

# One-Bit Quantized Constructive Interference Based Precoding for Massive Multiuser MIMO Downlink

Alireza Haqiqatnejad<sup>1</sup>, Farbod Kayhan<sup>1</sup>, Shahram Shahbazpanahi<sup>1,2</sup>, and Björn Ottersten<sup>1</sup>

<sup>1</sup>Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg  
Email: {alireza.haqiqatnejad, farbod.kayhan, bjorn.ottersten}@uni.lu

<sup>2</sup>Department of Electrical, Computer, and Software Engineering, University of Ontario Institute of Technology  
Email: shahram.shahbazpanahi@uoit.ca

**Abstract**—We propose a one-bit symbol-level precoding method for massive multiuser multiple-input multiple-output (MU-MIMO) downlink systems using the idea of constructive interference (CI). In particular, we adopt a max-min fair design criterion which aims to maximize the minimum instantaneous received signal-to-noise ratio (SNR) among the user equipments (UEs), while ensuring a CI constraint for each UE and under the restriction that the output of the precoder is a vector of binary elements. This design problem is an NP-hard binary quadratic programming due to the one-bit constraints on the elements of the precoder’s output vector, and hence, is difficult to solve. In this paper, we tackle this difficulty by reformulating the problem, in several steps, into an equivalent continuous-domain biconvex form. Our final biconvex reformulation is obtained via an exact penalty approach and can efficiently be solved using a standard block coordinate ascent algorithm. We show through simulation results that the proposed design outperforms the existing schemes in terms of (uncoded) bit error rate. It is further shown via numerical analysis that our solution algorithm is computationally-efficient as it needs only a few tens of iterations to converge in most practical scenarios.

**Index Terms**—Biconvex optimization, constructive interference, massive multiuser multiple-input multiple-output downlink, one-bit quantized precoding.

## I. INTRODUCTION

Massive multiple-input multiple-output (MIMO) is regarded as a promising, or perhaps even indispensable technology for future generation wireless communication networks, providing high spectral/energy efficiency and high reliability [1]. In a massive multiuser MIMO (MU-MIMO) system, the base station (BS) employs a large-scale (e.g., in the order of hundreds) antenna array to serve a much smaller number of user equipments (UEs) via spatial multiplexing. This multitude of antennas offers a vast number of spatial degrees of freedom to each UE, potentially leading to high data rates and some other favorable properties such as robustness to channel fading [1], [2].

While dirty paper coding (DPC), in the noncausal presence of perfect channel state information at the BS, achieves the sum-rate capacity of the MU-MIMO downlink [3], simpler linear precoding schemes such as maximum ratio transmission (MRT) and (regularized) zero-forcing (ZF) have been shown to be (asymptotically) near-optimal in the large system limit [4]. However, the benefits of such easy-to-implement precoding strategies under a massive MIMO setting come with a prohibitively high hardware complexity and cost as well as an extensively increased power consumption. This is primarily due to the need for an ideal radio frequency (RF) chain, including

highly linear power amplifiers and high-resolution digital-to-analog converter (DAC), dedicated to each antenna element. Thus, a more limited use of RF hardware components is of interest for practical realizations of massive MU-MIMO systems, but it requires in turn the precoding scheme to be properly designed by taking the hardware limitations into account.

A wide variety of approaches have been proposed addressing the hardware-limited multiuser precoding problem in a massive MU-MIMO downlink, among which we refer to finite-precision quantized precoding techniques where each antenna element is equipped with finite-precision DACs. In particular, the use of low-resolution DACs can hugely decrease the amount of power consumption, simplifies the hardware design, and reduces the associated cost. Within this line of work, there have been some research towards revisiting conventional linear precoding strategies while the effect of quantization distortion has been taken into account for low-to-moderate resolution (up to 5 bits) DACs, commonly referred to as linear-quantized precoding [5]–[8]. These linear precoders, however, mostly suffer from an unfavorably high error floor in the moderate-to-high SNR regime [8], and perform reasonably only in systems with extremely large number of BS antennas. Recently, the case with one-bit DACs has become an attractive research direction due to its simplicity and the dramatic reduction it can provide in circuit power consumption and hardware cost; see, e.g., [9]–[12]. Most of the work in this direction considers a non-linear precoding design based on a symbol-by-symbol approach. The superiority of these nonlinear quantized approaches over linear-quantized precoding is demonstrated in [9].

The idea of designing the precoder in a per-symbol manner has been studied in [13], and then elaborated in [14] and [15] where the concept of constructive interference (CI) is introduced. This type of precoding, which is also termed in the literature as symbol-level precoding (SLP), is based on the notion that a noise-free received signal can be decoded correctly not necessarily when it is close enough to the intended symbol, rather, as long as it lies within the correct decision region even far away from the intended symbol. This has been the underlying motivation in defining a variety of CI regions; see, e.g., [14]–[17]. In designing a one-bit quantized precoder for massive MU-MIMO downlink, one can utilize the CI concept to achieve lower bit error rates (BER) for the UEs. This approach has been used in [18] and [19] for PSK and QAM signaling.

In this paper, we propose a novel one-bit CI-based precoding method for massive MU-MIMO downlink. Unlike [18] and [19], we do not restrict ourselves to PSK or QAM signal-

The authors are supported by the Luxembourg National Research Fund (FNR) under CORE Junior project: C16/IS/11332341 Enhanced Signal Space Optimization for satellite communication Systems (ESSTIMS).

ing, but consider a generic modulation scheme. The adopted precoding design approach aims to maximize the minimum (instantaneous) SNR among all the UEs while ensuring the UEs' symbols to be received within the correct CI region. Due to the finite-alphabet domain of the design variable, the CI constraints may result in an empty feasible set; therefore, to avoid infeasibility, we first provide an unconstrained CI-penalized reformulation of the original problem. The new formulation is a binary quadratic programming, and thus, is NP-hard. We transform the binary constraints into an equivalent continuous-domain biconvex form. The resulting design formulation, after recasting via the exact penalty method, is reducible to a standard block coordinate ascent (BCA) problem under certain conditions which will be shown to be met in our design. This is different from [18] and [19] where the one-bit constraints are dealt via simple convex relaxations. Note also that in [20], a similar technique is used to treat the one-bit constraints; however, our design objective and constraints differ from those in [20], as will be explained later in the paper. Our simulation results indicate an improved uncoded bit error rate performance for the proposed method compared to the existing one-bit precoding schemes. Remarkably, the BCA algorithm used to solve the proposed design formulation converges (with a reasonable accuracy) in a few tens of iteration cycles for moderately-sized systems, which makes the proposed method attractive for practical use.

