Optimized Precoded Spatio-Temporal Partial-Response Signaling over Frequency-Selective MIMO Channels

Jelle Bailleul, *Student Member, IEEE*, Lennert Jacobs, *Member, IEEE*, Mamoun Guenach, *Senior Member, IEEE*, Marc Moeneclaey, *Fellow, IEEE*

***Abstract*—Due to the continuous demand for higher bit rates, the management of the spatio-temporal intersymbol interference in frequency-selective multiple-input multiple-output (MIMO) channels becomes increasingly important. For single-input single-output channels, equalized precoded partial-response signaling is capable of handling a large amount of intersymbol interference, but, to date, no equalization scheme with general partial-response signaling has been presented for the frequency-selective MIMO channel. Not only does this contribution extend partial-response signaling to the MIMO channel by proposing a general *spatio-temporal* partial-response precoder, but it also develops a minimum mean-squared-error optimization framework in which the equalization coefficients and the spatio-temporal target response are jointly optimized. Three iterative optimization algorithms are discussed, which update (part of) a row of the target impulse response matrix in each iteration. In particular, the third algorithm reformulates this row optimization as a lattice decoding problem. Numerical simulations confirm that the general partial-response signaling clearly outperforms the traditional full-response signaling in terms of the mean squared error and the bit error rate. The third optimization algorithm has a better performance but a higher complexity, compared to the first and the second algorithm.**

***Index Terms*—decision-feedback equalization, lattice decoding, MIMO frequency-selective channel, minimum mean-squared-error equalization, partial-response signaling, precoding.**

I. Introduction

BECAUSE of the continuous growth in bit rate and associated bandwidth, the spatio-temporal (ST) intersymbol interference (ISI) caused by frequency-selective (FS) multiple-input multiple-output (MIMO) channels becomes an increasingly prominent challenge in the design of high-speed communication systems. To mitigate this ST ISI, several equalization schemes have been proposed in the literature. The optimal detector for uncoded transmission performs maximum likelihood sequence detection (MLSD), which involves a bank of matched filters and the application of the Viterbi algorithm [1], [2]. Unfortunately, its exponential complexity in both the number of data streams and the channel memory is intolerably large for most applications, therefore limiting the practicability of this detector. A more practically feasible alternative is the MIMO linear equalization scheme, in which linear equalization filters first reduce the ST ISI, and a symbol-by-symbol detector then recovers the transmitted data [3]. In contrast to the MLSD, this equalization scheme requires only limited complexity, but suffers from a (sometimes considerable) performance degradation.

Besides the linear equalizers, several nonlinear equalization schemes for continuous transmission over FS MIMO channels have been investigated, that combine adequate performance with low complexity. For example, a decision-feedback equalizer (DFE) [3], [4] with the previously detected symbols at the receiver (RX) as input can completely eliminate the postcursor ISI provided that the previous symbols were correctly detected. However, its performance may be severely degraded by the propagation of erroneous decisions. Error propagation can be overcome by transferring the feedback structure from the RX to the transmitter (TX), resulting in an equalization scheme called Tomlinson-Harashima precoding (THP). The THP scheme has originally been introduced in the context of FS SISO channels [5], [6], and has also been described for flat and FS MIMO channels in [7] and [8], respectively. However, THP requires a modulo operator at both the TX and the RX, inducing both a power and a modulo loss.

The equalization performance can also be enhanced by partial-response signaling (PRS) [9], in which the channel is equalized to a target impulse response (TIR) with integer coefficients. In this TIR, a controlled amount of ISI is allowed, thereby considerably facilitating the equalization task. At the RX, the data detection is executed based on the equalized channel output in the extended signal set, after which the original data is retrieved from these decision by means of a feedback filter at the RX. However, this feedback structure is prone to error propagation. Alternatively, the feedback filter can be omitted when employing a precoder at the TX. Contrary to THP, this precoder does interestingly not induce any power loss. A common example of PRS is duo-binary signaling, where the TIR in z-transform notation is $1 + z^{-1}$ [10], [11].

Interestingly, a relation between PRS and lattice-reduction-aided (LRA) equalization has been established in [12], [13], [14]. LRA techniques have been applied to flat fading MIMO channels with only spatial ISI [15], [16], [17], [18]. More precisely, a lattice reduction algorithm, e.g., the LLL algorithm [19] or the element-based reduction algorithm [18], first computes a reduced channel matrix that is better conditioned than

J. Bailleul, L. Jacobs, and M.Moeneclaey are with the Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium (email: jelle.bailleul, lennert.jacobs,marc.moeneclaey@ugent.be).

M.Guenach is both with Imec, Leuven, Belgium and the Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium (email: guenach@ieee.org). Part of the research work was conducted at the Nokia Bell Labs, Antwerp.

the original channel matrix, i.e., closer to being orthogonal. Next, the equalizer with respect to this reduced channel is obtained by either the zero-forcing (ZF) or the minimum mean-squared-error (MMSE) criterion. After decision, the data is then recovered by means of the inverse base transformation. This procedure has been applied to block transmission over FS channels for single-input single-output (SISO) [20] and MIMO [21] systems and to mitigate the spatial ISI with low complexity in MIMO orthogonal frequency division multiplexing systems [22]. Additionally, [16] and [23] have proved that optimal diversity is reached by the LRA equalizers for a flat fading MIMO channel and a FS MIMO channel, respectively. However, continuous transmission, e.g., PRS signaling, might be preferred over block transmission to avoid guard times and to allow for time-invariant equalization filters. In [12], generalized PRS for the flat fading MIMO channel has been derived from the LRA equalizer by implementing the inverse base transformation at the TX instead of the RX. The similarity between PRS and LRA has been emphasized for a FS SISO channel in [13], [14], in which the channel's transfer function is decomposed as the product of a target response and a reduced channel. The ZF infinite-length equalizer is then designed for completely equalizing the latter and the target response inducing the least noise enhancement is considered as optimal. This optimization is done by means of a look-up table in [13], but a more advanced algorithm has been proposed in [14] by reformulating this problem as a lattice decoding problem [24].

Although PRS has already been described in the literature for the FS SISO channel [9] and for the flat-fading MIMO channel [12], no general ST PRS equalization scheme has been investigated for the FS MIMO channel. Indeed, the TIR in [12] and [10] is confined to spatial-only components and temporal-only components, respectively. Moreover, most contributions assume that the TIR is given [9], [10], [11], and no optimization with respect to this TIR is performed at all, resulting in an inferior trial-and-error selection of the target response.

This contribution focuses on PRS applied to a symbol-spaced, complex-valued, discrete-time, and FS MIMO channel. To address the problem of the ISI in the continuous transmission over the FS channel, a general PRS architecture is presented, consisting of a ST PRS precoder at the TX and a ST-DFE (encompassing a forward and a decision-feedback filter) at the RX. Interestingly, the TIR of the PRS contains both spatial and temporal components that are complex-valued as the baseband channel can be complex-valued. Contrary to the trial-and-error TIR selection often found in literature [9], [10], [11], the ST-TIR is jointly optimized with the ST-DFE according to the MMSE criterion. First, the optimal equalization coefficients can readily be expressed as a function of the TIR. The resulting MSE is then minimized with respect to the TIR by three different algorithms. Interestingly, the third algorithm reformulates the TIR optimization as a sequence of lattice decoding problems, which can be solved by the algorithm presented in [24]. In [14], a similar algorithm is proposed to compute a real-valued TIR in the context of a FS SISO channel with a whitened-matched filter front-end.

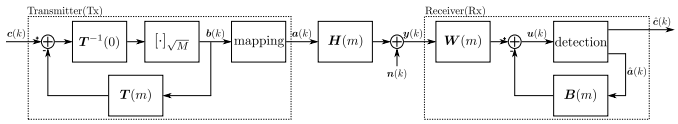


Figure 1. Model of a PRS-ST communication link. The TX consists of the partial-response precoder, whereas the RX aims to reduce the ISI by means of a linear feedforward and linear feedback filter.

However, the present contribution significantly differs from [14], since the real-valued TIR in [14] has been derived by minimizing a lower bound on the noise enhancement of the infinite-length ZF equalizer for a SISO channel, whereas this contribution jointly optimizes the complex-valued TIR coefficients and more practical, finite-length filters according to the MMSE-criterion for a MIMO channel.

