

RESEARCH

Open Access



# SNR maximization and modulo loss reduction for Tomlinson-Harashima precoding

Erica Debels\* and Marc Moeneclaey

## Abstract

Compared to linear precoding, Tomlinson-Harashima precoding (THP) requires less transmit power to eliminate the spatial interference in a multi-user downlink scenario involving a multi-antenna transmitter and geographically separated receivers. However, THP gives rise to certain performance losses, referred to as modulo loss and power loss. Based on the observation that part of the users can omit the modulo operation at the receiver during an entire frame, we present an alternative detector, which reduces the modulo loss compared to the conventional detector. In addition, this contribution compares several existing and novel algorithms for selecting the user ordering and the rotation of the constellations at the transmitter, to increase the SNR at the detector and decrease the modulo loss for the alternative detector. Compared to the better of linear precoding and THP with conventional detector, the optimized alternative detector achieves significant gains (up to about 4 dB) for terrestrial wireless communication, whereas smaller gains (up to about 1 dB) are obtained for multi-beam satellite communication.

**Keywords:** Multi-user multiple-input single-output (MU-MISO), Tomlinson-Harashima precoding (THP), Interference cancellation, Flat fading terrestrial communication, Multi-beam satellite communication, SNR maximization, Modulo loss (MoL) reduction

## 1 Introduction

When using spatial multiplexing in a communication system, consisting of a multi-antenna transmitter (TX), a frequency-flat channel, and a number of user terminals, each user receives a linear combination of the signals sent by the different TX antennas. When the TX has full channel information, the interference is known at the TX, in which case dirty paper coding (DPC) is capacity-achieving [1]. As the implementation of DPC is rather complex, it is of interest to consider simpler precoding schemes for reducing the interference among the different information streams. Linear precoding (LP) is the most basic scheme, where each antenna transmits a linear combination of the data symbols to be sent to the different users. LP is able to completely eliminate the interference at each user terminal, but has an important drawback: depending on the channel realization, the interference pre-subtraction term to be generated at the TX can be very large, which causes a substantial TX power penalty.

The power penalty associated with LP can be significantly reduced by applying a form of nonlinear precoding, referred to as Tomlinson-Harashima precoding (THP), which involves a modulo operation at both the TX and the receiver (RX).

Originally, THP has been introduced to avoid inter-symbol interference on frequency-selective single-input single-output channels [2, 3], but THP has also been applied to spatial multiplexing systems operating over flat channels. Depending on the scenario, one distinguishes between single-user (SU) multiple-input multiple-output (MIMO) THP, where both the TX and the RX are equipped with multiple antennas [4–6] multi-user (MU) multiple-input single-output (MISO) THP, where a multi-antenna TX sends to several single-antenna RXs [7–18]; and MU-MIMO THP, involving a multi-antenna TX sending to several multi-antenna RXs [19–21]. Moreover, THP is envisaged also in communication systems providing physical-layer security [22], and for simultaneous wireless information and power transfer (SWIPT) [23]. To implement THP, the TX requires channel information; whereas in many contributions, perfect channel information is assumed, the design of THP in the

\*Correspondence: erica.debels@ugent.be

Ghent University, Dept. of Telecommunications and Information Processing,  
Sint-Pietersnieuwstraat 41, B-9000 Gent, Belgium

presence of imperfect channel information is considered in [4, 6, 11, 16, 20, 23].

THP makes use of a modulo operation at the pre-coder to reduce its TX power compared to LP. However, as the modulo boundaries are larger than the constellation boundaries, the resulting TX power is larger than in the case without interference pre-subtraction; the latter difference is referred to as the power loss (PoL) [9]. At the receiver side, the conventional detector again applies a modulo operation, which comes with an additional performance penalty, referred to as the modulo loss (MoL) [9].

Several approaches have been proposed to improve the performance of the THP communication system; however, most of them leave the MoL unaltered, because the modulo operation at the RX is maintained. The PoL is minimized in [14] by applying constellation rotations, and in [15] by rotating and scaling the symbols of the first user. The performance of THP can be improved by an appropriate reordering of the users before precoding. Because of the upstream-downstream duality, the precoding order for THP can be derived from the decoding order in an upstream system with successive interference cancellation at the RX, referred to as V-BLAST [24]. To reduce the complexity of the V-BLAST ordering presented in [24], a low-complexity ordering algorithm based on a sorted QR decomposition is proposed in [25]. A precoding order derived from the V-BLAST ordering is considered in [12]. In [8, 13], algorithms based on a Cholesky decomposition with symmetric permutation are introduced, for optimum and suboptimum ordering in the upstream (interference cancellation at RX) and the downstream (precoding at TX) directions. While these orderings do not depend on the transmitted data symbols, a data-dependent precoding ordering algorithm is considered in [10]. The above ordering algorithms for THP pertain to MU-MISO; orderings for SU-MIMO and MU-MIMO are developed in [5] and [19, 21], respectively.

Based on the observation that, depending on the channel realization, part of the users can omit the modulo operation at the RX during an entire frame, an alternative detector is introduced in [17, 18], which applies the modulo operation only when the TX instructs the RX to do so; this way, the MoL is reduced compared to the conventional detector, for which the modulo operation is always active. At the same time, in [17, 18], the constellations are rotated to maximize the number of users which can discard the modulo operation.

In this contribution, we consider MU-MISO THP and we assume that perfect channel information is available at the TX and the RX. We investigate how the constellation rotations and the reordering of the users affect the SNR and the MoL of the alternative detector from [17, 18]. Besides applying some existing and some new

algorithms which optimize over either the user ordering or the constellation rotations, this contribution also jointly maximizes the SNR at the detector and reduces the MoL, by optimizing over both the ordering and the rotations. Instead of performing an exhaustive search over the rotations and/or the user orderings, we also investigate reduced-complexity algorithms, among which most are novel as well.

Since a broad range of SNR values is explored, the uncoded error performance is not a suitable performance measure, because it does not provide a proper indication of the performance at low SNR (where coding should be used). Instead, we consider the mutual information (MI) for the specific constellation, averaged over the users and the channel realizations; this average MI represents the information-theoretical achievable spectral efficiency (information bits per channel use) per user when the communication system makes use of THP on the considered channel. The alternative detector (non-optimized and optimized) will be compared to the conventional detector (non-optimized and optimized) for THP and to the LP, in terms of average MI.

Numerical results are presented for a terrestrial Rayleigh fading channel, and for a multi-beam satellite channel with full frequency reuse. Whereas THP on Rayleigh fading channels has already been intensively studied, the interest in applying THP in multi-beam satellite systems is rather new. Because of the continuously increasing demand for higher throughput, full frequency reuse in multi-beam satellite systems recently gained attention. Due to the closer proximity of users operating in the same bandwidth, co-channel interference is higher and becomes a limiting factor when not properly mitigated. Therefore, interference cancellation (in the upstream direction) and precoding (in the downstream direction) techniques suited for terrestrial Rayleigh fading channels are currently envisaged for multi-beam satellite systems as well [26–33].

We briefly summarize the main original contributions of this paper:

- Whereas other works focus on optimizing over either the constellation rotations or the user ordering, we also investigate the gain obtained from optimizing over both.
- We optimize the transmitter to jointly reduce the modulo loss and increase the SNR for the alternative detector introduced in [17].
- Whereas in literature the optimizations over the rotations typically use exhaustive searches, we present several novel reduced-complexity algorithms.
- Whereas in literature the optimizations over the user ordering do not aim to reduce the MoL, we present novel user ordering optimizations for reducing the MoL of the alternative detector.

- We present the results for both a terrestrial Rayleigh fading channel and a multi-beam satellite channel. Whereas the former channel has already been studied extensively in the context of THP, optimization results related to the latter channel are rather scarce in literature.

This contribution is organized as follows. Section 3 outlines the system and channel model of both the terrestrial wireless link and multi-beam satellite system. Section 4 describes the MU-MISO system with zero forcing (ZF) THP, and Section 5 identifies the modulo and power losses inherent to THP. Section 6 considers an alternative detector which reduces the modulo loss compared to the conventional detector, and in Section 7, various algorithms for optimizing the user ordering and the constellation rotations at the TX are presented. Section 8 provides numerical performance results, pertaining to the conventional and the alternative detector for THP (without and with optimizations) and to LP. Section 9 concludes the paper.

Throughout this paper the following notations are used. Lowercase bold letters refer to a column vector, while uppercase bold letters denote a matrix.  $\mathbf{I}_N$  refers to the  $N \times N$  identity matrix;  $(\cdot)^T$  and  $(\cdot)^H$  stand for the transpose and the hermitian transpose, respectively. We denote by  $\text{diag}(\mathbf{X})$  a diagonal matrix with the same diagonal elements as the square matrix  $\mathbf{X}$ .  $\|\cdot\|$  is the norm of a vector.  $\mathbb{E}[\cdot]$  stands for the statistical expectation, and  $I(x; y) = \mathbb{E}_{x,y}[\log_2(p(y|x)/p(y))]$  denotes the mutual information (MI) between the random variables  $x$  and  $y$ , with joint distribution  $p(x,y) = p(y|x)p(x) = p(x|y)p(y)$ . The notation  $\mathbf{x} \sim \mathcal{N}_c(0, \mathbf{R})$  is used to indicate that  $\mathbf{x}$  is a vector of complex-valued zero-mean circular symmetric Gaussian random variables with autocorrelation matrix  $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{R}$ . The real and imaginary part of  $x$  are denoted by  $\Re(x)$  (or  $x_R$ ) and  $\Im(x)$  (or  $x_I$ ), respectively. The cardinality of a set  $S$  is denoted  $|S|$ .

## 2 Methods

This study originates from the need for a more efficient use of the increasingly crowded radio spectrum, which can be achieved by spatial multiplexing. Our scenario consists of a MU-MISO system, where a multi-antenna base-station sends information to several single-antenna user terminals, using the same carrier frequency. To avoid interference between the different users, the information destined to the users is precoded before transmission. We adopt the common assumption that perfect channel state information (CSI) is present at the transmitter.

To avoid the potentially large power penalty caused by an ill-conditioned channel in the case of linear precoding, we restrict our attention to a type of nonlinear precoding, referred to as THP. The latter applies a modulo operation

at the transmitter and the receiver, thereby reducing the power loss but at the same time introducing a modulo loss.

Several TX optimizations for improving the performance of THP have already been presented in literature. In this contribution, we compare several of those techniques and some combinations thereof, and we propose some novel algorithms as well. These methods and algorithms are validated on two specific channels, i.e., the flat Rayleigh fading channel and the multi-beam satellite channel, for which we use statistical models available from literature.

As a performance indicator, we take the MI between the transmitted symbol and the resulting detector input. For given SNR,  $10^4$  channel realizations are generated independently. For each realization, the corresponding MI for the different users is evaluated by means of numerical integration; the average MI for the considered SNR is obtained as an arithmetical average over the users and the channel realizations. This computation is performed for SNR ranging from  $-15$  to  $15$  dB, with a 1 dB step.

## 3 MIMO channel model

We consider a TX equipped with  $N_T$  antennas, sending to  $N_U$  single-antenna users, all operating at the same carrier frequency. Assuming a frequency-flat channel, the received signals  $\mathbf{y}(k)$  associated with the  $k$ th symbol interval can be written as:

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k) + \mathbf{w}(k), \quad (1)$$

where  $\mathbf{x}(k) = (x_1(k), \dots, x_{N_T}(k))^T$  and  $\mathbf{y}(k) = (y_1(k), \dots, y_{N_U}(k))^T$  represent the signals transmitted by the  $N_T$  antennas and the corresponding signals received by the  $N_U$  users, respectively; we assume  $N_T \geq N_U$ . The channel is represented by the  $N_U \times N_T$  channel matrix  $\mathbf{H}$ , which is constant over a frame of  $K$  symbol intervals. The elements of the additive white Gaussian noise vector  $\mathbf{w}(k) \sim \mathcal{N}_c(0, N_0 \mathbf{I}_{N_U})$  have a variance equal to  $N_0$ . The quantity  $E_{\text{tr}} = \frac{1}{N_U} \mathbb{E}[|\mathbf{x}(k)|^2]$  denotes the average TX energy per symbol interval and per user.

This channel model will be used to describe two types of links, i.e., a terrestrial wireless link and a multi-beam satellite link. For notational convenience, the dependence on the symbol index  $k$  will be dropped.

### 3.1 Terrestrial wireless link

In the case of a terrestrial wireless link, we consider the downlink transmission from an  $N_T$ -antenna base station to  $N_U$  single-antenna terminals over a non-dispersive, Rayleigh block-fading channel, where  $N_U \leq N_T$ . The elements of the  $N_U \times N_T$  channel matrix  $\mathbf{H}$  are independent identically distributed (i.i.d.) with  $H_{n,m} \sim \mathcal{N}_c(0, 1)$  for  $m = 1, \dots, N_T$  and  $n = 1, \dots, N_U$ .

### 3.2 Multi-beam satellite link

Here, we describe a multi-beam geostationary satellite system operating in the Ka-band (20 GHz), with  $N_T$  beams transmitting data to  $N_U$  single-antenna user terminals. We assume that each user is operating under line-of-sight (LOS) conditions and that each user is located in a different beam; the latter implies  $N_U \leq N_T$ . Due to the sidelobes in the antenna radiation pattern, the signal transmitted in a beam induces co-channel interference (CCI) to the users located in the other beams [26, 31, 32, 34]. The level of the CCI depends on the users' relative position to the centers of the other beams and the TX power in those beams. We assume full frequency reuse, so that all users operate in the same frequency band. The satellite antenna is a tapered-aperture antenna with normalized power gain  $B(\theta)$  given by [26, 31]

$$B(\theta) = \left( \frac{J_1(u)}{2u} + 36 \frac{J_3(u)}{u^3} \right)^2, \quad (2)$$

where  $\theta$  is the angle between the spot beam center and the RX, seen from the satellite,  $u = 2.07123 \cdot \sin(\theta)/\sin(\theta_{3dB})$ ,  $\theta_{3dB}$  is the one-sided half-power beamwidth, and  $J_i(\cdot)$  denotes the Bessel function of the first kind and  $i$ th order; hence,  $B(0) = 1$  and  $B(\theta_{3dB}) = 1/2$ . The circle on which the received power is 3 dB below the power at the beam center has a diameter which is referred to as the beam diameter. The beam centers are located on a hexagonal grid; the distance between beam centers equals the beam diameter. From (2), the normalized amplitude gain  $A_{n,m} = \sqrt{B(\theta_{n,m})}$  from the  $m$ th antenna to the  $n$ th user is obtained, with  $\theta_{n,m}$  determined by the positions of the  $n$ th user and the beam center from the  $m$ th antenna.

