Massive MIMO 1-Bit DAC Transmission: A Low-Complexity Symbol Scaling Approach

Ang Li, Student Member, IEEE, Christos Masouros, Senior Member, IEEE, Fan Liu, Student Member, IEEE, and A. Lee Swindlehurst, Fellow, IEEE

Abstract—We study the multi-user massive multiple-input single-output (MISO) systems and focus on the downlink transmission, where the base station (BS) employs a large antenna array with low-cost 1-bit digital-to-analog converters (DACs). It is observed that the direct combination of existing beamforming schemes with 1-bit DACs will lead to an error floor at medium-to-high SNR regime, due to the coarse quantization of the DACs with limited precision. In this paper, based on the concept of constructive interference we consider both a quantized linear beamforming scheme where we analytically obtain the optimal beamforming matrix, and a non-linear mapping scheme where we directly design the transmit signal vector. Due to the 1-bit quantization, the formulated optimization for the non-linear mapping scheme is shown to be non-convex. To solve this problem, the non-convex constraints of 1-bit DACs are firstly relaxed, followed by an element-wise normalization to satisfy the 1-bit DAC transmission. We further propose a low-complexity symbol scaling scheme that consists of three stages, in which the quantized transmit signal on each antenna element is selected sequentially. Numerical results show that the proposed symbol scaling scheme achieves a comparable performance to the optimization-based non-linear mapping approach, while its corresponding complexity is negligible compared to that of the non-linear scheme.

Index Terms—Massive MIMO, 1-bit quantization, beamforming, constructive interference, Lagrangian, low-complexity scheme.

I. INTRODUCTION

MULTIPLE-input multiple-output (MIMO) techniques have been widely acknowledged in the field of wireless communications, due to the significant performance gains over single-input single-output (SISO) systems [1]. When the channel state information (CSI) is known at the base station (BS), dirty-paper coding (DPC) proposed in [2] can reach the sum-rate capacity. Nevertheless, DPC is difficult to implement due to the impractical assumption of infinite length of codewords. Its practical non-linear counterparts in the form of Tomlinson-Harashima precoding (THP) [3] and vector perturbation (VP) [4] still result in high computational complexity. Therefore, linear precoding schemes based on zero-forcing (ZF) have received extensive research attention due to their favourable low complexity [5]-[7]. In addition, optimization-based transmit beamforming schemes have also been a popular topic for MIMO downlink systems, where schemes targeting at signal-to-interference-plus-noise ratio (SINR) balancing [8]-[10], power minimization [11]-[13] and constructive interference are studied [14]-[21]. Towards the fifth generation (5G) and future wireless communication systems, massive MIMO systems [22] have been shown to greatly improve the spectral efficiency, and therefore have received increasing research attention in recent years. While the performance of non-linear precoding schemes is superior, they are difficult to implement in massive MIMO systems due to their prohibitive complexity. Therefore, linear precoding schemes are more favourable for massive MIMO, where it has further been shown that linear schemes such as ZF [9] and RZF [6] are near-optimal.

Nevertheless, with a large number of antennas employed at the BS, the hardware complexity and the resulting power consumption have also increased dramatically, which hinders the practical implementation of massive MIMO. To achieve a compromise between performance and hardware complexity and the consequent power consumption, one potential technique for massive MIMO is to reduce the number of radio frequency (RF) chains at the BS, which is achieved by employing the hybrid analog-digital structure [23], [24], where the transmit beamforming is divided into the analog domain and the digital domain. In the analog domain, low-cost phase shifters are employed to offer high-dimensional phase-only controls, while low-dimensional digital beamforming schemes are applied in the digital domain to manage the interference. Hybrid beamforming has been studied for single-user transmission in [23], and is extended to the case of multi-user scenarios in [26], [27].

In addition to the hybrid structures, another potential technique, which is the focus of this paper, is to reduce the cost and power consumption per RF chain by employing very low-resolution digital-to-analog converters (DACs) instead of high-precision DACs. It has been shown in [28] that DACs are one of the dominant power-consuming hardware components in the downlink, whose power consumption grows exponentially with the resolution and linearly with the bandwidth. In the traditional MIMO downlink transmission, each transmit signal is generated by a pair of high-resolution (usually more than 8-
bit) DACs that are connected to the RF chain. However, in the case of massive MIMO with hundreds of antennas employed at the BS, a large number of DACs are required and the resulting power consumption of the DACs will be prohibitively high. Therefore, employing low-resolution DACs, especially 1-bit DACs, can greatly reduce the power consumption per RF chain and the resulting total power consumed at the BS. When 1-bit DACs are employed, the output signal at each antenna element is equivalent to the constant-envelope symbol from a QPSK constellation, which enables the use of low-cost power amplifiers (PAs) and can further reduce the hardware complexity. In the existing literature, most recent studies however, have focused on the performance analysis for massive MIMO uplink with low-resolution analog-to-digital converters (ADCs), especially for the 1-bit case [29]-[31], where it is shown that the number of quantization bits can be reduced while a comparable performance is still achievable. For the case of downlink transmission with 1-bit DACs, there have been an increasing number of studies due to the benefits mentioned above [32]-[36]. In [32], a simple quantized ZF scheme is considered, where the transmit signal vector is obtained by a direct quantization on the ZF-precoded signals. The authors further analyze the performance of the quantized ZF scheme, and show that it outperforms the maximum likelihood (ML) encoder in the low-to-medium SNR regime. In [33], the quantized linear beamforming schemes based on minimum-mean squared error (MMSE) are proposed, whose performance is shown to be superior to the quantized ZF scheme in [32]. In [35], a non-linear symbol perturbation technique is introduced in 1-bit massive MIMO downlink for QPSK modulation, while in [36] an iterative non-linear beamforming scheme named ‘Pokemon’ is introduced via a biconvex relaxation approach, where the proposed scheme directly designs the transmit signal vector based on the MMSE criterion. Nevertheless, these MMSE-based schemes may be sub-optimal as they ignore the fact that interference can be exploited on an instantaneous basis in [7], [14]-[19] and [37]. Moreover, while there have been studies on the downlink beamforming schemes with 1-bit DACs [32]-[36], these existing schemes either suffer a severe performance degradation compared to the fully-digital case, or require sophisticated optimizations and iterative algorithms that are computationally inefficient.

In this paper, we focus on the massive MIMO downlink systems and consider multi-user beamforming schemes with 1-bit DACs. For massive MIMO systems with 1-bit DACs, symbol-level operations are required as the output signals are dependent on the data symbols, which creates the opportunity to exploit the formulation of constructive interference. It is this aspect of the transmission that allows us to observe the interference from an instantaneous point of view, and exploit it constructively [7], [14]-[19]. We firstly consider a quantized linear beamforming scheme by constructing a beamforming matrix before quantization. Based on the concept of constructive interference, the optimization aims to maximize the distance between the received symbols and the detection thresholds, where the optimization problem is formulated based on the geometry of the modulation constellations. By mathematically analyzing the optimization problem with Lagrangian and the Karush-Kuhn-Tucker (KKT) conditions, it is shown that the optimality is achieved by applying a strict phase rotation for the constructed problem in the case of massive MIMO. Due to the operation of the 1-bit quantization, the above quantized linear scheme based on constructive interference is analytically shown to be equivalent to the quantized ZF scheme, which suffers an error floor at high SNR. To improve the performance, we then propose a non-linear mapping scheme where we directly design the quantized transmit signal vector. Nevertheless, due to the constraint on the output signals of 1-bit DACs, the resulting optimization problem is shown to be non-convex. To solve this problem, we firstly apply a relaxation on the mathematical constraint resulting from the use of 1-bit DACs, such that the optimization problem becomes convex. Then, we apply an element-wise normalization on the signal vector obtained from the relaxed optimization to meet the constraint on the output signals of 1-bit DACs.

Nevertheless, since the variable of the non-linear optimization approach is the transmit signal vector, whose dimension is equal to the number of transmit antennas, the computational complexity of the resulting optimization is high in the case of massive MIMO. Therefore, to enable the practical implementation of 1-bit DACs, we further propose a low-complexity symbol scaling scheme based on a coordinate transformation on the problem formulation of the constructive interference, where we directly select the output signal of 1-bit DACs for each antenna element on a sequential basis, and a relaxation-normalization process is therefore no longer needed. The proposed symbol scaling approach consists of three stages: an initialization stage where we decide the output signals for some antenna elements whose channel coefficients satisfy certain requirements, an allocation stage where we sequentially select the output signals for the residual antenna elements, and a refinement stage where we check whether the performance with the obtained signal vector can be further improved based on the greedy algorithm. Both the ‘Sum-Max’ and the ‘Max-Min’ criteria are considered in the allocation stage, and the output signal vector that returns the best performance is then obtained within the above two criteria. We further study the computational costs of the proposed optimization-based and symbol scaling schemes in terms of the floating operations required. Numerical results show that in the case of small-scale MIMO systems, the proposed symbol scaling scheme is shown to achieve the best performance. In the case of massive MIMO, the optimization-based non-linear scheme achieves an improved performance over existing schemes and better approaches the fully-digital scheme, while the proposed symbol scaling scheme can achieve a comparable performance. In terms of the computational complexity, it is demonstrated that the complexity of the symbol scaling scheme is negligible compared to that of the non-linear mapping approach, while the performance of the symbol scaling scheme is superior to ‘Pokemon’ when their computational costs are similar, which favours its usefulness in practice.

For reasons of clarity, we summarize the contributions of this paper as:

1) We propose the downlink beamforming schemes for massive MIMO with 1-bit DACs based on the constructive
interference formulation. We firstly consider a quantized linear beamforming scheme, where it is analytically proven that, in the massive MIMO region, the optimality is achieved by employing a strict phase rotation due to the favourable propagation conditions.