**Notations:** We use bold-faced uppercase and lowercase letters to represent matrices and vectors, respectively. For matrices and vectors,  $\|\cdot\|$  denotes, respectively, the Frobenius norm and the  $\ell_2$  norm. For vectors,  $\preceq$  and  $\succeq$  denote elementwise inequality. Operators  $\text{diag}(\cdot)$  and  $\text{blkdiag}(\cdot)$  represent diagonal and block-diagonal matrices. We use  $\mathbf{I}$ ,  $\mathbf{1}$  and  $\mathbf{0}$  to represent, respectively, the identity matrix, the all-ones vector, and the all-zeros matrix (or vector, depending on the context) of appropriate dimensions. The operator  $\otimes$  stands for the Kronecker product.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a narrowband single-cell downlink MU-MIMO wireless system where a BS, equipped with an array of  $n_t$  antennas, serves  $n_r$  single-antenna UEs through the same time-frequency resource block, where  $n_t \gg n_r$ <sup>1</sup>. We further assume that each BS's antenna is equipped with a pair of one-bit DACs (where a single one-bit DAC simply operates as a one-bit scalar quantizer) followed by an individual RF chain. The BS aims to transmit  $n_r$  independent streams of data symbols, each intended for one UE. Let  $s_i$  denote the discrete-time complex symbol intended for the  $i$ th UE. It is assumed that  $s_i$  is drawn from a finite equiprobable constellation set with unit average power for all  $i = 1, \dots, n_r$ . As shown in Fig. 1, the UEs' data symbols are spatially multiplexed at the BS via a (non-linear) multiuser symbol-level precoder such that the  $n_t \times 1$  complex-valued precoded signal  $\bar{\mathbf{u}}$  is directly and instantaneously obtained every symbol period as the solution to an optimization problem. Then, each complex element of

<sup>1</sup>The assumption  $n_t \gg n_r$  rationalizes the use of one-bit DACs at the BS, however, it is not strictly necessary for the subsequent derivations in this paper.

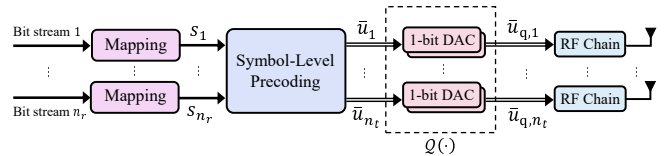


Fig. 1. The considered system model: the output of the symbol-level precoder,  $\bar{\mathbf{u}}$ , undergoes one-bit quantization, yielding the quantized baseband signal  $\bar{\mathbf{u}}_q$ .

the precoder's output is passed through a pair of one-bit scalar quantizers, each operating independently on the real/imaginary part of the input signal. The entire vector quantization operation  $\mathcal{Q}(\cdot)$  can be described as

$$\bar{\mathbf{u}}_q \triangleq \mathcal{Q}(\bar{\mathbf{u}}) = \text{sgn}(\text{Re}\{\bar{\mathbf{u}}\}) + j \text{sgn}(\text{Im}\{\bar{\mathbf{u}}\}), \quad (1)$$

where  $\bar{\mathbf{u}}_q$  is the resulting signal vector after DACs,  $\text{sgn}(\cdot)$  represents the element-wise sign function, and  $j = \sqrt{-1}$ . Next, the quantized baseband signal  $\bar{\mathbf{u}}_q$  is fed into a set of  $n_t$  RF chains, each dedicated to an antenna element, for up-conversion to the carrier frequency. The up-converted signal is then sent over uncorrelated quasi-static flat-fading channels towards the UEs. Under the above described settings, the overall downlink channel corresponding to the  $i$ th UE can be modeled as

$$r_i = \sqrt{\rho} \mathbf{h}_i^T \bar{\mathbf{u}}_q + w_i, \quad i = 1, \dots, n_r, \quad (2)$$

where  $r_i$  is the signal received by the  $i$ th UE,  $\rho = p_{\max}/(2n_t)$  is a fixed gain ensuring a total transmission power of  $p_{\max}$  (note that  $\|\bar{\mathbf{u}}_q\|^2 = 2n_t$ ),  $\mathbf{h}_i$  denotes the complex-valued  $n_t \times 1$  vector of the  $i$ th downlink channel coefficients and  $w_i$  represents the additive noise at the  $i$ th UE's receiver front-end which is modeled as a zero-mean complex Gaussian random variable with variance  $\sigma_i^2/2$  per real dimension. At the receiver side, it is assumed that each UE employs an infinite-precision analog-to-digital converter (ADC) and the optimal single-user maximum-likelihood (ML) criterion is used for symbol detection.