The remainder of the paper is organized as follows. Section II mathematically describes the communication link, consisting of a PRS precoder at the TX, a FS MIMO channel, and a nonlinear ST-DFE at the RX, the latter comprising a linear feedforward filter and a linear feedback filter operating on the previously detected symbols. The filters and the ST-TIR are calculated by minimizing the MSE in section III. Following the simple optimization with respect to the filters of the ST-DFE in subsection III-B, the computation of the optimized TIR is more challenging (subsection III-B). In total, three iterative optimization algorithms, each performing a row-by-row optimization of the TIR, are discussed (subsections III-B1-III-B4). To assess the performance of these algorithms, the bit error rate (BER) expression is derived in Section IV, after which numerical results are given in Section V. Finally, conclusions are drawn in Section VI.

In the following, upper-case (lower-case) boldface letters are used for matrices (vectors). The notations $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $\mathbb{E}[\cdot]$, $\|\cdot\|$, $\mathrm{Tr}(\cdot)$, $\mathrm{Re}[\cdot]$, $\mathrm{Im}[\cdot]$, $\det(\cdot)$, and $|\cdot|$ represent the complex conjugate, transpose, Hermitian transpose, the statistical expectation, the Euclidean vector norm, the trace, the real part, the imaginary part, the determinant, and the absolute value, respectively. Additionally, $\boldsymbol{I}$, $\mathbf{0}$, and $\mathbf{1}$ denote the identity matrix, the all-zero matrix, and the all-ones column. The elements of the single-entry matrix $\boldsymbol{J}_{l,q}$ are zero, except for the (l,q)th element which equals 1. The set of all integers ranging from a to b is written as $\{a,\ldots,b\}$, whereas $X \cup Y$, $X \cap Y$, and $X \backslash Y$ denote the union and intersection of the sets X and Y, and the complement of Y relative to X, respectively. Finally, an optimized variable is indicated by $\cdot^\star$.

II. System Model

Fig. 1 displays the system model of an equalized complex-valued MIMO $N_r \times N_t$ baseband channel with spatio-temporal PRS (PRS-ST). At the TX, the input consists of N_t continuous complex-valued data symbol streams, i.e., $\boldsymbol{c}(k) = [c_1(k), ..., c_l(k), ..., c_{N_t}(k)]^T$ with $c_l(k)$ denoting the kth symbol from the lth stream. All symbols are independently and uniformly drawn from the complex symbol set $C = \{0, 1, \ldots, \sqrt{M}-1\} + j\{0, 1, \ldots, \sqrt{M}-1\}$, where M is assumed to be an integer power of 4. The binary labels associated with the elements from C are according to the binary-reflected

Gray mapping. Next, the symbol vectors $\boldsymbol{c}(k)$ are applied to the MIMO partial-response precoder at the symbol rate $1/T_s$. This precoder is characterized by the target polynomial $\boldsymbol{T}_{\text{tar}}(D) = \sum_{m=0}^{L_\text{T}} \boldsymbol{T}(m)D^m$ with Gaussian integer $N_t \times N_t$ matrices $\boldsymbol{T}(m)$, $m = 0, ..., L_\text{T}$, i.e., $\boldsymbol{T}(m) \in \mathbb{Z}[j]^{N_t \times N_t}$ with $\mathbb{Z}[j] = \{a + jb \,|\, a, b \in \mathbb{Z}\}$. In contrast to previous work on PRS that focused mainly on SISO channels ($N_t = N_r = 1$), on strictly temporal partial response ($\boldsymbol{T}(m)$ is diagonal), or on strictly spatial partial response ($\boldsymbol{T}(m) = \boldsymbol{0}_{N_t}$ for $m > 0$), the more general complex-valued MIMO precoder presented here allows ST components (with spatial-only and temporal-only components as particular cases). The input sequence $\{\boldsymbol{c}(k)\}$ is converted by the precoder into the precoded output sequence $\{\boldsymbol{b}(k)\}$ according to

$$\boldsymbol{b}(k) = \left[\boldsymbol{T}^{-1}(0)\left(\boldsymbol{c}(k) - \sum_{m=1}^{L_\text{T}} \boldsymbol{T}(m)\boldsymbol{b}(k-m)\right)\right]_{\sqrt{M}}, \quad (1)$$

where $[.]_X$ symbolizes the element-wise modulo-X reduction to the interval $[0, X)$ of both the real and the imaginary part. To avoid any power loss, PRS requires that all entries of the precoded sequence $\boldsymbol{b}(k)$, just as the elements of $\boldsymbol{c}(k)$, are independently and uniformly drawn from the symbol set C. This property is acquired when all entries of $\boldsymbol{T}(m)$ are Gaussian integers and $\boldsymbol{T}(0)$ is a (complex-valued) unimodular matrix (Appendix A), since one can easily prove that $\boldsymbol{T}^{-1}(0) \in \mathbb{Z}[j]^{N_t \times N_t}$ if and only if the determinant of the Gaussian integer matrix $\boldsymbol{T}(0)$ is restricted to the set $\{1, -1, j, -j\}$ ([25] for real-valued matrix).

The components from $\boldsymbol{b}(k)$ are mapped to the M-QAM constellation, i.e.,

$$\boldsymbol{a}(k) = 2\Delta\boldsymbol{b}(k) + (1 + j)\Delta(-\sqrt{M} + 1)\boldsymbol{1}, \quad (2)$$

where $\Delta = \sqrt{\frac{3}{2(M-1)}}$ such that the symbol energy is normalized, i.e., $\mathbb{E}[\boldsymbol{a}(k)\boldsymbol{a}^H(k)] = \boldsymbol{I}_{N_t}$. The sequence $\boldsymbol{a}(k)$ is then transmitted over the discrete-time FS $N_r \times N_t$ MIMO channel which encompasses the complex-valued impulse response matrices $\boldsymbol{H}(m)$ and the zero-mean circular symmetric complex Gaussian noise $\boldsymbol{n}(k)$. The channel input-output relationship is given by

$$\boldsymbol{y}(k) = \sum_{m=-L_\text{H}^{(1)}}^{L_\text{H}^{(2)}} \boldsymbol{H}(m)\boldsymbol{a}(k-m) + \boldsymbol{n}(k), \quad (3)$$

in which the channel impulse response is assumed to be finite, i.e., $\boldsymbol{H}(m) = \boldsymbol{0}_{N_r \times N_t}$ $\forall m \notin \{-L_\text{H}^{(1)}, ..., L_\text{H}^{(2)}\}$. The RX first applies $\boldsymbol{y}(k)$ to the linear MIMO feedforward equalization filter characterized by the $N_t \times N_r$ impulse response matrices $\boldsymbol{W}(m)$. Next, the output of the linear MIMO feedback filter with impulse response matrices $\boldsymbol{B}(m)$, operating on the previously detected $\hat{\boldsymbol{a}}(k)$, is subtracted from the feedforward filter output. When the decisions $\hat{\boldsymbol{a}}(k-m)$ are correct, the decision variable $\boldsymbol{u}(k)$ can be formulated as

$$\boldsymbol{u}(k) = \sum_{m_2=-L_\text{W}^{(1)}}^{L_\text{W}^{(2)}} \sum_{m_1=-L_\text{H}^{(1)}}^{L_\text{H}^{(2)}} \boldsymbol{W}(m_2)\boldsymbol{H}(m_1)\boldsymbol{a}(k-m_1-m_2) + \sum_{m=-L_\text{W}^{(1)}}^{L_\text{W}^{(2)}} \boldsymbol{W}(m)\boldsymbol{n}(k-m) - \sum_{m\in\Phi_\text{B}} \boldsymbol{B}(m)\boldsymbol{a}(k-m), \quad (4)$$

where both filters are assumed to be finite, i.e., $\boldsymbol{W}(m) = \boldsymbol{0}_{N_t \times N_r}$ $\forall m \notin \{-L_\text{W}^{(1)}, \ldots, L_\text{W}^{(2)}\}$ and $\boldsymbol{B}(m) = \boldsymbol{0}_{N_t}$ $\forall m \notin \Phi_\text{B}$ with $\Phi_\text{B} = \{\Phi_\text{B}(1), ..., \Phi_\text{B}(L_\text{B})\}$. The L_B active time instants do not necessarily coincide with the time-delay indices $1, \ldots, L_\text{B}$, but the only restriction, imposed by causality, is $\Phi_\text{B}(i) \geq 1$ for $i = 1, ..., L_\text{B}$. According to (1), the target response vector

$$\boldsymbol{u}_\text{T}(k) = \sum_{m=0}^{L_\text{T}} \boldsymbol{T}(m)\boldsymbol{a}(k-m) \quad (5)$$

and the original sequence $\boldsymbol{c}(k)$ are related by

$$\boldsymbol{c}(k) = \left[\frac{\boldsymbol{u}_\text{T}(k)}{2\Delta} + \boldsymbol{c}_{\boldsymbol{v}}\right]_{\sqrt{M}}, \quad (6)$$

where

$$\boldsymbol{c}_{\boldsymbol{v}} = \frac{(1 + j)(\sqrt{M} - 1)}{2} \sum_{m=0}^{L_\text{T}} \boldsymbol{T}(m)\boldsymbol{1}. \quad (7)$$