2) We then consider a non-linear mapping scheme where we directly optimize the transmit signal vector. The resulting non-convex optimization is solved in two steps: we firstly relax the non-convex constraints of 1-bit DACs, followed by the normalization on the obtained signal vector to satisfy the 1-bit DAC transmission.

3) Based on a coordinate transformation of the constructive interference formulation, we further propose a low-complexity symbol scaling scheme where we directly select the quantized signal on each antenna element via a three-stage process. It is shown that the symbol scaling scheme can achieve a comparable performance to the optimization-based non-linear mapping scheme.

4) We further study and compare the computational costs of the optimization-based non-linear mapping scheme and the symbol scaling schemes in terms of the floating-point operations required, where it is shown mathematically and numerically that compared to the non-linear mapping approach, the complexity of the proposed symbol scaling approach is negligible.

The remainder of this paper is organized as follows. Section II introduces the system model. Both the proposed optimization-based quantized linear beamforming scheme and non-linear mapping scheme by exploiting the constructive interference are presented in Section III. The proposed low-complexity scheme that consists of three stages is presented in Section IV. Section V includes the analysis of the computational complexity for both schemes, and the numerical results are shown in Section VI. Section VII concludes the paper.

Notations: \( a, \mathbf{a}, \) and \( \mathbf{A} \) denote scalar, vector and matrix, respectively. \( (\cdot)^T \) and \( (\cdot)^H \) denote transposition and conjugate transposition of a matrix, respectively. \( \text{card}(\cdot) \) denotes the cardinality of a set. \( j \) denotes the imaginary unit, and \( \text{vec}(\cdot) \) denotes the vectorization operation. \( \mathbf{a}(k) \) denotes the \( k \)-th entry in vector \( \mathbf{a} \). \( |\cdot| \) denotes the modulus of a complex number or the absolute value of a real number. \( ||\cdot||_F \) denotes the Frobenius norm, and \( ||\cdot||_1 \) denotes the 1-norm. \( \mathbb{C}^{n \times n} \) represents an \( n \times n \) matrix in the complex set, and \( \mathbf{I} \) denotes the identity matrix. \( \Re(\cdot) \) and \( \Im(\cdot) \) denote the real and imaginary part of a complex number, respectively.

II. System Model

We consider a multi-user massive MIMO downlink, where 1-bit DACs are employed at the BS, as depicted in Fig. 1. As we focus on the transmit-side processing, ideal ADCs with infinite precision are assumed to be employed at each receiver. The BS with \( N_t \) transmit antennas is communicating with \( K \) single-antenna users simultaneously in the same time-frequency resource, where \( K \ll N_t \). We focus on the transmit beamforming designs and perfect CSI is assumed, while we also numerically study the performance of the proposed schemes with imperfect CSI in Section VI. Following the closely-related literature [32]-[34], the symbol vector is assumed to be from a normalized PSK constellation. We denote the data symbol vector as \( \mathbf{s} \in \mathbb{C}^{K \times 1} \), and the unquantized signal vector that is formed based on \( \mathbf{s} \) as \( \hat{\mathbf{x}}_T \in \mathbb{C}^{N_t \times 1} \). Then, the unquantized signal vector \( \hat{\mathbf{x}}_T \) can be expressed as

\[
\hat{\mathbf{x}}_T = \mathcal{B}(\mathbf{s}),
\]

where \( \mathcal{B} \) denotes a general transformation. When a linear beamforming scheme is employed, \( \mathcal{B} \) represents the linear beamforming matrix that is multiplied to \( \mathbf{s} \) before quantization, while in the case of non-linear beamforming, \( \mathcal{B} \) refers to a non-linear mapping scheme to form the transmit signals dependent on \( \mathbf{s} \). With 1-bit DACs employed, the output signal vector is then obtained as

\[
\mathbf{x}_T = \mathcal{Q}(\hat{\mathbf{x}}_T).
\]

In (2), \( \mathcal{Q} \) denotes the 1-bit quantization on both the real and imaginary part of each entry in \( \hat{\mathbf{x}}_T \). We denote \( x_n, n \in \{1, 2, \ldots, N_t\} \) as the \( n \)-th entry in \( \mathbf{x}_T \), and in this paper each \( x_n \) is normalized to satisfy

\[
x_n \in \left\{ \pm \frac{1}{\sqrt{2N_t}} \pm \frac{1}{\sqrt{2N_t}} j \right\}, \quad \forall n \in \mathcal{N}, \tag{3}
\]

where \( \mathcal{N} = \{1, 2, \ldots, N_t\} \). The above normalization guarantees that \( ||\mathbf{x}_T||_F^2 = 1 \), and we can then express the received signal vector as

\[
\mathbf{y} = \sqrt{P} \cdot \mathbf{H} \mathbf{x}_T + \mathbf{n}, \tag{4}
\]
where $\mathbf{H} \in \mathbb{C}^{K \times N_t}$ denotes the flat-fading Rayleigh channel with each entry following a standard complex Gaussian distribution. $\mathbf{n} \in \mathbb{C}^{K \times 1}$ denotes the additive Gaussian distributed noise vector with zero mean and covariance $\sigma^2 \cdot \mathbf{I}$. $P$ is the total available transmit power per antenna, and for simplicity in this paper we assume uniform power allocation for the antenna array.

III. 1-Bit Transmission Scheme Based on Constructive Interference

In this section, we present the proposed optimization-based schemes by exploiting the constructive interference. The concept of constructive interference and constructive region is first briefly reviewed, followed by the description of the proposed linear and non-linear 1-bit transmission schemes.

A. Constructive Interference and Constructive Region

Constructive interference is defined as the interference that pushes the received signals away from the detection thresholds of the modulation constellation [14]. The exploitation of constructive interference was firstly introduced in [7] to improve the performance of ZF beamforming scheme, and was more recently applied to optimization-based approaches [14], [18] and [19]. It is demonstrated that, as long as the resulting interfered signal is located in the constructive region of the modulation constellation, this increases the distance between the received signal and the detection thresholds, and an improved performance can be obtained. To illustrate this intuitively, in Fig. 2 we depict the constructive region for QPSK, where for simplicity and without loss of generality we focus on one quarter of the normalized QPSK constellation that corresponds to the constellation point $\left(\frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2} \cdot j\right)$. As can be observed, as long as the interfered signal ($\overrightarrow{OB}$ in Fig. 2) is located in the constructive region, the distance to the detection thresholds is increased, and an improved performance can be expected. The formulation of the optimization problem based on the constructive region will be introduced in the following.

B. 1-Bit Transmission Scheme - Linear Beamforming

When a linear beamforming scheme is considered, the unquantized transmit signal vector can be expressed as

$$\hat{x}_T = \mathbf{W}s.$$  (5)

To introduce the proposed scheme, we firstly decompose the channel matrix into

$$\mathbf{H} = \left[\mathbf{h}_1^T, \mathbf{h}_2^T, \cdots, \mathbf{h}_K^T\right]^T,$$  (6)

where each $\mathbf{h}_k \in \mathbb{C}^{1 \times N_t}$ denotes the channel vector of the $k$-th user. Then, the received signal for user $k$ can be obtained as

$$y_k = \sqrt{P} \cdot \mathbf{h}_k x_T + n_k$$
$$= \sqrt{P} \cdot \mathbf{h}_k \mathbf{Q}(\mathbf{W}s) + n_k,$$  (7)

where $n_k$ is the $k$-th entry in $\mathbf{n}$. In this paper, we propose to maximize the distance between the constructive region and the detection thresholds such that the received signals are pushed as far as possible away from the detection thresholds. For the proposed quantized linear scheme, the fully-digital beamforming matrix $\mathbf{W}$ assuming infinite-precision DACs is firstly obtained, followed by the quantization on the resulting transmit signal vector $\hat{x}_T$ to satisfy the constraint on the output signals of 1-bit DACs.

To formulate the desired optimization problem, let us firstly study the analytical constructive interference conditions. In Fig. 2, without loss of generality we denote $\overrightarrow{OA} = t \cdot s_k$ and $t = |\overrightarrow{OA}|$ is the objective to be maximized. We assume the node ‘B’ denotes the noiseless received signal ($\mathbf{h}_k \mathbf{W}s$) that is located in the constructive region, and we further denote $\overrightarrow{OB} = \lambda_k s_k$, where $\lambda_k$ is an introduced complex variable with $|\overrightarrow{OB}| = |\lambda_k|$. We can then obtain that

$$\overrightarrow{OB} = \mathbf{h}_k \mathbf{W}s = \lambda_k s_k.$$  (8)

Based on the fact that $\overrightarrow{OC}$ and $\overrightarrow{CB}$ are perpendicular, we can further obtain $\overrightarrow{OC}$ and $\overrightarrow{CB}$, expressed as

$$\overrightarrow{OC} = \Re(\lambda_k) s_k, \quad \overrightarrow{CB} = j \cdot \Im(\lambda_k) s_k,$$  (9)

where geometrically the imaginary unit ‘$j$’ denotes a phase rotation of $90^\circ$ along the anti-clockwise direction. As the nodes ‘O’, ‘A’, and ‘C’ are co-linear, we can then express $\overrightarrow{AC}$ as

$$\overrightarrow{AC} = |\Re(\lambda_k) - t| s_k.$$  (10)