For the sake of convenience, we define the following equivalent real-valued notations:  $\mathbf{u} \triangleq [\text{Re}(\bar{\mathbf{u}})^T, \text{Im}(\bar{\mathbf{u}})^T]^T$ ,  $\mathbf{u}_q \triangleq [\text{Re}(\bar{\mathbf{u}}_q)^T, \text{Im}(\bar{\mathbf{u}}_q)^T]^T$ , and for all  $i = 1, \dots, n_r$ , we denote  $\mathbf{s}_i \triangleq [\text{Re}(s_i), \text{Im}(s_i)]^T$  and

$$\mathbf{H}_i \triangleq \begin{bmatrix} \text{Re}(\mathbf{h}_i^T) & -\text{Im}(\mathbf{h}_i^T) \\ \text{Im}(\mathbf{h}_i^T) & \text{Re}(\mathbf{h}_i^T) \end{bmatrix}.$$

Using the above real-valued notations, the real-valued quantized signal vector can simply be expressed as

$$\mathbf{u}_q = \text{sgn}(\mathbf{u}). \quad (3)$$

When processing the UE's intended data symbols via a CI-based symbol-level precoder, each noise-free received signal  $\sqrt{\rho} \mathbf{H}_i \mathbf{u}_q$ ,  $i = 1, \dots, n_r$ , is allowed to lie within a specific region corresponding to the symbol vector  $\mathbf{s}_i$  which is referred to as CI region. In this work, we adopt the so-called distance-preserving CI regions [17], as illustrated in Fig. 2, in which any two points belonging to two distinct CI regions are distanced by at least the distance between the corresponding constellation points. Thereby, we aim to design a CI precoder maximizing the minimum instantaneous (per-symbol) quality-of-service (QoS) level among the UEs, while ensuring the CI constraint for every UE. The precoded signal  $\mathbf{u}$  is taken from a finite-alphabet set dictated by the set of one-bit DACs so that  $\mathbf{u} = \text{sgn}(\mathbf{u}) = \mathbf{u}_q$ ,

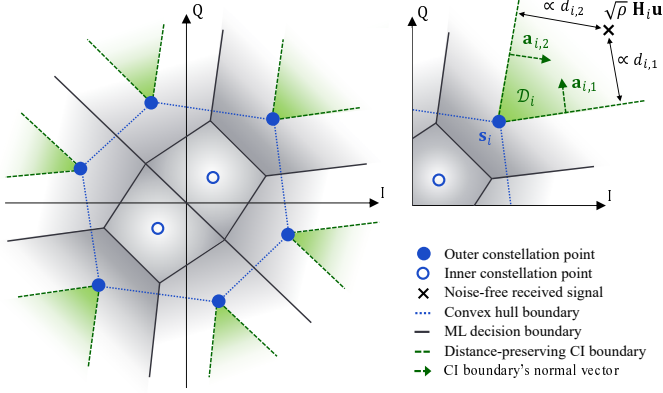


Fig. 2. The distance-preserving CI regions (depicted as green areas) associated with the optimized 8-ary constellation.

thereby the quantization distortion can be avoided. Denoting the distance-preserving CI region associated with symbol  $s_i$  by  $\mathcal{D}_i$ , we can obtain the optimal (one-bit constrained) precoded signal  $\mathbf{u}$  by solving the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{u} \in \mathbb{Q}^{2n_t}}{\text{maximize}} && \min_i \|\mathbf{H}_i \mathbf{u}\|^2 / \sigma_i^2 \\ & \text{s.t.} && \sqrt{\rho} \mathbf{H}_i \mathbf{u} \in \mathcal{D}_i, \quad i = 1, \dots, n_r, \end{aligned} \quad (4)$$

where  $\mathbb{Q} \triangleq \{-1, +1\}$ , implying the use of one-bit quantizers. It should be noted that in formulating (4), we implicitly assume that the downlink channels  $\{\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_{n_r}\}$  are perfectly and instantaneously known to the BS.

**Remark 1:** Unlike conventional linear precoding schemes, the nonlinear-precoded signal  $\mathbf{u}$  may not be explicitly decomposed as a linear combination of distinct UEs' signatures (i.e., precoding vectors). Hence, signal-to-interference-plus-noise ratio (SINR) could no longer be an applicable measure. Instead, we consider the UEs' received SNRs  $\{\|\mathbf{H}_i \mathbf{u}\|^2 / \sigma_i^2\}_{i=1}^{n_r}$  as the measure of QoS, as it is typical in the context of CI precoding; see, e.g., [14]–[16], [21].

The CI constraints of problem (4), as it has been shown in [22], can be written in an explicit compact form as  $\sqrt{\rho} \mathbf{H} \mathbf{u} = \mathbf{\Sigma} \mathbf{s} + \mathbf{A}^{-1} \mathbf{W} \mathbf{d}$ , where the following definitions are used:  $\mathbf{H} \triangleq [\mathbf{H}_1^T, \mathbf{H}_2^T, \dots, \mathbf{H}_{n_r}^T]^T$ ;  $\mathbf{A} \triangleq \text{blkdiag}(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_{n_r})$ ;  $\mathbf{A}_i = [\mathbf{a}_{i,1}, \mathbf{a}_{i,2}]^T$  is a  $2 \times 2$  real-valued matrix containing the normal vectors of the CI boundaries<sup>2</sup>;  $\mathbf{\Sigma} \triangleq \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_{n_r}) \otimes \mathbf{I}_2$ ;  $\mathbf{s} \triangleq [s_1, s_2, \dots, s_{n_r}]^T$ ; and  $\mathbf{W}$  denotes a diagonal weighting matrix with a diagonal element being one if the corresponding symbol is an outer constellation point<sup>3</sup> and zero otherwise. Furthermore, the  $2n_r \times 1$  non-negative slack vector variable  $\mathbf{d} \triangleq [d_1, d_2, \dots, d_{n_r}]^T$ , where  $\mathbf{d}_i = [d_{i,1}, d_{i,2}]^T$ , collects the orthogonal distances of the received symbol  $\sqrt{\rho} \mathbf{H}_i \mathbf{u}$  to its corresponding CI boundaries for all  $i = 1, \dots, n_r$ . An illustration of the CI regions and their characterizing parameters/variables is shown in Fig. 2 for the optimized 8-ary constellation [23]. We then proceed by reformulating (4) into an equivalent form, provided in [21], as

<sup>2</sup>The sub-matrices  $\{\mathbf{A}_1, \dots, \mathbf{A}_{n_r}\}$  can always be formed as full-rank square (hence invertible) matrices, so is  $\mathbf{A}$ .

