

Modulus Zero-Forcing Detection for MIMO Channels

Sha Hu and Fredrik Rusek

Abstract

We propose a modulus arithmetic based zero-forcing (MZF) detector for multi-input multi-output (MIMO) channels. Traditionally, a ZF detector completely eliminates interference from other symbol layers when detecting a particular symbol layer, which results in suboptimal performance due to noise-enhancement. The only constraint for application of our proposed MZF detector is that the transmitter must employ a finite cardinality M quadrature-amplitude-modulation (QAM) alphabet. With that, the modulus operand of the MZF is to allow for integer-valued interference and then remove it by modulus arithmetic operations.

Index Terms

Multi-input multi-output (MIMO), modulus, zero-forcing (ZF), quadrature-amplitude-modulation (QAM), pulse-amplitude-modulation (PAM), linear minimum-mean-square-error (LMMSE), sphere-decoding (SD), lattice-reduction (LR), Lenstra-Lenstra-Lovász (LLL).

I. INTRODUCTION

We consider a standard multi-input multi-output (MIMO) channel model with a received signal $\tilde{\mathbf{y}}$ expressed as

$$\tilde{\mathbf{y}} = \tilde{\mathbf{H}}\tilde{\mathbf{x}} + \tilde{\mathbf{n}}, \quad (1)$$

where $\tilde{\mathbf{H}}$, $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{n}}$ are the complex-valued MIMO channel, transmitted symbols and Gaussian noise, respectively.

The authors are with the Department of Electrical and Information Technology, Lund University, Lund, Sweden (email: {firstname.lastname}@eit.lth.se).

Given received signal model (1), detecting $\tilde{\mathbf{x}}$ is referred to as a MIMO detection problem, which has a history that can be traced back about half a century and a review on it can be found in e.g., [1]. In general, maximum likelihood (ML) detection [2] yields optimal performance but with prohibitive complexity when the MIMO dimension and/or the input alphabet has large cardinality. Effective implementations of ML detection, such as sphere-decoding (SD) [3], [4] can significantly reduce the complexity, but not overcome an exponential complexity in the number of symbol layers CITE JALDEN'S PAPER. On the other hand, linear detectors [2] such as zero-forcing (ZF) and linear minimum-mean-square-error (LMMSE), have low complexities, but also suboptimal performances. One direction for improving linear detectors is lattice-aided-reduction (LAR) [5] based approaches, which use lattice-reduction (LR) algorithms, e.g., Lenstra-Lenstra-Lovász (LLL), to find a short and nearly orthogonal basis for the lattice induced by the MIMO channel [6].

Other than the existing approaches [1], as the transmitted symbols are drawn from finite alphabets such as quadrature-amplitude-modulation (QAM) and pulse-amplitude-modulation (PAM) symbols, the modulus can also be used in MIMO detection for improving the detection performance. The modulus operation has been used in Tomlinson-Harashima precoding (THP) [7], [8] as suboptimal approximation for dirty-paper coding (DPC) [9], and recently it has also been considered in the designs of integer-forcing (IF) receivers for MIMO channels [13], [15], [16]. The IF scheme in [13] requires the transmitter to employ a same lattice code [14] for each transmitted layer, which does not apply in most of current communication systems. Besides, when higher-order modulations such as M -PAM are used, designing lattice code over \mathbb{Z}_M is challenging [14]. Simpler IF receivers dealing with linear binary codes such as turbo and LDPC are proposed in [15], [16], and the designs follow the same as in [13]. One disadvantage of the IF receivers in [13], [16] is that, each transmit antenna needs a separate encoding/decoding process, which is not the case in practical LTE systems where one codeword is split among transmit-antennas. Moreover, the IF design in [15] needs a separate encoding/decoding process per transmit antenna and per bit-layer in higher-order modulations. Another advantage is that, the receiver has to detect the linear combinations of codewords for all transmit-antennas first, followed by a matrix inversion (over a finite-field) process to recover the original codeword on each layer.

To overcome these disadvantages in previous IF receiver designs, we consider a new approach to improve linear detection with modulus operation, namely, the proposed modulus ZF (MZF)

detection. Note that, MZF is conceptually different from previous IF receivers, although they share quite some similarities. The fundamental difference is that, with MZF, there is no encoding/decoding process needed which are implemented on a finite-filed with IF receivers to recover the transmit symbols (i.e., the linear combinations of codewords across all transmit-antennas). Alternatively, we design MZF detector such that the transmit symbols on each transmit-antenna can be recovered directly by modulating away the interferences from the remaining transmit-antennas, and the symbol detection on different transmit-antenna is independent and fully in parallel. Such a principle simplifies the operation and can be well cooperated into practical systems such as LTE. To achieve such a nice property, with MZF the modulus matrix is carefully designed and optimized according to the specific modulation-order.

II. PRELIMINARIES

We start with reviewing the standard ZF detection. Before proceeding, without loss of generality, the matrix $\tilde{\mathbf{H}}$ is always assumed to be a square matrix, obtained by a QR factorization or padding zero rows to the matrix if necessary. With the following definitions,

$$\mathbf{y} = \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{y}}\} \\ \mathcal{I}\{\tilde{\mathbf{y}}\} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{x}}\} \\ \mathcal{I}\{\tilde{\mathbf{x}}\} \end{bmatrix}, \mathbf{n} = \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{n}}\} \\ \mathcal{I}\{\tilde{\mathbf{n}}\} \end{bmatrix}, \mathbf{H} = \begin{bmatrix} \mathcal{R}\{\tilde{\mathbf{H}}\} & -\mathcal{I}\{\tilde{\mathbf{H}}\} \\ \mathcal{I}\{\tilde{\mathbf{H}}\} & \mathcal{R}\{\tilde{\mathbf{H}}\} \end{bmatrix}, \quad (2)$$

we can rewrite (1) as a real-valued model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (3)$$

where the $K \times K$ channel matrix \mathbf{H} is known to the receiver, $\mathbf{x} = [x_1 \dots x_K]^T$ contains PAM symbols from an alphabet $\mathcal{A} = \{\pm 1, \pm 3, \dots, \pm(\sqrt{M} - 1)\}$, and \mathbf{n} is random Gaussian noise with a covariance matrix $(N_0/2)\mathbf{I}$. As the transmit power depends on M , the signal-to-noise (SNR) is defined as

$$\text{SNR} = 2\mathbb{E}[|x_k|^2]/N_0. \quad (4)$$

The ZF detector is given by

$$\hat{\mathbf{x}} = \mathcal{Q}_{\mathcal{A}}(\mathbf{H}^+ \mathbf{y}), \quad (5)$$

where $\mathcal{Q}_{\mathcal{A}}(\cdot)$ denotes entry-wise quantization to the nearest point in \mathcal{A} . This can be slightly rewritten as

$$\hat{x}_k = \mathcal{Q}_{\mathcal{A}}(r_k), \quad 1 \leq k \leq K$$

with

$$r_k = \boldsymbol{\delta}_k \mathbf{H}^+ \mathbf{y} \quad (6)$$

and

$$\boldsymbol{\delta}_k = [\underbrace{0 \dots 0}_{k-1} \ 1 \ \underbrace{0 \dots 0}_{K-k}].$$

For later use, we note that we may just as well replace the “1” with any arbitrary scalar value and equivalently work with

$$r_k = \tau x_k + w_k, \quad (7)$$

where w_k is zero-mean complex Gaussian noise with variance $\tau^2(N_0/2)\|\boldsymbol{\delta}_k \mathbf{H}^+\|^2$. Accordingly, the post-processing SNR, which is independent of τ , becomes

$$\gamma_k = \frac{\text{SNR}}{\|\boldsymbol{\delta}_k \mathbf{H}^+\|^2}, \quad (8)$$

where SNR is defined in (4).