We also include rain fading [28, 33] in the satellite link model. We use the common assumption that rain attenuation, in dB, can be modeled by a log-normal distribution [31, 33]. The power loss (in dB) experienced by the  $n$ th user due to rain attenuation is expressed as  $R_n = e^{\nu_n}$  [dB], where  $\nu_n$  has a Gaussian distribution with mean  $\mu_n$  and variance  $\sigma_n^2$ . The parameters  $\mu_n$  and  $\sigma_n^2$  depend on the user's geographical position and on the carrier frequency. We assume that the rain attenuation from a satellite antenna to a specific user is the same for all  $N_T$  antennas, while the attenuation experienced by different users is uncorrelated. Indeed, in [30], it is explained that the rain correlation decreases steeply with increasing distance and can be neglected for most beam sizes of interest [33]. The channel gain  $r_n$  (on a linear scale) associated with the rain fading affecting the  $n$ th user is then given by  $r_n = 10^{-\frac{R_n}{20}}$ .

The phase  $\phi_{n,m}$  of the channel gain from the  $m$ th antenna feed to the  $n$ th user is considered independent of the antenna index  $m$ , so that  $\phi_{n,m} = \phi_n$  for all  $m$ ; this is because the satellite antenna spacing is very small

compared to the communication distance [31, 33]. On the other hand, the phases  $\phi_n$  and  $\phi_{n'}$  corresponding to different users are assumed independent, and uniformly distributed in  $[0, 2\pi]$ .

The free space loss (FSL) is given by [34]:

$$L = \left( \frac{4\pi}{\lambda} \right)^2 D^2, \quad (3)$$

with  $D$  the distance between the satellite and the considered user, and  $\lambda$  the wavelength associated with the carrier frequency. As  $D$  is quasi-independent of the user position in the considered scenario, we take  $D = D_0$  with  $D_0$  the distance between the satellite and the earth, which equals 35,786 km for a geostationary satellite; hence, the FSL is assumed to be the same for all users.

We construct the  $N_U \times N_T$  channel matrix  $\mathbf{H}$  as  $(\mathbf{H})_{n,m} = \sqrt{G/L} e^{j\phi_n} \cdot r_n \cdot A_{n,m}$ , which combines the FSL ( $L$ ), the product ( $G$ ) of the maximum TX and RX antenna power gains, the phases ( $\phi_n$ ), the rain fading ( $r_n$ ), and the normalized antenna amplitude gains ( $A_{n,m}$ ). For the numerical results, we use  $\theta_{3dB} = 0.2^\circ$ , which corresponds to a beam diameter of 250 km. The user's positions are uniformly distributed within the respective beams, and we consider moderate rain fading, with  $\mu = -2.6$  and  $\sigma^2 = 1.63$  for all users.

### 4 Tomlinson-Harashima precoding

Let us consider the symbol vector  $\mathbf{a} = (a_1, \dots, a_{N_U})^T$  corresponding to a generic symbol interval, with  $a_n$  denoting the symbol destined to the  $n$ th user. In the following, all symbols  $a_n$  belong to the same constellation, which is either  $M$ -PAM or  $M^2$ -QAM. In the former case,  $a_n \in \mathcal{A}_M = \{-(M-1), -(M-3), \dots, (M-1)\}$ ; in the latter case, we have  $a_n = a_{R,n} + ja_{I,n} \in \mathcal{A}_M + j\mathcal{A}_M$ . We assume that all constellation points are equiprobable, and define  $\sigma_a^2 = \mathbb{E}[|a_n|^2]$ ; we have  $\sigma_a^2 = \frac{M^2-1}{3}$  for  $M$ -PAM and  $\sigma_a^2 = 2\frac{M^2-1}{3}$  for  $M^2$ -QAM.

Interference among the data symbols at each user's antenna is avoided when the transmitted signal vector  $\mathbf{x}$  in (1) depends on  $\mathbf{a}$  in such a way that  $y_n$  is a function of  $a_n$ , but not of  $\{a_i | i \neq n\}$ , for  $n = 1, \dots, N_U$ . In the case of LP, this is simply accomplished by taking  $\mathbf{x} = A\mathbf{H}^+\mathbf{a}$ , with  $\mathbf{H}^+ = \mathbf{H}^H(\mathbf{H}\mathbf{H}^H)^{-1}$  and  $A$  denoting a positive scaling factor setting the TX energy for the considered frame, yielding  $\mathbf{y} = A\mathbf{a} + \mathbf{w}$ . However, depending on the channel realization, the entries of  $\mathbf{H}^+$  can become quite large, so that a very high TX energy  $E_{tr}$  is needed to achieve a given SNR at the RX. This problem is mitigated by using THP, which is briefly outlined below.

We introduce the LQ decomposition  $\mathbf{H} = \mathbf{L}\mathbf{Q}$  of the channel matrix, where  $\mathbf{L}$  is an  $N_U \times N_U$  lower triangular matrix with positive diagonal elements and  $\mathbf{Q}$  is an  $N_U \times N_T$  matrix with orthogonal rows:  $\mathbf{Q}\mathbf{Q}^H = \mathbf{I}_{N_U}$ . As  $\mathbf{H}$

is constant over a frame of  $K$  symbol intervals, the matrices  $\mathbf{L}$  and  $\mathbf{Q}$  must be determined only once per frame. Using the decomposition

$$\mathbf{L} = \mathbf{L}_d(\mathbf{I}_{N_U} + \mathbf{B}) \quad (4)$$

where  $\mathbf{L}_d = \text{diag}(\mathbf{L})$  and  $\mathbf{B} = \mathbf{L}_d^{-1}\mathbf{L} - \mathbf{I}_{N_U}$  is lower triangular with zero diagonal, the resulting block diagram of the communication system with THP is shown in Fig. 1, assuming an  $M^2$ -QAM constellation. The precoder computes  $\mathbf{u} = \mathbf{L}_d^{-1}\mathbf{D}(\mathbf{a} - \mathbf{v})_{\text{mod}}$  where  $\mathbf{v} = \mathbf{D}^{-1}\mathbf{L}_d\mathbf{B}\mathbf{u}$ . The modulo operator  $(.)_{\text{mod}}$  acts on each element of the vector  $\mathbf{a} - \mathbf{v}$ , with the elements of  $\mathbf{v}$  denoting the interference pre-subtraction terms. The  $N_U \times N_U$  matrix  $\mathbf{D}$  is diagonal with elements  $(\mathbf{D})_{n,n} = d_n = e^{j\theta_n}$ , which are constant over a frame of  $K$  symbol intervals. We introduce the vector  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{N_U})$ ; in a conventional THP system, one takes  $\boldsymbol{\theta} = 0$  (or, equivalently,  $\mathbf{D} = \mathbf{I}_{N_U}$ ), but in this contribution, we will select  $\boldsymbol{\theta}$  such that a better performance is achieved compared to the case where  $\boldsymbol{\theta} = 0$ . Equivalently,

$$u_n = \frac{e^{j\theta_n}}{L_{n,n}}(a_n - v_n)_{\text{mod}}, \quad (5)$$

where the interference pre-subtraction terms  $v_n$  are given by:

$$v_n = \begin{cases} 0 & n = 1 \\ e^{-j\theta_n} \sum_{i=1}^{n-1} L_{n,i} u_i & n > 1 \end{cases}. \quad (6)$$

In (5), the operation  $(c+jd)_{\text{mod}} = (c)_{\text{mod}2M} + j(d)_{\text{mod}2M}$  represents the complex modulo operation, with  $(.)_{\text{mod}2M}$  denoting the modulo  $2M$  reduction of a real variable to the interval  $(-M, M]$ . The modulo operation can also be represented using the decomposition  $(a_n - v_n)_{\text{mod}} = a_n + 2Mk_n - v_n$ , where the real and imaginary part of  $k_n$  are the unique integers such that both  $\Re(a_n + 2Mk_n - v_n)$  and  $\Im(a_n + 2Mk_n - v_n)$  are in the interval  $(-M, M]$ .

The transmitted signal is  $\mathbf{x} = A\mathbf{Q}^H\mathbf{u}$ , where the positive factor  $A$  is constant over a frame of  $K$  symbol intervals;  $A$  sets the TX energy  $E_{\text{tr}}$  for the considered frame. For given  $\mathbf{H}$ , the quantities  $A^2$  and  $E_{\text{tr}}$  are related by:

$$E_{\text{tr}} = \frac{A^2}{N_U} \sum_{n=1}^{N_U} \frac{\sigma_{\text{mod},n}^2}{|L_{n,n}|^2} \quad (7)$$

with  $\sigma_{\text{mod},n}^2 = \mathbb{E}[|(a_n - v_n)_{\text{mod}}|^2]$ , where the expectation is over all  $M^{N_U}$  possible symbol vectors  $\mathbf{a}$ . Since  $v_1 = 0$ , we have  $\sigma_{\text{mod},1}^2 = \sigma_a^2$ . For  $n > 1$ ,  $\sigma_{\text{mod},n}^2$  is a function of  $(\theta_1, \dots, \theta_n)$  and the channel realization. In the following, we take  $A^2$  inversely proportional to  $\frac{1}{N_U} \sum_{n=1}^{N_U} \frac{\sigma_{\text{mod},n}^2}{|L_{n,n}|^2}$ , such that  $E_{\text{tr}}$  from (7) is a fixed value, irrespective of  $\mathbf{H}$  and  $\boldsymbol{\theta}$ . In a practical implementation,  $A$  can be computed from (7), with  $\sigma_{\text{mod},n}^2$  approximated by an arithmetical average of  $|(a_n(k) - v_n(k))_{\text{mod}}|^2$  over the  $K$  symbol intervals of the considered frame.

It can be verified from Fig. 1 that the signal received by the  $n$ -th user in the case of THP can be written as  $y_n = Ae^{j\theta_n}(a_n + 2Mk_n) + w_n$ , which indicates that  $a_n$  is scaled by  $A$  and rotated by  $\theta_n$ . The  $n$ -th user performs the operation  $\tilde{y}_n = y_n e^{-j\theta_n}/A$ , which yields:

$$\tilde{y}_n = (a_n + 2Mk_n) + \tilde{w}_n, \quad (8)$$

where the noise contribution has a variance  $\mathbb{E}[|\tilde{w}_n|^2] = N_0/A^2$  which is independent of the user index  $n$ ; the presence of  $k_n$  in (8) is a consequence of the ambiguity introduced by the modulo operation at the TX.

In order to remove the effect of  $k_n$  from (8), the conventional RX in a THP communication system applies a modulo operation to  $\tilde{y}_n$ , yielding:

$$(\tilde{y}_n)_{\text{mod}} = (a_n + \tilde{w}_n)_{\text{mod}} \quad (9)$$

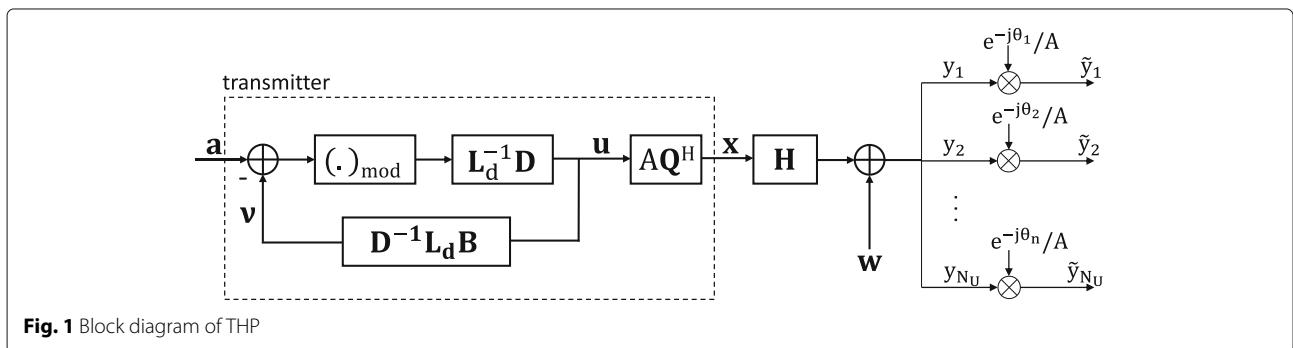
which is fed to the detector. The corresponding SNR of the detector is defined as  $\text{SNR}_{\text{det}} = \sigma_a^2/\mathbb{E}[|\tilde{w}_n|^2]$ , which reduces to  $\text{SNR}_{\text{det}} = \sigma_a^2 A^2 / N_0$ . As  $\text{SNR}_{\text{det}}$  does not depend on the user index  $n$ , all users yield the same performance when the detection is based on  $(\tilde{y}_n)_{\text{mod}}$  from (9). Alternatively, using (7),  $\text{SNR}_{\text{det}}$  can be expressed as:

$$\text{SNR}_{\text{det}} = \left( \frac{1}{N_U} \sum_{n=1}^{N_U} \frac{\sigma_{\text{mod},n}^2 / \sigma_a^2}{|L_{n,n}|^2} \right)^{-1} \frac{E_{\text{tr}}}{N_0} \quad (10)$$

which indicates that  $\text{SNR}_{\text{det}}$  depends on  $\mathbf{H}$  and  $\boldsymbol{\theta}$  through the factor multiplying  $E_{\text{tr}}/N_0$  in (10).

In the case of THP for  $M$ -PAM, (5) is replaced by:

$$u_n = \frac{e^{j\theta_n}}{L_{n,n}}(a_n - \Re(v_n))_{\text{mod}2M} \quad (11)$$



with  $v_n$  given by (6), and the detection is based on  $\Re(\tilde{y}_n)$ , with:

$$\Re(\tilde{y}_n) = (a_n + 2Mk_{R,n}) + \Re(\tilde{w}_n) \quad (12)$$

and  $k_{R,n}$  the unique integer for which  $-M < a_n + 2Mk_{R,n} - \Re(v_n) \leq M$ . Again, the conventional THP detector removes the effect of  $k_{R,n}$  from (12) by applying a modulo operation to  $\Re(\tilde{y}_n)$ , i.e.,  $(\Re(\tilde{y}_n))_{\text{mod}2M} = (a_n + \Re(\tilde{w}_n))_{\text{mod}2M}$ . The Eqs. (7) and (10) remain valid, where now  $\sigma_{\text{mod},n}^2$  is defined as  $\sigma_{\text{mod},n}^2 = \mathbb{E}[(a_n - \Re(v_n))_{\text{mod}2M}]^2$ .

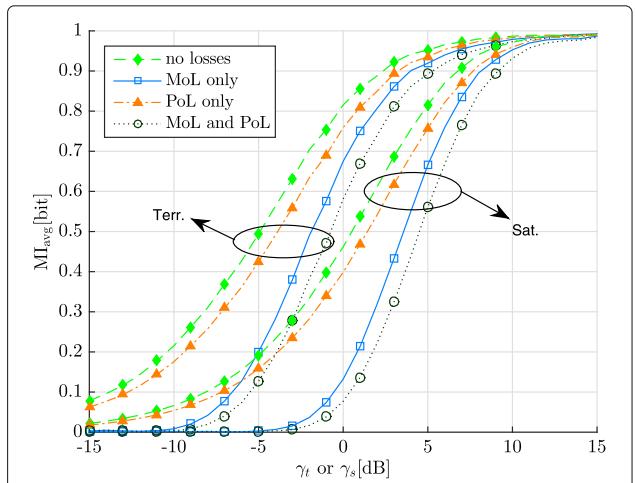
As the spectral efficiency of the THP scheme is proportional to the number of users, the maximum efficiency is achieved for  $N_U = N_T$ .