<sup>3</sup>Any constellation point residing on the convex hull of the constellation set is referred to as outer constellation point.

$$\underset{\mathbf{u} \in \mathbb{Q}^{2n_t}, \mathbf{d} \geq \mathbf{0}}{\text{maximize}} \quad \min(\mathbf{d}) \quad \text{s.t.} \quad \sqrt{\rho} \mathbf{H} \mathbf{u} = \mathbf{\Sigma} \mathbf{s} + \mathbf{A}^{-1} \mathbf{W} \mathbf{d}, \quad (5)$$

where  $\min(\cdot)$  denotes elementwise minimum. We further introduce a slack variable  $\gamma$  which enables us to recast problem (5), in an easier-to-handle form, as

$$\underset{\mathbf{u} \in \mathbb{Q}^{2n_t}, \mathbf{d} \geq \gamma \mathbf{1}, \gamma \geq 0}{\text{maximize}} \quad \gamma \quad \text{s.t.} \quad \sqrt{\rho} \mathbf{H} \mathbf{u} = \mathbf{\Sigma} \mathbf{s} + \mathbf{A}^{-1} \mathbf{W} \mathbf{d}. \quad (6)$$

Two difficulties arise with problem (6) as described below:

- i. The optimization problem (6) could have an empty feasible region since  $\mathbf{u}$  has to be chosen from the finite set  $\mathbb{Q}^{2n_t}$ . In fact, there could be situations where one (or more) CI constraint(s) cannot be satisfied for any  $\mathbf{u} \in \mathbb{Q}^{2n_t}$ .
- ii. Due to the finite-alphabet variable  $\mathbf{u}$ , problem (6) is a combinatorial optimization, and hence, is NP-hard. To be more specific, in order to find the exact solution to (6), one needs to solve a linear programming (LP) for every single vector  $\mathbf{u} \in \mathbb{Q}^{2n_t}$  and then pick the best solution for  $\mathbf{u}$  which results in the largest value of  $\gamma$ . The finite set  $\mathbb{Q}^{2n_t}$  has a cardinality of  $2^{2n_t}$ . Keeping in mind that  $n_t$  refers to the number of elements of a large-scale antenna array, such an approach requires solving an exponentially-growing number of LPs, for each UEs' symbol combination (i.e., for all possible realizations of vector  $\mathbf{s}$ ), which might be quite impractical.

In order to address the above challenges, we need to take a few steps to modify the original problem, as will be explained in the next section.

### III. PROPOSED ONE-BIT QUANTIZED CI PRECODING

We start off by addressing the first challenge highlighted in the previous section. To avoid infeasibility, we consider a new (not necessarily equivalent) design formulation by introducing soft CI constraints added as a penalizing term to the objective function, which results in the following problem:

$$\underset{\mathbf{u} \in \mathbb{Q}^{2n_t}, \mathbf{d} \geq \gamma \mathbf{1}, \gamma \geq 0}{\text{maximize}} \quad \gamma - \|\sqrt{\rho} \mathbf{H} \mathbf{u} - \mathbf{\Sigma} \mathbf{s} - \mathbf{A}^{-1} \mathbf{W} \mathbf{d}\|^2. \quad (7)$$

By applying the change of variable  $\mathbf{d} \rightarrow \mathbf{t} + \gamma \mathbf{1}$ , the optimization problem in (7) can equivalently be rewritten as

$$\underset{\mathbf{u} \in \mathbb{Q}^{2n_t}, \mathbf{t} \geq \mathbf{0}, \gamma \geq 0}{\text{maximize}} \quad \gamma - \|\sqrt{\rho} \mathbf{H} \mathbf{u} - \mathbf{\Sigma} \mathbf{s} - \mathbf{A}^{-1} \mathbf{W} (\mathbf{t} + \gamma \mathbf{1})\|^2. \quad (8)$$

It is worth noting that the design in [20] attempts to minimize the maximum (among the UEs) distance between a received signal and its intended constellation point up to a scaling factor, while our design formulation (8) aims to minimize the average distances (due to the  $\ell_2$  norm) from the boundaries of the CI regions. In fact, the scaling factor in [20] can be viewed as a special case of CI regions with strict phase constraints.

Given  $\mathbf{u}$  and  $\mathbf{t}$ , the maximization over  $\gamma$  is now amenable to a closed-form *provably positive* solution given by

$$\gamma^* = \frac{1}{2\eta} + \frac{(\sqrt{\rho} \mathbf{H} \mathbf{u} - \mathbf{\Sigma} \mathbf{s} - \mathbf{A}^{-1} \mathbf{W} \mathbf{t})^T \mathbf{A}^{-1} \mathbf{W} \mathbf{1}}{\eta}, \quad (9)$$

where  $\eta \triangleq \mathbf{1}^T \mathbf{W} \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{W} \mathbf{1}$ . By plugging  $\gamma^*$  into (8) and taking some straightforward algebraic steps, the design problem of interest boils down to a binary linearly-constrained quadratic programming (LCQP) as

$$\underset{\mathbf{u} \in \mathbb{Q}^{2n_t}, \mathbf{t} \succeq \mathbf{0}}{\text{maximize}} \quad \sqrt{\rho} \mathbf{q}^T \mathbf{u} - \mathbf{p}^T \mathbf{t} - \left\| \sqrt{\rho} \Theta \mathbf{H} \mathbf{u} - \Theta \mathbf{A}^{-1} \mathbf{W} \mathbf{t} - \mathbf{z} \right\|^2, \quad (10)$$

where  $\mathbf{q} \triangleq (1/\eta) \mathbf{H}^T \mathbf{A}^{-1} \mathbf{W} \mathbf{1}$ ,  $\mathbf{p} \triangleq (1/\eta) \mathbf{W} \mathbf{A}^{-T} \mathbf{A}^{-1} \mathbf{W} \mathbf{1}$ ,  $\Theta \triangleq \mathbf{I} - (1/\eta) (\mathbf{1}^T \mathbf{W} \mathbf{A}^{-T} \otimes \mathbf{A}^{-1} \mathbf{W} \mathbf{1})$ , and  $\mathbf{z} \triangleq \Theta \Sigma \mathbf{s} + (\mathbf{A}^{-1} \mathbf{W} \mathbf{1}) / (2\eta)$  are all non-variables.