III. DESCRIPTION OF THE PROPOSED MZF DETECTION

A main issue with ZF is that $\|\boldsymbol{\delta}_k \mathbf{H}^+\|^2$ in (8) is typically large and results in noise-enhancement. To combat that, we make use of the underlying idea of THP but apply it to detection, *without any involvement of the transmitter*. We propose to replace (6) with

$$r_k = (\tau \boldsymbol{\delta}_k + \mathbf{q}_k) \mathbf{H}^+ \mathbf{y} \quad (9)$$

where $\mathbf{q}_k = [q_{k1}, q_{k2}, \dots, q_{kK}]$ and $q_{k\ell} \in 2\mathbb{Z}$, i.e., the even integers. With that,

$$r_k = \tau x_k + \sum_{\ell=1}^K q_{k\ell} x_\ell + w_k, \quad (10)$$

and the post processing SNR changes to

$$\gamma_k = \frac{\tau^2 \text{SNR}}{\|(\tau \boldsymbol{\delta}_k + \mathbf{q}_k) \mathbf{H}^+\|^2}, \quad (11)$$

rather than (8).

Note that (10) coincides with the received signal per user in a vector perturbation (VP) system [10], [11]. Therefore, further processing of (10) follows the same steps as those for VP. Based on (10), hard-output detection of x_k from r_k can be obtained based on the following property.

Property 1. *Let*

$$y = z + \alpha \sum_{m=1}^M p_m b_m,$$

where $\alpha \geq 1$, $|z| < 2$, and $p_m, (b_m - 1) \in 2\mathbb{Z}$. Then,

$$z = \begin{cases} (y \bmod 4\alpha) - 2\alpha, & \text{if } \frac{1}{2} \sum_{m=1}^M p_m \text{ is odd,} \\ ((y + 2\alpha) \bmod 4\alpha) - 2\alpha, & \text{otherwise.} \end{cases} \quad (12)$$

Proof. See Appendix A. ■

Let us now consider $\alpha = 1$ (other values are considered in Section IV-A). In view of Property 1, we see that $q_{k\ell}$ and x_ℓ in (10) qualify as p_m and b_m . Further, from the condition $|z| < 2$ in Property 1, we have that τ must be selected so that

$$\tau \max_{a \in \mathcal{A}} |a| = \tau(\sqrt{M} - 1) < 2. \quad (13)$$

To finalize the detector, we let

$$z_k = \begin{cases} (r_k \bmod 4) - 2, & \text{if } \frac{1}{2} \sum_{\ell=1}^K q_{k\ell} \text{ is odd,} \\ ((r_k + 2) \bmod 4) - 2, & \text{otherwise,} \end{cases} \quad (14)$$

which can be expressed as

$$z_k = \tau x_k + \tilde{w}_k,$$

where \tilde{w}_k has a complicated distribution due to the modulus operation. The detected symbol \hat{x}_k can now be obtained as

$$\hat{x}_k = \mathcal{Q}_{\tau\mathcal{A}}(z_k), \quad (15)$$

where the quantization is implemented on $\tau\mathcal{A}$, with the definition $\tau\mathcal{A} = \{\tau x : x \in \mathcal{A}\}$.

We remark that, the choice $\tau = 2/(\sqrt{M} - 1)$ is not suitable in (13). This is so since if $\tau x_k + w_k = 2 + \epsilon$, for some small $\epsilon > 0$, then $z_k = -2 + \epsilon$. However, provided that $\tau \ll 2/(\sqrt{M} - 1)$, such wrap seldomly happens at high SNR and $\tilde{w}_k = w_k$ with high probability. Observe that for constellation points x_k with small magnitude, then $\tilde{w}_k = w_k$ with much higher probability than for constellation points x_k of large magnitude. To ensure equal error probability for all constellation points, we design τ such that:

The distance from 2 to the largest constellation point in $\tau\mathcal{A}$ is half the distance between two points in $\tau\mathcal{A}$.

Following this rule results in

$$\tau = 2^{(1-\log_2(\sqrt{M}))}, \quad (16)$$

where with out loss of generality we assume that $\log_2(\sqrt{M})$ is an integer in the rest of the paper. With that, we have that \tilde{w}_k is “nearly” Gaussian at high SNR; see VP-PAPER for further details.

To optimize the receiver, we should solve

$$\begin{aligned} \mathbf{q}_k^{\text{opt}} &= \arg \max_{\mathbf{q}_k} \gamma_k \\ &= \arg \min_{\mathbf{q}_k} \|(\tau \boldsymbol{\delta}_k + \mathbf{q}_k) \mathbf{H}^+\|^2 \end{aligned} \quad (17)$$

where elements of \mathbf{q}_k are even integers. We rewrite (17) as

$$\begin{aligned} \mathbf{q}_k^{\text{opt}} &= \arg \min_{\mathbf{q}_k} \|\tau \boldsymbol{\delta}_k \mathbf{H}^+ - \mathbf{q}_k (-\mathbf{H}^+)\|^2 \\ &= \arg \min_{\mathbf{q}_k} \|\mathbf{b}_k - \mathbf{q}_k \mathbf{B}\|^2, \end{aligned} \quad (18)$$

which is an instance of sphere detection over the integers [4], and

$$\mathbf{b}_k = \tau \boldsymbol{\delta}_k \mathbf{H}^+, \quad (19)$$

$$\mathbf{B} = -\mathbf{H}^+. \quad (20)$$

Without any further extensions, pseudo-code for implementation of the MZF detector is given in Algorithm 1. We remind the reader that the inputs \mathbf{H} and \mathbf{y} to the algorithm are assumed to be real-valued, while M denotes the cardinality of the complex-valued QAM constellation. We provide a worked out example of MZF detection in Appendix B to illustrate the process.

A. Some Remarks on the MZF Detection

With the principle of MZF detection introduced, we have a few important remarks as follows.

Remark 1. *The MZF is an extension of the ZF, where the latter is the special case of the former when*

$$\|\mathbf{q}_k^{\text{opt}} - \boldsymbol{\delta}_k \odot \mathbf{q}_k^{\text{opt}}\|^2 = 0, \quad (21)$$

where \odot denotes the Hadamard product. Therefore, from the perspective of post-processing SNR γ_k , MZF is always superior to ZF.

Remark 2. *Following Remark 1, when (21) holds, the modulus operation degrades performance and should be removed.*

Remark 3. *In general the minimum value achieved by $\mathbf{q}_k^{\text{opt}}$ in (18) increases as τ decreases. That is, for alphabets with large cardinality, the gain of MZF decreases.*

To resolve the issue in Remark 3 and to further improve performance, some useful extensions will be introduced in Sec. IV.

B. Discussion on the Diversity-Multiplexing Trade-off (DMT)

In the designs of IF receivers [13], [15], [16], the target is to directly optimize

$$\mathbf{p}_k^{\text{opt}} = \arg \min_{\mathbf{p}_k} \|\mathbf{p}_k \mathbf{H}^+\|^2, \quad (22)$$

given the constraint that $\mathbf{P} \in \mathbb{Z}^{K \times K}$ is full-rank¹ over \mathbb{R} , where \mathbf{P} is the matrix comprising all vectors \mathbf{p}_k . Comparing to (17), it can be seen that, with MZF we constrain \mathbf{p}_k in (22) to be

$$\mathbf{p}_k = \boldsymbol{\delta}_k + \frac{1}{\tau} \mathbf{q}_k, \quad (23)$$

where \mathbf{q}_k comprises even integers. With (23) it holds that

$$\mathbf{P} = \mathbf{I} + \frac{1}{\tau} \mathbf{Q}, \quad (24)$$

where $\tau < 1$ and the elements of \mathbf{Q} are even integers. Therefore, with MZF, the degrees-of-freedom (DoFs) in designing \mathbf{P} is less than that of an IF receiver². Nevertheless, we have the following Property 2 (which is an extension of [13, Theorem 5]) that shows that, the MZF also achieves the optimal DMT [18] as IF receiver does. Therefore, from an information-theoretic perspective, the MZF detection does not sacrifice much performance compared to an IF receiver, while the latter one has much higher encoding/decoding complexity.

¹The rationale behind is that, in order to have $\mathbf{P} \bmod p$ to be full-rank over \mathbb{Z}_p (to recover the codewords from their linear combinations after decoding on \mathbb{Z}_p), it suffices to check whether \mathbf{P} is full-rank over \mathbb{R} , given that the magnitudes of the elements of \mathbf{P} are upper-bounded by a constant [17, Th. 11].

²However, with MZF matrix \mathbf{P} is not required to be full-rank.