## 5 Power loss and modulo loss

As  $v_1 = 0$  (see (6)), we always have  $\sigma_{\text{mod},1}^2 = \sigma_a^2$ , irrespective of the channel realization; for  $n > 1$ , the presence of the interference pre-subtraction terms  $v_n$  gives rise to  $\sigma_a^2 \leq \sigma_{\text{mod},n}^2 \leq \frac{M^2+2}{M^2-1}\sigma_a^2$  for both  $M$ -PAM and  $M^2$ -QAM [35]. Hence, when for some  $n > 1$  we have  $\sigma_{\text{mod},n}^2 > \sigma_a^2$ ,  $\text{SNR}_{\text{det}}$  from (10) is reduced compared to the case where  $v_n = 0$  for all  $n$  (implying  $\sigma_{\text{mod},n}^2 = \sigma_a^2$  for all  $n$ ); this reduction of  $\text{SNR}_{\text{det}}$  represents the *power loss* described in [9].

When using THP, the conventional detector (denoted CD) operates on  $\tilde{y}_{CD,n} = (\tilde{y}_n)_{\text{mod}} = (a_n + \tilde{w}_n)_{\text{mod}}$ . Let us also consider a genie-aided detector (denoted GD), which knows  $k_n$  from (8); the GD operates on  $\tilde{y}_{GD,n} = \tilde{y}_n - 2Mk_n = a_n + \tilde{w}_n$ . We introduce the corresponding mutual information (MI)  $I_{GD} = I(a_n; \tilde{y}_{GD,n})$  for the GD and  $I_{CD} = I(a_n; \tilde{y}_{CD,n})$  for the CD, which do not depend on the user indices. It follows from the data processing theorem [36] that  $I_{CD} \leq I_{GD}$  for given  $\text{SNR}_{\text{det}}$ , indicating that the CD is affected by a performance penalty compared to the GD; this penalty, which is associated with the modulo operation at the RX, is referred to as the *modulo loss*.

For 2-PAM,  $N_U = N_T = 7$  and  $\theta = 0$ , Fig. 2 shows  $\text{MI}_{\text{avg}}$ , the MI averaged over the channel statistics, as a function of  $\gamma_t = E_{\text{tr}}/N_0$  for the Rayleigh fading terrestrial wireless channel, and as a function of  $\gamma_s = (E_{\text{tr}}/N_0) \cdot (G/L)$  for the multi-beam satellite channel; the entries “PoL only” and “PoL and MoL” correspond to the GD and CD, respectively. Also shown is  $\text{MI}_{\text{avg}}$  for the cases “no losses” and “MoL only”; these correspond to the GD and CD, respectively, but with  $\text{SNR}_{\text{det}}$  obtained from (10) with  $\sigma_{\text{mod},n}^2$  replaced by its lower bound  $\sigma_a^2$ , so that the PoL is removed. The horizontal shift between the curves “no losses” and “PoL only” is about 1 dB for  $\text{MI}_{\text{avg}}$  in the interval (0.1, 0.9); considering that the upper bound on the PoL for 2-PAM amounts to  $\frac{M^2+2}{M^2-1} = 2$  (3 dB), we conclude that this bound is rather conservative for  $M = 2$ . The MoL



**Fig. 2** MoL and PoL for the terrestrial and satellite channel (2-PAM,  $N_U = N_T = 7, \theta = 0$ )

is observed to be larger than the PoL, and increases with decreasing  $\text{MI}_{\text{avg}}$ .

## 6 Alternative detection strategy

For any channel realization, and irrespective of the symbol vector  $\mathbf{a}$ , the modulo operation at the TX has no effect for the user with index  $n = 1$ , because  $v_1 = 0$  and  $a_1$  is within the modulo boundary. Hence, we have  $k_1 = 0$  in (8) (for QAM) or  $k_{R,1} = 0$  in (12) (for PAM), so that this user can omit the modulo operation at the RX. Similarly, depending on the channel realization and the selection of  $\{\theta_n\}$  for the considered frame, a user with index  $n > 1$  can remove the modulo operation at the RX during the entire frame, irrespective of the symbol vector  $\mathbf{a}$ , provided that a certain condition  $C(n)$  holds for the channel matrix associated with the considered frame. The formulation of this condition depends on the constellation type; the condition  $C(n)$  is denoted  $C_{\text{PAM}}(n)$  and  $C_{\text{QAM}}(n)$  for  $M$ -PAM and  $M^2$ -QAM, respectively, with

- $C_{\text{PAM}}(n)$  is true  $\iff v_{R,n} \in [-1, 1]$  for all possible values of  $(a_1, \dots, a_{n-1})$
- $C_{\text{QAM}}(n)$  is true  $\iff v_{R,n} \in [-1, 1]$  and  $v_{I,n} \in [-1, 1]$  for all possible values of  $(a_1, \dots, a_{n-1})$

Indeed, when condition  $C(n)$  holds for given  $n$ , it is easily verified that  $(a_n - \Re(v_n))_{\text{mod}} = a_n - \Re(v_n)$  (for PAM) or  $(a_n - v_n)_{\text{mod}} = a_n - v_n$  (for QAM); in this case, the modulo operation at the TX has no effect, so that the modulo operation associated with the  $n$ th user can be discarded from the RX<sup>1</sup> during the considered frame.

We proposed in [17, 18] an alternative detection strategy to reduce the MoL compared to the CD. Let us denote by  $\mathcal{S}_C \subseteq \{1, \dots, N_U\}$  the set of all indices for which the corresponding users can dispose of the modulo operation, i.e.,  $n \in \mathcal{S}_C \iff C(n)$  holds. The alternative detector (denoted AD) for the  $n$ -th user is assumed to know whether  $n \in \mathcal{S}_C$  for the channel realization associated with the considered frame, and performs the detection accordingly: detection is based on  $\tilde{y}_{AD,n} = \tilde{y}_n = a_n + \tilde{w}_n$  when  $n \in \mathcal{S}_C$ , and on  $\tilde{y}_{AD,n} = (\tilde{y}_n)_{mod} = (a_n + \tilde{w}_n)_{mod}$  when  $n \notin \mathcal{S}_C$ . Hence, with the AD, only the users with  $n \notin \mathcal{S}_C$  are affected by MoL. For given  $\mathbf{H}$ , the average performance (average over the  $N_U$  users) of the AD improves with an increasing fraction,  $\rho_C = |\mathcal{S}_C|/N_U$ , of users for which  $C(n)$  is met. As the transmitted vector  $\mathbf{x}$  is not affected by which type of detector is used, the AD, CD, and GD exhibit the same PoL, for given  $\mathbf{H}$  and  $\theta$ .

The practical operation of the AD implies that, first, the TX verifies which user indices belong to  $\mathcal{S}_C$ . The verification for user  $n$  can be accomplished by checking whether  $a_n(k) - \Re(v_n(k))$  (for PAM) or  $a_n(k) - v_n(k)$  (for QAM) is within the modulo boundaries for all symbol indices  $k$  belonging to the frame; for growing frame size  $K$ , this condition becomes equivalent to  $C(n)$ . Next, the TX informs the users accordingly; this requires the transmission of only one (properly encoded) bit of channel information per user and per frame, which typically represents a very small overhead [18].

For 2-PAM,  $N_U = N_T = 7$  and  $\theta = 0$ , the MI averaged over the channel statistics and over the users (denoted  $MI_{avg}$ ) for the CD, GD, and AD is shown in Fig. 3, for both the terrestrial Rayleigh fading channel and the multi-beam satellite channel. We observe that the AD recovers a major part of the MoL; when  $MI_{avg}$  is in the interval (0.1, 0.8),

the AD provides a gain (compared to the CD) between 1 and 5 dB for the terrestrial channel and between 1.5 and 8 dB for the satellite channel. The MoL of the AD is smaller for the satellite channel than for the terrestrial channel. This is explained by noticing that  $\mathbb{E}[|H_{n,m}|^2]$  is the same for all  $(n, m)$  for the terrestrial channel, whereas for the satellite channel  $\mathbb{E}[|H_{n,m}|^2]$  is typically larger than  $\mathbb{E}[|H_{i,m}|^2]$  with  $i \neq n$  when the  $n$ -th user is located in the  $m$ -th beam; hence, on average, the interference on the terrestrial channel is larger than on the satellite channel, yielding a larger fraction  $\rho_C$  for the multi-beam satellite channel and, hence, a smaller MoL for the AD on the latter channel.

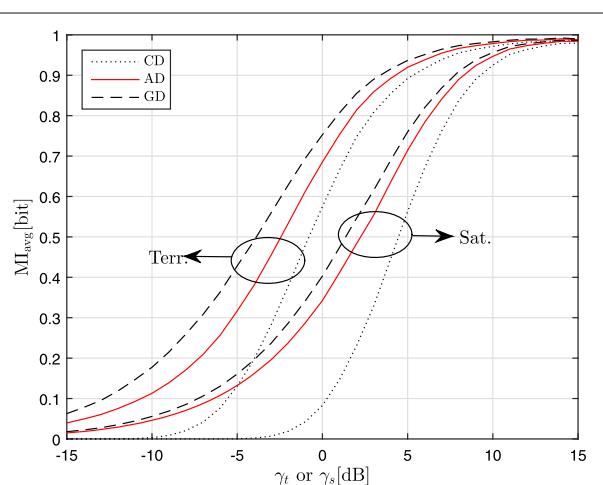
## 7 Transmitter optimization

In this section, we optimize the TX for given  $\mathbf{H}$ , by a proper selection of (i) the angles  $\theta_n$  in  $d_n = e^{j\theta_n}$ , corresponding to a rotation of the symbols  $a_n$ , and (ii) the permutation of the rows of  $\mathbf{H}$ , which corresponds to a reordering of the users. As  $\mathbf{H}$  is constant over an entire frame, the optimization must be carried out only once per frame. The selection of the rotations and the permutation both affect  $SNR_{det}$  and  $\rho_C$ , and, therefore, the performance of the CD (which depends only on  $SNR_{det}$ ) and the AD (which depends on both  $SNR_{det}$  and  $\rho_C$ ). In the following, when the row permutation is not optimized, we take for each frame a same fixed permutation irrespective of  $\mathbf{H}$ ; when the rotations are not optimized, we take  $\theta = 0$ .

As the  $M^2$ -QAM constellation has a symmetry angle of  $\pi/2$ , we limit  $\theta_n$  to the interval  $[0, \pi/2)$  for  $n = 1, \dots, N_U$ , without loss of generality; we will restrict  $\theta_n$  to the finite set  $\Theta = \left\{0, \frac{\pi}{2Q}, \frac{2\pi}{2Q}, \dots, \frac{(Q-1)\pi}{2Q}\right\}$  with size  $Q$ . Similarly, in the case of  $M$ -PAM,  $\theta_n$  will be restricted to the set  $\Theta = \left\{0, \frac{\pi}{Q}, \frac{2\pi}{Q}, \dots, \frac{(Q-1)\pi}{Q}\right\}$ , because the constellation symmetry angle equals  $\pi$ .

Finding the best rotations and permutation requires a search over  $Q^{N_U}$  angle vectors  $\theta$  and over  $N_U!$  row permutations of  $\mathbf{H}$ , which becomes computationally prohibitive for a large number of users. Therefore, some reduced-complexity searches will be explored, at the expense of a performance penalty.

The algorithms will be denoted by a label of the type (X, Y), where X refers to the performance indicator to be optimized (X = MoL, X = SNR, and X = MoL+SNR, for minimizing the MoL, maximizing the SNR, or jointly maximizing SNR and minimizing MoL, respectively), and Y indicates over which parameters the optimization is conducted (Y = R and Y = P for optimization over the rotation angles and the row permutations respectively). The reduced-complexity version of the algorithm (X, Y) is denoted (X, Y)<sub>RC</sub>. In Sections 7.1 and 7.2, we restrict our attention to algorithms which optimize over the rotation



**Fig. 3** Comparison of detection strategies for the terrestrial and satellite channel (2-PAM,  $N_U = N_T = 7$ ,  $\theta = 0$ )

angles and over the row permutations, respectively; in Section 7.3, these algorithms are combined to jointly optimize over the rotation angles and the row permutations.

For THP using the CD, the MoL cannot be reduced. However, algorithms that reduce the MoL of the AD might also increase  $\text{SNR}_{\text{det}}$  as a secondary effect, in which case THP with CD would also benefit from these algorithms. However, THP with CD will gain more from algorithms directly aiming at maximizing  $\text{SNR}_{\text{det}}$ . Therefore, the former algorithms will be considered only for THP with AD, whereas the latter will be applied to THP with CD and to THP with AD.

## 7.1 Optimization over rotations

It can be verified from (5), (6), and (11) that when replacing all  $\theta_n$  by  $\theta_n + \phi$  for  $n = 1, \dots, N_U$ , neither the interference pre-subtraction terms  $v_n$  nor the magnitudes  $|u_n|$  are affected by the choice of  $\phi$ . Hence, without loss of optimality, we choose  $\theta_1 = 0$  when selecting the rotations.

Below, we present the following rotation optimization algorithms: the exhaustive-search algorithm (SNR, R), which is similar<sup>2</sup> to the algorithm from [14]; the tree-search algorithm (MoL, R) from [17]; and (MoL+SNR, R), which is a novel extension of (MoL, R). In addition, the corresponding reduced-complexity algorithms are derived, by limiting the search space in the tree; this requires turning (SNR, R) into a tree-search algorithm.

### 7.1.1 Maximizing the SNR

The selection of  $\theta$  affects  $\text{SNR}_{\text{det}}$  only through  $\sigma_{\text{mod},n}^2$  with  $n > 1$ ; as a consequence, the maximization of  $\text{SNR}_{\text{det}}$  is equivalent to the minimization of the PoL. The (SNR, R) algorithm selects  $\theta = (0, \theta_2, \dots, \theta_{N_U})$  that maximizes  $\text{SNR}_{\text{det}}$ . For given  $\mathbf{H}$ , this optimization involves the computation of  $\text{SNR}_{\text{det}}$  from (10) for all  $Q^{N_U-1}$  possible  $\theta$ , followed by the selection of the vector  $\theta$  which yields the largest  $\text{SNR}_{\text{det}}$  for a given  $E_{\text{tr}}/N_0$ . Taking into account that  $\sigma_{\text{mod},n}^2 = \mathbb{E}[|(a_n - v_n)_{\text{mod}}|^2]$  for  $n > 1$  depends on  $(\theta_2, \dots, \theta_n)$ , and represents an expectation over  $(a_1, \dots, a_n)$ , the computational complexity associated with the maximization of  $\text{SNR}_{\text{det}}$  becomes prohibitively large for large  $N_U$ .