Based on the expression of $\overrightarrow{AC}$ and $\overrightarrow{CB}$, $\tan \theta_{AB}$ is obtained as

$$\tan \theta_{AB} = \frac{|\overrightarrow{CB}|}{|\overrightarrow{AC}|} = \frac{|\Im(\lambda_k) s_k|}{|\Re(\lambda_k) - t| s_k} = \frac{|\Im(\lambda_k)|}{\Re(\lambda_k) - t}.$$  (11)
In Fig. 2, it is geometrically observed that to have ‘B’ located in the constructive region is equivalent to the following condition:

\[
\theta_{AB} \leq \theta_t \Rightarrow \tan \theta_{AB} \leq \tan \theta_t \Rightarrow \frac{\Im(\lambda_k)}{\Re(\lambda_k) - t} \leq \tan \theta_t \Rightarrow [\Re(\lambda_k) - t] \tan \theta_t \geq |\Im(\lambda_k)|. 
\]

(12)

For M-PSK modulation, based on the geometry of the modulation constellation it is easy to obtain the threshold angle \( \theta_t \), given by

\[
\theta_t = \pi M. 
\]

(13)

With the obtained conditions to achieve constructive interference in (8) and (12), the unquantized linear beamforming optimization that maximizes the distance between the constructive region and the detection thresholds can be formulated as

\[
P_1 : \max_w t \quad \text{s.t.} \quad h_k w_s = \lambda_k s_k, \forall k \in K \\quad [\Re(\lambda_k) - t] \tan \theta_t \geq |\Im(\lambda_k)|, \forall k \in K \\
\|w_s\|_F \leq \sqrt{p_0} \\
t \geq 0
\]

(14)

where \( K = \{1, 2, \ldots, K\} \), and \( \|w_s\|_F \leq \sqrt{p_0} \) is the instantaneous power constraint on the beamformer as the beamforming is dependent on the data symbols. Due to the existence of the subsequent 1-bit quantization operation, \( p_0 \) in \( P_1 \) can be any positive value, and this will not have an impact on the final obtained quantized signal vector \( x_r \). \( P_1 \) is a second-order cone programming (SOCP) optimization problem and can be efficiently solved with convex optimization tools, where we can further obtain the following proposition in the case of massive MIMO.

**Proposition:** In the case of massive MIMO, the optimality conditions for each \( \lambda_k \) and \( t \) of the optimization problem \( P_1 \) are obtained as

1) \( |\Im(\lambda_k)| = 0, \forall k \in K; \)
2) \( t^* = \lambda_1^* = \lambda_2^* = \cdots = \lambda_K^* = \sqrt{\frac{N_r p_0}{R}}. \)

Proof: We prove the above proposition by analyzing the optimization problem \( P_1 \) with Lagrangian and KKT conditions. We firstly transform \( P_1 \) into a standard minimization problem, given by

\[
P_2 : \min_{w} -t \quad \text{s.t.} \quad h_k \sum_{i=1}^{K} w_i s_i - \lambda_k s_k = 0, \forall k \in K \\
|\Im(\lambda_k)| - |\Re(\lambda_k) - t| \tan \theta_t \leq 0, \forall k \in K \\
\sum_{i=1}^{K} s_i^H w_i^H w_i s_i - p_0 \leq 0
\]

(15)

where we note that the constraint on \( t \) in \( P_1 \) can be omitted in the above formulation, and the beamforming matrix \( W \) is decomposed into \( W = [w_1, w_2, \ldots, w_K] \). We can then express the Lagrangian of \( P_2 \) as

\[
\mathcal{L}(w_i, t, \delta_k, \mu_k, \mu_0) = -t + \sum_{k=1}^{K} \delta_k (h_k \sum_{i=1}^{K} w_i s_i - \lambda_k s_k) + \mu_0 \left( \sum_{i=1}^{K} s_i^H w_i^H w_i s_i - p_0 \right),
\]

(16)

where \( \mu_0, \delta_k \) and \( \mu_k \) are the dual variables, and \( \mu_0 \geq 0, \mu_k \geq 0, \forall k \in K \). Based on the Lagrangian in (16), the KKT conditions for optimality are then obtained as

\[
\frac{\partial \mathcal{L}}{\partial w_i} = -1 + \sum_{k=1}^{K} \mu_k = 0
\]

(17a)

\[
\frac{\partial \mathcal{L}}{\partial t} = \sum_{k=1}^{K} \delta_k \cdot h_k = 0
\]

(17b)

\[
\mu_k [\Im(\lambda_k) - |\Re(\lambda_k) - t| \tan \theta_t + t \cdot \tan \theta_t] = 0, \forall k \in K
\]

(17c)

\[
0 = \mu_0 \left( \sum_{i=1}^{K} s_i^H w_i^H w_i s_i - p_0 \right)
\]

(17d)

Based on (17b), firstly it is easily obtained that \( \mu_0 \neq 0 \) which with the fact that \( \mu_0 \geq 0 \) further leads to \( \mu_0 > 0 \). Then, we can obtain \( w_i^H \) as

\[
w_i^H = -\frac{1}{\mu_0} \cdot \sum_{k=1}^{K} \delta_k h_k s_i, \forall i \in K.
\]

(18)

By denoting

\[
a_k = -\frac{\delta_k}{\mu_0}, \forall k \in K,
\]

(19)

\( w_i \) can be obtained from (18) and expressed as

\[
w_i = \sum_{k=1}^{K} a_k h_k^H s_i, \forall i \in K.
\]

(20)

Then, with the expression of each \( w_i \), the beamforming matrix \( W \) is obtained in a compact form as

\[
W = [w_1, w_2, \ldots, w_K] = \left[ \sum_{k=1}^{K} a_k h_k^H \right] \left[ s_1^H, s_2^H, \ldots, s_K^H \right] = [h_1^H, h_2^H, \ldots, h_K^H] [a_1, a_2, \ldots, a_K]^T \left[ s_1^H, s_2^H, \ldots, s_K^H \right] = H^H A s^H. 
\]

(21)

In order to obtain \( A \), we firstly rewrite (8) in a compact form, which is expressed as

\[
HWs = \text{diag}(\lambda_k)s.
\]

(22)

Then, by substituting (21) into (22), the matrix \( A \) can be obtained based on \( \lambda_k \), given by

\[
HH^H A s^H = \text{diag}(\lambda_k)s
\]

\[
\Rightarrow A = \frac{1}{K} \cdot (HH^H)^{-1} \text{diag}(\lambda_k)s.
\]

(23)
The beamforming matrix $\mathbf{W}$ is then obtained as
\[
\mathbf{W} = \frac{1}{K} \cdot \mathbf{H}^H \left( \mathbf{H} \mathbf{H}^H \right)^{-1} \text{diag}(\lambda) \mathbf{s} \mathbf{s}^H.
\] (24)

Based on the fact that $\mu_0 \neq 0$, it is obtained from (17e) that the power constraint of the optimization problem $\mathcal{P}_1$ is strictly active, which further leads to
\[
\|\mathbf{W}s\|_F = \sqrt{p_0} \\
\Rightarrow tr \left\{ \mathbf{W} \mathbf{s} \mathbf{s}^H \mathbf{W} \right\} = p_0 \quad (25) \\
\Rightarrow \mathbf{s}^H \mathbf{W} \mathbf{W}^H \mathbf{s} = p_0.
\]

Then, by substituting (24) into (25), we obtain that
\[
s^H \text{diag}(\lambda^H) \left( \mathbf{HH}^H \right)^{-1} \text{diag}(\lambda) \mathbf{s} = p_0 \\
\Rightarrow \text{vec}^T \left( \lambda^H \right) \text{diag}(s^H) \left( \mathbf{HH}^H \right)^{-1} \text{diag}(s) \text{vec}(\lambda) = p_0 \\
\Rightarrow [\lambda_1^H, \lambda_2^H, ..., \lambda_K^H] \cdot \mathbf{T} \cdot [\lambda_1, \lambda_2, ..., \lambda_K]^T = p_0,
\]
where $\mathbf{T}$ is defined as
\[
\mathbf{T} = \text{diag}(s^H) \left( \mathbf{HH}^H \right)^{-1} \text{diag}(s).
\] (27)

In the case of massive MIMO, as $N_t \to \infty$, the favourable propagation property gives us that
\[
\mathbf{HH}^H \approx N_t \cdot \mathbf{I} \Rightarrow \left( \mathbf{HH}^H \right)^{-1} \approx \frac{1}{N_t} \cdot \mathbf{I},
\] (28)

based on which $\mathbf{T}$ is further transformed into
\[
\mathbf{T} \approx \frac{1}{N_t} \cdot \text{diag}(s^H) \text{diag}(s) = \frac{1}{N_t} \cdot \mathbf{I}.
\] (29)

From the result in (29), (26) can be expanded and further transformed into
\[
\frac{1}{N_t} \cdot \left| \lambda_1 \right|^2 + \left| \lambda_2 \right|^2 + \cdots + \left| \lambda_K \right|^2 = p_0.
\] (30)

To maximize $t$, as per (12) and (30) it is then easily obtained that the optimality is achieved when each $\lambda_k^*$ is real and identical, given by
\[
t^* = \lambda_1^* = \cdots = \lambda_K^* = \sqrt{\frac{N_t \cdot p_0}{K}},
\] (31)

which completes the proof.