Property 2. For a complex-valued MIMO channel with K transmit-antennas, $N \geq K$ receive-antennas, and independent and identically distributed (i.i.d.) Rayleigh fading, the achievable DMT with the MZF detection³ is

$$d_{\text{MZF}} = N \left(1 - \frac{r}{K}\right), \quad 0 \leq r \leq K. \quad (25)$$

Proof. See Appendix C. ■

IV. EXTENSIONS OF THE MZF DETECTION

In this section we introduce some extensions to the basic MZF detector for further improving its performance. While Extension 1 and 4 are generalizations of the basic algorithm, Extension 2 is to resolve the issue mentioned for large cardinality alphabets and improve the performance for weak bit-layers, while Extension 3 is a decision feedback version of Extension 2.

A. Extension 1: A scaled modulus

This first extension arises from a slight relaxation of $\alpha = 1$ in the MZF detector. From Property 1, we can replace (14) as

$$z_k = \begin{cases} (r_k \bmod 4\alpha) - 2\alpha, & \text{if } \frac{1}{2} \sum_{\ell=1}^K q_{k\ell} \text{ is odd,} \\ ((r_k + 2\alpha) \bmod 4\alpha) - 2\alpha, & \text{otherwise.} \end{cases} \quad (26)$$

This requires us to optimize, instead of (17),

$$(\mathbf{q}_k^{\text{opt}}, \alpha^{\text{opt}}) = \arg \min_{|\alpha| \geq 1, \mathbf{q}_k} \|(\tau \boldsymbol{\delta}_k + \alpha \mathbf{q}_k) \mathbf{H}^+\|^2. \quad (27)$$

Solving (27) is harder than solving (17) since it can be regarded as an instance of non-coherent sphere detection. Instead, we solve (17) first, and then plug the optimal solution into (27) and solve for the optimal α . That is, $\mathbf{q}_k^{\text{opt}}$ is obtained with (17), and

$$\alpha^{\text{opt}} = \arg \min_{|\alpha| \geq 1} \|(\tau \boldsymbol{\delta}_k + \alpha \mathbf{q}_k^{\text{opt}}) \mathbf{H}^+\|^2, \quad (28)$$

where we slightly abused notation since the pair $(\mathbf{q}_k^{\text{opt}}, \alpha^{\text{opt}})$ is in general not *jointly* optimal in the sense of (27). Although Extension 3 is intuitive, the gain seems marginal according to numerical results.

³In this case, the size of \mathbf{H} corresponding to the real-valued model in (3) is $2N \times 2K$.

B. Extension2: Bit-wise Modulus Zero-Forcing

An underlying assumption of this extension is that the bit-mapping to the symbols in \mathcal{A} is such that the constellation has an additive structure. By this we mean that a PAM symbol x_k should be of the form

$$x_k = \sum_{b=1}^{\log_2(\sqrt{M})} u_{kb} 2^{b-1}, \quad (29)$$

where $u_{kb} \in \{\pm 1\}$ correspond to information bits.

Using Algorithm 1, the bits u_{kb} are determined by the output $\hat{x}_k = \mathcal{Q}_{\tau\mathcal{A}}(z_k)$, with setting $\tau = 2^{(1-\log_2(\sqrt{M}))}$. As M increases, τ decreases and so are the gains of MZF detection. To resolve this for large values of M , we extend the symbol-based MZF detector in Algorithm 1 to a bit-wise version.

Note that we can rewrite (10) as

$$r_k = \tau \sum_{b=1}^{\log_2(\sqrt{M})} u_{kb} 2^{b-1} + \sum_{\ell=1}^K q_{k\ell} x_{\ell} + w_k. \quad (30)$$

Assuming that we are interested in the n -th bit u_{kn} , we let

$$\tilde{x}_k = \sum_{b=1}^n u_{kb} 2^{b-1}, \quad (31)$$

which belongs to a 2^n -PAM alphabet. Setting $\tau(n) = 2^{1-n}$ in (30) yields

$$r_k = 2^{1-n} \tilde{x}_k + \sum_{b=n+1}^{\log_2(\sqrt{M})} u_{kb} 2^{b-n} + \sum_{\ell=1}^K q_{k\ell} x_{\ell} + w_k. \quad (32)$$

It can be easily seen that $\frac{1}{2} \sum_{b=n+1}^{\log_2(\sqrt{M})} u_{kb} 2^{b-n}$ is an odd integer so it qualifies as a valid value of $q_{k\ell}$, and u_{kn} can be detected as

$$\hat{u}_{kn} = \text{sign}(z_k). \quad (33)$$

Therefore, for each bit-layer, a different value of τ is used and only a sign operation is needed for detection. Extension 2 has a complexity increment over Algorithm 1 since an optimization to find q_k is needed for each bit-layer.

Note that, according to Remark 2, when detecting the last bit-layer and if (21) holds, the ZF estimate shall be used for detection, while for detecting the other layers, modulus operations are still needed to cancel the transmitted bits corresponding to higher bit-layers. Psuedo-code for the MZF with this extension is summarized in Algorithm 2.

C. Extension 3: A decision feedback version of Extension 2

An obstacle with Extension 2 is that $\tau(n)$ decreases as n grows, and as previously mentioned, performance deteriorates. Small values of n correspond to weak bit-layers, and large n correspond to strong bit-layers. Thus, with Extension 2, predominantly the weak bit-layers can gain by the MZF, while the gain could be minuscule for strong bit-layers. A gain for weak bit-layers is important since it is typically these bit-layers that limit ultimate performance. However, we can also harvest a gain for strong bit-layers via a decision feedback mechanism. To prevent error propagation in decision feedback equalization, strong bits are typically detected first and then canceled. That option is not available for MZF, rather we detect the weakest bit-layer first and then move on to stronger ones.

The method works as follows. First set $n = 1$ and follow Extension 2 verbatim to obtain $\hat{\mathbf{u}}_1 = [\hat{u}_{11} \dots \hat{u}_{K1}]^T$. For notational convenience, define $\mathbf{y}_1 = \mathbf{y}$. Now construct

$$\mathbf{y}_2 = \frac{1}{2} (\mathbf{y}_1 - \mathbf{H} \hat{\mathbf{u}}_1). \quad (34)$$

Provided that $\hat{\mathbf{u}}_1$ is correct, \mathbf{y}_2 is described with the same MIMO channel as \mathbf{y}_1 , but with $\sqrt{M}/2$ -PAM rather than \sqrt{M} -PAM inputs. Next, move on to $n=2$ and keep $\tau(2)=1$. Since neither the value of $\tau(2)$ nor the channel \mathbf{H} has changed, the optimal vector \mathbf{q}_k for $n=2$ coincides with that already found for $n=1$. We then have that for \mathbf{y}_2

$$r_k = u_{k2} + \sum_{b=3}^{\log_2(\sqrt{M})} u_{kb} 2^{b-2} + \sum_{\ell=1}^K q_{k\ell} \frac{1}{2} (x_\ell - \hat{u}_{k1}) + w_k, \quad (35)$$

and \hat{u}_{k2} is obtained by taking the sign of z_k as in (33). We proceed by

$$\mathbf{y}_3 = \frac{1}{2} (\mathbf{y}_2 - \mathbf{H} \hat{\mathbf{u}}_2), \quad (36)$$

and continue the process until all bit-layers have been detected.

Similarly, according to Remark 2, whenever (21) holds, the ZF estimate shall be used. Pseudo-code is provided in Algorithm 3.

Extension 3 is similar to Extension 2 in the sense that, the detection for all bit-layers only needs to take the signs of z_k as in (33), but it has less complexity since only one optimization of (17) is needed which is shared for all bit-layers. A drawback with Extension 3 is that, as for all decision-feedback based detectors, the processing of the bit-layers cannot be parallelized, which is however, possible with Extension 2. Another drawback is potential error-propagation at low SNRs.