The equalization coefficients are therefore selected such that the decision variable $\boldsymbol{u}(k)$ from (4) approaches the target response vector $\boldsymbol{u}_\text{T}(k)$ as close as possible. Consequently, the decision on $\boldsymbol{c}(k)$, $\hat{\boldsymbol{c}}(k)$, follows from replacing the quantity $\frac{\boldsymbol{u}_\text{T}(k)}{2\Delta}$ in the right-hand side of (6) by the Gaussian integer which is nearest to $\frac{\boldsymbol{u}(k)}{2\Delta}$, after which the modulo operator maps this decision to the original M-QAM constellation. The input of the feedback equalizers, i.e., the decision on $\boldsymbol{a}(k)$, $\hat{\boldsymbol{a}}(k)$, is constructed by substituting $\boldsymbol{c}(k)$ for $\hat{\boldsymbol{c}}(k)$ in the precoder equation (1), yielding the decision on the vectors $\boldsymbol{b}(k)$, $\hat{\boldsymbol{b}}(k)$, which are then mapped to M-QAM symbols.

To simplify the notations in (4), the following matrices are introduced:

$$\boldsymbol{W} = \left[\begin{array}{ccc} \boldsymbol{W}(-L_\text{W}^{(1)}) & \cdots & \boldsymbol{W}(L_\text{W}^{(2)}) \end{array}\right], \quad (8)$$

$$\boldsymbol{B} = \left[\begin{array}{ccc} \boldsymbol{B}(\Phi_\text{B}(1)) & \cdots & \boldsymbol{B}(\Phi_\text{B}(L_\text{B})) \end{array}\right], \quad (9)$$

$$\boldsymbol{G}(m) = \left[\begin{array}{ccc} \boldsymbol{H}^H(m + L_\text{W}^{(1)}) & \cdots & \boldsymbol{H}^H(m - L_\text{W}^{(2)}) \end{array}\right]^H, \quad (10)$$

and

$$\bar{\boldsymbol{n}}(k) = \left[\begin{array}{ccc} \boldsymbol{n}^H(k + L_\text{W}^{(1)}) & \cdots & \boldsymbol{n}^H(k - L_\text{W}^{(2)}) \end{array}\right]^H. \quad (11)$$

Based on these shorthand notations and $\Phi_\text{G} = \{-L_\text{H}^{(1)} - L_\text{W}^{(1)}, \ldots, L_\text{H}^{(2)} + L_\text{W}^{(2)}\}$, the decision variable $\boldsymbol{u}(k)$ from (4) can be rewritten as

$$\boldsymbol{u}(k) = \sum_{m\in\Phi_\text{G}} \boldsymbol{W}\boldsymbol{G}(m)\boldsymbol{a}(k-m) + \boldsymbol{W}\bar{\boldsymbol{n}}(k) - \sum_{m\in\Phi_\text{B}} \boldsymbol{B}(m)\boldsymbol{a}(k-m), \quad (12)$$

III. MMSE EQUALIZATION

In this contribution, the MMSE criterion is employed to jointly optimize the feedforward filter $\boldsymbol{W}$, the feedback filter $\boldsymbol{B}$, and the TIR $\boldsymbol{T} = \left[\begin{array}{ccc} \boldsymbol{T}(0) & \cdots & \boldsymbol{T}(L_\text{T}) \end{array}\right]$. With $\boldsymbol{e}(k) = \boldsymbol{u}(k) - \boldsymbol{u}_\text{T}(k)$ denoting the difference between the

decision variable $\boldsymbol{u}(k)$ and the target response vector $\boldsymbol{u}_{\mathrm{T}}(k)$, the corresponding normalized MSE is defined as

$$\mathrm{MSE} \triangleq \frac{\mathbb{E}\left[||\boldsymbol{e}(k)||^2\right]}{\mathbb{E}\left[\|\boldsymbol{a}(k)\|^2\right]} = \frac{\mathrm{Tr}\left(\boldsymbol{R}_e\right)}{\mathbb{E}\left[\|\boldsymbol{a}(k)\|^2\right]} = \frac{1}{N_t}\sum_{l=1}^{N_t}\mathrm{MSE}^{(l)}, \quad (13)$$

where the expectation is taken over the data symbols and the noise, and the error covariance matrix $\boldsymbol{R}_e$ is given by $\boldsymbol{R}_e \triangleq \mathbb{E}\left[\boldsymbol{e}(k)\boldsymbol{e}^H(k)\right]$. More precisely, the lth diagonal element of $\boldsymbol{R}_e$, denoted by $\mathrm{MSE}^{(l)}$, equals the MSE corresponding to the lth data stream. The MSE from (13) equals the normalized sum of all $\mathrm{MSE}^{(l)}$.

First, a simplified notation is developed. The vector $\boldsymbol{y}_W$ contains all channel output samples contributing to the decision variable $\boldsymbol{u}(k)$. Vectors $\boldsymbol{a}_{\mathrm{T}}$ and $\boldsymbol{a}_{\mathrm{B}}$ comprise all symbols contributing to $\boldsymbol{u}_{\mathrm{T}}(k)$ and to the input of the feedback filter, respectively. All remaining data symbols are collected in the vector $\boldsymbol{a}_{\mathrm{N}}$, i.e.,

$$\boldsymbol{y}_W = \left[\begin{array}{ccc} \boldsymbol{y}^H(k+L_{\mathrm{W}}^{(1)}) & \cdots & \boldsymbol{y}^H(k-L_{\mathrm{W}}^{(2)}) \end{array}\right]^H, \quad (14)$$

$$\boldsymbol{a}_{\mathrm{B}} = \left[\begin{array}{ccc} \boldsymbol{a}^H(k-\Phi_{\mathrm{B}}(1)) & \cdots & \boldsymbol{a}^H(k-\Phi_{\mathrm{B}}(L_{\mathrm{B}})) \end{array}\right]^H, \quad (15)$$

$$\boldsymbol{a}_{\mathrm{T}} = \left[\begin{array}{ccc} \boldsymbol{a}^H(k) & \cdots & \boldsymbol{a}^H(k-L_{\mathrm{T}}) \end{array}\right]^H, \quad (16)$$

$$\boldsymbol{a}_{\mathrm{N}} = \left[\begin{array}{ccc} \cdots & \boldsymbol{a}^H(k-m) & \cdots \end{array}\right]^H \ \forall m \in \Phi_{\mathrm{G}}\backslash\{\Phi_{\mathrm{T}}\cup\Phi_{\mathrm{B}}\}. \quad (17)$$