**Method 1 – Convex Relaxation:** In order to tackle the second difficulty, one simple approach is to solve a convex relaxation of (10) obtained by replacing the strict equality constraints on the elements of  $\mathbf{u}$  (imposed by  $\mathbb{Q}$ ) with appropriate box constraints. This relaxed problem can be expressed as a standard LCQP, i.e.,

$$\text{P1: } \underset{-1 \preceq \mathbf{u} \preceq 1, \mathbf{t} \succeq \mathbf{0}}{\text{maximize}} \quad \sqrt{\rho} \mathbf{q}^T \mathbf{u} - \mathbf{p}^T \mathbf{t} - \left\| \sqrt{\rho} \Theta \mathbf{H} \mathbf{u} - \Theta \mathbf{A}^{-1} \mathbf{W} \mathbf{t} - \mathbf{z} \right\|^2. \quad (11)$$

It is worth noting that solving P1 results in an upper-bound for the objective function of (10). Our numerical experiments show that a relatively noticeable number of the elementwise box constraints  $-1 \leq u_l \leq 1$ , for  $l = 1, \dots, 2n_t$ , are not active at the optimum of (11). These observations suggest the possibility of further improvement of the method. In the sequel, the goal is to achieve a more accurate solution, while considering P1 as a benchmark for comparison purposes.

**Method 2–Equivalent Biconvex Formulation:** We first restate from [24] an equivalent implication of the binary constraint  $\mathbf{u} \in \mathbb{Q}^{2n_t}$ , which is given in a biconvex form.

**Lemma 1.** *Let  $\mathbf{u}$  and  $\mathbf{v}$  be two real vectors of equal length  $2n_t$ . Then, provided that  $-1 \preceq \mathbf{u} \preceq 1$  and  $\|\mathbf{v}\|^2 \leq 2n_t$ , the condition  $\mathbf{u}^T \mathbf{v} = 2n_t$  implies that  $\mathbf{u} = \mathbf{v}$  and  $\mathbf{u} \in \{-1, +1\}^{2n_t}$ .*

The biconvex implication of the binary constraint on  $\mathbf{u}$ , provided by Lemma 1, allows us to cast an equivalent *continuous-domain* reformulation of the original binary LCQP (10), which can be expressed as

$$\underset{-1 \preceq \mathbf{u} \preceq 1, \mathbf{t} \succeq \mathbf{0}, \mathbf{v}}{\text{maximize}} \quad \sqrt{\rho} \mathbf{q}^T \mathbf{u} - \mathbf{p}^T \mathbf{t} - \left\| \sqrt{\rho} \Theta \mathbf{H} \mathbf{u} - \Theta \mathbf{A}^{-1} \mathbf{W} \mathbf{t} - \mathbf{z} \right\|^2, \\ \text{s.t.} \quad \mathbf{u}^T \mathbf{v} = 2n_t, \quad \mathbf{v}^T \mathbf{v} \leq 2n_t, \quad (12)$$

where  $\mathbf{u}^T \mathbf{v} = 2n_t$  is referred to as the equilibrium constraint. The reformulation (12) is not yet a convex problem; however, there are well-known efficient approaches to solve (12) such as the exact penalty method (EPM) or the alternating direction method of multipliers (ADMM). Here, we adopt an EPM-based approach due to its simplicity. The accuracy and convergence characteristics of the EPM are studied in, e.g., [25].

In the EPM, the equilibrium constraint is enforced by adding a penalizing term to the objective function. The added penalty function appears with a positive multiplier  $\lambda$  which has to be monotonically increased until a certain threshold is reached. Denoting the objective function of (12) by  $f(\mathbf{u}, \mathbf{t})$ , we write

$$\text{P2: } \underset{-1 \preceq \mathbf{u} \preceq 1, \mathbf{t} \succeq \mathbf{0}, \mathbf{v}}{\text{maximize}} \quad f(\mathbf{u}, \mathbf{t}) + \lambda (\mathbf{u}^T \mathbf{v} - 2n_t), \\ \text{s.t.} \quad \mathbf{v}^T \mathbf{v} \leq 2n_t, \quad (13)$$

which is our final formulation for the one-bit quantized CI precoding problem of interest. Even though, generally speaking, problem P2 is not equivalent to (12), it can be verified that if  $f(\mathbf{u}, \mathbf{t})$  is an  $L$ -Lipschitz continuous concave function on  $-1 \preceq \mathbf{u} \preceq 1$ , problem P2 has the same local and global maxima

as those of (12) for  $\lambda \geq 2L$ , where  $L$  is the Lipschitz constant of  $f(\mathbf{u}, \mathbf{t})$  with respect to  $\mathbf{u}$ ; see [24, Th. 1]. The following lemma indicates that the Lipschitz continuity condition holds for  $f(\mathbf{u}, \mathbf{t})$  within the domain  $-1 \preceq \mathbf{u} \preceq 1$ .

**Lemma 2.** *Given  $\mathbf{t}$ , function  $f(\mathbf{u}, \mathbf{t})$  is  $L$ -Lipschitz continuous on  $-1 \preceq \mathbf{u} \preceq 1$  with Lipschitz constant*

$$L = \sqrt{\rho} \|\mathbf{q}\| + 2\mathbf{H}^T \Theta^T (\Theta \mathbf{A}^{-1} \mathbf{W} \mathbf{t} + \mathbf{z}) + 2\rho \sqrt{2n_t} \|\Theta \mathbf{H}\|^2. \quad (14)$$

*Sketch of the proof.* Given  $\mathbf{t}$ , function  $f(\mathbf{u}, \mathbf{t})$  is composed of affine and  $\ell_2$  norm terms in  $\mathbf{u}$  and hence is continuously differentiable everywhere. Then, plugging  $\mathbf{u}_1 \in \mathbf{R}^{2n_t \times 1}$  and  $\mathbf{u}_2 \in \mathbf{R}^{2n_t \times 1}$  with  $-1 \preceq \mathbf{u}_1, \mathbf{u}_2 \preceq 1$  as two distinct inputs into  $f(\mathbf{u}, \mathbf{t})$ , we obtain an upper bound on  $|f(\mathbf{u}_1, \mathbf{t}) - f(\mathbf{u}_2, \mathbf{t})|$  in terms of  $\|\mathbf{u}_1 - \mathbf{u}_2\|$ . This leads us to an upper bound on the absolute value of the derivative of  $f(\mathbf{u}, \mathbf{t})$ , i.e., the Lipschitz constant, implying the Lipschitz continuity property for  $f(\mathbf{u}, \mathbf{t})$ . The details are omitted due to space limitation, but will appear in an extended version of this paper.  $\square$