D. Extension 4: Replacing ZF by LMMSE

So far we have introduced modulus arithmetic detection using ZF, however, \mathbf{H}^+ can also be replaced by other linear detectors⁴ such as LMMSE, which sets

$$\mathbf{H}^+ = \mathbf{H}^H (\mathbf{H}\mathbf{H}^H + N_0\mathbf{I})^{-1}. \quad (37)$$

In vector form, and with the introduction of a matrix \mathbf{T} , the received signal after equalization is

$$\mathbf{T}\mathbf{H}^+\mathbf{y} = \mathbf{T}\mathbf{x} + \mathbf{T}(\mathbf{H}^+\mathbf{H} - \mathbf{I})\mathbf{x} + \mathbf{T}\mathbf{H}^+\mathbf{n},$$

where $\mathbf{T} = (\tau\mathbf{I} + \mathbf{Q})$. The target of optimizing \mathbf{q}_k in this case, is to minimize the interference plus noise power that equals

$$\tilde{\mathbf{q}}_k = \arg \min_{\mathbf{q}_k} \|(\tau\boldsymbol{\delta}_k + \mathbf{q}_k)\mathbf{E}\|^2, \quad (38)$$

where

$$\mathbf{E} = [\mathbf{H}^+\mathbf{H} - \mathbf{I}, N_0\mathbf{H}^+]. \quad (39)$$

Note that, when \mathbf{H}^+ is the pseudo-inverse of \mathbf{H} such as with ZF, \mathbf{E} degrades to \mathbf{B} , which shows the generalization of the MZF concept. The reason for introducing Extension 4 is that, the ZF detector is suboptimal to LMMSE at low SNRs, in which case the modulus operation based on LMMSE can improve performance. Therefore, it is beneficial to use LMMSE instead of ZF. Since only \mathbf{H}^+ is replaced by LMMSE equalization in Extension 4, all Algorithms 1-3 still apply verbatim.

There are also many other possible variations of the MZF, but which we will not pursue any further. Next we put an interest on comparing the MZF to a traditional LAR detector. The reason is that, solving (17) involves significant complexity, and we put forth an approximated solution based on LR with less computational efforts.

V. A SOLUTION BASED ON, AND A COMPARISON TO, LATTICE REDUCTION

Except for approximately solving (17) with LR, another reason for comparing MZF with LR detection is that, the obtained MZF allows for a direct comparison to LAR detectors. In LAR as well as the MZF, the most burdening task is to execute the LLL algorithm (or other similar algorithms), thus the complexities of LAR and MZF become virtually identical. As we will demonstrate, the detection-performance of MZF is superior in some cases.

⁴This is also known as regularized perturbation in VP [10].

A. A quick review of LAR

Given (3), LAR starts by performing the LLL algorithm on \mathbf{H} , so that we obtain $\bar{\mathbf{H}} = \mathbf{H}\mathbf{T}$ where \mathbf{T} is unimodular and $\bar{\mathbf{H}}$ is nearly orthogonal. With $\mathbf{z} = \mathbf{T}^{-1}\mathbf{x}$ we have

$$\mathbf{y} = \bar{\mathbf{H}}\mathbf{z} + \mathbf{n}. \quad (40)$$

Performing ZF based on $\bar{\mathbf{H}}$ and quantizing to the nearest integers gives

$$\hat{\mathbf{z}} = \mathcal{Q}_{\mathbb{Z}}(\bar{\mathbf{H}}^{-1}\mathbf{y}) \quad (41)$$

from which one can obtain

$$\hat{\mathbf{x}} = \mathcal{Q}_{\mathcal{A}}(\mathbf{T}\hat{\mathbf{z}}).$$

Clearly, once $\bar{\mathbf{H}}$ has been established, the remaining steps are of miniscule complexity.

At this point, a reasonable question is, what the relation between LAR and MZF is, and whether they are equivalent? The answers to these questions are that, they are closely related, but not equivalent. Prior to quantization in (41), we can write

$$\begin{aligned} \mathbf{r} &= \bar{\mathbf{H}}^{-1}\mathbf{y} \\ &= \mathbf{T}^{-1}\mathbf{x} + \mathbf{w}. \end{aligned} \quad (42)$$

Since \mathbf{T} is unimodular, so is \mathbf{T}^{-1} .

On the other hand, written in vector form, (9) equals

$$\begin{aligned} \mathbf{r} &= (\tau\mathbf{I} + \mathbf{Q})\mathbf{H}^{-1}\mathbf{y} \\ &= (\tau\mathbf{I} + \mathbf{Q})\mathbf{x} + \mathbf{w}. \end{aligned} \quad (43)$$

Comparing (42) and (43) with $\tau = 1$, we see that in both cases \mathbf{r} equals an integer-valued matrix multiplied with the data symbols, plus noise. However, the matrix \mathbf{T}^{-1} in (42) has no particular structure (besides being unimodular) so the modulus operation in (14) is not available. This makes LAR, i.e., (42) and MZF, i.e., (43) fundamentally different, as the structure of (42) requires further processing in the form of (41) while (43) allows for further processing via (14).

B. An approximate solution to (17) based on LLL

In (17) we have the following problem to solve

$$\mathbf{q}^{\text{opt}} = \arg \min_{\mathbf{q}} \|\mathbf{b} - \mathbf{q}\mathbf{B}\|^2, \quad (44)$$

where we removed the subscript k , and the vectors are row-vectors. Perform the LLL algorithm to \mathbf{B}^T so that we have

$$\bar{\mathbf{B}} = \mathbf{B}^T \mathbf{T}.$$

Since $\mathbf{B} = -\mathbf{H}^+$ the LLL algorithm needs, similar to LAR, to be executed only once per coherence interval. We can now proceed as in the LAR case,

$$\hat{\mathbf{z}} = \mathcal{Q}_{\mathbb{Z}}(\bar{\mathbf{B}}^{-1} \mathbf{b}^T)$$

followed by

$$\mathbf{q}^{\text{opt}} = [\mathcal{Q}_{2\mathbb{Z}}(\mathbf{T}\hat{\mathbf{z}})]^T. \quad (45)$$

Note that, the optimization (44) itself is also a MIMO detection problem (but only needs to run once per coherence-interval), therefore, there are also other low-complexity suboptimal algorithms to solve (44), such as using ZF or partial marginalization [12]. In the simulations, we will focus on the optimal SD and the suboptimal LLL solutions for (44), respectively.

VI. NUMERICAL RESULTS

In this section, we show some numerical results of the proposed MZF detector, as well as its extensions. In all tests, we test with $K \times K$ real-valued MIMO channels (each element is an independent and identically distributed Gaussian variable with a zero-mean and unit-variance) with \sqrt{M} -PAM modulated symbols that are transferred from $K/2 \times K/2$ complex-valued MIMO channels and M -QAM modulated symbols. We simulate 50,000 channel realizations for each of the tests.

A. SINR improvements

In Fig. 1 we show the post-processing SNR improvements with the MZF detector using Algorithm 1, and compare to a traditional ZF detector with different PAM modulations (i.e., τ values). As can be seen, the SNRs are greatly improved, especially for low-order modulations (or the weak bit-layers of high-order modulations with Extension 2). When τ decreases, the gains become smaller. We also test the MZF with Extension 1, where we can observe only marginal gains (not shown in Fig. 1). Therefore, in the remaining tests we set $\alpha=1$.

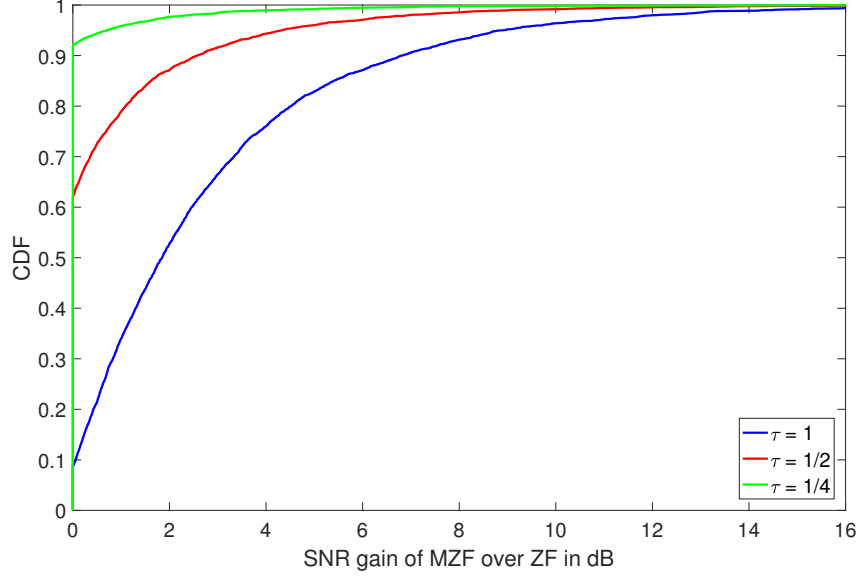


Fig. 1. SNR gains under real-valued 12×12 MIMO with 2-PAM, 4-PAM and 8-PAM modulations, respectively.