Since all symbols in $\boldsymbol{a}(k)$ are spatially and temporally uncorrelated, all cross correlations between $\boldsymbol{a}_{\mathrm{T}}$, $\boldsymbol{a}_{\mathrm{B}}$, $\boldsymbol{a}_{\mathrm{N}}$ are zero and the autocorrelation matrix of these vectors is equal to the identity matrix, e.g., $\mathbb{E}[\boldsymbol{a}_{\mathrm{T}}\boldsymbol{a}_{\mathrm{T}}^H] = \boldsymbol{I}$. In this contribution, the intersection of $\Phi_{\mathrm{T}} = \{0,...,L_{\mathrm{T}}\}$ and Φ_{B} is assumed to be empty by design. Otherwise, the target response matrices $\boldsymbol{T}(m)$ $\forall m \in \Phi_{\mathrm{T}} \cap \Phi_{\mathrm{B}}$ would not influence the communication link's quality, as the feedback equalizer is capable of removing the contribution from $\boldsymbol{a}(k-m)$ to $\boldsymbol{e}(k)$, for any $m \in \Phi_{\mathrm{B}}$. Based on (3) and the notations in (14)-(17), the input-output relationship between $\boldsymbol{y}_W$ and the data symbols can be rewritten as

$$\boldsymbol{y}_W = \boldsymbol{G}_T\boldsymbol{a}_{\mathrm{T}} + \boldsymbol{G}_B\boldsymbol{a}_{\mathrm{B}} + \boldsymbol{G}_N\boldsymbol{a}_{\mathrm{N}} + \bar{\boldsymbol{n}}(k), \quad (18)$$

where

$$\boldsymbol{G}_T = \left[\begin{array}{ccc} \boldsymbol{G}(0) & \cdots & \boldsymbol{G}(L_{\mathrm{T}}) \end{array}\right], \quad (19)$$

$$\boldsymbol{G}_B = \left[\begin{array}{ccc} \boldsymbol{G}(\Phi_{\mathrm{B}}(1)) & \cdots & \boldsymbol{G}(\Phi_{\mathrm{B}}(L_{\mathrm{B}})) \end{array}\right], \quad (20)$$

$$\boldsymbol{G}_N = \left[\begin{array}{ccc} \cdots & \boldsymbol{G}(m) & \cdots \end{array}\right] \ \forall m \in \Phi_{\mathrm{G}}\backslash\{\Phi_{\mathrm{T}}\cup\Phi_{\mathrm{B}}\}. \quad (21)$$

Consequently, the MSE defined in (13) can be expressed as

$$\mathrm{MSE} = \frac{1}{N_t}\mathbb{E}\Big[\Big\|\boldsymbol{W}\left(\boldsymbol{G}_T\boldsymbol{a}_{\mathrm{T}} + \boldsymbol{G}_B\boldsymbol{a}_{\mathrm{B}} + \boldsymbol{G}_N\boldsymbol{a}_{\mathrm{N}} + \bar{\boldsymbol{n}}(k)\right) - \boldsymbol{B}\boldsymbol{a}_{\mathrm{B}} - \boldsymbol{T}\boldsymbol{a}_{\mathrm{T}}\Big\|^2\Big]. \quad (22)$$

In the subsections below, the optimal $\boldsymbol{W}$ and $\boldsymbol{B}$ are first formulated for a given $\boldsymbol{T}$ in subsection III-A, after which three algorithms are presented in subsection III-B to optimize the matrix $\boldsymbol{T}$.

A. Optimization over $\boldsymbol{W}$ and $\boldsymbol{B}$

The optimal coefficients $\boldsymbol{W}^\star$ and $\boldsymbol{B}^\star$, for given $\boldsymbol{T}$, are straightforwardly obtained by equating the derivatives of the MSE in (22) with respect to $\boldsymbol{W}$ and $\boldsymbol{B}$ to zero, yielding [4], [10]

$$\boldsymbol{W}^\star = \boldsymbol{T}\boldsymbol{G}_T^H\left(\boldsymbol{G}_T\boldsymbol{G}_T^H + \boldsymbol{G}_N\boldsymbol{G}_N^H + \boldsymbol{R}_{\mathrm{n}}\right)^{-1} \quad (23)$$

$$\boldsymbol{B}^\star = \boldsymbol{W}^\star\boldsymbol{G}_B \quad (24)$$

where $\boldsymbol{R}_{\mathrm{n}} \triangleq \mathbb{E}\left[\bar{\boldsymbol{n}}(k)\bar{\boldsymbol{n}}^H(k)\right]$. The $\mathrm{MSE}_{\boldsymbol{W}^\star,\boldsymbol{B}^\star}$ is then given by plugging (23) and (24) into (22), yielding

$$\mathrm{MSE}_{\boldsymbol{W}^\star,\boldsymbol{B}^\star} = \frac{1}{N_t}\mathrm{Tr}\left[\boldsymbol{T}\boldsymbol{G}\boldsymbol{T}^H\right], \quad (25)$$

where $\boldsymbol{G} = \boldsymbol{I}_{N_t(L_{\mathrm{T}}+1)} - \boldsymbol{G}_T^H\left(\boldsymbol{G}_T\boldsymbol{G}_T^H + \boldsymbol{G}_N\boldsymbol{G}_N^H + \boldsymbol{R}_{\mathrm{n}}\right)^{-1}\boldsymbol{G}_T$. For full-response signaling (FRS), i.e., $\boldsymbol{T} = \left[\boldsymbol{I}_{N_t} \quad \boldsymbol{0}_{N_t\times L_{\mathrm{T}}N_t}\right]$, the expressions in (23) and (24) are verified to simplify to the standard expressions for the MMSE equalizer [3], [4].

B. Optimization over $\boldsymbol{T}$

Instead of considering a channel-independent TIR $\boldsymbol{T}$ such as FRS or duo-binary, this contribution optimizes $\boldsymbol{T}$ by minimizing (25) subjected to two constraints: (i) all entries of $\boldsymbol{T}$ must belong to $\mathbb{Z}[j]$, and (ii) $\det(\boldsymbol{T}(0))$ must belong to $\{-1,1,j,-j\}$. TIRs satisfying these constraints are called feasible, and ensure that $\boldsymbol{b}(k)$ and $\boldsymbol{c}(k)$ possess identical statistics (see Appendix A).

Equation (25) reveals that the lth row of $\boldsymbol{T}$ only influences $\mathrm{MSE}^{(l)} = \left(\boldsymbol{T}\boldsymbol{G}\boldsymbol{T}^H\right)_{l,l}$, the MSE of the lth data stream. Hence, when the constraint on $\det(\boldsymbol{T}(0))$ is satisfied, constructing a new target impulse response matrix $\boldsymbol{T}'$ by multiplying one row of $\boldsymbol{T}$ with the complex conjugate of $\det(\boldsymbol{T}(0))$ results in $\det(\boldsymbol{T}'(0)) = 1$, while the MSE from (25) remains unaltered. Therefore, without loss of generality, the second constraint will be replaced by $\det(\boldsymbol{T}(0)) = 1$ in the sequel.

Although the lth row of $\boldsymbol{T}$ affects only $\mathrm{MSE}^{(l)}$, the rows of $\boldsymbol{T}$ cannot be chosen independently from each other due to the constraint $\det(\boldsymbol{T}(0)) = 1$, impeding the optimization of each row individually. This subsection instead discusses three iterative algorithms to compute an optimized TIR $\boldsymbol{T}^\star$. In the ith iteration of each algorithm, the TIR is incremented as follows:

$$\boldsymbol{T}_{i+1} = \boldsymbol{T}_i + \boldsymbol{T}_{\mathrm{inc}}, \quad (26)$$

where the increment matrix $\boldsymbol{T}_{\mathrm{inc}}$ is a function of $\boldsymbol{T}_i$. The corresponding MSE_{i+1} is given by

$$\mathrm{MSE}_{i+1} = \frac{1}{N_t}\mathrm{Tr}\left(\boldsymbol{T}_{i+1}\boldsymbol{G}\boldsymbol{T}_{i+1}^H\right) \quad (27)$$

$$= \mathrm{MSE}_i + \frac{1}{N_t}\mathrm{Tr}\left(\boldsymbol{T}_i\boldsymbol{G}\boldsymbol{T}_{\mathrm{inc}}^H + \boldsymbol{T}_{\mathrm{inc}}\boldsymbol{G}\boldsymbol{T}_i^H + \boldsymbol{T}_{\mathrm{inc}}\boldsymbol{G}\boldsymbol{T}_{\mathrm{inc}}^H\right). \quad (28)$$

In the ith iteration, $\boldsymbol{T}_{\mathrm{inc}}$ is restricted to consist of only one nonzero row, namely the row with index $l_i^\star$. Hence, the ith iteration focuses entirely on the reduction of $\mathrm{MSE}_i^{(l_i^\star)}$, which is chosen to be the largest among the *reducible* $\mathrm{MSE}_i^{(l)}$: $\mathrm{MSE}_i^{(l_i^\star)}$ is defined as the largest diagonal element of the error

covariance matrix that can be reduced by an increment $\boldsymbol{T}_{\text{inc}}$ computed by a specific algorithm. More specifically, when the sequence $\left(\text{MSE}_i^{(m_1)},...,\text{MSE}_i^{(m_{N_t})}\right)$ is sorted from large to small, i.e., $\text{MSE}_i^{(m_n)} \geq \text{MSE}_i^{(m_{n+1})}$ for $n = 1,...,N_t-1$, then $l_i^\star = m_n$ if and only if $\text{MSE}_i^{(l)}$ is not reducible for $l = m_1,...,m_{n-1}$ and $\text{MSE}_i^{(m_n)}$ is reducible. Further, to guarantee the feasibility of $\boldsymbol{T}_{i+1}$, each entry of $\boldsymbol{T}_{\text{inc}}$ must be Gaussian integer, i.e., $(\boldsymbol{T}_{\text{inc}})_{l,q} \in \mathbb{Z}[j]$ and, when updating a row of $\boldsymbol{T}_i(0)$, the first N_t elements of the increment must additionally be a linear combination of the other rows of $\boldsymbol{T}_i(0)$ such that $\det(\boldsymbol{T}(0))$ does not alter. Initializing $\boldsymbol{T}_0$ as any feasible TIR results then obviously in an optimized $\boldsymbol{T}^\star$, since convergence to a (local) optimum is guaranteed as the MSE is lowered in each iteration and bounded below by 0. Unfortunately, this optimization method does not guarantee convergence to the global optimum, making the initialization and the computation of $\boldsymbol{T}_{\text{inc}}$ crucial for the quality of $\boldsymbol{T}^\star$. Initializing $\boldsymbol{T}_0$ as FRS is the most logical choice because the resulting MSE is then upper bounded by the MSE of FRS.