By substituting (31) into (24), the optimal beamforming matrix $\mathbf{W}^*$ can be expressed as
\[
\mathbf{W}^* = \sqrt{\frac{N_t \cdot p_0}{K^3}} \cdot \mathbf{H}^H \left( \mathbf{HH}^H \right)^{-1} \mathbf{s} \mathbf{s}^H.
\] (32)

Then, with $\mathbf{W}^*$ obtained, the output signal vector that satisfies 1-bit DAC transmission is given as
\[
\mathbf{x}_T = \mathbf{Q} \left( \mathbf{W}^* \mathbf{s} \right) \\
= \mathbf{Q} \left( \sqrt{\frac{N_t \cdot p_0}{K^3}} \cdot \mathbf{H}^H \left( \mathbf{HH}^H \right)^{-1} \mathbf{s} \mathbf{s}^H \mathbf{s} \right) \\
= \mathbf{Q} \left( \sqrt{\frac{N_t \cdot p_0}{K^3}} \cdot \mathbf{H}^H \left( \mathbf{HH}^H \right)^{-1} \mathbf{s} \right).
\] (33)

The intuition from the above proposition and (33) is that the quantized linear scheme based on the constructive interference is equivalent to the conventional quantized ZF scheme in the case of massive MIMO with 1-bit quantization, which suffers an error floor at high SNR regime [32]. This then motivates the proposed non-linear mapping scheme that achieves an improved performance in the following.

\section*{C. 1-Bit Transmission Scheme - Non-linear Mapping}

We proceed to introduce the optimization-based non-linear mapping scheme for massive MIMO with 1-bit DACs. The non-linear mapping scheme refers to the approach where, instead of designing a linear beamforming matrix that is multiplied to $s$ before quantization, we directly design the quantized transmit signal vector $\mathbf{x}_T$. Then, the optimization based on the constructive interference can be formulated as
\[
\mathcal{P}_3 : \max_{\mathbf{x}_T} t \\
s.t. \ h_k \mathbf{x}_T = \lambda_k s_k, \ \forall k \in K \\
\left| \Re \left( \lambda_k \right) - t \right| \tan \theta_t \geq \left| \Im \left( \lambda_k \right) \right|, \ \forall k \in K \\
x_n \in \left\{ \pm \frac{1}{\sqrt{2N_t}} \pm \frac{1}{\sqrt{2N_t}} \right\}, \ \forall n \in \mathcal{N}, \\
t \geq 0
\]

It is observed that the optimization problem $\mathcal{P}_3$ is non-convex due to the output signal constraint for the 1-bit DACs in (34). To solve the above non-convex optimization, we adopt a two-step approach.

1) \textit{Relaxation:} In the first step, we relax the strict modulus constraint on each $x_n$ for both the real and imaginary part, and the resulting relaxed constraint can be expressed as
\[
\left| \Re \left( x_n \right) \right| \leq \frac{1}{\sqrt{2N_t}}, \ \left| \Im \left( x_n \right) \right| \leq \frac{1}{\sqrt{2N_t}}, \ \forall n \in \mathcal{N}.
\] (35)

The optimization problem $\mathcal{P}_3$ is then reformulated into a relaxed version $\mathcal{P}_4$, given by
\[
\mathcal{P}_4 : \max_{\mathbf{x}_T} t \\
s.t. \ h_k \hat{x}_T = \lambda_k s_k, \ \forall k \in K \\
\left| \Re \left( \lambda_k - t \right) \right| \tan \theta_t \geq \left| \Im \left( \lambda_k \right) \right|, \ \forall k \in K \\
\left| \Re \left( \hat{x}_n \right) \right| \leq \frac{1}{\sqrt{2N_t}}, \ \forall n \in \mathcal{N}, \\
\left| \Im \left( \hat{x}_n \right) \right| \leq \frac{1}{\sqrt{2N_t}}, \ \forall n \in \mathcal{N}, \\
t \geq 0
\]

where we denote $\hat{x}_n$ as the $n$-th entry in the relaxed transmit signal vector $\mathbf{x}_T$. The resulting $\mathcal{P}_4$ is convex and can be solved with convex optimization tools.

2) \textit{Normalization:} The solution obtained from the relaxed optimization $\mathcal{P}_4$ cannot always guarantee the equality on both the real and imaginary part of $\hat{x}_n$. To force the constraint of 1-bit transmission, the elements of the 1-bit DAC output $\mathbf{x}_T$ are obtained as
\[
x_n = \frac{\Re \left( \hat{x}_n \right)}{\sqrt{2N_t}} + \frac{\Im \left( \hat{x}_n \right)}{\sqrt{2N_t}}, \ \forall n \in \mathcal{N}.
\] (37)

We further note that, while we perform a relaxation on the constant modulus constraint on each $x_n$ in $\mathcal{P}_3$, it turns out that most entries of the obtained $\mathbf{x}_T$ from the relaxed problem
\begin{table}  
\centering  
\begin{tabular}{|c|c|c|c|c|}  
\hline  
Antenna number $N_t$ & 16 & 32 & 48 & 64 \\  
\hline  
Ratio $\eta$ & 20.52\% & 10.8\% & 7.28\% & 5.46\% \\  
\hline  
\end{tabular}  
\caption{Value of $\eta$ with respect to the number of transmit antennas, $K = 4$, 500 channel realizations}  
\end{table}

$P_4$ already meet the requirement for 1-bit quantization, i.e. strict equality is satisfied for these entries in the case of massive MIMO. Accordingly, there exist only a small number of $\hat{x}_n$ that need to be normalized to satisfy the constraint on the output signals for 1-bit DACs. Moreover, to evaluate the deviation of the relaxed optimization $P_4$ from the original problem $P_3$, we introduce $n_{iR}$ and $n_{i3}$ that are defined as the number of entries in the obtained $\hat{x}_T$ whose absolute values are smaller than $\frac{1}{\sqrt{2N_t}}$ for the real and imaginary part, respectively. We further introduce

$$\eta = \frac{n_{iR} + n_{i3}}{2N_t}$$

as the ratio of the number of entries that do not satisfy the 1-bit transmission to the total number of entries in $\hat{x}_T$, and this ratio therefore represents the deviation of the solution obtained by the relaxed problem from the original problem. We have $0 \leq \eta \leq 1$, and $P_4$ is equivalent to $P_3$ if $\eta = 0$. It is also observed that a smaller value of $\eta$ means that the relaxed optimization is closer to the original optimization.

To study this numerically, we present the value of $\eta$ with respect to the number of antennas in Table I, where we have assumed a total number of $K = 4$ users in the downlink system, and the result is based on 500 channel realizations. It is observed that the ratio $\eta$ decreases with the increase in the number of transmit antennas, which means that the solution obtained via the relaxed optimization problem $P_4$ can be regarded as asymptotically optimal with an increasing number of transmit antennas in the case of massive MIMO.

IV. PROPOSED LOW-COMPLEXITY SYMBOL SCALING SCHEME

While the above non-linear mapping scheme can be relaxed into a convex optimization problem, the corresponding computational complexity is still prohibitively high as the variable dimension is equal to the number of transmit antennas. We study this mathematically and numerically in Section V and VI, respectively. Therefore in this section, we propose a three-stage symbol scaling scheme, which requires much reduced complexity. It will be shown in the numerical results that for the small-scale MIMO systems, the low-complexity scheme even outperforms the optimization-based non-linear mapping scheme in Section III and achieves the best performance, since no relaxation or normalization is required for this scheme. In the case of massive MIMO, it also achieves a comparable performance to the proposed non-linear mapping scheme.

A. A New Look at the Constructive Interference Criteria

To introduce the proposed symbol scaling scheme, we firstly perform a coordinate transformation on the formulation of the constructive interference constraint. To be specific, we firstly decompose each data symbol $s_k$ along its two corresponding detection thresholds of the modulation constellation, given by

$$s_k = s_k^R + s_k^I,$$

where $s_k^R$ and $s_k^I$ are both complex values, and denoted as the two bases that are parallel to the two detection thresholds that correspond to the constellation point $s_k$. In the following, for simplicity we shall use $s_k^R$ and $s_k^I$ to denote the two bases. This is also shown geometrically in both Fig. 2 and Fig. 3 where we employ QPSK and 8-PSK modulation as examples, respectively. As observed in both figures, we decompose ‘OA’ that represents the data symbol $s_k$ along its detection thresholds into ‘OF’ and ‘OG’. For QPSK, based on Fig. 2 it is easy to observe that the real and imaginary axes are the detection thresholds, which leads to

$$\overrightarrow{OF} = s_k^R = \frac{1}{\sqrt{2}} \quad \overrightarrow{OG} = s_k^I = \frac{1}{\sqrt{2}} \cdot j$$

for the corresponding constellation point ‘A’. For 8-PSK, ‘OD’ and ‘OE’ in Fig. 3 are the detection thresholds for the constellation point ‘A’. Then, with $\theta_1 = \pi / 8$ for 8-PSK we can obtain the bases $s_k^R$ and $s_k^I$ that correspond to the constellation point ‘A’ as

$$\overrightarrow{OF} = s_k^R = \frac{e^{j \frac{\pi}{8}}}{e^{j \frac{\pi}{8}} + e^{j \frac{3\pi}{8}}} = a_k + b_k \cdot j,$$

$$\overrightarrow{OG} = s_k^I = \frac{e^{j \frac{3\pi}{8}}}{e^{j \frac{\pi}{8}} + e^{j \frac{3\pi}{8}}} = c_k + d_k \cdot j.$$

where $(a_k, b_k)$ and $(c_k, d_k)$ denote the coordinates of $s_k^R$ and $s_k^I$ in the conventional real-imaginary complex plane, respectively. The extension to other constellation points and higher order PSK modulations can be easily obtained in a similar way.