### 7.1.2 Minimizing the modulo loss

The MoL of the AD from Section 6 decreases when the condition  $C(n)$  is met for a larger number of users. We can increase  $|\mathcal{S}_C|$ , compared to the case where  $\theta = 0$ , by choosing the appropriate  $\theta$  for a given channel realization. Instead of exhaustively going through all  $Q^{N_U-1}$  possible vectors  $\theta = (0, \theta_2, \dots, \theta_N)$  for maximizing  $|\mathcal{S}_C|$ , a more efficient algorithm, referred to as (MoL, R), has been devised in [17]. The (MoL, R) algorithm finds the angles  $(\theta_2, \dots, \theta_{N_c})$ , such that  $C(n)$  holds for the largest number ( $N_C$ ) of consecutive user indices  $n = 1, 2, \dots, N_C$ ,

with  $N_C \leq N_U$ . Obviously,  $\max_{\theta} N_C \leq \max_{\theta} |\mathcal{S}_C|$ , but the (MoL, R) algorithm has a smaller complexity than the exhaustive algorithm. The (MoL, R) algorithm is explained below.

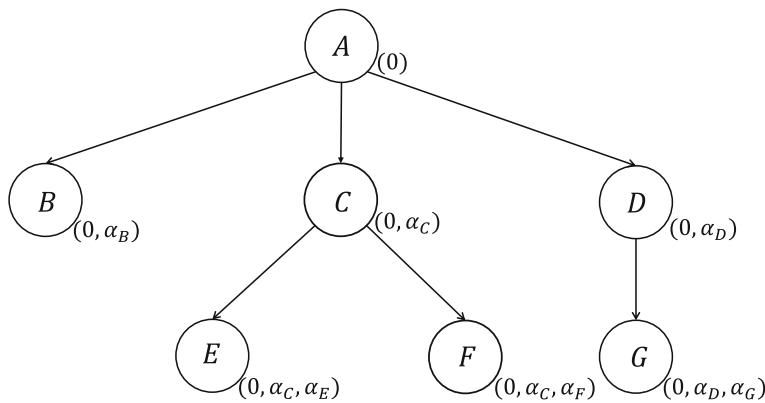
Let us introduce the notion of a *suitable n-tuple*:  $(0, \alpha_2, \dots, \alpha_n)$  is a suitable  $n$ -tuple if and only if the selection  $(\theta_1, \theta_2, \dots, \theta_n) = (0, \alpha_2, \dots, \alpha_n)$  yields condition  $C(i)$  for  $i = 1, \dots, n$  (in which case the modulo operation for the users with consecutive indices 1, ...,  $n$  can be omitted). It is easily verified that the first  $n - 1$  elements of a suitable  $n$ -tuple form a suitable  $(n - 1)$ -tuple; hence, the set of all suitable  $n$ -tuples for  $n = 1, \dots, N_C$  can be represented by a tree. The tree consists of  $N_C$  levels, with each node at level  $n$  denoting a suitable  $n$ -tuple. The children of a parent node  $(0, \alpha_2, \dots, \alpha_{n-1})$  are the suitable  $n$ -tuples of which the first  $n - 1$  elements equal the elements of the parent node. Level 1 of the tree contains the root node representing the 1-tuple  $(0)$ , which is the parent of all suitable 2-tuples. As an example, Fig. 4 shows a tree with three levels, i.e.,  $N_C = 3$ , along with the corresponding suitable  $n$ -tuples.

In order to determine the children of a parent node  $(0, \alpha_2, \dots, \alpha_{n-1})$ , we have to determine for which  $\theta_n \in \Theta$  the vector  $(0, \alpha_2, \dots, \alpha_{n-1}, \theta_n)$  is a suitable  $n$ -tuple. As  $(0, \alpha_2, \dots, \alpha_{n-1})$  is a suitable  $(n - 1)$ -tuple, the modulo operation at the TX has no effect for the users with indices  $1, \dots, n - 1$ , so that  $v_n$  from (6) can be decomposed as a linear combination of the symbols  $(a_1, \dots, a_{n-1})$ , i.e.,  $v_n = \sum_{i=1}^{n-1} \beta_{n,i} a_i$ , where the coefficients  $\beta_{n,i}$  depend on  $(\theta_1, \dots, \theta_i)$ , on the channel realization, and on the type (PAM or QAM) of constellation. More specifically, for QAM, these coefficients can be computed recursively as:

$$\beta_{n,i} = \begin{cases} \left( \frac{L_{n,i}}{L_{i,i}} e^{j\theta_i} - \sum_{l=i+1}^{n-1} L_{n,l} e^{j\theta_l} \beta_{l,i} \right) e^{-j\theta_n} & i = 1, \dots, n-2 \\ \frac{L_{n,n-1}}{L_{n-1,n-1}} e^{-j(\theta_n - \theta_{n-1})} & i = n-1 \end{cases} \quad (13)$$

The recursion for PAM is obtained by replacing in the first line of (13)  $\beta_{l,i}$  by  $\Re(\beta_{l,i})$ . The linear decomposition of  $v_n$  allows an efficient verification of whether  $C(n)$  is met: denoting by  $|v_{R,n}|_{\text{max}}$  and  $|v_{I,n}|_{\text{max}}$  the maximum values of  $|v_{R,n}|$  and  $|v_{I,n}|$  over all possible  $(a_1, \dots, a_{n-1})$ , the condition  $C(n)$  holds when  $|v_{R,n}|_{\text{max}} < 1$  (for  $M$ -PAM), or when  $|v_{I,n}|_{\text{max}} < 1$  and  $|v_{R,n}|_{\text{max}} < 1$  (for  $M^2$ -QAM); it can be verified that  $|v_{R,n}|_{\text{max}} = (M-1) \sum_{i=1}^{n-1} |\beta_{R,n,i}|$  for  $M$ -PAM and  $|v_{R,n}|_{\text{max}} = |v_{I,n}|_{\text{max}} = (M-1) \sum_{i=1}^{n-1} (|\beta_{R,n,i}| + |\beta_{I,n,i}|)$  for  $M^2$ -QAM, where  $\beta_{R,n,i} = \Re(\beta_{n,i})$  and  $\beta_{I,n,i} = \Im(\beta_{n,i})$ .

The (MoL, R) algorithm from [17] reduces the MoL of the AD, but disregards the effect of the symbol rotations on the PoL. The algorithm performs a depth-first search



**Fig. 4** Illustration of a tree representing all suitable  $n$ -tuples, with  $N_C = 3$

in the tree, until a suitable  $N_C$ -tuple is found. Two cases must be distinguished:

- When  $N_C = N_U$ , the search is ended when finding the first suitable  $N_U$ -tuple, which is used as vector  $\theta$ . In this case, none of the users requires the modulo operation at the RX.
- When  $N_C < N_U$ , the entire tree must be searched in order to find out that there exist no suitable  $n$ -tuples with  $n = N_C + 1$ , after which one of the suitable  $N_C$ -tuples is selected. The corresponding  $\theta$  is obtained by appending  $N_U - N_C$  zeroes to the selected suitable  $N_U$ -tuple. In this case, the modulo operation at the RX is needed only for the users with  $n = N_C + 1, \dots, N_U$ .

### 7.1.3 Minimizing modulo loss and power loss

The above (MoL, R) algorithm can be modified into the (MoL+SNR, R) algorithm, which takes, besides the MoL, also the PoL into account. The (MoL+SNR, R) algorithm minimizes the MoL of the AD, but in case several suitable  $N_C$ -tuples exist, the one yielding the higher SNR<sub>det</sub> (or, equivalently, the smaller PoL) is selected.

This algorithm finds *all* suitable  $N_C$ -tuples and constructs the corresponding vectors  $\theta$  by appending  $N_U - N_C$  zeroes if  $N_C < N_U$ ; for each of the resulting  $\theta$ , SNR<sub>det</sub> is computed according to (10), and the vector  $\theta$  yielding the largest SNR<sub>det</sub> is selected. As a modulo operation is required for  $n = N_C + 1, \dots, N_U$ , the evaluation of the corresponding  $\sigma_{n,\text{mod}}^2$  requires a numerical averaging over all possible  $(a_1, \dots, a_n)$ , involving a summation of  $M^n$  (for  $M$ -PAM) or  $M^{2n}$  (for  $M^2$ -QAM) terms. The resulting computational complexity is, however, less than with the (SNR, R) algorithm, because SNR<sub>det</sub> must be computed only for a number of vectors  $\theta$  equal to the number of

the suitable  $N_C$ -tuples, rather than for all  $Q^{N_U-1}$  possible vectors  $\theta$ .

### 7.1.4 Complexity reduction

The above (MoL, R) and (MoL+SNR, R) algorithms involve an exhaustive search in the tree of suitable  $n$ -tuples. We obtain novel algorithms with significantly reduced computational complexity, by restricting the number of nodes at each level of the tree to  $L$ , with  $L$  representing a design parameter. Let us denote by  $\#(n)$  the number of suitable  $n$ -tuples, or, equivalently, the number of nodes at level  $n$  in the original tree, with  $n = 1, \dots, N_C$ ; note that  $\#(1) = 1$ . Suppose that  $\#(n) \leq L$  for  $n = 1, \dots, i-1$ , but  $\#(i) > L$ . The reduced-complexity algorithms keep at level  $i$  only the best  $L$  nodes, i.e., those yielding the smallest  $L$  values of  $|v_{R,i}|_{\max}$  (for the (MoL,R)<sub>RC</sub> algorithm) or the largest  $L$  values of SNR<sub>det</sub> (for the (MoL+SNR,R)<sub>RC</sub> algorithm). All children, issuing from the  $L$  remaining parent nodes at level  $i$ , are determined, and when the total number of children at level  $i+1$  exceeds  $L$ , again only the best  $L$  are kept. This procedure is continued until at most  $L$  nodes at level  $N_C$  are obtained.

A similar complexity reduction can be applied to the (SNR, R) algorithm, yielding the (SNR,R)<sub>RC</sub> algorithm. The set of all  $Q^{N_U-1}$  vectors  $\theta = (0, \theta_2, \dots, \theta_{N_U})$  can also be represented by a tree, where the level  $n$  has  $Q^{n-1}$  nodes ( $n = 1, \dots, N_U$ ) and each node at levels  $1, \dots, N_U - 1$  has exactly  $Q$  children. When the number of nodes at level  $i$  exceeds  $L$ , the (SNR,R)<sub>RC</sub> algorithm keeps only the best  $L$  nodes, i.e., those yielding the smallest  $L$  values of  $\sum_{n=1}^i \frac{\sigma_{\text{mod},n}^2}{|L_{n,n}|^2}$ , for  $i = 2, \dots, N_U$ .

The upper bound on the complexity of these algorithms is proportional to  $(N_U - 1)LQ$ , instead of exponential in  $N_U$ ; in Section 8, we point out that reduced-complexity

algorithms yield only a small performance loss, caused by not searching the entire tree.

## 7.2 Optimization over row permutations

Denoting by  $\mathbf{P}$  an  $N_U \times N_U$  permutation matrix, the matrix  $\mathbf{L}$ , which results from the LQ decomposition of  $\mathbf{PH}$ , depends on  $\mathbf{P}$ . Hence, applying a row permutation to  $\mathbf{H}$  (which corresponds to selecting a precoding order for the users) affects both the value of  $\text{SNR}_{\text{det}}$  for given  $E_{\text{tr}}/N_0$  and the fraction  $\rho_C$  of users for which the condition  $C(n)$  holds. In principle, both the maximization of  $\text{SNR}_{\text{det}}$  and of  $\rho_C$  for given  $\mathbf{H}$  can be achieved by means of an exhaustive search over all possible  $\mathbf{P}$ . However, the associated computational complexity becomes prohibitively large for large  $N_U$ , because  $N_U!$  possible permutation matrices exist.

Below, we present the (SNR, P) and (MoL, P) algorithms, which straightforwardly perform a full search over all row permutations to achieve the maximum value of  $\text{SNR}_{\text{det}}$  and  $N_C$ , respectively. As these algorithms have a high computational cost, we also consider reduced-complexity algorithms. The (SNR, P)<sub>RC</sub> algorithm performs a sorted LQ decomposition; this algorithm is a straightforward adaptation from [25], where a sorted QR decomposition is presented as a low-complexity alternative to the user ordering in V-BLAST [24], at the expense of only a small performance loss. As the user ordering algorithms from literature do not aim at reducing the MoL, the (MoL, P)<sub>RC</sub> algorithm and its extension (MoL+SNR, P)<sub>RC</sub> are entirely novel.

### 7.2.1 Maximizing the SNR

When aiming at maximizing  $\text{SNR}_{\text{det}}$  from (10) for given  $\mathbf{H}$  and  $\theta$ , one has to compute for each row permutation the quantities  $\sigma_{\text{mod},n}^2$  and  $|L_{n,n}|$  for all  $N_U$  users; note that  $\sigma_{\text{mod},n}^2$  depends on  $\theta$  and represents an expectation over  $n$  data symbols  $(a_1, \dots, a_n)$ , involving a summation of  $M^n$  ( $M$ -PAM) or  $M^{2n}$  ( $M^2$ -QAM) terms. To avoid the numerical complexity associated with the evaluation of  $\sigma_{\text{mod},n}^2$  for all  $N_U$  users, the (SNR, P) algorithm maximizes  $\sum_{n=1}^{N_U} |L_{n,n}|^{-2}$  instead of  $\text{SNR}_{\text{det}}$  without a significant performance loss; this optimization does not depend on  $\theta$ .

In order to avoid the high complexity of the exhaustive search associated with the (SNR, P) algorithm, we consider instead a suboptimum low-complexity algorithm which performs a sorted LQ decomposition. This algorithm, referred to as (SNR, P)<sub>RC</sub>, is outlined in Algorithm 1 and aims at maximizing  $\min_n (|L_{n,n}|)$ . When the first  $i$  rows of  $\mathbf{PH}$  have been determined, the (SNR, P)<sub>RC</sub> algorithm selects among the  $N_U - i$  remaining rows from  $\mathbf{H}$  the row for which the projection, on the subspace orthogonal to the first  $i$  rows of  $\mathbf{PH}$ , is the smallest, and makes the selected row the  $(i+1)$ th row of  $\mathbf{PH}$ ; the magnitude of this projection equals  $L_{i+1,i+1}$ . The algorithm is initialized

by selecting as the first row of  $\mathbf{PH}$  the row from  $\mathbf{H}$  with the smallest magnitude. The algorithm stops when  $i = N_U - 1$ , at which point the one remaining row from  $\mathbf{H}$  becomes the last row from  $\mathbf{PH}$ . The row permutation of  $\mathbf{H}$  resulting from (SNR, P)<sub>RC</sub> is such that  $\min_n (|L_{n,n}|)$  cannot be further increased by swapping any two consecutive rows from  $\mathbf{PH}$ . The complexity of (SNR, P)<sub>RC</sub> is proportional to  $(N_U - 1)N_U$  rather than  $N_U!$ .