The first algorithm A1 updates the element of the $l_i^\star$th row of $\boldsymbol{T}_i$ inducing the largest decrease in $\text{MSE}_i^{(l_i^\star)}$, yielding a fast and low-complexity algorithm. The second algorithm A2 is similar to A1, but updates the element of the $l_i^\star$th row of $\boldsymbol{T}_i$ yielding the largest *guaranteed* decrease, which is defined in subsection III-B2 and does not only depend on the current $\boldsymbol{T}_i$, but also on the potential of future TIR increments. Due to this extra information, convergence to a better (local) optimum is expected at the cost of a slightly larger complexity. In the last algorithm A3, even better performance is expected as the entire $l_i^\star$-th row of $\boldsymbol{T}_i$ is updated by solving a relatively complex lattice decoding problem. Below, a detailed description of A1, A2, and A3 is included.

1) Algorithm 1 (A1): In algorithm A1, $\boldsymbol{T}_{\text{inc}}$ is restricted to the set $\left\{\boldsymbol{T}_{\text{inc}}^{(l,q)},\ l \in \{1,\ldots,N_t\},\ q \in \{1,...,(L_{\text{T}}+1)N_t\}\right\}$. When $q > N_t$, $\boldsymbol{T}_{\text{inc}}^{(l,q)}$ increments the (l,q)th element of $\boldsymbol{T}_i$, i.e.,

$$\boldsymbol{T}_{\text{inc}}^{(l,q)} = \lambda_{l,q}\boldsymbol{J}_{l,q} \quad q > N_t. \tag{29}$$

When $q \leq N_t$, $\boldsymbol{T}_{\text{inc}}^{(l,q)}$ adds a multiple of row q of $\boldsymbol{T}_i(0)$ to row l of $\boldsymbol{T}_i(0)$, i.e.,

$$\boldsymbol{T}_{\text{inc}}^{(l,q)} = [\ \ \lambda_{l,q}\boldsymbol{J}_{l,q}\boldsymbol{T}_i(0) \quad \boldsymbol{0}\ \ ] \quad q \leq N_t,\ l \neq q. \tag{30}$$

In (29) and (30), $\lambda_{l,q}$ must belong to $\mathbb{Z}[j]$, and its optimum value, $\lambda_{l,q}^\star$, is determined by maximizing the decrease $\delta_{l,q} = \text{MSE}_i^{(l)} - \text{MSE}_{i+1}^{(l)}$, where MSE_{i+1} is obtained by plugging (29) and (30) into (28). For given $\lambda_{l,q}$, this decrease is verified to be

$$\delta_{l,q} = -2\text{Re}\left[\lambda_{l,q}A\right] - \left|\lambda_{l,q}\right|^2 B \tag{31}$$

with

$$A = \begin{cases} (\boldsymbol{G}\boldsymbol{T}_i^H)_{q,l} & q > N_t \\ \left(\boldsymbol{T}_i(0)\boldsymbol{G}_{N_t}\boldsymbol{T}_i^H\right)_{q,l} & q \leq N_t,\ l \neq q \end{cases} \tag{32}$$

and

$$B = \begin{cases} (\boldsymbol{G})_{q,q} & q > N_t \\ \left(\boldsymbol{T}_i(0)\boldsymbol{G}_{N_t,N_t}\boldsymbol{T}_i^H(0)\right)_{q,q} & q \leq N_t,\ l \neq q \end{cases}, \tag{33}$$

Table I
PSEUDOCODE OF ALGORITHM A1

1: $\boldsymbol{T}_0 = [\boldsymbol{I}_{N_t}\ \boldsymbol{0}]$, $i = 0$
2: **while** $\exists l : \text{MSE}_i^{(l)}$ reducible **do**
3: Select $l_i^\star$ as largest reducible $\text{MSE}_i^{(l)}$
4: $q_i^\star = \arg\max_q \delta_{l_i^\star,q}^\star$
5: Compute $\boldsymbol{T}_{i+1} = \boldsymbol{T}_i + \boldsymbol{T}_{\text{inc}}^{(l_i^\star,q_i^\star)}$ with $\boldsymbol{T}_{\text{inc}}^{(l_i^\star,q_i^\star)}$ given by (29)-(30)
6: $i \leftarrow i+1$
7: **end**
8: $\boldsymbol{T}^\star = \boldsymbol{T}_i$

where $\boldsymbol{G}_{N_t}$ and $\boldsymbol{G}_{N_t,N_t}$ are constructed by keeping the first N_t rows from $\boldsymbol{G}$ and the first N_t columns of $\boldsymbol{G}_{N_t}$, respectively. The optimal Gaussian integer $\lambda_{l,q}^\star$ that maximizes $\delta_{l,q}$ is given by

$$\lambda_{l,q}^\star = \left\lfloor \frac{-A^*}{B} \right\rceil, \tag{34}$$

where $\lfloor\cdot\rceil$ rounds both the real and the imaginary part to the nearest integer. For given (l,q), the largest decrease of $\text{MSE}_i^{(l)}$ corresponding to the increment $\boldsymbol{T}_{\text{inc}}^{(l,q)}$ with $\lambda_{l,q} = \lambda_{l,q}^\star$ is then given by inserting (34) into (31), which yields

$$\delta_{l,q}^\star = -2\text{Re}\left[\lambda_{l,q}^\star A\right] - \left|\lambda_{l,q}^\star\right|^2 B. \tag{35}$$

For given l, $\text{MSE}_i^{(l)}$ is called reducible when the largest reduction of $\text{MSE}_i^{(l)}$ resulting from an increment matrix $\boldsymbol{T}_{\text{up}}^{(l,q)}$, i.e., $\max_q \delta_{l,q}^\star$, is larger than 0. Otherwise, $\text{MSE}_i^{(l)}$ is irreducible. In the ith iteration, A1 determines the index $l_i^\star$ such that $\text{MSE}_i^{(l_i^\star)}$ is the largest of the reducible $\text{MSE}_i^{(l)}$. The algorithm then increments $\boldsymbol{T}_i$ with $\boldsymbol{T}_{\text{up}}^{(l_i^\star,q_i^\star)}$, where $q_i^\star = \arg\max_q \delta_{l_i^\star,q}$, and reduces $\text{MSE}_i^{(l_i^\star)}$ by an amount $\delta_{l_i^\star,q_i^\star}$. Algorithm A1 terminates when none of the $\text{MSE}_i^{(l)}$ can be further reduced. Due to the greedy nature of this algorithm, the resulting $\boldsymbol{T}^\star$ is not guaranteed to be globally optimal. For example, several small decreases may be better than one large decrease, or one small decrease in the current iteration could enable a large decrease in future iterations. Both these events are completely ignored by this greedy algorithm. Finally, Table I lists the pseudocode of A1.