The objective function of P2, i.e.,  $f(\mathbf{u}, \mathbf{t}) + \lambda (\mathbf{u}^T \mathbf{v} - 2n_t)$  is a biconvex quadratic function in  $\mathbf{u}$  and  $\mathbf{v}$ , i.e., fixing either  $\mathbf{u}$  or  $\mathbf{v}$  results in a concave function in the other variable. From Lemma 2, it then follows that finding at least a locally optimal solution to problem (12) is equivalent to solving P2 via, e.g., a standard block coordinate ascent (BCA) algorithm, where a coordinate block refers to either of the vector variables  $\mathbf{t}$ ,  $\mathbf{u}$  or  $\mathbf{v}$ . More specifically, the objective function  $f(\mathbf{u}, \mathbf{t}) + \lambda (\mathbf{u}^T \mathbf{v} - 2n_t)$  can be maximized over  $\mathbf{t}$ ,  $\mathbf{u}$ , and  $\mathbf{v}$  by solving the respective sub-problems in an alternating manner. The penalty multiplier  $\lambda$  can also be updated in every  $K$  outer iterations until the threshold  $2L$  is achieved. Based on this approach, the BCA algorithm solving P2 runs in the  $k$ th iteration as follows:

**First step – Updating  $\mathbf{t}$ :** Given  $\mathbf{u}$ , maximizing  $f(\mathbf{u}, \mathbf{t})$  over  $\mathbf{t}$  is equivalent to a standard LCQP. Hence, the value of  $\mathbf{t}$  is updated as the solution to the following maximization problem:

$$\mathbf{t}^{(k)} = \underset{\mathbf{t} \succeq \mathbf{0}}{\text{argmax}} \quad -\mathbf{p}^T \mathbf{t} - \left\| \sqrt{\rho} \Theta \mathbf{H} \mathbf{u}^{(k-1)} - \Theta \mathbf{A}^{-1} \mathbf{W} \mathbf{t} - \mathbf{z} \right\|^2. \quad (15)$$

**Second step – Updating  $\mathbf{u}$ :** For given  $\mathbf{t}$  and  $\mathbf{v}$ , the value of  $\mathbf{u}$  in the  $k$ th iteration can be updated by solving the following box-constrained QP:

$$\mathbf{u}^{(k)} = \underset{-1 \preceq \mathbf{u} \preceq 1}{\text{argmax}} \quad f(\mathbf{u}, \mathbf{t}^{(k)}) + \lambda (\mathbf{u}^T \mathbf{v}^{(k-1)} - 2n_t). \quad (16)$$

**Third step – Updating  $\mathbf{v}$ :** The  $k$ th update of  $\mathbf{v}$  can be obtained as the optimal solution to the following problem:

$$\mathbf{v}^{(k)} = \underset{\|\mathbf{v}\|^2 \leq 2n_t}{\text{argmax}} \quad \mathbf{v}^T \mathbf{u}^{(k)}. \quad (17)$$

This norm-constrained inner product maximization has a simple closed-form solution which is given by

$$\mathbf{v}^{(k)} = \sqrt{2n_t} \mathbf{u}^{(k)} / \|\mathbf{u}^{(k)}\|. \quad (18)$$

**Forth step – Updating  $\lambda$ :** In every  $K$  iterations, we update the penalty parameter  $\lambda$  as

$$\lambda^{(k)} = \min\{2L, \lambda^{(k-1)} \mu\}, \quad (19)$$

where  $\mu > 1$  is an arbitrary constant.

**Remark 2.** No closed-form solution is known in general for the inner sub-problems (15) and (16). Nonetheless, they both can be solved using standard algorithms such as (accelerated) projected gradient methods [26], or quasi-Newton methods, e.g., L-BFGS-B [27]. In particular, for a Lipschitz smooth (not strongly) concave objective function, all these algorithms converge superlinearly at a rate of  $\mathcal{O}(1/k^2)$ .

#### IV. SIMULATION RESULTS

We consider a downlink massive MU-MIMO with multiuser precoding and subsequent one-bit quantization at the BS, where independent QPSK symbols are intended for the UEs. At the UEs, identical noise distributions  $w_i \sim \mathcal{CN}(0, \sigma^2)$  with  $\sigma^2 = 1$  are assumed for all  $i = 1, \dots, n_r$ . We further assume a Rayleigh block fading channel, where i.i.d. vectors  $\{\mathbf{h}_i\}_{i=1}^{n_r}$  are randomly generated for each fading block following the standard circularly symmetric complex Gaussian distribution, i.e.,  $\mathbf{h}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ . As for the EPM-based formulation, we set  $\mu = 1.5$  and initialize the penalty parameter as  $\lambda^0 = 0.4$ , where the values are chosen empirically. All the presented results have been averaged over 200 fading block realizations, each of 200 symbols. Throughout this section, the one-bit quantized precoding schemes of interest are referred to as:

- *MSM*: Maximum safety margin [19]
- *SQUID*: Squared-infinity norm Douglas-Rachford splitting [9]
- *QSLP-CRE*: Quantized SLP via convex relaxation (P1)
- *QSLP-BIC*: Quantized SLP via biconvex formulation (P2)

The inner sub-problems within the outer cycle of QSLP-BIC are solved via the accelerated projected gradient ascent algorithm [26]. The results are also compared to those obtained from the conventional matched filter (MF), ZF, and Wiener filter (WF) precoding techniques [28] with one-bit quantized outputs. We further consider the unquantized WF precoding as the benchmark for the infinite-resolution case.

