E. G. Larsson, and T. Eriksson, "Waveforms for the Massive MIMO Downlink: Amplifier Efficiency, Distortion, and Performance," *IEEE Transactions on Communications*, vol. 64, no. 12, pp. 5050–5063, Dec 2016.
- [3] H. Jedda and J. A. Nossek, "On PAPR in Single-Carrier Massive MIMO Systems," in *WSA 2019; 23rd International ITG Workshop on Smart Antennas*, 2019.
- [4] A. Mezghani, R. Ghiaï, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *16th IEEE International Conference on Electronics, Circuits, and Systems 2009 (ICECS 2009)*, Hammamet, Tunisia, 2009.
- [5] 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)*, 2016, pp. 3381–3385.
- [6] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of One-Bit Quantized Precoding for the Multiuser Massive MIMO Downlink," *IEEE Transactions on Signal Processing*, vol. 65, no. 17, pp. 4624–4634, 2017.
- [7] O. De Candido, H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Reconsidering Linear Transmit Signal Processing in 1-Bit Quantized Multi-User MISO Systems," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 254–267, 2019.
- [8] Y. Li, C. Tao, A. Lee Swindlehurst, A. Mezghani, and L. Liu, "Downlink Achievable Rate Analysis in Massive MIMO Systems With One-Bit DACs," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1669–1672, 2017.
- [9] A. K. Saxena, A. Mezghani, and R. W. Heath, "Linear CE and 1-bit quantized precoding with optimized dithering," *IEEE Open Journal of Signal Processing*, vol. 1, pp. 310–325, December 2020.
- [10] A. Kakkavas, J. Munir, A. Mezghani, H. Brunner, and J. A. Nossek, "Weighted Sum Rate Maximization for Multi-User MISO Systems with Low Resolution Digital to Analog Converters," in *WSA 2016; 20th International ITG Workshop on Smart Antennas*, March 2016.
- [11] A. Mezghani, R. Ghiaï, and J. A. Nossek, "Tomlinson Harashima Precoding for MIMO Systems with Low Resolution D/A-Converters," in *ITG/IEEE Workshop on Smart Antennas*, Berlin, Germany, February 2008.
- [12] 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, 2016.
- [13] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized Precoding for Massive MU-MIMO," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 4670–4684, 2017.
- [14] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Nonlinear 1-bit precoding for massive MU-MIMO with higher-order modulation," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 763–767.
- [15] A. L. Swindlehurst, A. K. Saxena, Amine Mezghani, and I. Fijalkow, "Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink," in *Proc. ICASSP 2017*.
- [16] C. Studer and E. G. Larsson, "PAR-Aware Large-Scale Multi-User MIMO-OFDM Downlink," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 303–313, 2013.
- [17] M. Shao, Q. Li, W. Ma, and A. M. So, "A Framework for One-Bit and Constant-Envelope Precoding Over Multiuser Massive MISO Channels," *IEEE Transactions on Signal Processing*, vol. 67, no. 20, pp. 5309–5324, 2019.
- [18] E. S. P. Lopes and L. T. N. Landau, "Optimal Precoding for Multiuser MIMO Systems With Phase Quantization and PSK Modulation via Branch-and-Bound," *IEEE Wireless Communications Letters*, vol. 9, no. 9, pp. 1393–1397, 2020.
- [19] Hui Zou and Trevor Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 2, pp. 301–320, 2005.
- [20] A.Y. Yang, S.S. Sastry, A. Ganesh, and Y. Ma, "Fast  $\ell_1$ -minimization algorithms and an application in robust face recognition: A review," in *2010 IEEE International Conference on Image Processing*, 2010, pp. 1849–1852.
- [21] H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Precoding under instantaneous per-antenna peak power constraint," in *2017 25th European Signal Processing Conference (EUSIPCO)*, 2017, pp. 863–867.