### 7.2.2 Minimizing the modulo loss

For given  $\mathbf{H}$  and fixed  $\theta$ , the MoL resulting from the AD can be minimized by selecting the permutation matrix  $\mathbf{P}$  yielding the largest fraction ( $\rho_C$ ) of users for which the condition  $C(n)$  is met. For each of the  $N_U!$  row permutations of  $\mathbf{H}$ , the corresponding  $\rho_C$  is obtained by verifying for which  $n \in \{1, \dots, N_U\}$  the condition  $C(n)$  holds. Checking whether  $C(n)$  is met for a given  $n$  requires the evaluation of the interference pre-subtraction term  $v_n$  for all  $M^{n-1}$  (for  $M$ -PAM) or  $M^{2(n-1)}$  (for  $M^2$ -QAM) possible  $(a_1, \dots, a_{n-1})$ , which involves a high computational complexity when  $N_U$  is large. Instead, the (MoL, P) algorithm determines among all  $N_U!$  permutations the permutation which maximizes  $N_C$ , the number of consecutive user indices  $1, 2, \dots, N_C$  for which  $C(n)$  holds. As explained in Section 7.1.2, when  $C(i)$  is met for  $i = 1, \dots, n - 1$ , then  $C(n)$  holds if and only if  $(M - 1) \sum_{i=1}^{n-1} |\beta_{R,n,i}| < 1$  (for  $M$ -PAM) or  $(M - 1) \sum_{i=1}^{n-1} (|\beta_{R,n,i}| + |\beta_{I,n,i}|)$  (for  $M^2$ -QAM). Hence, checking whether  $C(n)$  holds is far less complex when maximizing the amount of consecutive users  $N_C$  instead of the total amount  $|\mathcal{S}_C|$ .

The complexity of (MoL, P), which performs an exhaustive search over all  $N_U!$  row permutations of  $\mathbf{H}$ , can be avoided by using the algorithm described in Algorithm 2, and referred to as (MoL, P)<sub>RC</sub>. The (MoL, P)<sub>RC</sub> algorithm consists of  $N_U - 1$  steps. During the  $i$ th step, we select  $\mathbf{P}_i \mathbf{H}$  yielding the largest  $N_C$ , from a set of  $N_U - i + 1$  row permutations of  $\mathbf{H}$ ; this set contains the matrix  $\mathbf{P}_{i-1} \mathbf{H}$ , resulting from the previous step, and the  $N_U - i$  permutations of  $\mathbf{P}_{i-1} \mathbf{H}$  obtained by swapping the  $i$ th row from  $\mathbf{P}_{i-1} \mathbf{H}$  with a row having a row index larger than  $i$ ; as these  $N_U - i + 1$  matrices have the first  $i - 1$  rows in common, they give rise to the same elements  $L_{m,n}$  with  $m \leq n$  and  $n = 1, \dots, i - 1$  and, therefore, to the same interference pre-subtraction terms  $(v_1, \dots, v_{i-1})$ ; as a consequence,  $N_C$  can never be lower than in a previous step. The algorithm starts with  $i = 1$ , taking  $\mathbf{P}_0 = \mathbf{I}_{N_U}$ . When during the  $i$ th step of Algorithm 2, the largest  $N_C$  is achieved for more than one of the possible row permutations, we select the permutation which maximizes  $|L_{i,i}|$ , so that  $|\beta_{R,n,i}|$  and  $|\beta_{I,n,i}|$  are minimized; this selection contributes to minimizing  $|\nu_{R,n}|_{\max}$  and  $|\nu_{I,n}|_{\max}$ , with  $n > i$ . In case  $N_C < i$  in the  $i$ th step, we stop our search because we cannot further increase the number of consecutive users for which condition  $C(n)$  holds.

### 7.2.3 Trade-off between minimizing the modulo loss and maximizing $\text{SNR}_{\text{det}}$

When more than one permutation achieves the maximal  $N_C$  during the  $i$ -th step of the (MoL, P)<sub>RC</sub> algorithm from the previous section, we selected the permutation which maximizes  $|L_{i,i}|$ , in order to increase the chance that  $N_C$  will be larger in the next step. However, this choice tends to decrease  $\text{SNR}_{\text{det}}$  (as follows from the numerical results in Section 8.2.4).

Instead, we consider the (MoL + SNR, P)<sub>RC</sub> algorithm, which makes a trade-off between minimizing the MoL and maximizing  $\sum_{n=1}^{N_U} |L_{n,n}|^{-2}$ . More specifically, when more than one permutation achieves the maximal  $N_C$  during the  $i$ -th step, we now select the permutation which minimizes  $|L_{i,i}|$ . This algorithm thus selects in the  $i$ th step, among all permutations achieving the largest  $N_C$  in that step, the one that maximizes  $\sum_{n=1}^{N_U} |L_{n,n}|^{-2}$ .

---

### Algorithm 1 (SNR,P)<sub>RC</sub> algorithm

---

```

Initialization
 $\mathbf{P} = [\mathbf{p}_{ij}] = \mathbf{0}$ 
 $\mathbf{u}_i = \mathbf{h}_i \quad \forall i = 1, \dots, N_U$ 
 $\mathbf{S} = \{\mathbf{u}_1, \dots, \mathbf{u}_{N_U}\}$ 
 $\mathbf{index} = (1, 2, \dots, N_U)$ 
for  $i = 1$  to  $N_U$  do
     $j = \arg \min_m (\|\mathbf{u}_m\|^2) \quad \forall \mathbf{u}_m \in \mathbf{S}$ 
     $k = \mathbf{index}(j)$ 
     $p_{i,k} = 1$ 
    remove  $\mathbf{u}_j$  from  $\mathbf{S}$ 
    remove  $j$ th element from  $\mathbf{index}$ 
     $\mathbf{u}_n = \mathbf{u}_n - \text{proj}_{\mathbf{u}_j} \mathbf{u}_n \quad \forall \mathbf{u}_n \in \mathbf{S}$ 
end do

```

---

### 7.3 Joint optimization over rotations and permutations

For a further improvement of the performance of the AD, a rotation optimization algorithm from Section 7.1 can be combined with a user ordering optimization algorithm from Section 7.2.

The rotation angle vectors  $\theta$  resulting from the (SNR, R) and (MoL, R) algorithms depend on the considered row permutation matrix  $\mathbf{P}$ . The (MoL, P) algorithm yields a permutation matrix  $\mathbf{P}$  which depends on the considered  $\theta$ , whereas the permutation matrix  $\mathbf{P}$  resulting from the (SNR, P) algorithm is independent of  $\theta$ . Hence, when envisioning the joint optimization over  $\mathbf{P}$  and  $\theta$  involving (SNR, P), one can perform a *consecutive* optimization, where first (SNR, P) is applied, followed by algorithm  $U$ , with  $U \in \{( \text{SNR}, \text{R}), ( \text{MoL}, \text{R}), ( \text{MoL+SNR}, \text{R})\}$ . In the case of a joint optimization involving algorithm  $V$ , with  $V \in \{( \text{MoL}, \text{P}), ( \text{MoL+SNR}, \text{P})\}$ , we consider a nested optimization, where  $V$  is the outer algorithm, and  $W$ , with  $W \in \{( \text{SNR}, \text{R}), ( \text{MoL}, \text{R}), ( \text{MoL+SNR}, \text{R})\}$  is used as

---

### Algorithm 2 (MoL,P)<sub>RC</sub> algorithm

---

```

Initialization
 $\mathbf{P} = \mathbf{I}_N$ 
 $\mathbf{u}_n = \mathbf{h}_n \quad \forall i = 1, \dots, N$ 
 $\mathbf{S} = \{\mathbf{u}_1, \dots, \mathbf{u}_{N_U}\}$ 
for  $i = 1$  to  $N$  do
    for  $k = i$  to  $N$  do
         $\mathbf{H}_{\text{temp}} = \mathbf{P}\mathbf{H}$ 
        swap  $\mathbf{h}_k$  and  $\mathbf{h}_i$  of  $\mathbf{H}_{\text{temp}}$ 
        calculate  $N_{C,k}(\mathbf{H}_{\text{temp}})$ 
    end do
    if  $(\max_k(N_C, k) < i)$  break
     $\mathbf{k}_{\max} = \arg \max_k \{N_{C,k}\}$ 
     $k_{\max} = \arg \max_m \{\mathbf{u}_m | m \in \mathbf{k}_{\max}\}$ 
     $\mathbf{u}_j = 0$ 
    remove  $\mathbf{u}_n$  from  $\mathbf{S}$ 
    swap rows  $k_{\max}$  and  $i$  of  $\mathbf{P}$ 
     $\mathbf{u}_n = \mathbf{u}_n - \text{proj}_{\mathbf{h}_{k_{\max}}} \mathbf{u}_n \quad \forall n=i, \dots, N$ 
end do

```

---

the inner algorithm. The same considerations are valid in case reduced-complexity algorithms are applied. To the best of our knowledge, the algorithms performing joint optimization over  $\mathbf{P}$  and  $\theta$  have not been presented in literature.

## 8 Numerical results

We assess the performance of the various optimization algorithms and detectors considered, in terms of their SNR gain compared to the CD without TX optimization. When, for a given constellation, a particular configuration  $X$  (consisting of optimization algorithm and detector) and the CD without TX optimization give rise to  $\text{MI}_{\text{avg}} = \text{MI}_{\text{ref}}$  at  $E_{\text{tr}}/N_0 = (E_{\text{tr}}/N_0)_X$  and  $E_{\text{tr}}/N_0 = (E_{\text{tr}}/N_0)_{\text{CD}}$ , respectively, the SNR gain (in dB) of configuration  $X$  at  $\text{MI}_{\text{avg}} = \text{MI}_{\text{ref}}$  equals  $G_{\text{dB}} = 10 \log_{10} \left( \frac{(E_{\text{tr}}/N_0)_{\text{CD}}}{(E_{\text{tr}}/N_0)_X} \right)$ . This means that to obtain  $\text{MI}_{\text{avg}} = \text{MI}_{\text{ref}}$ , the TX power for configuration  $X$  is  $G_{\text{dB}}$  dB less than for the CD without optimization.

The SNR gains resulting from the configurations for THP considered above are displayed in tabular form in Sections 8.2 to 8.4, allowing a direct comparison of the various existing and novel configurations. In Section 8.5, the CD and the AD, both without optimization and with the best performing optimization, are compared in terms of the average MI versus  $\gamma_t$  (terrestrial channel) or  $\gamma_s$  (satellite channel), while in Section 8.6, these configurations for THP are compared to LP.

### 8.1 Full-search versus reduced-complexity optimization

The complexity of the algorithms for optimizing  $\theta$  is reduced by limiting the search tree to at most  $L$  nodes

at each level; the resulting complexity is proportional to  $LQ$ . We consider different  $(L, Q)$  yielding a fixed  $LQ$  and select the combination obtaining the highest average (over the range of MI) SNR gain of the AD for  $(\text{MoL} + \text{SNR}, R)_{\text{RC}}$  in the case of 2-PAM; the selected  $(L, Q)$  are shown in Table 1 for  $LQ = 2, 4, 8, 16, 32$ . We have verified (results not shown for conciseness) that, with increasing  $LQ$ , the SNR gain of the AD resulting from the  $(\text{MoL} + \text{SNR}, R)$  algorithm increases, but the SNR gain increments get smaller. For terrestrial communication, we select  $(L, Q) = (4, 4)$ , because the gain increment is smaller than 0.1 dB when moving from  $(L, Q) = (4, 4)$  to  $(L, Q) = (8, 4)$ . For satellite communication, we select  $(L, Q) = (1, 8)$ , because the SNR gains for  $(L, Q) = (1, 8), (1, 16), (1, 32)$  are nearly the same. These selections are applied to all reduced-complexity algorithms involving an optimization over  $\theta$ , and for all constellations. We have verified that the reduction of the SNR gain, caused by applying the reduced-complexity optimization (with  $(L, Q) = (4, 4)$  for terrestrial communication and  $(L, Q) = (1, 8)$  for satellite communication) rather than the full-search optimization (with  $Q = 4$ ), is limited to only about 0.6 dB and 0.2 dB for the terrestrial channel and the satellite channel, respectively, for the algorithms involving an optimization over  $\theta$  only, in the case of 2-PAM.

When optimizing over the row permutation, an exhaustive search is avoided by using the reduced-complexity optimizations. We have verified in the case of 2-PAM that, for both satellite and terrestrial channels, the loss is limited to less than 0.25 dB, when replacing the  $(\text{SNR}, P)$  and  $(\text{MoL}, P)$  algorithms by  $(\text{SNR}, P)_{\text{RC}}$  and  $(\text{MoL}, P)_{\text{RC}}$ , respectively.

## 8.2 CD and AD performance for 2-PAM

Tables 2 and 3 show the SNR gains pertaining to the CD and the AD, respectively, for 2-PAM. Results are given for the case without TX optimization ( $\theta = 0$ ,  $\mathbf{P} = \mathbf{I}_{N_u}$ ), and for the reduced-complexity optimization algorithms from Sections 7.1.4 and 7.2. For each category (i.e., optimization over rotations only, optimization over permutation only, joint optimization over rotations and permutation), the SNR gain of the best algorithm for the considered MI value is displayed in italics; the entries in bold refer to the overall best algorithm for the considered MI. The SNR gains are discussed below.

**Table 1** Best  $(L, Q)$  for different  $LQ$

$LQ$	Terrestrial					Satellite				
	2	4	8	16	32	2	4	8	16	32
$(L, Q)$	(1,2)	(1,4)	(2,4)	(4,4)	(8,4)	(1,2)	(1,4)	(1,8)	(1,16)	(1,32)

### 8.2.1 CD performance

Table 2 displays the SNR gains for the CD. As the CD always performs a modulo operation, its MoL cannot be reduced by any of the optimization algorithms. Consequently, as far as the optimization over  $\theta$  only, over  $\mathbf{P}$  only and over  $(\mathbf{P}, \theta)$  is concerned, the largest SNR gains are obtained for  $(\text{SNR}, R)_{\text{RC}}$ ,  $(\text{SNR}, P)_{\text{RC}}$  and  $((\text{SNR}, P), (\text{SNR}, R))_{\text{RC}}$ , respectively; for conciseness, the results for the other algorithms are not shown. Comparing  $(\text{SNR}, R)_{\text{RC}}$  and  $(\text{SNR}, P)_{\text{RC}}$ , the former outperforms the latter for terrestrial communication, whereas the opposite holds for satellite communication. The SNR gains resulting from  $((\text{SNR}, P), (\text{SNR}, R))_{\text{RC}}$  are approximately the sum of the gains provided by  $(\text{SNR}, R)_{\text{RC}}$  and  $(\text{SNR}, P)_{\text{RC}}$  individually and are in the range (1.4 dB, 3.2 dB) for terrestrial communication and (0.6 dB, 1.0 dB) for satellite communication. Compared to the better of  $(\text{SNR}, R)_{\text{RC}}$  and  $(\text{SNR}, P)_{\text{RC}}$ , the  $((\text{SNR}, P), (\text{SNR}, R))_{\text{RC}}$  algorithm provides an additional gain ranging from about 0.2 to 0.5 dB.

### 8.2.2 AD performance without optimization

The non-optimized AD already provides a substantial SNR gain over the non-optimized CD. This gain is in the range (1.0 dB, 5.1 dB) for terrestrial communication and (1.0 dB, 6.7 dB) for satellite communication. The largest gains occur at low MI, where the MoL of the non-optimized CD is the largest.