2) Algorithm 2 (A2): The structure of algorithm A2 is identical to the structure of algorithm A1 described in subsection III-B1 except that the selection criterion (line 4 in Table I) exercises more caution. Instead of opting for the TIR increment $\boldsymbol{T}_{\text{inc}}^{(l,q)}$ inducing the largest decrease in the largest reducible $\text{MSE}_i^{(l)}$, a more promising TIR increment is chosen. More precisely, this algorithm selects the TIR increment $\boldsymbol{T}_{\text{inc}}^{(l,q)}$ inducing the largest *guaranteed* decrease rather than the largest current decrease. This guaranteed decrease $d_{l,q}^\star$ is defined as the sum of two terms:

$$d_{l,q}^\star = \delta_{l,q}^\star + \delta_{l,q,\text{LB}}^\star \ \ \forall(l,q). \tag{36}$$

The first term, $\delta_{l,q}^\star$, is defined in (35) and represents the optimal decrease achieved by the increment $\boldsymbol{T}_{\text{inc}}^{(l,q)}$ in iteration i. The second term, $\delta_{l,q,\text{LB}}^\star$, is a lower bound on the maximal

Table II
CALCULATION OF $\delta^{\star}_{l,q,\mathrm{LB}}$

Input: $\boldsymbol{T}_i$, $\boldsymbol{T}_{\mathrm{inc}}^{(l,q)}$
$\boldsymbol{T}_{i+1} = \boldsymbol{T}_i + \boldsymbol{T}_{\mathrm{inc}}^{(l,q)}$, $\delta^{\star}_{\mathrm{LB}} = 0$, $m = 0$
$\forall l_{\mathrm{inc}}, q_{\mathrm{inc}}$: calculate $\Delta_{l_{\mathrm{inc}},q_{\mathrm{inc}}} = \delta^{\star}_{l_{\mathrm{inc}},q_{\mathrm{inc}}}$ using (35) with $\boldsymbol{T}_i$ replaced by $\boldsymbol{T}_{i+1}$
while $m < N_t$ **do**
    $l_{i+1+m}, q_{i+1+m} = \arg\max_{l_{\mathrm{inc}},q_{\mathrm{inc}}} \Delta_{l_{\mathrm{inc}},q_{\mathrm{inc}}}$
    $\delta^{\star}_{\mathrm{LB}} \leftarrow \delta^{\star}_{\mathrm{LB}} + \Delta_{l_{i+1+m},q_{i+1+m}}$
    $\Delta_{l_{i+1+m},|} = \mathbf{0}$
    if $q_{i+1+m} \leq N_t$ **then**
        $\Delta_{|,q_{i+1+m}} = \mathbf{0}$
    end
    $m \leftarrow m + 1$
end

realizable decrease achievable by TIR increments in future iterations $i+1, i+2$, etc. For given $\boldsymbol{T}_{i+1}$, the reduction of the MSE realized during iteration $i+1+m$, with $m > 0$, usually depends on the increments previously performed during iterations $i+1, ..., i+m$. To compute the lower bound $\delta^{\star}_{l,q,\mathrm{LB}}$, the increments are *constrained* in iterations $i+2$, $i+3$, etc. such that, for all $m > 0$, the decrease achieved in iteration $i+1+m$ is influenced only by $\boldsymbol{T}_{i+1}$ and by the increment made in iteration $i+1+m$, but not by the increments in iterations $i+1, ..., i+m$. This way, the sum of the decreases related to the increments in the iterations $i+1, i+2$, etc. represents a lower bound on the maximum possible decrease, which would be accomplished by unconstrained increments.

When $\boldsymbol{T}_{\mathrm{inc}}^{(l_j,q_j)}$ denotes the increment in iteration j, equations (32) and (33) infer that the MSE reduction achieved in iteration j depends on the l_jth row of $\boldsymbol{T}_j$ and, if $q_j \leq N_t$, also on the q_jth row of $\boldsymbol{T}_j(0)$. Hence, with the purpose that the MSE reduction in iteration $i+1+m$ is not influenced by the increments made in iterations $i+1, ..., i+m$ for any $m > 0$, the quantities l_{i+1+m} and q_{i+1+m} must be outside the sets $\{l_{i+n} \mid n = 1, ..., m\}$ and $\{l_{i+n} \mid q_{i+n} \leq N_t, n = 1, ..., m\}$, respectively. Consequently, the largest value of m to be considered cannot exceed $N_t - 1$. When $\{(l_{i+1+m}, q_{i+1+m}) \mid m > 0\}$ is constructed this way, the MSE reduction corresponding to iteration $i+1+m$ can be computed from (35) by substituting the matrices $\boldsymbol{T}_{i+1}$ and $\boldsymbol{T}_{i+1}(0)$ (rather than $\boldsymbol{T}_{i+1+m}$ and $\boldsymbol{T}_{i+1+m}(0)$) for $\boldsymbol{T}_i$ and $\boldsymbol{T}_i(0)$ in (32) and (33). For given $\boldsymbol{T}_i$ and $\boldsymbol{T}_{\mathrm{up}}^{(l_i,q_i)}$, the suboptimal greedy algorithm from Table II selects in each iteration $i+1, i+2, ..., i+N_t$ the constrained increment that achieves the largest decrease, and sums these decreases to obtain $\delta^{\star}_{l_i,q_i,\mathrm{LB}}$. Subsequently, the guaranteed decrease $d^{\star}_{l_i,q_i}$ corresponding to $\boldsymbol{T}_{\mathrm{up}}^{(l_i,q_i)}$ then follows from (36).

Algorithm A2 opts in the ith iteration for the increment matrix that induces the largest guaranteed reduction, denoted by $d^{\star}_{l_i^{\star},q_i^{\star}}$, of $\mathrm{MSE}_i^{(l_i^{\star})}$, where $l_i^{\star}$ equals the index of the largest of the reducible $\mathrm{MSE}_i^{(l)}$ (as in A1, $\mathrm{MSE}_i^{(l)}$ is called reducible when $\max_{q_i} \delta^{\star}_{l,q_i} > 0$), and $q_i^{\star} = \arg\max_{q_i} d^{\star}_{l_i^{\star},q_i}$. Consequently, the pseudocode of A2 differs from A1 only in the selection criterion in line 4 from Table I. Although this criterion more carefully selects the increment $\boldsymbol{T}_{\mathrm{inc}}^{(l,q)}$ in A2, the obtained $\boldsymbol{T}^{\star}$ is not guaranteed to be globally optimal. Still, A2 is expected to yield a better performance than A1, as it can only benefit from the additional information in the increment selection. Moreover, the convergence condition for A1 and A2 is the same, i.e., both algorithms terminate when none of the $\mathrm{MSE}_i^{(l)}$ can be further reduced. Hence, neither algorithm is able to further lower the MSE when one algorithm is initialized with the TIR obtained by the other algorithm.

3) Algorithm 3 (A3): In algorithm A3, the ith iteration determines the optimized increment for an entire row of $\boldsymbol{T}_i$ rather than a part of row $\boldsymbol{T}_i$ as in the algorithms above. In this regard, the Hermitian positive-definite matrix $\boldsymbol{G}$ is first factorized according to its Cholesky decomposition, i.e., $\boldsymbol{G} = \boldsymbol{L}\boldsymbol{L}^H$, where $\boldsymbol{L}$ is a lower triangular matrix. Equation (27) can thus be rewritten as

$$\mathrm{MSE}_{i+1} = \frac{1}{N_t}\mathrm{Tr}\left(\boldsymbol{T}_{i+1}\boldsymbol{L}\boldsymbol{L}^H\boldsymbol{T}_{i+1}^H\right) = \frac{1}{N_t}\sum_{l=1}^{N_t}\left\|\boldsymbol{t}_{l,i+1}\boldsymbol{L}\right\|^2, \quad (37)$$

where $\boldsymbol{t}_{l,i+1}$ equals the lth row of $\boldsymbol{T}_{i+1}$. When l_i denotes the index of the single row that is updated in the ith iteration, the updated row $\boldsymbol{t}_{l_i,i+1}$ can be obtained by

$$\boldsymbol{t}_{l_i,i+1} = \boldsymbol{t}_{l_i,i} + \boldsymbol{t}_{\mathrm{inc}}, \quad (38)$$

while the other rows remain unaltered. To satisfy the additional constraints, all components of the row increment $\boldsymbol{t}_{\mathrm{inc}}$ must belong to $\mathbb{Z}[j]$, and its first N_t components must be a linear combination of the first N_t components of the other rows $\boldsymbol{t}_{l,i}$ with $l \neq l_i$. For a more mathematical description, $\boldsymbol{t}_{\mathrm{inc}}$ and $\boldsymbol{L}$ are first decomposed as

$$\boldsymbol{t}_{\mathrm{inc}} = \begin{bmatrix}\boldsymbol{t}_{\mathrm{inc}}^{(0)} & \boldsymbol{\lambda}_{\mathrm{inc}}^{(1)}\end{bmatrix} \quad (39)$$