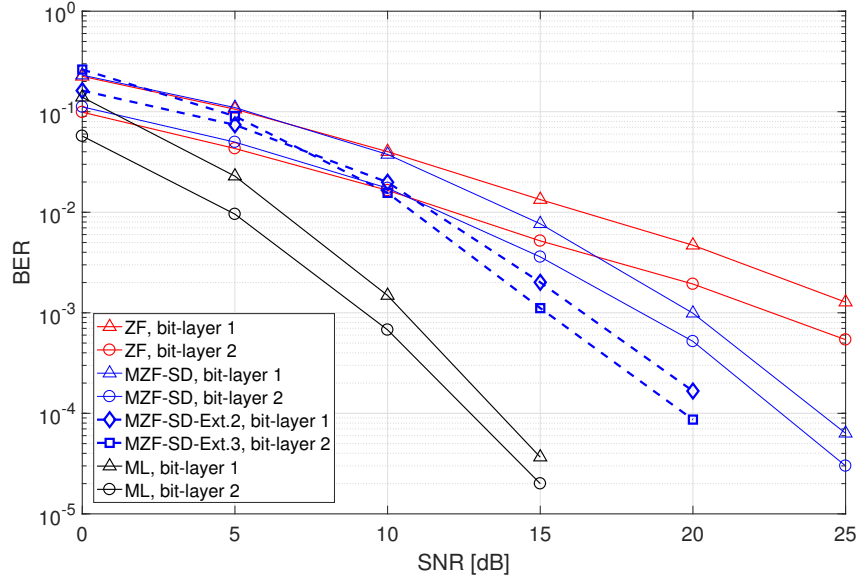


Fig. 2. Uncoded BER under real-valued 6×6 MIMO with 4-PAM modulation.

B. Uncoded bit-error-rate (BER)

Next we show uncoded BER performance. In Fig. 2 we compare MZF with ZF and ML under 6×6 MIMO with 4-PAM modulation. The MZF uses SD to find the optimal \mathbf{q}_k . As can be seen, the MZF without extensions outperforms the ZF with more than 2 dB at 0.1% BER. With Extension 2, the BER of the first bit-layer (weaker layer) is greatly improved by more than 4 dB at 0.1% BER and outperforms the second bit-layer, which justifies the application of Extension

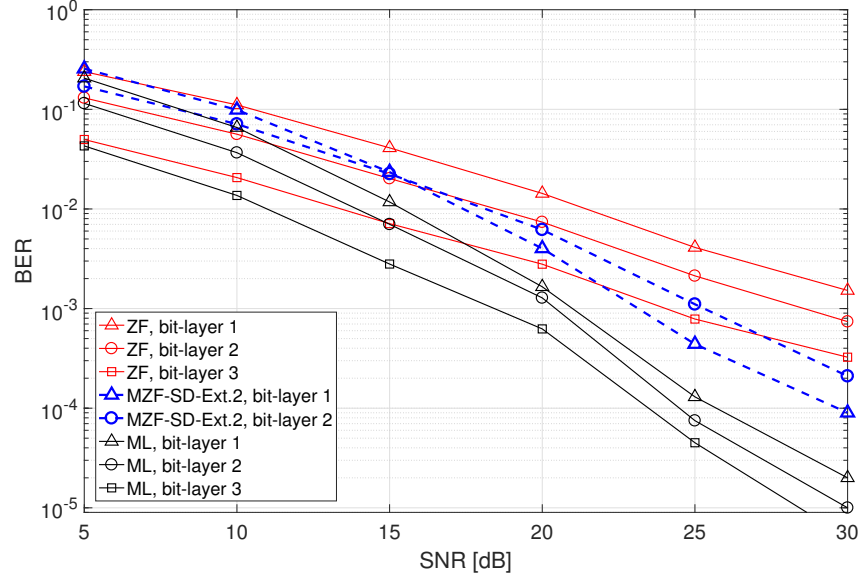


Fig. 3. Uncoded BER under real-valued 4×4 MIMO with 8-PAM modulation.

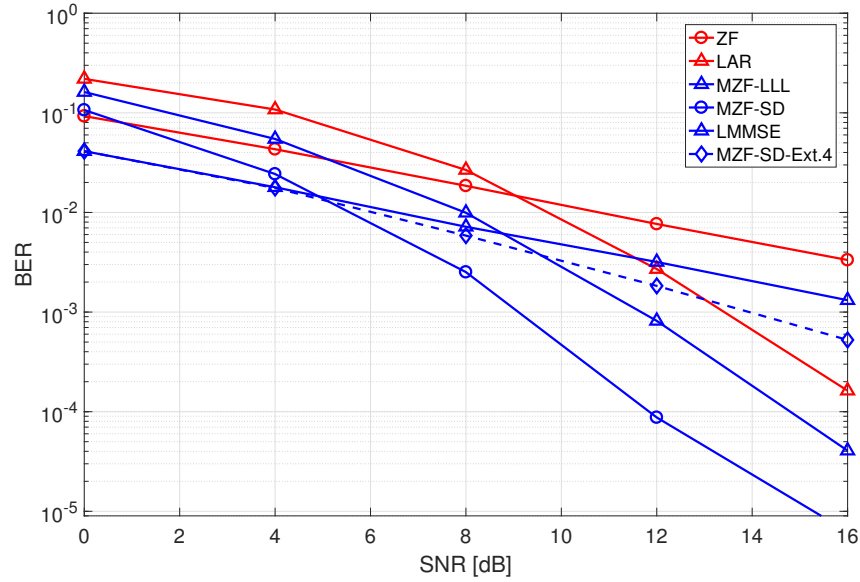


Fig. 4. Uncoded BER under real-valued 8×8 MIMO with 2-PAM modulation.

3. With Extension 3, where the feedbacks of the first bit-layer are used, the BER of the second bit-layer is also improved by more than 3 dB at 0.1% BER compared to MZF with Extension 2. The gaps between ZF and ML are significantly reduced by the MZF, and the slopes of BER with MZF are also much steeper than the ZF, and close to those with the ML. However, as also can be observed, the MZF has only marginal gains at low SNRs, and the decision-feedback approach performs even worse due to inaccurate feedbacks. This obstacle can be relieved by

using LMMSE based approaches, i.e., Extension 4.

In Fig. 3 we repeat the tests in Fig. 2 under 4×4 MIMO with 8-PAM modulation, that is, three bit-layers are considered. The MZF with Extension 2 using SD is compared to ZF and ML. As already shown in Fig. 1, setting $\tau = 1/4$ for detecting the third-layer (strongest layer) only has small gains, and the BER performance is also close to ZF and therefore are not shown in Fig. 3. Nevertheless, the BER of the first and second bit-layers are significantly boosted by MZF. As can be seen, the MZF performs around 3 dB better than the ZF at 0.1% BER for the second bit-layer, and 7 dB better for the first bit-layer. Since the weakest bit-layer usually has a stronger impact on the decoding performance, the gains for the first bit-layer is of importance.

C. Comparison with LAR

In Fig. 4 we compare the MZF with the LAR under 8×8 MIMO with 2-PAM modulation. The MZF uses both SD and the approximate LLL method to find the optimal \mathbf{q}_k . As can be seen, MZF outperforms LAR with more than 1.5 dB at 0.1% BER, with a similar complexity for running LLL algorithm for LR.

Moreover, with Extension 4 (the LMMSE based detection), the BERs at low SNRs are also improved with the MZF which is inferior to the original ZF based MZF at high SNRs. Nevertheless, with Extension 4, the MZF is 4 dB better than a normal ZF, and more than 2dB better than a normal LMMSE detector at 0.1% BER. Another observation is that, the SD based MZF is more than 2 dB better than the LLL based MZF, which shows that optimal selection of \mathbf{q}_k is important.