We compare in Fig. 3 and Fig. 4 the BER performances achieved by the one-bit quantized precoding schemes of interest versus transmit SNR, i.e.,  $p_{\max}/\sigma^2$  for two practical systems with  $(n_t, n_r) = (16, 4)$  and  $(n_t, n_r) = (64, 8)$ . It can be seen that the proposed method, QSLP-BIC, outperforms both the MSM and the SQUID one-bit precoders. The gain is around 1 dB in the applicable range of SNR (i.e., 5-10 dB) for an uncoded QPSK signaling. Furthermore, the QSLP-BIC approach performs superior to our naive precoding formulation QSLP-CRE, which exploits CI in the design but simply treats the one-bit constraints via convex relaxations. From Fig. 3, we can further observe that the one-bit precoders MSM, SQUID and QSLP-CRE all experience an error floor at high SNRs (i.e., above 15 dB). This indicates that these one-bit precoders require more degree of freedoms (either more transmit antennas or higher resolution bits) to perform well for the multiuser system with  $(n_t, n_r) = (16, 4)$ .

Next, we numerically evaluate the complexity of the QSLP-BIC method in Fig. 5, where the numbers of inner and outer iterations required for a normalized squared error of  $10^{-4}$  are separately plotted. It can be verified from the simulation results that all the iteration numbers grow linearly with transmit power

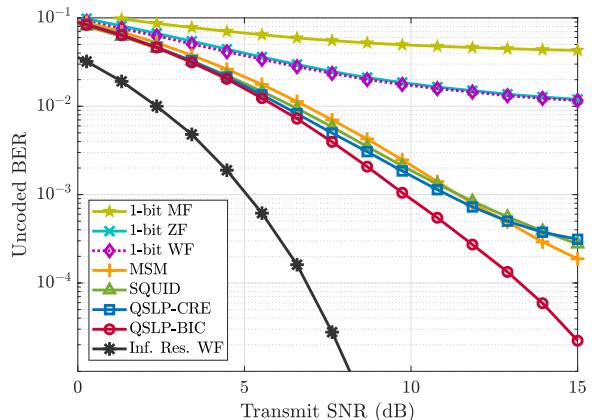


Fig. 3. BER versus transmit SNR over a massive MU-MIMO downlink system with  $(n_t, n_r) = (16, 4)$ .

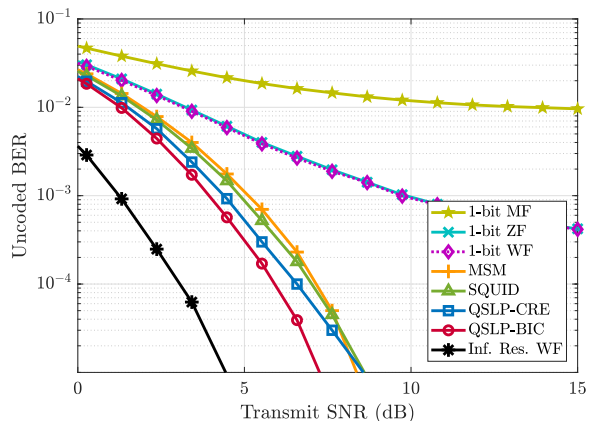


Fig. 4. BER versus transmit SNR over a massive MU-MIMO downlink system with  $(n_t, n_r) = (64, 8)$ .

for large values of  $p_{\max}$ , while the number of inner iterations to solve the sub-problem on  $\mathbf{t}$  shows a relatively faster growth with  $p_{\max}$ . However, we remark that the update iterations on  $\mathbf{t}$  are of dimension  $2n_r$ , whereas the dominant complexity order comes from the even larger dimension  $2n_t$ . For this reason, in our evaluations, only those  $2n_t$ -dimensional iterations updating  $\mathbf{u}$  are accounted for the complexity cost of QSLP-BIC.

Finally, to evaluate the convergence behavior of QSLP-BIC as a function of system parameters, i.e.,  $n_t$  and  $n_r$ , we report in Table I the average number of required iterations for different values of  $p_{\max}/\sigma^2$  within the effective SNR range associated with QPSK signaling. With reference to Table I, the proposed QSLP-BIC method offers a favorably fast converges speed, in the order of tens of iterations, even for large system parameters. For instance, at  $p_{\max}/\sigma^2 = 3.4 \approx 5.4$  dB, the QSLP-BIC algorithm needs only  $\sim 48$  and  $\sim 56$  iterations on average to achieve those uncoded BER performances as shown in Fig. 3 and Fig. 4. Table I, on the other hand, indicates that the complexity of QSLP-BIC (in terms of the number of iterations till convergence) scales linearly with  $n_t$ , which is an attractive feature for implementation purposes.

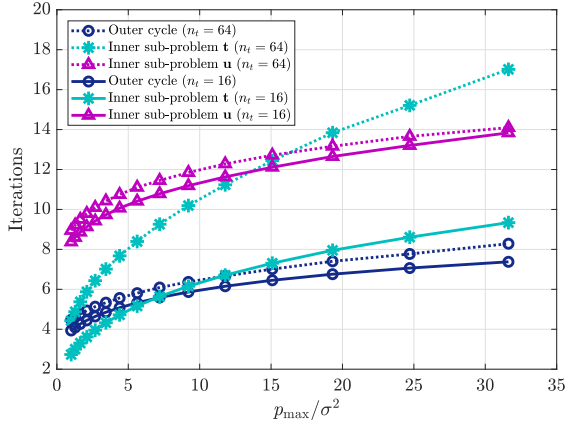


Fig. 5. The QSLP-BIC method: average number of outer and inner iterations to reach a squared error of  $10^{-4}$  as a function of transmit SNR in linear scale.

TABLE I  
AVERAGE NUMBER OF ITERATIONS OF DIMENSION  $2n_t$  TILL CONVERGENCE OF THE QSLP-BIC ALGORITHM.