### 8.2.3 AD performance when optimizing $\theta$ only

When only the rotations are optimized, we observe from Table 3 that for most values of the MI,  $(\text{MoL}, R)_{\text{RC}}$  performs slightly better than  $(\text{SNR}, R)_{\text{RC}}$ , and that both of these algorithms are outperformed by  $(\text{MoL} + \text{SNR}, R)_{\text{RC}}$ . Compared to the non-optimized AD,  $(\text{MoL} + \text{SNR}, R)_{\text{RC}}$  provides an additional gain in the range (0.8 dB, 1.7 dB) for the terrestrial link and (0.8 dB, 1.2 dB) for the satellite link.

The value of  $\theta$  affects both  $\rho_C$  and  $\text{SNR}_{\text{det}}$ . Table 4 shows  $\mathbb{E}[\text{SNR}_{\text{det}}]$  (in dB) and  $\mathbb{E}[\rho_C]$  (averages are over the channel realizations), for the non-optimized AD and for the AD with optimization over  $\theta$ . We observe that  $(\text{SNR}, R)_{\text{RC}}$  (which maximizes  $\text{SNR}_{\text{det}}$ ) not only provides that largest  $\text{SNR}_{\text{det}}$ , but also yields a value of  $\mathbb{E}[\rho_C]$  which is larger than for the non-optimized case. Similarly,  $(\text{MoL}, R)_{\text{RC}}$  (which maximizes  $N_C$ ) gives rise to the largest  $\mathbb{E}[\rho_C]$  and to a value of  $\mathbb{E}[\text{SNR}_{\text{det}}]$  which is larger than for the case without optimization. Hence, the maximization of one parameter ( $\text{SNR}_{\text{det}}$  or  $N_C$ ) also increases the other parameter, so that both parameters benefit from either maximization; therefore, all three algorithms provide a SNR gain which is larger than when no optimization is carried out. The  $(\text{MoL} + \text{SNR}, R)_{\text{RC}}$  algorithms achieves the same  $\mathbb{E}[\rho_C]$  as  $(\text{MoL}, R)_{\text{RC}}$ , and in addition provides

**Table 2** SNR gain (dB) of the CD for 2-PAM

	Terrestrial					Satellite				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
MI <sub>avg</sub>	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
$\theta = 0, \mathbf{P} = \mathbf{I}_{N_u}$	0	0	0	0	0	0	0	0	0	0
(SNR, R) <sub>RC</sub>	0.34	0.41	0.40	0.34	0.43	0.38	0.46	0.45	0.48	0.52
(SNR, P) <sub>RC</sub>	1.10	1.33	1.55	1.86	2.81	0.12	0.19	0.21	0.30	0.44
((SNR, P), (SNR, R)) <sub>RC</sub>	<b>1.45</b>	<b>1.78</b>	<b>1.96</b>	<b>2.31</b>	<b>3.22</b>	<b>0.58</b>	<b>0.68</b>	<b>0.71</b>	<b>0.77</b>	<b>1.00</b>

a higher  $\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$ . This explains the (slight) superiority of  $(\text{MoL} + \text{SNR}, \text{R})_{\text{RC}}$  over both  $(\text{MoL}, \text{R})_{\text{RC}}$  and  $(\text{SNR}, \text{R})_{\text{RC}}$ , observed in Table 3.

#### 8.2.4 AD performance when optimizing P only

When only the row permutation is optimized, Table 3 indicates that (i) for terrestrial communication  $(\text{SNR}, \text{P})_{\text{RC}}$  is the best algorithm for most values of the MI, whereas  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  is the best for small MI, and (ii) for satellite communication  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  is the best algorithm<sup>3</sup> for most values of the MI, whereas  $(\text{SNR}, \text{P})_{\text{RC}}$  is the best only for large MI. Compared to the non-optimized AD, the better of  $(\text{SNR}, \text{P})_{\text{RC}}$  and  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  provides an additional gain in the range (0.8 dB, 2.4 dB) for the terrestrial link and (0.3 dB, 0.6 dB) for the satellite link.

Table 3 indicates that for some values of the MI,  $(\text{SNR}, \text{P})_{\text{RC}}$  or  $(\text{MoL}, \text{P})_{\text{RC}}$  perform worse, compared to

the case of no optimization. The explanation follows from Table 5, which shows  $\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$  and  $\mathbb{E}[\rho_C]$  for the AD in the absence of optimization, and for  $(\text{SNR}, \text{P})_{\text{RC}}$ ,  $(\text{MoL}, \text{P})_{\text{RC}}$  and  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$ . We observe that  $(\text{SNR}, \text{P})_{\text{RC}}$  (which maximizes  $\text{SNR}_{\text{det,max}}$ ) yields the largest value of  $\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$ , but the corresponding  $\mathbb{E}[\rho_C]$  is smaller than for no optimization. Similarly,  $\mathbb{E}[\rho_C]$  is largest for  $(\text{MoL}, \text{P})_{\text{RC}}$  (which maximizes  $\rho_C$ ), but the corresponding  $\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$  is smaller than for no optimization. Hence, maximizing  $\text{SNR}_{\text{det,max}}$  or  $\rho_C$  automatically reduces  $\rho_C$  or  $\text{SNR}_{\text{det}}$ , respectively; the resulting performance is a combination of both effects, which in some cases can be worse than when no optimization is carried out. The algorithm  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  yields a value of  $\mathbb{E}[\rho_C]$  which is only slightly smaller than for  $(\text{MoL}, \text{P})_{\text{RC}}$ ; the corresponding  $\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$  is significantly larger than for  $(\text{MoL}, \text{P})_{\text{RC}}$ , and, for terrestrial communication, even larger than without optimization.

**Table 3** SNR gain (dB) of the AD for 2-PAM

	Terrestrial					Satellite				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
MI <sub>avg</sub>	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
$\theta = 0, \mathbf{P} = \mathbf{I}_{N_u}$	5.07	2.46	1.64	1.21	1.03	6.65	3.48	2.22	1.50	0.98
(MoL, R) <sub>RC</sub>	6.77	3.82	2.71	2.05	1.79	7.75	4.61	3.21	2.35	1.65
(SNR, R) <sub>RC</sub>	6.58	3.67	2.65	2.05	1.79	7.66	4.52	3.12	2.30	1.69
(MoL + SNR, R) <sub>RC</sub>	6.80	3.84	2.73	2.10	1.82	7.81	4.66	3.29	2.39	1.73
(MoL, P) <sub>RC</sub>	5.85	2.89	1.74	1.05	0.50	7.18	4.05	2.60	1.74	0.93
(SNR, P) <sub>RC</sub>	5.38	3.22	2.77	2.72	3.47	6.31	3.21	2.09	1.52	1.24
(MoL + SNR, P) <sub>RC</sub>	6.00	3.11	2.08	1.49	1.28	7.22	4.03	2.67	1.75	0.99
((SNR, P), (MoL, R)) <sub>RC</sub>	<b>7.51</b>	<b>4.70</b>	<b>3.86</b>	<b>3.54</b>	3.97	7.86	4.71	3.33	2.48	1.96
((MoL, P), (MoL, R)) <sub>RC</sub>	6.98	4.00	2.65	1.74	0.96	7.79	4.68	3.31	2.42	1.62
((MoL + SNR, P), (MoL, R)) <sub>RC</sub>	7.37	4.46	3.16	2.25	1.47	7.94	4.84	3.46	2.49	1.75
((SNR, P), (SNR, R)) <sub>RC</sub>	7.11	4.39	3.70	3.48	<b>4.00</b>	7.70	4.55	3.22	2.46	<b>2.04</b>
((MoL, P), (SNR, R)) <sub>RC</sub>	6.47	3.45	2.32	1.16	1.05	7.88	4.77	3.36	2.42	1.69
((MoL + SNR, P), (SNR, R)) <sub>RC</sub>	6.79	3.89	2.95	2.48	2.49	<b>8.00</b>	<b>4.89</b>	<b>3.49</b>	<b>2.61</b>	1.81
((SNR, P), (MoL + SNR, R)) <sub>RC</sub>	7.41	4.60	3.76	3.43	3.85	7.91	4.71	3.35	2.54	1.96
((MoL, P), (MoL + SNR, R)) <sub>RC</sub>	6.87	3.93	2.56	1.66	0.77	7.83	4.71	3.36	2.46	1.66
((MoL + SNR, P), (MoL + SNR, R)) <sub>RC</sub>	7.37	4.35	3.06	2.21	1.33	7.95	4.86	3.44	2.51	1.72

**Table 4**  $\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$  (dB) and  $\mathbb{E}[\rho_C]$  when optimizing rotation angles  $\theta$  for 2-PAM

	Terrestrial		Satellite	
	$\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$ (dB)	$\mathbb{E}[\rho_C]$	$\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$ (dB)	$\mathbb{E}[\rho_C]$
$\theta = 0$	1.14	0.62	-4.20	0.8
$(\text{SNR}, \text{R})_{\text{RC}}$	1.60	0.79	-3.75	0.91
$(\text{MoL}, \text{R})_{\text{RC}}$	1.49	0.84	-4.08	0.95
$(\text{MoL} + \text{SNR}, \text{R})_{\text{RC}}$	1.52	0.84	-3.81	0.95

### 8.2.5 AD performance when optimizing $(\mathbf{P}, \theta)$

Let us consider the SNR gains from Table 3, related to the joint optimization of the row permutation and the rotation angles.

For terrestrial communication, we observe that the combined algorithms involving  $(\text{SNR}, \text{P})_{\text{RC}}$  outperform those involving  $(\text{MoL}, \text{P})_{\text{RC}}$ . For a large range of MI, the best performance is obtained for  $((\text{SNR}, \text{P}), (\text{MoL}, \text{R}))_{\text{RC}}$ , whereas  $((\text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$  slightly outperforms  $((\text{SNR}, \text{P}), (\text{MoL}, \text{R}))_{\text{RC}}$  at high MI. Compared to the AD without TX optimization, additional gains in the range (2.2 dB, 3.0 dB) are achieved; these gains are about 0.5 dB to 1.1 dB larger than the maximum gains resulting from the optimization over only  $\mathbf{P}$  or only  $\theta$ .

In the case of satellite communication,  $((\text{MoL} + \text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$  slightly outperforms all other optimization strategies for a large range of MI, whereas  $((\text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$  performs better for the largest MIs. Compared to the AD without TX optimization, additional gains in the range (1.0 dB, 1.4 dB) are achieved; these gains are about 0.2 dB to 0.3 dB larger than the maximum gains resulting from the optimization over only  $\mathbf{P}$  or only  $\theta$ . However, in case one wants to avoid the nested algorithms because of their high computational complexity,  $((\text{MoL} + \text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$  can be replaced by  $((\text{SNR}, \text{P}), (\text{MoL} + \text{SNR}, \text{R}))_{\text{RC}}$ , giving rise to only a small reduction in gain, not exceeding 0.2 dB.

## 8.3 CD and AD performance for 4-PAM

### 8.3.1 CD performance

Similarly to 2-PAM, for 4-PAM, we need to consider for the CD only  $(\text{SNR}, \text{R})_{\text{RC}}$ ,  $(\text{SNR}, \text{P})_{\text{RC}}$ , and

**Table 5**  $\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$  (dB) and  $\mathbb{E}[\rho_C]$  when optimizing the row permutation matrix  $\mathbf{P}$  for 2-PAM

	Terrestrial		Satellite	
	$\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$ (dB)	$\mathbb{E}[\rho_C]$	$\frac{\mathbb{E}[\text{SNR}_{\text{det}}]}{E_{\text{tr}}/N_0}$ (dB)	$\mathbb{E}[\rho_C]$
$\mathbf{P} = \mathbf{I}_{N_u}$	1.14	0.62	-4.20	0.8
$(\text{SNR}, \text{P})_{\text{RC}}$	2.66	0.51	-3.97	0.72
$(\text{MoL}, \text{P})_{\text{RC}}$	0.64	0.82	-4.35	0.94
$(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$	1.54	0.81	-4.32	0.93

$((\text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$ . The resulting SNR gains are shown in Table 6. We observe that the gains resulting from  $(\text{SNR}, \text{R})_{\text{RC}}$  are much smaller than for 2-PAM; this is because without optimization, the power loss for 4-PAM is smaller than for 2-PAM (as indicated by the upper bound [35] on  $\sigma_{\text{mod},n}^2/\sigma_a^2$ ). The best algorithm is  $((\text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$ , which only slightly (by less than 0.07 dB) outperforms the (less complex)  $(\text{SNR}, \text{P})_{\text{RC}}$  algorithm. The resulting SNR gains are in the range (1.3 dB, 2.6 dB) for the terrestrial channel and (0.3 dB, 0.5 dB) for the satellite channel.

### 8.3.2 AD performance without optimization

Table 7 shows the results for the AD. Without optimization, the AD provides a gain compared to the non-optimized CD, in the range (0.3 dB, 1.8 dB) for terrestrial communication and (0.2 dB, 2.2 dB) for satellite communication. Note that these gains are smaller compared to the case of 2-PAM transmission: with 4-PAM, the interference pre-subtraction terms have larger peak values, causing the condition  $C_{\text{PAM}}(n)$  to hold less frequently.

### 8.3.3 AD performance when optimizing $\theta$ only

In the case of terrestrial communication,  $(\text{MoL}, \text{R})_{\text{RC}}$  yields the largest SNR gain for a large range of MI, whereas  $(\text{SNR}, \text{R})_{\text{RC}}$  performs only slightly better for high MI; the resulting gains compared to the non-optimized AD are in the range (0.2 dB, 1.0 dB).

For satellite communication,  $(\text{SNR}, \text{R})_{\text{RC}}$  performs best, yielding gains compared to the non-optimized AD in the range (0.2 dB, 1.6 dB).

### 8.3.4 AD performance when optimizing $\mathbf{P}$ only

For the terrestrial link, the largest SNR gain results from  $(\text{SNR}, \text{P})_{\text{RC}}$  for a large range of MI, whereas  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  performs better at small MI; compared to the non-optimized AD, gains in the range (1.2 dB, 2.5 dB) are achieved.

For the satellite link, the largest SNR gains at small, medium and large MI are obtained for  $(\text{MoL}, \text{P})_{\text{RC}}$ ,  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  and  $(\text{SNR}, \text{P})_{\text{RC}}$ , respectively; note that for small MI, the SNR gains resulting from  $(\text{MoL}, \text{P})_{\text{RC}}$  and  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  are nearly the same. Compared to the non-optimized AD, the gains are in the range (0.3 dB, 1.4 dB).