VII. SUMMARY

We have proposed a novel modulus arithmetic based zero-forcing (MZF) detector for multi-input multi-output (MIMO) channels, with a number of possible extensions of the basic algorithm. The MZF detector shows significant gains in terms of post-processing signal-to-noise-ratio (SNR) and bit-error-rate (BER) compared to traditional linear detectors, at medium and high SNR scenarios and in particular for weak bit-layers. At low SNRs and with large cardinalities of the input alphabet, we have provided several possible extensions to improve the performance of the MZF. Finding the optimal modulus matrix itself is a burdening MIMO detection problem, but it needs to be done only once per a coherence-interval of the MIMO channel using sphere-decoding (SD) or other suboptimal algorithms. In particular, with a similar complexity, MZF

with lattice-reduction (LR) based approaches outperforms the traditional lattice-aided-reduction (LAR) detector, which justifies its potential in MIMO detection.

APPENDIX A: PROOF OF PROPERTY 1

Since $p_m, (b_m - 1) \in 2\mathbb{Z}$, we let $p_m = 2\tilde{p}_m$ and $b_m = 2\tilde{b}_m + 1$, where $\tilde{p}_m, \tilde{b}_m \in \mathbb{Z}$. Then,

$$\begin{aligned} y &= z + \alpha \sum_{m=1}^M p_m b_m \\ &= z + 4\alpha \sum_{m=1}^M \tilde{p}_m \tilde{b}_m + 2\alpha \sum_{m=1}^M \tilde{p}_m. \end{aligned} \quad (46)$$

Since $|z| < 2$ and $\alpha > 1$, it holds that $z + 2\alpha > 0$. If $\frac{1}{2} \sum_{m=1}^M p_m = \sum_{m=1}^M \tilde{p}_m$ is odd, we have

$$y \bmod 4\alpha = z + 2\alpha; \quad (47)$$

Otherwise, if $\sum_{m=1}^M \tilde{p}_m$ is even, it also holds that

$$(y + 2\alpha) \bmod 4\alpha = z + 2\alpha. \quad (48)$$

Combing (47) and (48), z can be obtained as in (12).

APPENDIX B: A 4×4 EXAMPLE FOR APPLYING THE MZF DETECTION

Below we give a 4×4 real-valued MIMO example with 4-PAM modulation to illustrate the process of MZF detection, with assuming the channel, transmitted symbol vector and received signal vector as

$$\mathbf{H} = \begin{bmatrix} -6 & 0 & -1 & 5 \\ -3 & -2 & -1 & 1 \\ 1 & -5 & -6 & 0 \\ 1 & -1 & -3 & -2 \end{bmatrix}, \mathbf{x} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 3 \\ 1 \\ 15 \\ 11 \end{bmatrix},$$

respectively. Then it can be shown that

$$\mathbf{H}^+ = \frac{1}{185} \begin{bmatrix} -5 & -55 & 30 & -40 \\ 35 & -59 & -25 & 58 \\ -30 & 40 & -5 & -55 \\ 25 & -58 & 35 & -59 \end{bmatrix},$$

and the ZF estimate of \mathbf{x} equals

$$\tilde{\mathbf{x}}_{\text{ZF}} = \mathbf{H}^+ \mathbf{y} = \frac{1}{185} \begin{bmatrix} -60 \\ 309 \\ -730 \\ -107 \end{bmatrix},$$

where only the third symbol is correctly detected.

Next we use the basic MZF detection with Algorithm 1. Setting $\tau = 1$ and run SD for optimization (17) yields an optimal \mathbf{Q} as

$$\mathbf{Q} = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & -2 & 0 \\ 2 & 0 & 0 & -2 \end{bmatrix}.$$

We first see that, the MZF shall reuse the ZF estimates for the first and third layers based on (21). Then, we see that with \mathbf{Q} , the post-processing SNR (assuming the noise power equals 1) for the second bit-layer (which is identical to the fourth bit-layer) is increased from $1/\|\boldsymbol{\delta}_2 \mathbf{H}^+\| = 185/47$ to $1/\|(\boldsymbol{\delta}_2 + \mathbf{q}_2) \mathbf{H}^+\|^2 = 185/27$. Next we compute estimates with the MZF for the these two layers.

For the second layer, according to (14) we have

$$r_2 = (\boldsymbol{\delta}_2 + \mathbf{q}_2) \tilde{\mathbf{x}}_{\text{ZF}} = \frac{1}{185} [0 \ 1 \ 2 \ 0] \begin{bmatrix} -60 \\ 309 \\ -730 \\ -107 \end{bmatrix} = \frac{-1151}{185},$$

and

$$z_2 = (r_2 \bmod 4) - 2 = \frac{-7}{38}.$$

Similarly, for the fourth layer we have

$$r_4 = (\boldsymbol{\delta}_4 + \mathbf{q}_4) \tilde{\mathbf{x}}_{\text{ZF}} = \frac{1}{185} [2 \ 0 \ 0 \ -1] \begin{bmatrix} -60 \\ 309 \\ -730 \\ -107 \end{bmatrix} = \frac{-13}{185},$$

and

$$z_4 = (r_4 \bmod 4) - 2 = \frac{357}{185}.$$

As can be seen, the MZF corrects both detections for the second and the fourth layers where the ZF fails.

APPENDIX C: PROOF OF PROPERTY 2

In the case where each transmit antenna encodes an independent data stream, the optimal DMT [18] is

$$d_{\text{ML}}(r) = N \left(1 - \frac{r}{K}\right), \quad 0 \leq r \leq K, \quad (49)$$

which can be achieved by joint maximum likelihood (ML) decoding, while the ZF (and also the LMMSE) receiver attains a DMT

$$d_{\text{ZF}}(r) = (N - K + 1) \left(1 - \frac{r}{K}\right), \quad 0 \leq r \leq K. \quad (50)$$

Since $\tilde{\mathbf{H}}$ has i.i.d. Rayleigh entries, $\tilde{\mathbf{H}}$ is full column rank with probability 1, and so is $\mathbf{H} \in \mathbb{Z}^{2N \times 2K}$ which is the real-valued representation $\tilde{\mathbf{H}}$. To prove that the MZF detection can achieve the optimal DMT, we follow a similar proof in [13] for IF receiver, which builds on a result in [23] that showed that, uncoded signaling coupled with LR can achieve the full diversity and with a multiplexing gain of zero.

We define a lattice Λ and its dual Λ^* generated by $2N \times 2K$ matrix \mathbf{H} and $2K \times 2N$ matrix \mathbf{H}^+ as [19]

$$\Lambda = \{\mathbf{H}\mathbf{p}, \mathbf{p} \in \mathbb{Z}^{2K}\}, \quad (51)$$

$$\Lambda^* = \{(\mathbf{H}^T)^+ \mathbf{p}, \mathbf{p} \in \mathbb{Z}^{2K}\}, \quad (52)$$

respectively. From [19]–[21], the i -th successive minimum $\epsilon_i(\Lambda)$ of a lattice Λ is defined as the smallest length r (with respect to the Euclidean norm) such that, there are i vectors in Λ of length at most r that are linearly independent (with respect to \mathbb{R}). By this definition, it holds that [13]

$$\epsilon_1^2(\Lambda) = \min_{\mathbf{p} \in \mathbb{Z}^{2K}} \|\mathbf{p}\mathbf{H}^T\|^2, \quad (53)$$

$$\epsilon_{2K}^2(\Lambda^*) = \min_{\substack{\mathbf{P} \in \mathbb{Z}^{2K \times 2K} \\ \text{rank}(\mathbf{P})=2K}} \max_{1 \leq k \leq 2K} \|\mathbf{p}_k \mathbf{H}^+\|^2. \quad (54)$$

Then, we state the following Lemma 1, which follows the proof in [13, Appendix C].