$$\boldsymbol{L} = \begin{bmatrix}(\boldsymbol{L}^{(0)})^H & (\boldsymbol{L}^{(1)})^H\end{bmatrix}^H, \quad (40)$$

where $\boldsymbol{t}_{\mathrm{inc}}^{(0)}$ and $\boldsymbol{L}^{(0)}$ contain the first N_t elements of $\boldsymbol{t}_{\mathrm{inc}}$ and the first N_t rows of the matrix $\boldsymbol{L}$, respectively, whereas $\boldsymbol{\lambda}_{\mathrm{inc}}^{(1)}$ and $\boldsymbol{L}^{(1)}$ comprise the last $N_t L_{\mathrm{T}}$ elements of $\boldsymbol{t}_{\mathrm{inc}}$ and the last $N_t L_{\mathrm{T}}$ rows of $\boldsymbol{L}$, respectively. In (39), $\boldsymbol{t}_{\mathrm{inc}}^{(0)}$ must be a linear combination of the first N_t elements of the rows $\boldsymbol{t}_{l,i}$ with $l \neq l_i$, which can be expressed as

$$\boldsymbol{t}_{\mathrm{inc}}^{(0)} = \sum_{\substack{l=1\\ l\neq l_i}}^{N_t} \lambda_{l,\mathrm{inc}}^{(0)}\boldsymbol{t}_{l,i}^{(0)} = \boldsymbol{\lambda}_{\mathrm{inc}}^{(0)}\boldsymbol{T}_{l_i,i}^{(0)}, \quad (41)$$

where $\boldsymbol{\lambda}_{\mathrm{inc}}^{(0)}$ represents the row vector containing all $\lambda_{l,\mathrm{inc}}^{(0)}$, $\boldsymbol{t}_{l,i}^{(0)}$ consists of the first N_t elements of the lth row of $\boldsymbol{T}_i$, and $\boldsymbol{T}_{l_i,i}^{(0)}$ is obtained by removing the l_ith row from $\boldsymbol{T}_i(0)$. For given $\boldsymbol{T}_i$, the increment $\boldsymbol{t}_{\mathrm{inc}}$ in the ith iteration is selected to minimize $\mathrm{MSE}_{i+1}^{(l_i)}$. After the substitution of (38), (39), (40), and (41) into (37), this minimization problem can be reformulated as the following closest point search:

$$\boldsymbol{\lambda}_{\mathrm{inc}}^{\star} = \underset{\lambda\in\mathbb{Z}[j]^{(L_{\mathrm{T}}+1)N_t-1}}{\arg\min}\left\|\lambda\boldsymbol{G}_{\mathrm{lat}} - \boldsymbol{x}\right\|^2, \quad (42)$$

where $\boldsymbol{G}_{\mathrm{lat}} = \begin{bmatrix}\boldsymbol{T}_{l_i,i}^{(0)}\boldsymbol{L}^{(0)}\\ \boldsymbol{L}^{(1)}\end{bmatrix}$, $\boldsymbol{x} = -\boldsymbol{t}_{l_i,i}\boldsymbol{L}$, and $\lambda = \begin{bmatrix}\boldsymbol{\lambda}_{\mathrm{inc}}^{(0)} & \boldsymbol{\lambda}_{\mathrm{inc}}^{(1)}\end{bmatrix}$. To solve(42), the lattice decoding algorithm presented in [24] is applied after decomposing all complex-valued quantities in

Table III
PSEUDOCODE OF ALGORITHM A3

1: $\boldsymbol{T}_0 = [\boldsymbol{I}_{N_t}\ \boldsymbol{0}]$, $i = 0$
2: **while** $\exists\, l : \mathrm{MSE}_i^{(l)}$ reducible **do**
3: Select $l_i^\star$ as largest reducible $\mathrm{MSE}_i^{(l)}$
4: Compute update $\boldsymbol{t}_{\mathrm{inc}}$ by solving optimization problem (42)
5: Update $\boldsymbol{T}_{i+1}$ by computing $\boldsymbol{t}_{l_i^\star, i+1}$ using (38)
6: $i \leftarrow i + 1$
7: **end**
8: $\boldsymbol{T}^\star = \boldsymbol{T}_i$

(42) into their real and their imaginary parts. This algorithm searches for a point $\hat{\boldsymbol{x}}$ from a lattice with generator matrix $\boldsymbol{G}_{\mathrm{lat}}$ that is closest to $\boldsymbol{x}$, by recursively decomposing the lattice into lower-dimensional sublattices. Further, $\mathrm{MSE}_i^{(l_i)}$ is reducible when $\boldsymbol{\lambda}_{\mathrm{inc}}^\star$ has at least one nonzero element. The pseudocode of A3 (Table III) is similar to A1 and A2 except that A3 updates in each iteration an entire row of $\boldsymbol{T}_i$. The increments $\boldsymbol{T}_{\mathrm{inc}}$, allowed in A1 and A2, correspond to a vector $\boldsymbol{\lambda}_{\mathrm{inc}}^\star$ with exactly one nonzero element, thus forming a small subset of all increments allowed in A3. Hence, A3 could potentially further lower the MSE when A3 is initialized with the TIR derived by either A1 or A2, whereas the reverse is not true, implying that A3 has superior performance to A1 and A2.

4) Complexity considerations: This section provides some general remarks on the complexity per iteration of the proposed algorithms. For A1, at most $\left((L_{\mathrm{T}}+1)N_t^2 - N_t\right)$ optimal $\lambda_{l,q}^\star$ from (34) must be calculated per iteration, which is of course feasible in polynomial time. In A2, for each increment in the current iteration, all $\lambda_{l,q}^\star$ of the next iteration must be determined as well. Hence, at most $\left((L_{\mathrm{T}}+1)N_t^2 - N_t\right)^2$ different $\lambda_{l,q}^\star$ must be computed, yielding, compared to A1, a larger complexity that is still polynomial in N_t and L_{T}. Clearly, A3 is the most complex, as each iteration must solve a closest point problem. Unfortunately, this problem is NP-hard and no polynomial time algorithm is available (yet) to solve it. Indeed, the search time of the lattice decoding algorithm in [24] rises exponentially with the problem's dimensions, e.g., N_t, L_{T}.

IV. BER EXPRESSION FOR PRS

This section describes how the BER expression for the MIMO PRS-ST is derived. In [26], the symbol error rate in the case of SISO PRS with an M-PAM constellation has already been discussed in detail, and the present contribution extends this work to obtain the BER for MIMO PRS-ST with an M-QAM constellation using a two-dimensional binary reflected Gray mapping.

First, a scaled and translated version $\boldsymbol{v}(k)$ of the decision variable $\boldsymbol{u}(k)$ is introduced, i.e.,

$$\boldsymbol{v}(k) = \frac{\boldsymbol{u}(k)}{2\Delta} + \boldsymbol{c}_{\boldsymbol{v}} \tag{43}$$

with $\boldsymbol{c}_{\boldsymbol{v}}$ defined in (7). Based on the decomposition of $\boldsymbol{u}(k)$ from (12), the target response vector $\boldsymbol{u}_{\mathrm{T}}(k)$ from (5), and the mapping rule (2), $\boldsymbol{v}(k)$ can be expressed as

$$\boldsymbol{v}(k) = \sum_{m=0}^{L_{\mathrm{T}}} \boldsymbol{T}(m)\boldsymbol{b}(k-m) + \mathbf{isi}(k) + \boldsymbol{n}_v(k), \tag{44}$$

where $\boldsymbol{n}_v(k) = \frac{\boldsymbol{W}\tilde{\boldsymbol{n}}(k)}{2\Delta}$ and $\mathbf{isi}(k) = \sum_{m\in\Phi_{\mathrm{G}}} \boldsymbol{E}(m)\boldsymbol{a}(k-m)$ with

$$2\Delta\boldsymbol{E}(m) = \begin{cases} \boldsymbol{W}\boldsymbol{G}(m) - \boldsymbol{T}(m) & m \in \Phi_{\mathrm{T}} \\ \boldsymbol{W}\boldsymbol{G}(m) - \boldsymbol{B}(m) & m \in \Phi_{\mathrm{B}} \\ \boldsymbol{W}\boldsymbol{G}(m) & m \in \Phi_{\mathrm{G}} \setminus \{\Phi_{\mathrm{T}} \cup \Phi_{\mathrm{B}}\} \end{cases}. \tag{45}$$