Then for each $k$, instead of employing a complex scaling value $\lambda_k$ that is multiplied to $s_k$, with the above formulation (39)-(41) we introduce a symbol scaling approach where we decompose (8) along the two corresponding detection thresholds of $s_k$, given by

$$h_k x_T = \alpha_k^R s_k^R + \alpha_k^I s_k^I,$$
are two introduced scaling factors that are multiplied to the bases $s_k^R$ and $s_k^I$, respectively. We can then observe that a larger value of $\alpha_k^R$ or $\alpha_k^I$ therefore represents a larger distance to the other detection threshold, and we further denote $(\alpha_k^R, \alpha_k^I)$ as the coordinate of the node ‘B’ in the complex plane expanded by the bases $s_k^R$ and $s_k^I$. By expanding (42) using the coordinate transformation, we can obtain the generic expression of $\alpha_k^R$ and $\alpha_k^I$ as a function of the transmit signal vector, given by (see Appendix)

\[
\alpha_k^R = \frac{d_k \Re(h_k) - c_k \Im(h_k)}{a_k d_k - b_k c_k} x_T^R - \frac{d_k \Im(h_k) + c_k \Re(h_k)}{a_k d_k - b_k c_k} x_T^I, \\
\alpha_k^I = \frac{a_k \Re(h_k) - b_k \Im(h_k)}{a_k d_k - b_k c_k} x_T^R + \frac{a_k \Im(h_k) + b_k \Re(h_k)}{a_k d_k - b_k c_k} x_T^I.
\]

In (44), for simplicity we have employed the following denotations

\[
x_T^R = \Re(x_T^T), \quad x_T^I = \Im(x_T^T).
\]

By further denoting

\[
A_k = \frac{d_k \Re(h_k) - c_k \Im(h_k)}{a_k d_k - b_k c_k}, \quad B_k = -\frac{d_k \Im(h_k) + c_k \Re(h_k)}{a_k d_k - b_k c_k}, \\
C_k = \frac{a_k \Re(h_k) - b_k \Im(h_k)}{a_k d_k - b_k c_k}, \quad D_k = \frac{a_k \Im(h_k) + b_k \Re(h_k)}{a_k d_k - b_k c_k},
\]

the formulation of (44) is simplified into

\[
\alpha_k^R = A_k x_T^R + B_k x_T^I, \\
\alpha_k^I = C_k x_T^R + D_k x_T^I.
\]

By defining

\[
R_k = [ A_k \quad B_k ], \quad I_k = [ C_k \quad D_k ],
\]

and

\[
x = \begin{bmatrix} (x_T^R)^T \quad (x_T^I)^T \end{bmatrix}^T, \quad \Lambda = \begin{bmatrix} \alpha_1^R, \ldots, \alpha_K^R, \alpha_1^I, \ldots, \alpha_K^I \end{bmatrix}^T,
\]

(47) can be further expressed in a compact form as

\[
\Lambda = Mx,
\]

where $M$ is given by

\[
M = [ R_1^T \cdots R_K^T I_1^T \cdots I_K^T ]^T.
\]

With the above formulation, we can then construct the optimization problem as

\[
\mathcal{P}_5 : \max_{x} \min_{l} \alpha_l \\
\text{s.t.} \quad \Lambda = Mx \\
\alpha_l \geq 0, \quad \forall l \in \mathcal{L} \\
x_i^E \in \left\{ -\frac{1}{\sqrt{2N_t}}, \cdots, \frac{1}{\sqrt{2N_t}} \right\}, \quad \forall i \in \mathcal{I}
\]

where we have omitted $\Re$ and $\Im$ in the expression of the entries of $\Lambda$, and simply denote $\alpha_l$ as its $l$-th entry. In $\mathcal{P}_5$, $\mathcal{L} = \{1, 2, \cdots, 2K\}$, $x_i^E$ denotes the $i$-th entry in $x$ and $\mathcal{I} = \{1, 2, \cdots, 2N_t\}$. The above optimization problem $\mathcal{P}_5$ is interpreted as follows: we aim to maximize the minimum value of $\alpha_l$ by selecting each $x_i^E$ as either $-\frac{1}{\sqrt{2N_t}}$ or $\frac{1}{\sqrt{2N_t}}$. With the above problem formulation, the relaxation-normalization process on the transmit signals is no longer needed. The above formulation motivates us to propose the following low-complexity scheme, which consists of three stages: an initialization stage, an allocation stage, and a refinement stage, all presented in the following in detail.

B. Initialization Stage

In the initialization stage, we directly select the value of $x_i^E$ for some $i$ by simple observation. To achieve this, we firstly decompose (50) into

\[
\Lambda = \sum_{i=1}^{2N_t} M_i x_i^E,
\]

where we decompose $M$ into

\[
M = \begin{bmatrix} M_1 & M_2 & \cdots & M_{2N_t} \end{bmatrix},
\]

with each $M_i \in C^{2K \times 1}$. Then, we have the following observation.

Observation: As long as all the entries of $M_i$ share the same sign, then it is optimal to set the sign of the corresponding $x_i^E$ equal to that of $M_i$, as in this case the values of each entry in $\Lambda$ are guaranteed to increase.

Then, the corresponding $x_i^E$ is obtained as

\[
x_i^E = \frac{\text{sgn}(M_i)}{\sqrt{2N_t}}, \quad \forall i \in \mathcal{S},
\]

where $\text{sgn}(a)$ defines a sign function for vector and is only valid when each entry in the vector $a$ has the same sign. $\mathcal{S}$ denotes the set that consists of the column numbers of $M$ that satisfy the sign-identity condition. We further introduce a column vector $t$ that represents a temporary value of $\Lambda$, given by

\[
t = \sum_{i \in \mathcal{V}} M_i x_i^E,
\]

where the set $\mathcal{V}$ consists of the column numbers of $M$ whose corresponding $x_i^E$ have been allocated a value. We note that when $\text{card}(\mathcal{V}) = 2N_t$, we have $t = \Lambda$.

In the case that no column in $M$ satisfies the sign-identity condition, in the initialization stage we select only one column, i.e. $\text{card}(\mathcal{S}) = 1$, with the following criterion:

\[
i = \arg \max_{i \in \mathcal{L}} \|M_i\|_1,
\]

where we select the column that has the maximum effect on the value of $\Lambda$. Then, the value of the corresponding $x_i^E$ is set as

\[
x_i^E = \frac{\text{sgn}(\|M_i\|_1)}{\sqrt{2N_t}}.
\]

In the initialization stage, we have $\mathcal{V} = \mathcal{S}$ or $\text{card}(\mathcal{V}) = 1$. We summarize the algorithm for the initialization stage in Algorithm 1.
Algorithm 1 Initialization Stage

input : \( \mathbf{s}, \mathbf{H} \)  
output : \( \mathbf{t}, \mathbf{V} \)
Decompose each \( s_k = s_k^R + s_k^I \) based on modulation type;  
Obtain \( \mathbf{M} \) based on (42)-(51);  
Find \( \mathbf{M}_i \) that satisfies the sign-identity condition;  
Obtain \( \mathcal{S} \);  
if \( \mathcal{S} \neq \emptyset \) then  
\( x_i^E = \frac{\text{sgn}(\mathbf{M}_i)}{\sqrt{2N_t}}, \forall i \in \mathcal{S} \);  
\( \mathcal{V} = \mathcal{S} \);  
else  
Obtain \( i \) based on (57), \( x_i^E = \frac{\text{sgn}([\mathbf{M}_i]_i)}{\sqrt{2N_t}} \);  
\( \mathcal{V} = \{ i \} \);  
end if  
Calculate \( \mathbf{t} \) based on (56).

C. Allocation Stage

At this stage we allocate the value of each \( x_i^E \) for the residual \( i \) that belong to \( \mathcal{W} \), where we define the set \( \mathcal{W} \) as  
\[ \mathcal{W} = \{ i \mid i \in \mathcal{I} \text{ and } i \notin \mathcal{V} \}. \]  
(59)
\( \mathcal{W} \) consists of those \( x_i^E \) whose values have not been allocated in the initialization stage. In the following allocation stage, we consider both a ‘Sum-Max’ and a ‘Max-Min’ criteria for the allocation scheme.

1) Sum-Max: For the allocation scheme based on the ‘Sum-Max’ criterion, instead of considering a max-min optimization as in \( \mathcal{P}_3 \), we consider a sum-max optimization where the objective function is constructed as  
\[ \mathcal{F}(\mathbf{x}) = \text{sum}(\mathbf{A}), \]  
(60)
where \( \text{sum}(\mathbf{a}) \) returns the sum of the entries in a column vector \( \mathbf{a} \). Then, based on (50) the objective can be further transformed into  
\[ \mathcal{F}(\mathbf{x}) = \text{mx} = \sum_{i=1}^{2N_t} \mathbf{m}(i) x_i^E, \]  
(61)
where \( \mathbf{m} \in \mathbb{C}^{1 \times 2N_t} \) is the sum of entries in \( \mathbf{M} \) by row. Each \( \mathbf{m}(i) \) denotes the \( i \)-th entry in \( \mathbf{m} \), given by  
\[ \mathbf{m}(i) = \sum_{l=1}^{2K} \mathbf{M}_i(l). \]  
(62)
It is then easy to observe that \( \mathcal{F}(\mathbf{x}) \) is maximized when the sign of each \( x_i^E \) is the same as that of \( \mathbf{m}(i) \), and therefore the optimal \( x_i^E \) for the ‘Sum-Max’ criterion is given by  
\[ x_i^E = \frac{\text{sgn}([\mathbf{M}_i]_i)}{\sqrt{2N_t}}, \forall i \in \mathcal{W}. \]  
(63)
While the above solution guarantees that the sum of \( \alpha_l \) is maximized, it does not specifically consider each value of \( \alpha_l \), which may lead to performance loss. Indeed, it is possible that the value of one \( \alpha_l \) can be very small or even negative. This is the reason why the refinement in Section IV-D is further introduced. The algorithm for the allocation stage based on ‘Sum-Max’ is summarized in Algorithm 2.