Lemma 1. *Let Λ be the lattice generated by \mathbf{H} lattice generated by \mathbf{H} according to (51). Then, for a sufficiently large $s > 0$, the successive minimum $\epsilon_K^2(\Lambda)$ of Λ satisfies*

$$\mathbb{P}(\epsilon_{2K}^2(\Lambda^*) > s) \leq \mathbb{P}\left(\frac{K^2(2K+3)}{\epsilon_1^2(\Lambda)} > s\right) \quad (55)$$

$$= \mathbb{P}(\epsilon_1^2(\Lambda) < \tilde{s}) \quad (56)$$

$$\leq -\gamma \tilde{s}^{2N} (\ln \tilde{s})^{1+N}, \quad (57)$$

where $\epsilon_1^2(\Lambda)$ is the first successive minimum for Λ , and

$$\tilde{s} = K \sqrt{\frac{2K+3}{s}}. \quad (58)$$

The inequality (55) follows from [13, Lemma 4], and the equality (56) holds by switching $\epsilon_1^2(\Lambda)$ to the other side, and the inequality (57) is a result of [13, Lemma 5] when $\tilde{s} > 0$ and is sufficiently small.

With the post-processing SNR γ_k in MZF detection in (11), we can show that for a given target rate $r \ln \text{SNR}$, the outage probability equals

$$\begin{aligned} P(R < r \ln \text{SNR}) &= P\left(\max_{\mathbf{Q} \in 2\mathbb{Z}^{2K \times 2K}} \sum_{k=1}^{2K} \frac{1}{2} \ln(1 + \gamma_k) < r \ln \text{SNR}\right) \\ &\leq P\left(\max_{\substack{\mathbf{Q} \in 2\mathbb{Z}^{2K \times 2K} \\ \text{rank}(\mathbf{Q})=2K}} \sum_{k=1}^{2K} \frac{1}{2} \ln(1 + \gamma_k) < r \ln \text{SNR}\right) \\ &\leq P\left(\max_{\substack{\mathbf{Q} \in 2\mathbb{Z}^{2K \times 2K} \\ \text{rank}(\mathbf{Q})=2K}} \min_{1 \leq k \leq 2K} \gamma_k < \text{SNR}^{r/K}\right) \\ &= P\left(\max_{\substack{\mathbf{Q} \in 2\mathbb{Z}^{2K \times 2K} \\ \text{rank}(\mathbf{Q})=2K}} \min_{1 \leq k \leq 2K} \frac{\tau^2 \text{SNR}}{\|(\tau \boldsymbol{\delta}_k + \mathbf{q}_k) \mathbf{H}^+\|^2} < \text{SNR}^{r/K}\right) \\ &= P\left(\min_{\substack{\mathbf{Q} \in 2\mathbb{Z}^{2K \times 2K} \\ \text{rank}(\mathbf{Q})=2K}} \max_{1 \leq k \leq 2K} \|(\tau \boldsymbol{\delta}_k + \mathbf{q}_k) \mathbf{H}^+\|^2 > \tau^2 \text{SNR}^{1-r/K}\right). \quad (59) \end{aligned}$$

Using the inequality

$$\|(\tau \boldsymbol{\delta}_k + \mathbf{q}_k) \mathbf{H}^+\|^2 \leq \|\mathbf{q}_k \mathbf{H}^+\|^2 + \tau^2 \|\boldsymbol{\delta}_k \mathbf{H}^+\|^2, \quad (60)$$

and inserting (60) back into (59) yields

$$\begin{aligned}
P(R < r \ln \text{SNR}) &\leq P \left(\min_{\substack{\mathbf{Q} \in \mathbb{Z}^{2K \times 2K} \\ \text{rank}(\mathbf{Q})=2K}} \max_{1 \leq k \leq 2K} \|\mathbf{q}_k \mathbf{H}^+\|^2 > \tau^2 \left(\text{SNR}^{1-r/K} - \|\boldsymbol{\delta}_k \mathbf{H}^+\|^2 \right) \right) \\
&= P \left(\min_{\substack{\mathbf{P} \in \mathbb{Z}^{2K \times 2K} \\ \text{rank}(\mathbf{P})=2K}} \max_{1 \leq k \leq 2K} \|\mathbf{p}_k \mathbf{H}^+\|^2 > \frac{\tau^2}{4} \left(\text{SNR}^{1-r/K} - \|\boldsymbol{\delta}_k \mathbf{H}^+\|^2 \right) \right) \\
&= P \left(\epsilon_{2K}^2(\Lambda^*) > \frac{\tau^2}{4} \left(\text{SNR}^{1-r/K} - \|\boldsymbol{\delta}_k \mathbf{H}^+\|^2 \right) \right), \tag{61}
\end{aligned}$$

where the successive minimum $\epsilon_{2K}^2(\Lambda^*)$ is defined in (54). Comparing (61) to [13, Eq. (43)], it can be seen that the outage probability with MZF detection could be inferior to that with the IF receiver. However, the slope of the outage probability in high SNR regimes remains the same.

Using Lemma 1, it holds that

$$\begin{aligned}
d_{\text{MZF}}(r) &= \lim_{\text{SNR} \rightarrow \infty} \frac{-\ln P(R < r \ln \text{SNR})}{\ln \text{SNR}} \\
&= \lim_{\text{SNR} \rightarrow \infty} \frac{-\ln P \left(\epsilon_{2K}^2(\Lambda^*) > \frac{\tau^2}{4} \left(\text{SNR}^{1-r/K} - \|\boldsymbol{\delta}_k \mathbf{H}^+\|^2 \right) \right)}{\ln \text{SNR}} \\
&\geq \lim_{\text{SNR} \rightarrow \infty} \frac{-\ln \left(-\gamma \hat{s}^{2N} (\ln \hat{s})^{1+N} \right)}{\ln \text{SNR}}, \tag{62}
\end{aligned}$$

where

$$\begin{aligned}
\hat{s} &= \frac{2K}{\tau} \sqrt{\frac{2K+3}{\text{SNR}^{1-r/K} - \|\boldsymbol{\delta}_k \mathbf{H}^+\|^2}} \\
&\geq \frac{2K}{\tau} \sqrt{\frac{2K+3}{\text{SNR}^{1-r/K}}}. \tag{63}
\end{aligned}$$

Inserting (63) into (62), it holds that

$$d_{\text{MZF}}(r) \geq N \left(1 - \frac{r}{K} \right). \tag{64}$$

But as the MZF detection is inferior to the ML detection, so is the DMT. Therefore, the equality in (64) holds.

REFERENCES

- [1] S. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. Surveys & Tutorials*, vol. 17, no. 4, pp. 1941-1988, Sep. 2015.
- [2] S. M. Kay, "Fundamentals of statistical signal processing, volume I: Estimation theory," Prentice Hall signal processing series, 1993.