$p_{\max}/\sigma^2$	$(n_t, n_r)$			
	(8, 2)	(16, 4)	(64, 8)	(128, 16)
$3.4 \approx 5.4$ dB	35.3	47.3	55.7	64.9
$9.2 \approx 9.6$ dB	48.8	65.8	75.7	87.2

## V. CONCLUSIONS

We proposed a quantized symbol-level precoding method for massive MU-MIMO downlink system equipped with one-bit DACs, using the idea of constructive interference (CI). With the objective of maximizing the minimum instantaneous SNR among all the UEs with CI constraints, the design problem of interest is a binary (due to one-bit constraints) linearly-constrained quadratic programming where its solution requires a high computational complexity. Therefore, we used an equivalent implication of the binary one-bit constraints given in a continuous-domain biconvex form. We then dealt with the biconvex constraints via applying the exact penalty technique. The resulting design reformulation was reduced to a standard block coordinate ascent problem by proving the Lipschitz continuity property for the objective function. Comparing our proposed symbol-level approach with well-known one-bit precoding schemes showed a superior performance in terms of uncoded BER. Numerical analysis on the complexity indicated that the our proposed method converges in a few tens of iterations in a practical massive MU-MIMO system.

## REFERENCES

- [1] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.
- [2] L. Lu, G. Y. Li, A. L. 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.
- [3] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inform. Theory*, vol. 29, no. 3, pp. 439–441, May 1983.
- [4] 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.

- [5] A. Mezghani, R. Ghat, and J. A. Nossek, "Transmit processing with low resolution D/A-converters," in *2009 16th IEEE Int. Conf. Electronics, Circuits and Syst. (ICECS)*, Dec. 2009, pp. 683–686.
- [6] L. N. Ribeiro, S. Schwarz, M. Rupp, and A. L. F. de Almeida, "Energy efficiency of mmWave massive MIMO precoding with low-resolution DACs," *IEEE J. Sel. Topics in Signal Process.*, vol. 12, no. 2, pp. 298–312, May 2018.
- [7] 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 Commun. Lett.*, vol. 21, no. 7, pp. 1669–1672, Jul. 2017.
- [8] A. K. Saxena, I. Fijalkow, and A. L. Swindlehurst, "Analysis of one-bit quantized precoding for the multiuser massive MIMO downlink," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4624–4634, Sep. 2017.
- [9] S. Jacobsson, G. Durisi, M. Coldrey, T. Goldstein, and C. Studer, "Quantized precoding for massive MU-MIMO," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4670–4684, Nov. 2017.
- [10] F. Sohrabi, Y. Liu, and W. Yu, "One-bit precoding and constellation range design for massive MIMO with QAM signaling," *IEEE J. Sel. Topics in Signal Process.*, vol. 12, no. 3, pp. 557–570, Jun. 2018.
- [11] A. Swindlehurst, A. Saxena, A. Mezghani, and I. Fijalkow, "Minimum probability-of-error perturbation precoding for the one-bit massive MIMO downlink," in *2017 IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2017, pp. 6483–6487.
- [12] O. B. Usman, H. Jedda, A. Mezghani, and J. A. Nossek, "MMSE precoder for massive MU-MIMO using 1-bit quantization," in *IEEE Conf. Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2016, pp. 3381–3385.
- [13] C. Masouros and E. Alsusa, "Soft linear precoding for the downlink of DS/CDMA communication systems," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 203–215, Jan. 2010.
- [14] C. Masouros and G. Zheng, "Exploiting known interference as green signal power for downlink beamforming optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 14, pp. 3628–3640, Jul. 2015.
- [15] M. Alodeh, S. Chatzinotas, and B. Ottersten, "Constructive multiuser interference in symbol level precoding for the MISO downlink channel," *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2239–2252, May 2015.
- [16] A. Kalantari, M. Soltanalian, S. Maleki, S. Chatzinotas, and B. Ottersten, "Directional modulation via symbol-level precoding: A way to enhance security," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 8, pp. 1478–1493, Dec. 2016.
- [17] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Constructive interference for generic constellations," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 586–590, Apr. 2018.
- [18] A. Li, C. Masouros, F. Liu, and A. L. Swindlehurst, "Massive MIMO 1-bit DAC transmission: A low-complexity symbol scaling approach," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 11, pp. 7559–7575, Nov. 2018.
- [19] H. Jedda, A. Mezghani, A. L. Swindlehurst, and J. A. Nossek, "Quantized constant envelope precoding with PSK and QAM signaling," *IEEE Trans. Wirel. Commun.*, vol. 17, no. 12, pp. 8022–8034, Dec. 2018.
- [20] M. Shao, Q. Li, and W. Ma, "One-bit massive MIMO precoding via a minimum symbol-error probability design," in *IEEE Int. Conf. Acoust., Speech and Signal Process. (ICASSP)*, Apr. 2018, pp. 3579–3583.
- [21] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "Symbol-level precoding design based on distance preserving constructive interference regions," *IEEE Trans. Signal Process.*, vol. 66, no. 22, pp. 5817–5832, Nov. 2018.
- [22] A. Haqiqatnejad, F. Kayhan, and B. Ottersten, "An approximate solution for symbol-level multiuser precoding using support recovery," in *2019 IEEE 20th Int. Workshop on Signal Process. Advances in Wirel. Commun. (SPAWC)*, Jul. 2019, pp. 1–5.
- [23] F. Kayhan and G. Montorsi, "Constellation design for memoryless phase noise channels," *IEEE Trans. Wirel. Commun.*, vol. 13, no. 5, pp. 2874–2883, May 2014.
- [24] G. Yuan and B. Ghanem, "Binary optimization via mathematical programming with equilibrium constraints," *arXiv preprint:1608.04425*, 2016.
- [25] X. Hu and D. Ralph, "Convergence of a penalty method for mathematical programming with complementarity constraints," *Journal of Optimization Theory and Applications*, vol. 123, no. 2, pp. 365–390, 2004.
- [26] R. A. Polyak, "Projected gradient method for non-negative least square," *Contemp Math*, vol. 636, pp. 167–179, 2015.
- [27] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.
- [28] M. Joham, W. Utschick, and J. A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, Aug. 2005.