Algorithm 2 Allocation Stage - ‘Sum-Max’

input : \( \mathcal{V}, \mathbf{M} \)  
output : \( \mathbf{x}_{\text{sum-max}} \)
Calculate \( \mathcal{W} \) based on (59);  
Calculate \( \mathbf{m} \) and each \( \mathbf{m}(i) \) based on (61), (62);  
Allocate \( x_i^E = \frac{\text{sgn}[\mathbf{m}(i)]_{i}}{\sqrt{2N_t}}, \forall i \in \mathcal{W} \);  
Obtain \( \mathbf{x} \), denoted as \( \mathbf{x}_{\text{sum-max}} \).

2) Max-Min: For the allocation scheme with ‘Max-Min’ criterion, in each step we aim to improve the minimum value in \( \mathbf{A} \) as much as possible. Denoting \( q \) as the row number of the minimum entry in \( t \) obtained in the initialization stage, we have  
\[ t(q) = \min(t), \]  
(64)
where \( \min(t) \) returns the minimum value in \( t \). Subsequently, we iteratively select \( \mathbf{M}_i \) with the largest absolute value in the \( q \)-th row, given by  
\[ \hat{i} = \arg \max_{i \in \mathcal{W}} |\mathbf{M}_i(q)|, \]  
(65)
and the corresponding \( x_i^E \) is then obtained as  
\[ x_i^E = \frac{\text{sgn}([\mathbf{M}_i]_i)}{\sqrt{2N_t}}. \]  
(66)
Then, we update \( \mathcal{V} \) and \( t \), and based on the updated \( t \) we repeat the above procedure until \( \mathcal{V} = \mathcal{I} \). This means that each entry in \( x \) has been allocated, and the algorithm for the allocation stage based on ‘Max-Min’ is summarized in Algorithm 3.

Algorithm 3 Allocation Stage - ‘Max-Min’

input : \( \mathcal{V}, \mathbf{M}, \mathbf{t} \)  
output : \( \mathbf{x}_{\text{max-min}} \)
while \( \mathcal{V} \neq \mathcal{I} \) do  
Calculate \( \mathcal{W} \) based on (59);  
Obtain \( q \) that satisfies \( t(q) = \min(t) \);  
Find \( \hat{i} = \arg \max_{i \in \mathcal{W}} |\mathbf{M}_i(q)| \);  
Allocate \( x_i^E = \frac{\text{sgn}([\mathbf{M}_i]_i)}{\sqrt{2N_t}} \);  
Update \( \mathcal{V} \) and \( t \);  
end while  
Obtain \( \mathbf{x} \), denoted as \( \mathbf{x}_{\text{max-min}} \).

D. Refinement Stage

In the refinement stage, we check whether the performance based on the obtained signal vector in the allocation stage can be further improved based on the greedy algorithm. To introduce the refinement process, we denote the obtained expanded 1-bit signal vector after the allocation stage as \( \mathbf{x} \) (obtained based on either the ‘Sum-Max’ or the ‘Max-Min’ criterion). First, we sequentially change the sign of one entry (for example \( x_i^E \) in \( \mathbf{x} \)) at a time while fixing the signs of other entries in \( \mathbf{x} \), and denote the modified signal vector as \( \mathbf{x}(i) \). We then compare the minimum value in \( \mathbf{A} \) obtained by the modified \( \mathbf{x}(i) \) with the minimum value in the original \( \mathbf{A} \) obtained by \( \mathbf{x}(0) \). The sign of \( x_i^E \) is selected as the one that
returns a larger minimum value in $\mathbf{A}$. The refinement process is sequentially performed for each entry in $\mathbf{x}_{(0)}$. The algorithm for the refinement stage is then shown in Algorithm 4.

### Algorithm 4 Refinement Stage

**Input:** $\mathbf{x}_{\text{sum-max}}$ (or $\mathbf{x}_{\text{max-min}}$)  
**Output:** $\mathbf{x}_t$  
1. Denote $\mathbf{x}_{(0)} = \mathbf{x}_{\text{sum-max}}$ (or $\mathbf{x}_{\text{max-min}}$);  
2. for $i = 1 : 2N_t$ do  
   - Calculate $\Lambda_{(0)} = \mathbf{M}\mathbf{x}_{(0)}$;  
   - Obtain $\mathbf{x}_{(i)} = [x_{(i)}^1, \ldots, x_{(i)}^{E_{\text{max}}}, x_{(i)}^{E_{\text{max}} + 1}, \ldots, x_{(i)}^{E_{\text{max}} + 3}]^T$;  
   - Calculate $\Lambda_{(i)} = \mathbf{M}\mathbf{x}_{(i)}$;  
   - if $\min(\Lambda_{(i)}) > \min(\Lambda_{(0)})$ then  
     - $x_{(i)}^E \leftarrow -x_{(i)}^E$;  
     - Update $\mathbf{x}_{(0)}$;  
   - end if  
3. end for  
4. Obtain $\mathbf{x}_t$ based on the updated $\mathbf{x}_{(0)}$.

The refinement stage is performed for the signal vectors obtained by both the ‘Sum-Max’ and ‘Max-Min’ criteria independently. The final output signal vector of the proposed symbol scaling scheme that generates the best performance is then selected between the signal vectors obtained with these two criteria.

### E. Algorithm

Based on the above description, the algorithm for the three-stage symbol scaling scheme is summarized in Algorithm 5, where the final output signal vector of the proposed symbol scaling scheme that generates the best performance is selected within the signal vectors obtained by the ‘Sum-Max’ and ‘Max-Min’ criteria.

### Algorithm 5 The Proposed Symbol Scaling Scheme

**Input:** $\mathbf{s}$, $\mathbf{H}$  
**Output:** $\mathbf{x}_t$  
1. **Initialization Stage**  
   - Obtain $\mathcal{U}$, $\mathbf{M}$, and $t$ with Algorithm 1;  
2. **Allocation Stage**  
   - 1. ‘Sum – Max’:  
     - Obtain $\mathbf{x}_{\text{sum-max}}$ with Algorithm 2;  
   - 2. ‘Max – Min’:  
     - Obtain $\mathbf{x}_{\text{max-min}}$ with Algorithm 3;  
3. **Refinement Stage**  
   - Update both $\mathbf{x}_{\text{sum-max}}$ and $\mathbf{x}_{\text{max-min}}$ with Algorithm 4;  
   - Calculate $\Lambda_s = \mathbf{M}\mathbf{x}_{\text{sum-max}}$ and $\Lambda_m = \mathbf{M}\mathbf{x}_{\text{max-min}}$;  
   - if $\min(\Lambda_s) > \min(\Lambda_m)$ then  
     - $\mathbf{x} = \mathbf{x}_{\text{sum-max}}$;  
   - else  
     - $\mathbf{x} = \mathbf{x}_{\text{max-min}}$;  
   - end if  
   - Decompose $\mathbf{x} = \begin{bmatrix} (\mathbf{x}_1^q)^T & (\mathbf{x}_2^q)^T \end{bmatrix}^T$;  
   - Output $\mathbf{x}_t = \mathbf{x}_1^q + \mathbf{x}_2^q \cdot j$.

### V. Computational Complexity Analysis

In this section we study the computational costs of the proposed schemes in terms of the floating operations required. As a reference, we also study the complexity of the exhaustive search scheme and the non-linear ‘Pokemon’ scheme in [36].

#### A. Exhaustive Search

For massive MIMO transmission with 1-bit quantization, the output signal on each antenna element has 4 potential values, and therefore the complexity of the exhaustive search scheme is obtained as
\[
C_E = \mathcal{O}\left\{4^{N_t}\right\} = \mathcal{O}\left\{2^{2N_t}\right\}.
\]  
(67)

It is easy to conclude that in the case of massive MIMO, the exhaustive search scheme is inapplicable due to the overwhelmingly high computational cost.

#### B. Optimization-based Non-linear Mapping $\mathcal{P}_1$

For the proposed non-linear mapping scheme, in the relaxation stage the complexity is dominated from solving the relaxed convex problem $\mathcal{P}_5$ via the interior-point method [38]. It has been shown in [59] that the complexity of the interior-point method is given by
\[
C_1 = \mathcal{O}\left\{\sqrt{NM} \left( N^3M^2 + N^2M^3 \right) \right\},
\]  
(68)

where $N$ is the number of variables, and $M$ is the dimension of the variable. Based on our problem $\mathcal{P}_5$, we obtain $M = N_t$ and $N = 1$, which leads to
\[
C_{1,N} = \mathcal{O}\left\{\sqrt{N_t} \left( N_t^3 + N_t^2 \right) \right\} = \mathcal{O}\left\{N_t^{3.5} + N_t^{2.5}\right\}.
\]  
(69)

In the normalization stage, the dominant complexity comes from the search for the signals that do not satisfy the output constraint for the 1-bit transmission. There are a total number of $2N_t$ entries in $\hat{x}_T$ including both the real and imaginary part, and therefore a one-dimensional search of $2N_t$ entries is required. Then, the resulting complexity is obtained as
\[
C_{N}^2 = \mathcal{O}\left\{2N_t\right\},
\]  
(70)

which leads to the total computational cost for the optimization-based non-linear mapping scheme as
\[
C_N = C_{1,N} + C_{N}^2 = \mathcal{O}\left\{N_t^{3.5} + N_t^{2.5}\right\} + \mathcal{O}\left\{2N_t\right\}.
\]  
(71)

In the case of massive MIMO where $N_t$ is large, we have the following approximation:
\[
C_N \approx \mathcal{O}\left\{N_t^{3.5} + N_t^{2.5}\right\}.
\]  
(72)

#### C. Symbol Scaling Scheme

For the proposed symbol scaling scheme, in the following we calculate the computational cost for each stage. In the initialization stage, based on Algorithm 1 a one-dimensional search is required to find the columns that satisfy the sign-identity condition, and the resulting computational cost is given by
\[
C_{1,i}^I = \mathcal{O}\left\{2N_t\right\};
\]  
(73)
as M consists of 2Ni column vectors. In the allocation stage, the dominant complexity comes from the allocation scheme based on the ‘Max-Min’ criterion where we need to iteratively allocate the value for x^F_n. Since card(\mathcal{V}) is difficult to obtain analytically in the initialization stage, we consider a worst-case complexity where card(\mathcal{V}) = 1. This then leads to

\[ C_L^2 = \mathcal{O}\{(2N_t - 1)\} \approx \mathcal{O}\{2N_t\} \]  

(74)

in the case of massive MIMO. In the refinement stage, based on Algorithm 5 it is obtained that a one-dimensional search of 2N_t entries is required for both x_{sum-min} and x_{max-min}, which leads to the computational cost as

\[ C_L^2 = \mathcal{O}\{2N_t + 2N_t\} = \mathcal{O}\{4N_t\} \].  