- [3] A. Ghasemmehdi and E. Agrell, "Faster recursions in sphere decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3530-3536, Jun. 2011.
- [4] E. Agrell, T. Eriksson, A. Vardy and K. Zeger, "Closest point search in lattices," *IEEE Trans. Inf. Theory*, vol. 48, no. 8, pp. 2201-2214, Aug. 2002.
- [5] D. Wubben, R. Bohnke, V. Kuhn and K. Kammeyer, "Near-maximum-likelihood detection of MIMO systems using MMSE-based lattice reduction," *IEEE Int. Conf. Commun. (ICC)*, Jun. 2004, pp. 798-802.
- [6] A. K. Lenstra, H. W. Lenstra and L. Lovász, "Factoring polynomials with rational coefficients" *Mathematische Annalen*, vol. 261 no. 4, pp. 515-534, Jul. 1982.
- [7] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics Lett.*, vol. 7, no. 5, pp. 138-139, Mar. 1971.
- [8] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels With intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774-780, Aug. 1972.
- [9] M. H. Costa, "Writing on dirty paper," *IEEE Trans. Inf. Theory*, vol. 29, no. 2, pp. 439-441, May 1983.
- [10] B. M. Hochwald, C. B. Peel and A. L. Swindlehurst, "A vector-perturbation technique for near-capacity multiantenna multiuser communication - Part II: Perturbation," *IEEE Trans. Commun.*, vol. 53, no. 3, pp. 537-544, Mar. 2005.
- [11] J. Maurer, J. Jaldén, D. Seethaler and G. Matz, "Vector perturbation precoding revisited," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 315-328, Jan. 2011.
- [12] S. Hu and F. Rusek, "A soft-output MIMO detector with achievable information rate based partial marginalization," *IEEE Trans. Signal Processing*, vol. 65, no.6, pp. 1622-1637, Mar. 2017.
- [13] J. Zhan, B. Nazer, U. Erez, and M. Gastpar, "Integer-forcing linear receivers," *IEEE Trans. Inform. Theory*, vol. 60, no. 12, pp. 7661-7685, Dec. 2014.
- [14] S.-N. Hong and G. Caire, "Compute-and-forward strategies for cooperative distributed antenna systems," *IEEE Trans. Inform. Theory*, vol. 59, no. 9, pp. 5227-5243, Sep. 2013.
- [15] S. H. Chae, M. Jang, and S.-K. Ah, "Multilevel coding scheme for integer-forcing MIMO receivers with binary codes", *IEEE Trans. Wireless. Commun.*, vol. 16, no. 8, pp. 5428-5441, Aug. 2017.
- [16] O. Ordentlich and U. Erez, "Achieving the gains promised by Integer-Forcing equalization with binary codes", *IEEE Conv. Elect. and Electron. Eng. in Israel (IEEEI)*, Nov. 2010, pp. 000703-000707.
- [17] B. Nazer and M. Gastpar, "Compute-and-forward: Harnessing interference through structured codes", *IEEE Trans. Inform. Theory*, vol. 57, no. 10, pp. 6463-6486, Oct. 2011.
- [18] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels, *IEEE Trans. Inf. Theory*," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073-1096, May 2003.
- [19] J. C. Lagarias, H. W. Lenstra, Jr., and C. P. Schnorr, "KorkinZolotarev bases and successive minima of a lattice and its reciprocal lattice," *Combinatorica*, vol. 10, no. 4, pp. 333-348, Dec. 1990.
- [20] S. Lang, *Introduction to Diophantine approximations*, New York, NY, USA, Springer-Verlag, 1995.
- [21] I. Aliev and M. Henk, "Successive minima and best simultaneous Diophantine approximations," *Monatshefte Math.*, vol. 147, no. 2, pp. 95-101, 2006.
- [22] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication over MIMO broadcast channels using lattice-basis reduction", *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4567-4582, Dec. 2007.
- [23] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "LLL reduction achieves the receive diversity in MIMO decoding", *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4801-4805, Dec. 2007.

Algorithm 1 Modulus Zero-Forcing (MZF) Algorithm

\mathbf{H} is $K \times K$ real-valued

\mathbf{y} is $K \times 1$ real-valued

M is cardinality of QAM constellation

1: **function** $\hat{\mathbf{x}} = \text{MODULARZF}(\mathbf{H}, \mathbf{y}, M)$

2: $\tau = 2^{(1 - \log_2(\sqrt{M}))}$

3: $\mathbf{B} = -\mathbf{H}^+$

Preprocessing for each coherence interval

4: **for** $k = 1$ **to** K

5: $\mathbf{b}_k = \tau \boldsymbol{\delta}_k \mathbf{H}^+$

6: Solve : $\mathbf{q}_k^{\text{opt}} = \arg \min_{\mathbf{q}_k} \|\mathbf{b}_k - \mathbf{q}_k \mathbf{B}\|^2$

7: **end for**

Executed for every channel observation

8: **for** $k = 1$ **to** K

9: $r_k = (\tau \boldsymbol{\delta}_k + \mathbf{q}_k^{\text{opt}}) \mathbf{H}^+ \mathbf{y}$

10: **if** $\|\mathbf{q}_k^{\text{opt}} - \boldsymbol{\delta}_k \odot \mathbf{q}_k^{\text{opt}}\|^2 = 0$ **then**

11: $z_k = \mathbf{b}_k \mathbf{y}$

12: **else**

13: **if** $\frac{1}{2} \sum_{\ell=1}^K q_{k\ell}^{\text{opt}}$ is odd **then**

14: $z_k = (r_k \bmod 4) - 2$

15: **else**

16: $z_k = ((r_k + 2) \bmod 4) - 2$

17: **end if**

18: **end if**

19: $\hat{x}_k = \mathcal{Q}_{\tau\mathcal{A}}(z_k)$

20: **end for**

21: **end function**

Algorithm 2 MZF Algorithm with Extension 2

\mathbf{H} is $K \times K$ real-valued

\mathbf{y} is $K \times 1$ real-valued

M is cardinality of QAM constellation

1: **function** $\hat{\mathbf{x}} = \text{MODULARZFEXT2}(\mathbf{H}, \mathbf{y}, M)$

2: $N = \log 2(\sqrt{M})$

3: $\mathbf{B} = -\mathbf{H}^+$

4: **for** $n = 1$ **to** N

5: $\tau(n) = 2^{1-n}$

6: **end for**

Preprocessing for each coherence interval

7: **for** $k = 1$ **to** K

8: $\mathbf{b}_k = \delta_k \mathbf{H}^+$

9: **for** $n = 1$ **to** N

10: $\mathbf{b}_{kn} = \tau(n) \mathbf{b}_k$

11: Solve : $\mathbf{q}_{k,n}^{\text{opt}} = \arg \min_{\mathbf{q}_k} \|\mathbf{b}_{kn} - \mathbf{q}_k \mathbf{B}\|^2$

12: **end for**

13: **end for**

Executed for every channel observation

14: **for** $k = 1$ **to** K

15: **for** $n = 1$ **to** N

16: **if** $n = N$ **and** $\|\mathbf{q}_k^{\text{opt}} - \delta_k \odot \mathbf{q}_k^{\text{opt}}\|^2 = 0$, **then**

17: $z_k = \mathbf{b}_k \mathbf{y}$

18: **else**

19: $r_k = (\tau(n) \delta_k + \mathbf{q}_k^{\text{opt}}) \mathbf{H}^+ \mathbf{y}$

20: $z_k = (r_k \bmod 4) - 2$

21: **end if**

22: $\hat{u}_{kn} = \text{sign}(z_k)$

23: **end for**

24: **end for**

25: **end function**

Algorithm 3 MZF Algorithm with Extension 3

\mathbf{H} is $K \times K$ real-valued

\mathbf{y} is $K \times 1$ real-valued

M is cardinality of QAM constellation

1: **function** $\hat{\mathbf{x}} = \text{MODULARZFEXT3}(\mathbf{H}, \mathbf{y}, M)$

2: $\tau = 1$

3: $N = \log_2(\sqrt{M})$

4: $\mathbf{B} = -\mathbf{H}^+$

Preprocessing for each coherence interval

5: **for** $k = 1$ **to** K

6: $\mathbf{b}_k = \tau \delta_k \mathbf{H}^+$

7: Solve : $\mathbf{q}_k^{\text{opt}} = \arg \min_{\mathbf{q}_k} \|\mathbf{b}_k - \mathbf{q}_k \mathbf{B}\|^2$

8: **end for**

Executed for every channel observation

9: $\hat{\mathbf{y}} = \mathbf{y}$

10: **for** $n = 1$ **to** N

11: $\hat{\mathbf{u}}_n = \mathbf{0}$

12: **for** $k = 1$ **to** K

13: **if** $\|\mathbf{q}_k^{\text{opt}} - \delta_k \odot \mathbf{q}_k^{\text{opt}}\|^2 = 0$, **then**

14: $\mathbf{z}_k = \tau \delta_k \mathbf{H}^+ \hat{\mathbf{y}}$

15: **else**

16: $\mathbf{r}_k = (\tau \delta_k + \mathbf{q}_k^{\text{opt}}) \mathbf{H}^+ \hat{\mathbf{y}}$

17: $\mathbf{z}_k = (\mathbf{r}_k \bmod 4) - 2$

18: **end if**

19: $\hat{\mathbf{u}}_{kn} = \text{sign}(\mathbf{z}_k)$

20: **end for**

21: $\hat{\mathbf{u}}_n = [\hat{u}_{1n} \ \hat{u}_{2n} \ \dots \ \hat{u}_{Kn}]^T$

22: $\hat{\mathbf{y}} = (\hat{\mathbf{y}} - \mathbf{H} \hat{\mathbf{u}}_n)/2$

23: **end for**

24: **end function**