### 8.3.5 AD performance when optimizing $(\mathbf{P}, \theta)$

On the terrestrial channel,  $((\text{SNR}, \text{P}), (\text{MoL}, \text{R}))_{\text{RC}}$  yields the largest SNR gain for a large range of MI, but this algorithm is outperformed by  $((\text{MoL} + \text{SNR}, \text{P}), (\text{MoL} + \text{SNR}, \text{R}))_{\text{RC}}$  for small MI. The gain compared to the non-optimized AD is in the range (1.8 dB, 2.5 dB); these gains are about 0.2 dB to 1.2 dB larger than the maximum gains resulting from the optimization over only  $\mathbf{P}$  or only  $\theta$ . When one wants to avoid nested

**Table 6** SNR gain (dB) of the CD for 4-PAM

	Terrestrial					Satellite				
	0.2	0.6	1	1.4	1.8	0.2	0.6	1	1.4	1.8
MI <sub>avg</sub>	0.2	0.6	1	1.4	1.8	0.2	0.6	1	1.4	1.8
$\mathbf{P} = \mathbf{I}_{N_u}$	0	0	0	0	0	0	0	0	0	0
(SNR, R) <sub>RC</sub>	0.02	0.01	0.05	0.02	0.04	0.08	0.08	0.08	0.09	0.03
(SNR, P) <sub>RC</sub>	1.25	1.51	1.65	1.89	2.57	0.28	0.30	0.33	0.36	0.40
((SNR, P), (SNR, R)) <sub>RC</sub>	<b>1.28</b>	<b>1.55</b>	<b>1.69</b>	<b>1.91</b>	<b>2.57</b>	<b>0.35</b>	<b>0.37</b>	<b>0.40</b>	<b>0.43</b>	<b>0.47</b>

algorithms because of their high computational complexity,  $((\text{SNR}, \text{P}), (\text{MoL}, \text{R}))_{\text{RC}}$  can be used also at small MI, at the expense of a loss of about 0.6 dB.

When using the satellite channel,  $((\text{MoL} + \text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$  gives rise to the largest SNR gain for a large range of MI, whereas  $((\text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$  performs best for large MI. The gain compared to the non-optimized AD is in the range (0.6 dB, 3.0 dB); these gains are about 0.2 dB to 1.5 dB larger than the maximum gains resulting from the optimization over only  $\mathbf{P}$  or only  $\theta$ . When nested algorithms must be avoided for complexity reasons,  $((\text{SNR}, \text{P}), (\text{SNR}, \text{R}))_{\text{RC}}$  should be used for the entire range of MI, at the expense of a loss ranging from 0.2 to 1.5 dB.

#### 8.4 CD and AD performance for 4-QAM

Because of the  $\pi/4$  angular symmetry of the QAM constellation, optimizing over  $\theta$  provides only a negligible gain; therefore, for both the AD and the CD, only the optimization of the row permutation is considered.

As the CD always performs a modulo operation,  $(\text{SNR}, \text{P})_{\text{RC}}$  automatically yields the best performance among all row permutation optimizations considered. Table 8 shows a resulting SNR gain in the range (1.1 dB,

2.8 dB) for the terrestrial link and (0.2 dB, 0.5 dB) for the satellite link.

The SNR gains for the AD are displayed in Table 9. Without optimization, a gain compared to the non-optimized CD is achieved, in the range (0.6 dB, 2.9 dB) for terrestrial communication and (0.5 dB, 3.9 dB) for satellite communication. When optimizing  $\mathbf{P}$  for the terrestrial channel,  $(\text{SNR}, \text{P})_{\text{RC}}$  yields the largest SNR gain for a wide range of MI, whereas  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  has the best performance for small MI. For the satellite channel,  $(\text{MoL}, \text{P})_{\text{RC}}$  performs best for the lower range of MI, whereas for higher MI either  $(\text{SNR}, \text{P})_{\text{RC}}$  or  $(\text{MoL} + \text{SNR}, \text{P})_{\text{RC}}$  provides the large gain. The best performing algorithms yield additional gains compared to the non-optimized AD, in the range (1.0 dB, 2.5 dB) for terrestrial communication and (0.2 dB, 1.5 dB) for satellite communication.

#### 8.5 Performance comparison summary for THP

Here, we summarize the performances of the algorithms considered above. We present the average MI achieved by the CD and the AD, with and without TX optimization, as a function of  $\gamma_t$  (terrestrial channel) or  $\gamma_s$  (satellite channel). In the case of TX optimization, we always select

**Table 7** SNR gain (dB) of the AD for 4-PAM

	Terrestrial					Satellite				
	0.2	0.6	1	1.4	1.8	0.2	0.6	1	1.4	1.8
MI <sub>avg</sub>	0.2	0.6	1	1.4	1.8	0.2	0.6	1	1.4	1.8
$\theta = 0, \mathbf{P} = \mathbf{I}_{N_u}$	1.84	0.87	0.58	0.38	0.29	2.21	0.97	0.64	0.41	0.21
(MoL, R) <sub>RC</sub>	2.86	1.38	0.86	0.63	0.44	2.48	1.10	0.72	0.44	0.26
(SNR, R) <sub>RC</sub>	2.32	1.14	0.72	0.48	0.45	3.84	1.71	1.10	0.71	0.39
(MoL + SNR, R) <sub>RC</sub>	2.14	1.05	0.67	0.44	0.43	2.48	1.11	0.72	0.44	0.25
(MoL, P) <sub>RC</sub>	2.70	1.00	0.45	0.07	-0.29	3.62	1.48	0.85	0.44	0.06
(SNR, P) <sub>RC</sub>	2.62	2.16	2.07	2.16	2.74	2.09	1.14	0.85	0.69	0.60
(MoL + SNR, P) <sub>RC</sub>	3.05	1.57	1.11	0.85	0.80	3.61	1.58	0.94	0.54	0.22
((SNR, P), (MoL, R)) <sub>RC</sub>	3.69	<b>2.62</b>	<b>2.37</b>	<b>2.36</b>	<b>2.83</b>	2.11	1.15	0.88	0.69	0.62
((SNR, P), (SNR, R)) <sub>RC</sub>	3.50	2.55	2.33	2.34	2.82	3.63	1.79	1.29	<b>0.98</b>	<b>0.81</b>
((MoL + SNR, P), (SNR, R)) <sub>RC</sub>	3.10	1.83	1.45	1.29	1.24	<b>5.16</b>	<b>2.38</b>	<b>1.47</b>	0.87	0.37
((SNR, P), (MoL + SNR, R)) <sub>RC</sub>	3.47	2.53	2.31	2.32	2.80	2.11	1.15	0.88	0.68	0.62
((MoL, P), (MoL + SNR, R)) <sub>RC</sub>	3.90	1.57	0.77	0.22	-0.17	4.73	2.11	1.21	0.70	0.23
((MoL + SNR, P), (MoL + SNR, R)) <sub>RC</sub>	<b>4.28</b>	2.11	1.38	0.96	0.80	4.74	2.16	1.27	0.78	0.33

**Table 8** SNR gain (dB) of the CD for 4-QAM

	Terrestrial					Satellite					
	MI <sub>avg</sub>	0.2	0.6	1	1.4	1.8	0.2	0.6	1	1.4	1.8
	$\mathbf{P} = \mathbf{I}_{N_u}$	0	0	0	0	0	0	0	0	0	0
(SNR, P) <sub>RC</sub>	1.10	1.31	1.60	1.93	2.77	0.19	0.26	0.26	0.34	0.50	

the RC algorithm providing the largest MI for the considered value of  $\gamma_s$  or  $\gamma_t$ . As indicated in Tables 3 and 9, the AD in combination with an algorithm involving the novel (MoL, P)<sub>RC</sub> or (MoL + SNR, P)<sub>RC</sub> is often optimum at low MI (where the MoL of the CD without TX optimization is large).

Figures 5, 6, and 7 pertain to 2-PAM, 4-PAM, and 4-QAM, respectively. The following observations are made.

- For the satellite channel using the CD, TX optimization provides only a modest gain. The optimized CD performs worse than the non-optimized AD, except for large MI where the former is only marginally better than the latter.
- For the terrestrial channel using the CD, TX optimization provides a larger gain than on the satellite channel. The curves for the optimized CD and the non-optimized AD intersect, with the latter outperforming the former in the lower range of MI, where the MoL of the non-optimized CD is large.
- For both types of channel, the best performance is achieved by the optimized AD; the resulting gain compared to the non-optimized CD is largest for 2-PAM, because the fraction of the users that can remove the modulo operation is larger for 2-PAM than for 4-PAM and 4-QAM.

### 8.6 Comparison with LP

The LP is not affected by MoL and gives rise to a SNR at the detector given by:

$$\begin{aligned}\text{SNR}_{\text{det},\text{LP}} &= \left( \frac{1}{N_U} \text{tr} \left( (\mathbf{H}\mathbf{H}^H)^{-1} \right) \right)^{-1} \frac{E_{\text{tr}}}{N_0} \\ &= \left( \frac{1}{N_U} \text{tr} \left( (\mathbf{L}\mathbf{L}^H)^{-1} \right) \right)^{-1} \frac{E_{\text{tr}}}{N_0}\end{aligned}$$

**Table 9** SNR gain (dB) of the AD for 4-QAM

	Terrestrial					Satellite				
	MI <sub>avg</sub>	0.2	0.6	1	1.4	1.8	0.2	0.6	1	1.4
$\mathbf{P} = \mathbf{I}_{N_u}$	2.92	1.26	0.95	0.75	0.61	3.88	1.65	1.00	0.74	0.53
(MoL, P) <sub>RC</sub>	3.62	1.35	0.76	0.30	-0.09	5.33	2.43	1.40	0.89	0.22
(SNR, P) <sub>RC</sub>	3.31	2.28	2.31	2.41	3.14	3.13	1.47	1.01	0.88	0.85
(MoL + SNR, P) <sub>RC</sub>	3.95	1.90	1.40	1.12	0.92	5.15	2.37	1.37	0.94	0.48

Note that the performance resulting from LP is not affected by the user ordering nor the constellation rotations. If THP were without PoL, the corresponding SNR at the detector (10) would become

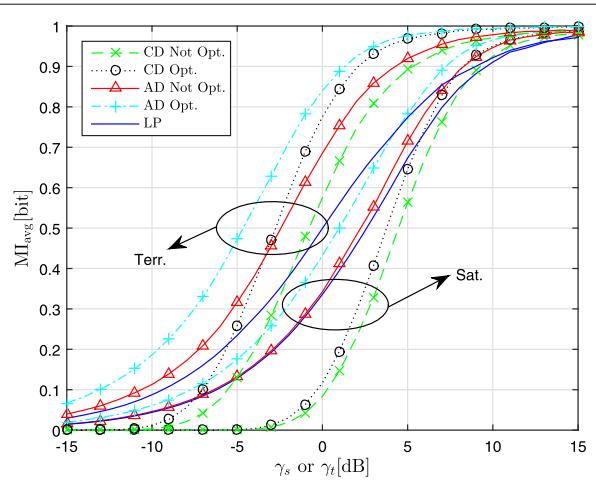
$$\text{SNR}_{\text{det},\text{THP,noPoL}} = \left( \frac{1}{N_U} \text{tr} \left( (\mathbf{L}_d \mathbf{L}_d^H)^{-1} \right) \right)^{-1} \frac{E_{\text{tr}}}{N_0}$$

We have  $\text{SNR}_{\text{det},\text{LP}} \leq \text{SNR}_{\text{det},\text{THP,noPoL}}$ , with equality if and only if the rows of  $\mathbf{H}$  are orthogonal. In spite of this inequality, LP can outperform THP, when the latter is affected by a large amount of PoL and MoL.

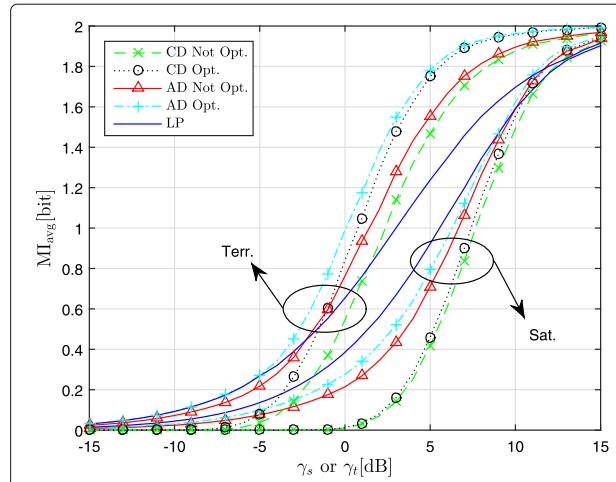
We have included in Figs. 5, 6, and 7 also the average MI resulting from LP. Comparing the LP with the non-optimized CD for a given constellation, we observe that at large MI the non-optimized CD performs better, whereas the opposite occurs at low MI; this is because the MoL increases with decreasing MI, favoring LP at low MI. The crossing point, of the curves for LP and the non-optimized CD, occurs at larger MI for the satellite channel, compared to the terrestrial channel: the rows of  $\mathbf{H}$  for the former channel tend to be more orthogonal (resulting in smaller interference), yielding a ratio  $\text{SNR}_{\text{det},\text{LP}}/\text{SNR}_{\text{det},\text{THP,noPoL}}$  closer to 1.

The optimized AD reduces the MoL and/or increases  $\text{SNR}_{\text{det},\text{THP,noPoL}}$  of the THP scheme. On the terrestrial channel, the optimized AD outperforms LP over the entire range of MI and for all constellations considered. On the satellite channel, only for 2-PAM the optimized AD is the best over the entire range of MI; for 4-PAM and 4-QAM, LP outperforms the optimized AD for  $\text{MI} < 0.8$  and  $\text{MI} < 1.5$ , respectively.

Figure 8 compares the different constellations considered, in terms of the maximum average MI over the LP and the optimized AD. For a given operating point, it can be seen from Figs. 5, 6, and 7 whether this maximum MI is achieved by the LP or by the optimized AD. The following observations can be made.