(75)

Based on the above, we can obtain the total computational cost for the proposed symbol scaling approach as

\[ C_L = C_L^1 + C_L^2 + C_L^3 \]
\[ = \mathcal{O}\{2N_t\} + \mathcal{O}\{(2N_t - 1)\} + \mathcal{O}\{4N_t\} \]
\[ = \mathcal{O}\{8N_t - 1\} \approx \mathcal{O}\{8N_t\} \].  

(76)

**D. Pokemon**

As comparison, we also include the complexity of the non-linear scheme ‘Pokemon’ proposed in [36]. The ‘Pokemon’ scheme is based on the biconvex relaxation approach and requires an iterative design, whose performance is dependent on the number of iterations. Based on [36], in each iteration an element-wise allocation is conducted for the transmit signal vector, which leads to the total computational cost as

\[ C_P = \mathcal{O}\{n_{\text{max}} \cdot 2N_t\} = \mathcal{O}\{2n_{\text{max}}^2N_t\} \],  

(77)

where n_{\text{max}} denotes the number of iterations for the ‘Pokemon’ scheme. Comparing the computational cost of ‘Pokemon’ with the proposed symbol scaling scheme, we have

\[ \frac{C_L}{C_P} = \frac{\mathcal{O}\{8N_t\}}{\mathcal{O}\{2n_{\text{max}}^2N_t\}} = \mathcal{O}\left\{\frac{4}{n_{\text{max}}}\right\} \].  

(78)

To numerically study the complexity gains of the proposed low-complexity scheme, in Table II we show the number of floating operations required with the increase in the number of transmit antennas, where for ‘Pokemon’ we employ n_{\text{max}} = 20 following [36]. As can be seen, the computational cost of the proposed non-linear mapping scheme is higher than that of the proposed low-complexity scheme and the ‘Pokemon’ scheme, while the number of operations required for the proposed symbol scaling scheme is only 20% of the number of operations for ‘Pokemon’.

| Antenna Number | Exhaustive Search | Proposed Non-linear Mapping P4 | Proposed Symbol Scaling Scheme | Non-linear Pokemon, n_{\text{max}} = 20 |
|----------------|------------------|--------------------------------|--------------------------------|-----------------------------------------|
| 64             | \mathcal{O}\{3.4 \times 10^{38}\} | \mathcal{O}\{2.13 \times 10^{40}\} | \mathcal{O}\{5.12 \times 10^{42}\} | \mathcal{O}\{2.56 \times 10^{42}\} |
| 96             | \mathcal{O}\{6.3 \times 10^{37}\} | \mathcal{O}\{8.76 \times 10^{39}\} | \mathcal{O}\{7.68 \times 10^{42}\} | \mathcal{O}\{3.84 \times 10^{42}\} |
| 128            | \mathcal{O}\{1.2 \times 10^{37}\} | \mathcal{O}\{2.39 \times 10^{39}\} | \mathcal{O}\{1.024 \times 10^{41}\} | \mathcal{O}\{5.12 \times 10^{41}\} |
| 256            | \mathcal{O}\{1.3 \times 10^{34}\} | \mathcal{O}\{2.69 \times 10^{39}\} | \mathcal{O}\{2.048 \times 10^{41}\} | \mathcal{O}\{1.02 \times 10^{41}\} |

**TABLE II** Comparison of the computational costs of different schemes

![Fig. 4: BER v.s. transmit SNR, N_t = 8, K = 2, n_{\text{max}} = 20, QPSK](chart)

**VI. Numerical Results**

In this section we present the numerical results of the proposed schemes based on Monte Carlo simulations. In each plot, the transmit SNR is defined as \( \rho = P/\sigma^2 \). Both QPSK and 8-PSK modulations are considered in the numerical results. We compare our proposed schemes with both the quantized linear schemes and the non-linear mapping schemes, and for clarity the following abbreviations are used throughout this section:

1) ‘ZF-FD’: Fully-digital ZF scheme with infinite-precision DACs;
2) ‘ZF 1-Bit’: Quantized ZF scheme with 1-bit DACs introduced in [32];
3) ‘MMSE’: MMSE-based quantized linear scheme in [33];
4) ‘Pokemon, n_{\text{max}} = K’: Non-linear Pokemon scheme proposed in [36] with K iterations;
5) ‘Constructive’: Proposed non-linear mapping scheme P4 in Section III-B;
6) ‘sum-max’: Proposed symbol scaling scheme based on the ‘sum-max’ allocation scheme with Algorithm 1, 2 and 4;
7) ‘max-min’: Proposed symbol scaling scheme based on the ‘max-min’ allocation scheme with Algorithm 1, 3 and 4;
8) ‘Symbol Scaling’: Proposed symbol scaling scheme obtained via Algorithm 5 where we select the best signal vector out of ‘sum-max’ or ‘max-min’ criteria.

In Fig. 4, we firstly consider a moderate scale MIMO with...
a total number of $N_t = 8$ transmit antennas at the BS and $K = 2$ single-antenna users in the system. For schemes with 1-bit quantization, we observe that the proposed symbol scaling scheme based on Algorithm 5 achieves the best BER performance, while both the proposed non-linear mapping scheme and ‘Pokemon’ achieve an inferior performance. This is because both the non-linear mapping scheme and the ‘Pokemon’ scheme involve the relaxation-normalization process. For small-scale MIMO systems, based on Table I we can infer that $\eta$ will be large in this case, which means that the deviation of the solution obtained by the relaxation-normalization approach from the solution of the original 1-bit optimization problem is large, and the normalization process may lead to further detection errors. For the proposed symbol scaling scheme, the performance is promising since we directly select the quantized signal for each antenna element and therefore no relaxation or quantization is needed.

We then consider a massive MIMO system with $N_t = 128$ transmit antennas and $K = 16$ users in Fig. 5. In the case of massive MIMO, all the schemes can achieve a lower BER thanks to the large number of antennas at the BS, and generally non-linear schemes outperform linear schemes. For schemes with 1-bit DACs, the proposed non-linear mapping scheme outperforms the non-linear ‘Pokemon’ scheme and achieves the best BER performance. As for the proposed low-complexity symbol scaling scheme, by comparing Fig. 4 and Fig. 5, we can observe that the ‘Max-Min’ criterion is most suitable for small-scale MIMO systems, while the ‘Sum-Max’ criterion is more favourable for massive MIMO systems. Moreover, while we have observed around a 2dB SNR loss compared to the ‘Pokemon’ scheme in the case of massive MIMO, its computational cost is 20% of that for Pokemon in this scenario, which is shown mathematically in Table II and will be shown numerically in Fig. 7.

In Fig. 6, we show the performance of different schemes for 8-PSK modulation with $N_t = 128$ and $K = 8$. For quantized beamforming schemes, it is observed that the proposed optimization-based non-linear scheme achieves the best BER performance. For the symbol scaling approach, it is observed that in the case of 8-PSK, only 1dB SNR loss is observed compared to the non-linear iterative ‘Pokemon’ scheme, and therefore the proposed low-complexity symbol scaling scheme is more favourable in terms of the performance and complexity tradeoff.

In Fig. 7, we compare the computational complexity of each scheme in terms of the execution time required per 10 channel realizations. It is not surprising to observe that the computational cost of the proposed non-linear scheme is the highest. Compared to the non-linear ‘Pokemon’ scheme, the execution time required for the proposed symbol scaling scheme is much less, especially for the ‘sum-max’ case. For ‘Symbol Scaling’ that returns the best performance based on Algorithm 5, the execution time required is similar to that of the ‘max-min’, which validates our analysis in Section V-C that most of the computational cost in the allocation stage
comes from the ‘Max-Min’ criterion. Moreover, it is observed that the execution time of ‘Symbol Scaling’ is approximately 20% of that of the ‘Pokemon’ scheme in Fig. 7. This matches our analysis in (78) (C_{sc} = \frac{1}{3} when n_{max} = 20), and the above complexity gains of the proposed symbol scaling approach therefore favour its practical application.

To further compare the proposed schemes with ‘Pokemon’, in Fig. 8 we present the BER performance with different iteration numbers for Pokemon. This iteration number does not have an effect on other schemes and therefore the BER for other schemes stay constant. It is observed that the performance of Pokemon is improved with the increase in n_{max}. Nevertheless, we note that the improvement becomes less significant with a larger n_{max} and the Pokemon achieves its best performance when the iteration number n_{max} is around 25. An important observation is when the iteration number is equal to n_{max} = 4, where the computational cost of Pokemon and our proposed scheme is similar, as shown by (78). In this case, our proposed symbol scaling approach is shown to achieve an improved performance, which validates the superiority of the proposed scheme.

To demonstrate the performance-complexity tradeoff directly, in Fig. 9 we depict the BER with respect to the analytical floating operations required for a range of transmit antennas from N_t = 32 to N_t = 128, where the number of users is fixed as K = 8. It can be observed that the proposed optimization-based scheme achieves the best performance at the cost of the highest complexity. An important comparison is between the proposed ‘Symbol Scaling’ approach and the ‘Pokemon’ scheme with n_{max} = 4, where we observe a significant performance gain of our proposed scheme for the same computational complexity, especially when the number of antennas is large. Moreover, while the performance of the proposed low-complexity based on ‘sum-max’ achieves an inferior performance to the ‘Symbol Scaling’ approach when N_t is large, it indeed achieves a better BER performance with a lower computational cost compared to Pokemon. Both of

\[ \hat{H} = H + Q. \]  

(79)

In (79), \( \hat{H} \) is the obtained CSI at BS. Q denotes an error matrix with \( Q \sim \mathcal{CN}(0, \delta \cdot I) \), where \( \delta \) denotes the variance of the channel error. \( \delta \) is modelled as inversely proportional to the transmit SNR and is expressed as \( \delta = \beta / \rho \), where \( \beta \) denotes the error coefficient [18]. The BER result with imperfect CSI
is depicted in Fig. 10, where a similar trend can be observed. We can further observe that the proposed non-linear mapping scheme still achieves the best performance among the schemes with 1-bit quantization in the case of imperfect CSI, while the proposed low-complexity symbol scaling approach can achieve a comparable performance with a greatly reduced computational cost.

VII. CONCLUSION

In this paper, we propose the transmit beamforming schemes for the massive MIMO downlink with 1-bit DACs based on the formulation of constructive interference, and we consider both a quantized linear scheme and a non-linear mapping scheme. With the analysis of Lagrangian and KKT conditions, the linear scheme is mathematically proven to be equivalent to the quantized ZF scheme. For the proposed non-linear mapping scheme, it is shown to be non-convex and solved by firstly relaxing the constraint on 1-bit quantization, followed by a normalization. We further propose a low-complexity symbol scaling scheme, where the quantized transmit signals are directly obtained. Numerical results reveal the superiority of the proposed symbol scaling scheme in small-scale MIMO systems. In the case of massive MIMO, the performance advantage of the proposed non-linear mapping scheme is validated, while the proposed symbol scaling scheme achieves a better performance-complexity tradeoff, which favours its usefulness in practical systems.

APPENDIX

COORDINATE TRANSFORMATION

We employ 8-PSK modulation in Fig. 3 as the example to demonstrate the coordinate transformation, where we focus on the constellation point ‘A’ in Fig. 3. Then, in the conventional real-imaginary complex plane, for node ‘B’ in Fig. 3, we have

\[
OB = h_k x_T = B_r \cdot 1 + B_i \cdot j,
\]

(80)

where 1 and \(j\) are the bases, and we denote \((B_r, B_i)\) as the corresponding coordinate. Based on (8), \(B_r\) and \(B_i\) are obtained as

\[
\begin{align*}
B_r &= \Re (h_k x_T) = \Re (h_k) x_T^R - \Im (h_k) x_T^\Im, \\
B_i &= \Im (h_k x_T) = \Im (h_k) x_T^R + \Re (h_k) x_T^\Im.
\end{align*}
\]

(81)

In the plane expanded by the two detection thresholds that correspond to the constellation point ‘A’, following (42) \(OB\) is decomposed into

\[
OB = h_k x_T = \alpha_k^R s_k^R + \alpha_k^\Im s_k^\Im.
\]

(82)

Based on (41) and the fact that \(\alpha_k^R\) and \(\alpha_k^\Im\) are real numbers, (82) is further transformed into

\[
\begin{align*}
h_k x_T &= \alpha_k^R (a_k + b_k \cdot j) + \alpha_k^\Im (c_k + d_k \cdot j) \\
&= (a_k \alpha_k^R + c_k \alpha_k^\Im) + (b_k \alpha_k^R + d_k \alpha_k^\Im) \cdot j.
\end{align*}
\]

(83)

By substituting (81) into (83), we obtain

\[
\begin{align*}
B_r &= \Re (h_k) x_T^R - \Im (h_k) x_T^\Im = a_k \alpha_k^R + c_k \alpha_k^\Im, \\
B_i &= \Im (h_k) x_T^R + \Re (h_k) x_T^\Im = b_k \alpha_k^R + d_k \alpha_k^\Im.
\end{align*}
\]

(84)

which leads to the expression of \(\alpha_k^R\) and \(\alpha_k^\Im\), given by

\[
\begin{align*}
\alpha_k^R &= \frac{d_k B_r - c_k B_i}{a_k d_k - b_k c_k} \\
&= \frac{d_k [\Re (h_k) x_T^R - \Im (h_k) x_T^\Im] - c_k [\Im (h_k) x_T^R + \Re (h_k) x_T^\Im]}{a_k d_k - b_k c_k}, \\
\alpha_k^\Im &= \frac{d_k \Re (h_k) - c_k \Im (h_k) x_T^R}{a_k d_k - b_k c_k} \\
&= \frac{d_k \Re (h_k) - c_k \Im (h_k) x_T^R}{a_k d_k - b_k c_k} + \frac{d_k \Im (h_k) + c_k \Re (h_k)}{a_k d_k - b_k c_k} x_T^\Im,
\end{align*}
\]

(85)

and

\[
\begin{align*}
\alpha_k^\Im &= \frac{a_k \Im (h_k) - b_k \Re (h_k) x_T^R}{a_k d_k - b_k c_k} \\
&= \frac{a_k \Im (h_k) - b_k \Re (h_k) x_T^R}{a_k d_k - b_k c_k} + \frac{a_k \Re (h_k) + b_k \Im (h_k)}{a_k d_k - b_k c_k} x_T^\Im.
\end{align*}
\]

(86)

The extension to other constellation points of 8-PSK and other PSK modulations can be similarly obtained and is omitted for brevity.

REFERENCES

[1] L. Zheng and D. N. C. Tse, “Diversity and Multiplexing: A Fundamental Tradeoff in Multiple-Antenna Channels,” IEEE Trans. Inf. Theory, vol. 49, no. 5, pp. 1073–1096, May 2003.
[2] M. Costa, “Writing on Dirty Paper,” IEEE Trans. Inf. Theory, vol. IT-29, no. 3, pp. 439–441, May 1983.
[3] L. Sun and M. Lei, “Quantized CSI-based Tomlinson-Harashima Precoding in Multiuser MIMO Systems,” IEEE Trans. Wireless Commun., vol. 12, no. 3, pp. 1118–1126, Mar. 2013.
[4] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, “A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiscale Communication-part II: Perturbation,” IEEE Trans. Commun., vol. 53, no. 3, pp. 537–544, Mar. 2005.
[5] T. Haustein, C. von Helmholt, E. Jorswieck, V. Jungnickel, and V. Pohl, “Performance of MIMO Systems with Channel Inversion,” in Vehicular Technology Conference. IEEE 55th Vehicular Technology Conference. VTC Spring 2002 (Cat. No.02CH37367), vol. 1, 2002, pp. 35–39.
[6] C. B. Peel, B. M. Hochwald, and A. L. Swindlehurst, “A Vector-Perturbation Technique for Near-Capacity Multiantenna Multiscale Communication-part I: Channel Inversion and Regularization,” IEEE Trans. Commun., vol. 53, no. 1, pp. 195–202, Jan. 2005.
[7] C. Masouros, “Correlation Rotation Linear Precoding for MIMO Broadcast Communications,” IEEE Trans. Sig. Process., vol. 59, no. 1, pp. 252–262, Jan. 2011.
[8] A. Wiesel, Y. C. Eldar, and S. Shamai (Shitz), “Linear Precoding via Conic Optimization for Fixed MIMO Receivers,” IEEE Trans. Sig. Process., vol. 54, no. 1, pp. 161–176, Jan. 2006.
[9] M. Schubert and H. Boche, “Solution of the Multiuser Downlink Beamforming Problem with Individual SINR Constraints,” IEEE Trans. Veh. Tech., vol. 53, no. 1, pp. 18–28, Jan. 2004.
[10] M. F. Hanif, L.-N. Tran, A. Tolli, and M. Juntti, “Computationally Efficient Robust Beamforming for SINR Balancing in Multicell Downlink with Applications to Large Antenna Array Systems,” IEEE Trans. Commun., vol. 62, no. 6, pp. 1908–1920, June 2014.
[11] F. Wang, X. Wang, and Y. Zhu, “Transmit Beamforming for Multiuser Downlink with Per-Antenna Power Constraints,” in 2014 IEEE International Conference on Communications (ICC), Sydney, NSW, 2014, pp. 4692–4697.
[12] M. Bengtsson and B. Ottersten, “Optimal and Suboptimal Transmit Beamforming,” Handbook of Antennas in Wireless Communications, Jan. 2001.
[13] E. Karipidis, N. D. Sidoropoulos, and Z.-Q. Luo, “Quality of Service and Max-Min Fair Transmit Beamforming to Multiple Ccochannel Multicast Groups,” IEEE Trans. Sig. Process., vol. 56, no. 3, pp. 1286–1279, Mar. 2008.
[14] C. Masouros and G. Zheng, “Exploiting Known Interference as Green Signal Power for Downlink Beamforming Optimization,” IEEE Trans. Sig. Process., vol. 63, no. 14, pp. 3628–3640, July 2015.