**Fig. 5** Performance of CD, AD and LP for 2-PAM on terrestrial and satellite channel

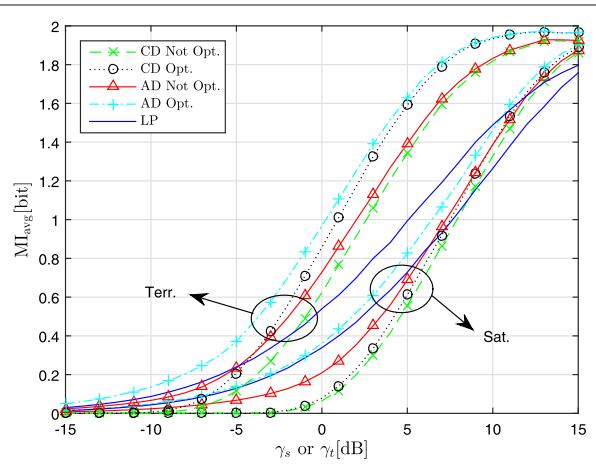


**Fig. 7** Performance of CD, AD, and LP for 4-QAM on terrestrial and satellite channel

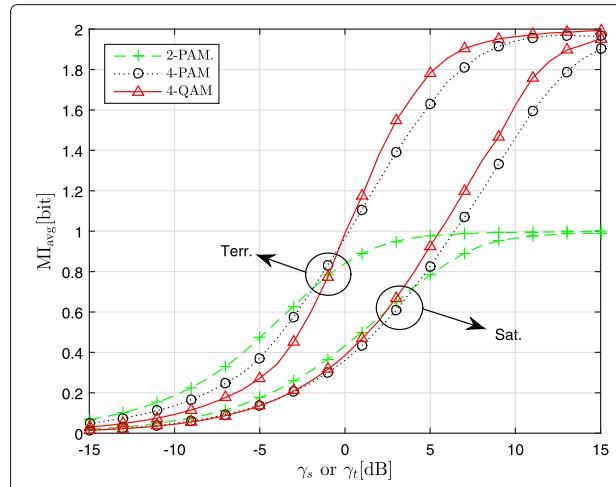
- For the terrestrial channel, the best selection of constellations is (i) 2-PAM (with optimized AD) for  $MI < 0.7$ ; (ii) 4-PAM (with optimized AD) for  $0.7 < MI < 0.9$ ; and (iii) 4-QAM (with optimized AD) for  $MI > 0.9$ . In the interval  $0.7 < MI < 0.9$ , 4-PAM is only marginally better than either of 2-PAM and 4-QAM, so that only a small loss (less than 0.5 dB) is incurred when selecting 2-PAM for  $MI < 0.8$  and 4-QAM for  $MI > 0.8$  instead. According to Figs. 5, 6, and 7, the latter selection of constellations achieves a significant gain of up to about 4 dB, compared to the better of non-optimized CD and LP.
- For the satellite channel, the best selection of constellations is (i) 2-PAM (with optimized AD) for  $MI < 0.6$ ; (ii) 4-QAM (with LP) for  $0.6 < MI < 1.5$ ; and (iii) 4-QAM (with optimized AD) for  $MI > 1.5$ . According to Figs. 5, 6, and 7, this optimum selection

achieves only a moderate gain (up to about 1 dB) over the better of non-optimized CD and LP. Considering the higher complexity of the optimized AD, one might prefer to always use LP (with 2-PAM for  $MI < 0.15$  and 4-QAM for  $MI > 0.15$ ) instead.

From the above comparison, it follows that the channel statistics have a major impact on which algorithm and which constellation are optimum at a given MI: for the terrestrial channel, the optimized AD (combined with the proper constellation) provides the best performance over the entire range of MI, whereas the LP with 4-QAM performs best for the satellite channel over the medium range ( $0.6 < MI < 1.5$ ). This different behavior is attributed to the smaller interference on the satellite channel.



**Fig. 6** Performance of CD, AD and LP for 4-PAM on terrestrial and satellite channel



**Fig. 8** Best performance (over LP and optimized AD) for the different constellations on terrestrial and satellite channel

## 9 Conclusions

In this contribution, we consider a MU-MISO communication system with THP. The receiver uses a CD (always performing a modulo operation) or an AD (performing a modulo operation only when needed). We investigate the effect of reduced-complexity algorithms that select the symbol rotations at the TX and the ordering of the users, to reduce the MoL and to increase the SNR at the detector; several of these algorithms are novel.

Taking the average MI at the detector as a performance measure, results are presented for a terrestrial wireless channel and for a multi-beam satellite channel; we consider 2-PAM, 4-PAM, and 4-QAM constellations, since the PoL and MoL are the largest for small constellation sizes.

For the CD, the largest MI is obtained when the SNR at the detector is maximized, by selecting first the user ordering and then the rotations of the constellations; the latter step brings a substantial additional gain only for 2-PAM.

For the AD, no single algorithm is optimum for the entire range of MI; as the MoL is larger for small MI, the better algorithms are those which reduce the MoL for small MI and increase the SNR at the detector for large MI. When optimizing the TX, the gains resulting from the AD are considerably larger than those from the CD, especially at small MI (where the MoL of the CD is large); hence, the AD with TX optimization outperforms the CD with TX optimization. These gains are larger for 2-PAM than for 4-PAM and 4-QAM, mainly because the optimization of the constellation rotations has a larger effect for 2-PAM.

When selecting the best constellation for the optimized AD and for the LP, it is found that, on the terrestrial channel, the optimized AD outperforms the LP; compared to the better of LP and non-optimized CD, the optimized AD achieves a significant gain (up to about 4 dB). In contrast, on the satellite channel, the optimized AD actually performs worse than the LP for the medium range of MI and achieves only a modest gain (up to about 1 dB) for small and large MI. Hence, the application of the optimized AD is quite promising for wireless terrestrial communication, whereas its usefulness for multi-beam satellite communication is rather limited.

Although the focus of this contribution is on MU-MISO THP, the concepts are easily extended to SU-MIMO and MU-MIMO scenarios.

## Endnotes

<sup>1</sup> When condition  $C(n)$  holds, it does not matter whether or not the modulo operation at the TX is active for the  $n$ -th user. Hence, from a practical point of view,

having the modulo operation at the TX active for all  $N_U$  users is the simplest choice.

<sup>2</sup> The algorithm from [14] maximizes  $\text{SNR}_{\text{det}}$  for the CD, with  $\sigma_{\text{mod},n}^2$  computed as an arithmetical average over a block of transmitted data, instead of an expectation over the data. Both approaches yield similar results for long frames.

<sup>3</sup> We make abstraction of the 0.02 dB higher SNR gain at  $\text{MI} = 0.3$  for  $(\text{MoL}, P)_{\text{RC}}$  compared to  $(\text{MoL} + \text{SNR}, P)_{\text{RC}}$ , which is negligible and might be caused by limited numerical accuracy.

## Abbreviations

AD: Alternative detector; CCI: Co-channel interference; CD: Conventional detector; CSI: Channel state information; DPC: Dirty paper coding; FSL: Free space loss; LOS: Line of sight; LP: Linear precoding; MI: Mutual information; MIMO: Multiple-input multiple-output; MISO: Multiple-input single-output; MoL: Modulo loss; MU: Multi-user; PAM: Phase amplitude modulation; PoL: Power loss; QAM: Quadrature amplitude modulation; RC: Reduced-complexity; RX: Receiver; SNR: Signal-to-noise ratio; SU: Single-user; THP: Tomlinson-Harashima precoding; TX: Transmitter; ZF: Zero-forcing

## Funding

This work was supported by the Research Foundation Flanders (FWO) under grant no. EOS-30452698.

## Authors' contributions

ED conceived the study, designed the novel algorithms, performed the simulation experiments, and drafted the manuscript. MM reviewed and edited the manuscript. Both authors read and approved the final manuscript.

## Authors' information

E. Debels is currently pursuing the PhD degree at Ghent University. M. Moeneclaey is Full Professor at Ghent University.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 7 May 2018 Accepted: 1 October 2018

Published online: 06 November 2018

## References

1. M. H. M. Costa, Writing on dirty paper. *IEEE Trans. Inform. Theory*. **IT-29**, 439–441 (1983)
2. M. Tomlinson, New automatic equalizer employing modulo arithmetic. *Lett.* **7**(5), 138–139 (1971)
3. H. Harasima, H. Miyakawa, Matched-Transmission Technique for Channels with Intersymbol Interference. *IEEE Trans. Commun.* **COM-20**, 774–780 (1972)
4. R. F. H. Fischer, C. Windpassinger, A. Lampe, J. B. Huber, Tomlinson-Harashima precoding in space-time transmission for low-rate backward channel. *Int. Zurich Semin. Broadband Commun. Access Transm. Netw.* 1–6 (2002)
5. C. Windpassinger, T. Vencel, R. F. H. Fischer. Precoding and loading for BLAST-like systems. *IEEE Int. Conf. Commun. (ICC)*, 3061–3065 (2003)
6. F. S. Tseng, B. G. Sun, Sensitivity Analysis for RVQ-Based Tomlinson-Harashima Precoded MIMO Systems. *IEEE Trans. Veh. Technol.* **65**(2), 978–985 (2016)
7. R. F. H. Fischer, C. A. Windpassinger, M.I.M.O. Improved, precoding for decentralized receivers resembling concepts from lattice reduction. *IEEE Glob. Telecommun. Conf. (GLOBECOM)*, 1852–1856 (2003)

8. K. Kusume, M. Joham, W. Utschick, G. Bauch, Efficient Tomlinson-Harashima Precoding for Spatial Multiplexing on Flat MIMO Channel. *IEEE Int. Conf. Commun.* **3**, 2021–2025 (2005)
9. W. Yu, D. P. Varodayan, J. M. Cioffi, Trellis and convolutional precoding for transmitter-based interference presubtraction. *IEEE Trans. Commun.* **53**(7), 1220–1230 (2005)
10. R. Habendorf, G. Fettweis, On Ordering Optimization for MIMO Systems with Decentralized Receivers. *Vehicular Technology Conf (VTC)*, 1844–1848 (2006)
11. F. A. Dietrich, P. Breun, W. Utschick, Robust Tomlinson-Harashima precoding for the wireless broadcast channel. *IEEE Trans. Signal Proc.* **55**(2), 631–644 (2007)
12. J. Liu, W. A. Krzymien, A Novel Nonlinear Precoding Algorithm for the Downlink of Multiple Antenna Multi-User Systems. *Wirel. Pers. Commun.* 207–223 (2007)
13. K. Kusume, M. Joham, W. Utschick, G. Bauch, Cholesky factorization with symmetric permutation applied to detecting precoding spatially multiplexed data streams. *IEEE Trans. Signal Proc.* **55**(6), 3089–3103 (2007)
14. J. Kang, H. Ku, D. S. Kwon, C. Lee, Tomlinson-Harashima precoder with tilted constellation for reducing the transmission power. *IEEE Trans. Wirel. Comm.* **8**(7), 3658–3667 (2009)
15. C. Masouros, M. Sellathurai, T. Ratnarajah, Interference Optimization for Transmit Power Reduction in Tomlinson-Harashima Precoded MIMO Downlinks. *IEEE Trans. Sig. Process.* **60**(5), 2470–2481 (2012)
16. L. Sun, M. R. McKay, Tomlinson-Harashima Precoding for Multiuser MIMO Systems With Quantized CSI Feedback and User Scheduling. *IEEE Trans. Sig. Process.* **62**(16), 4077–4090 (2014)
17. E. Debels, A. Suls, M. Moeneclaey, in *IEEE Symposium on Commun. and Vehicular Techn. in the Benelux (SCVT)*. Modulo loss reduction in spatial multiplexing systems with Tomlinson-Harashima precoding (IEEE, Luxembourg, 2015)
18. E. Debels, A. Suls, M. Moeneclaey, Modulo loss reduction for Tomlinson-Harashima precoding in a multi-beam satellite forward link. *IEEE Int. Work. Sign. Process. Adv. Wirel. Commun. (SPAWC)*, 5 (2016)
19. K. Zu, R. C. de Lamare, M. Haardt, Multi-Branch Tomlinson-Harashima Precoding Design for MU-MIMO Systems: Theory and Algorithms. *IEEE Trans. Commun.* **62**(3), 939–951 (2014)
20. X. Geng, B. An, F. Liu, F. Cao, Robust THP Transceiver Design for MIMO Interference Channel. *IEEE Commun. Lett.* **19**(9), 1640–1643 (2015)
21. L. Sun, J. Wang, V. Leung, Joint Transceiver, Data Streams, and User Ordering Optimization for Nonlinear Multiuser MIMO Systems. *IEEE Trans. Commun.* **63**(10), 3686–3701 (2015)
22. X. Lu, R. C. de Lamare, K. Zu, Successive optimization Tomlinson-Harashima precoding strategies for physical-layer security in wireless networks. *J. Wirel. Com. Netw.* **259**(2016), 1–12 (2016)
23. Q. Li, Q. Zhang, J. Qin, Robust Tomlinson-Harashima Precoding With Gaussian Uncertainties for SWIPT in MIMO Broadcast Channels. *IEEE Trans. Signal Process.* **65**(6), 1399–1411 (2017)
24. P. W. Wolniansky, G. J. Foschini, G. D. Golden, R. A. Valenzuela, V-BLAST: an architecture for realizing very high data rates over the rich-scattering wireless channel. *IEEE Int. Symp. Signals Syst. Electron. (ISSE)*, 295–300(1998)
25. D. Wubben, J. Rinas, R. Bohnke, V. Kuhn, K. D. Kammeyer, in *4th International ITG Conference on Source and Channel Coding*. Efficient Algorithm for Detecting Layered Space-Time Codes (IEEE, Berlin, 2002)
26. M. Diaz, N. Courville, C. Mosquera, G. Liva, G. Corazza, in *IEEE Intern. Workshop on Satellite and Space Commun. (IWSSC)*. Non-Linear Interference Mitigation for Broadband Multimedia Satellite Systems (IEEE, Salzburg, 2007), pp. 61–65
27. M. Poggioni, M. Berioli, P. Banelli, in *IEEE International Conf. on Commun.* BER performance of Multibeam Satellite Systems with Tomlinson-Harashima Precoding (IEEE, Dresden, 2009), pp. 1–6
28. P. Arapoglou, K. Liolios, M. Bertinelli, A. Panagopoulos, P. Cottis, R. De Gaudenzi, MIMO over satellite: A Review. *IEEE Commun. Surv. Tutorials.* **13**(1), 27–50 (2011)
29. F. Lombardo, E. A. Candreva, I. Thibault, A. Vanelli-Coralli, G. E. Corazza, in *2011 Conference Record of the Forty-Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. Multi-user interference mitigation techniques for broadband multi-beam satellite systems (IEEE, Pacific Grove, 2011), pp. 1805–1809
30. J. Arnau, B. Devillers, C. Mosquera, A. Perez-Neira, *Performance study of multiuser interference mitigation schemes for hybrid broadband multibeam satellite architectures*. (Springer International Publishing, 2012). 05 April 2012, <https://doi.org/10.1186/1687-1499-2012-132>
31. G. Zheng, S. Chatzinotas, B. Ottersten, Generic Optimization of Linear Precoding in Multibeam Satellite Systems. *IEEE Trans. Wirel. Commun.* **11**(6), 2308–2320 (2012)
32. V. Boussemart, L. Marini, M. Berioli, in *6th Advanced Sat. Multimedia Syst. Conf. (ASMS) and 12th Signal Proc. for Space Commun. Workshop (SPSC)*. Multi-Beam Satellite MIMO Systems: BER Analysis of Interference Cancellation and Scheduling (IEEE, Baiona, 2012), pp. 197–204
33. J. Arnau, C. Mosquera, in *IEEE 6th Advanced Sat. Multimedia Systems Conf. (ASMS) and 12th Signal Proc. for Space Comm. Workshop (SPSC)*. Performance Analysis of Multiuser Detection for Multibeam Satellites Under Rain Fading (IEEE, Baiona, 2012), pp. 212–219
34. G. Maral, M. Bosquet, *Satellite communications systems*, 4th ed. (Wiley, 2002). ISBN 0471496545, 9780471496540
35. J. Mazo, J. Salz, On the Transmitted Power In Generalized Partial Response. *IEEE Trans. Commun.* **24**(3), 348–352 (1976)
36. T. M. Cover, J. A. Thomas, *Elements of information theory*. (Wiley-Interscience, New York, 1991)

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](http://springeropen.com)