Because (1) yields $\sum_{m=0}^{L_{\mathrm{T}}} \boldsymbol{T}(m)\boldsymbol{b}(k-m) = \boldsymbol{c}(k) + \boldsymbol{T}(0)\boldsymbol{d}(k)\sqrt{M}$, where the components of $\boldsymbol{d}(k)$ are Gaussian integers, $\boldsymbol{v}(k)$ from (44) reduces to $\boldsymbol{v}(k) = \boldsymbol{c}_{\mathrm{ex}}(k) + \mathbf{isi}(k) + \boldsymbol{n}_v(k)$ with $\boldsymbol{c}_{\mathrm{ex}}(k) = \boldsymbol{c}(k) + \boldsymbol{T}(0)\boldsymbol{d}(k)\sqrt{M}$. At the Rx, the symbol-by-symbol detector makes a decision $\hat{\boldsymbol{c}}_{\mathrm{ex}}(k)$ of $\boldsymbol{c}_{\mathrm{ex}}(k)$ in the extended symbol set by rounding $\boldsymbol{v}(k)$ to the closest Gaussian integer. As this procedure is performed for each data stream individually, the decision of the received symbol at instant k in the lth stream, $\hat{c}_{\mathrm{ex}}^{(l)}(k)$ is based on

$$v^{(l)}(k) = c_{\mathrm{ex}}^{(l)}(k) + \mathrm{isi}^{(l)}(k) + n_v^{(l)}(k), \tag{46}$$

where $v^{(l)}(k)$, $c_{\mathrm{ex}}^{(l)}(k)$, $\mathrm{isi}^{(l)}(k)$, and $n_v^{(l)}(k)$ equal the lth component of the associated vectors $\boldsymbol{v}(k)$, $\boldsymbol{c}_{\mathrm{ex}}(k)$, $\mathbf{isi}(k)$, and $\boldsymbol{n}_v(k)$, respectively. Afterwards, the modulo operator is applied to $\hat{c}_{\mathrm{ex}}^{(l)}(k)$, yielding the decision $\hat{c}^{(l)}(k) \in C$ of the symbol $c^{(l)}(k)$ transmitted in the lth data stream at instant k. As the error performance does not depend on the symbol index k, this index is dropped in the sequel. The BER for the lth data stream, $\mathrm{BER}^{(l)}$, can be expressed as

$$\mathrm{BER}^{(l)} = \sum_{(c,\hat{c})\in C^2} \frac{N_{\mathrm{bit}}(c,\hat{c})}{\log_2(M)} \Pr(\mathrm{c}^{(l)} = c, \hat{c}^{(l)} = \hat{c}), \tag{47}$$

in which $N_{\mathrm{bit}}(c,\hat{c})$ represents the number of bits by which the binary labels of c and $\hat{c}$ differ. In Appendix B, an elaborate discussion is given on how the following simple but accurate approximation for $\mathrm{BER}^{(l)}$ is derived from (47):

$$\mathrm{BER}^{(l)} \approx \frac{1}{\log_2(M)} \mathbb{E}_{\boldsymbol{a}}\left[4Q\left(\frac{0.5 - \mathrm{Re}\left[\mathrm{isi}^{(l)}\right]}{\sigma_{n_v^{(l)}}}\right)\right], \tag{48}$$

where $Q(\cdot)$ represents the tail distribution function of the standard normal distribution, $\sigma_{n_v^{(l)}}$ equals the standard deviation of the real part of $n_v^{(l)}(k)$, and $\mathbb{E}_{\boldsymbol{a}}[\cdot]$ denotes the expectation over all symbols contributing to $\mathrm{Re}\left[\mathrm{isi}^{(l)}\right]$. The computational complexity in (48) increases exponentially with channel and filter length, thus rapidly becoming prohibitively large in numerical simulations. To circumvent this complexity problem, a large number, N_{large}, of realizations of $\mathrm{Re}\left[\mathrm{isi}^{(l)}\right]$ is generated, and the expectation in (48) is replaced by the arithmetical average.

V. NUMERICAL RESULTS

This section characterizes the performance of the optimization algorithms (A1, A2, and A3) discussed above by numerical simulations for three scenarios, whose simulation settings

Table IV
SIMULATION SETTINGS

• 4-QAM constellation	• $\Phi_{\mathrm{B}} = \{4, 5\}$
• 3500 channel realizations	• $L_{\mathrm{H}}^{(1)} = 0$, $L_{\mathrm{H}}^{(2)} = 25$
• PRS settings:	• Scenarios:
FRS ($L_{\mathrm{T}} = 0$, $\boldsymbol{T}(0) = \boldsymbol{I}$)	S1: $\mu = 2$, $L_{\mathrm{w}}^{(1)} + L_{\mathrm{w}}^{(2)} + 1 = 13$
PRS-S ($L_{\mathrm{T}} = 0$; A1, A2 and A3)	S2: $\mu = 2$, $L_{\mathrm{w}}^{(1)} + L_{\mathrm{w}}^{(2)} + 1 = 21$
PRS-ST($L_{\mathrm{T}} = 3$; A1, A2 and A3)	S3: $\mu = 5$, $L_{\mathrm{w}}^{(1)} + L_{\mathrm{w}}^{(2)} + 1 = 13$

are summarized in Table IV. In all scenarios, the channel $\boldsymbol{H}(m)$ is a FS Rayleigh-fading 4×4 MIMO channel with an exponentially decaying power delay profile with base μ, i.e., $\mathbb{E}\left[\left|(\boldsymbol{H}(m))_{(p,q)}\right|^2\right] = \mu^{-(m+L_{\mathrm{H}}^{(1)})}$, $p, q \in \{1, ..., 4\}$, and $m \in \{-L_{\mathrm{H}}^{(1)}, \ldots, L_{\mathrm{H}}^{(2)}\}$. All channel taps are further assumed to be spatially and temporally uncorrelated. Moreover, all components of the circular symmetric Gaussian noise $\boldsymbol{n}(k)$ are also spatially and temporally uncorrelated and possess variance N_0, whereas the components from $\boldsymbol{a}(k)$ belong to a Gray-mapped 4-QAM constellation. In the first scenario S1, a severely FS channel ($\mu = 2$) is equalized using a DFE consisting of a 13-tap feedforward filter and a 2-tap feedback filter with $\Phi_{\mathrm{B}} = \{4, 5\}$. Extending the feedforward filter to 21 taps in the second scenario S2 evidently improves the performance at the cost of a larger complexity. The third scenario S3 is identical to S1, except that the FS channel is less severe ($\mu = 5$). In all scenarios, a comparison is made between: (i) PRS-ST, whose TIR consists of both spatial and temporal components ($L_{\mathrm{T}} = 3$); (ii) spatial-only PRS (PRS-S), whose TIR possesses only spatial components ($L_{\mathrm{T}} = 0$); and (iii) traditional FRS.

This discussion principally focuses on the performance in terms of MSE and BER ($N_{\mathrm{large}} = 4 \cdot 10^7$), because the former is the objective function of the optimization, whereas the latter is an important performance measure in practice. In the case of S1, Fig. 2, depicts the 1/MSE and the BER performances averaged over 3500 channel realizations as a function of SNR = $\frac{E_b}{N_0}$ with E_b the transmitted energy per bit. Furthermore, Table V lists the SNR in dB needed to achieve a BER of 10^{-8} in all scenarios. What is immediately apparent is the significant improvement achieved by PRS compared to FRS. Indeed, the smaller MSE accomplished by PRS is a direct consequence of the additional minimization over $\boldsymbol{T}$, but, more interestingly, Fig. 2 and Table V reveal that also the BER of the PRS is significantly lower compared to FRS. One noteworthy example is the drastic reduction of the error floor by considering PRS instead of FRS in S1 (Fig. 2). Moreover, the longer feedforward filter in S2 allows PRS even to completely remove the error floor encountered by FRS or at least to lower it to below 10^{-8}. Finally, PRS accomplishes a reduction up to nearly 16 dB in SNR to reach a target BER of 10^{-8} in S3 (Table V).

Significant differences in average performance occur between the algorithms when the TIR has both spatial and temporal components ($L_{\mathrm{T}} = 3$). In this configuration, A3 not only yields an average MSE that is smaller than the MSEs of A1 and A2, but, more importantly, also drastically lowers the

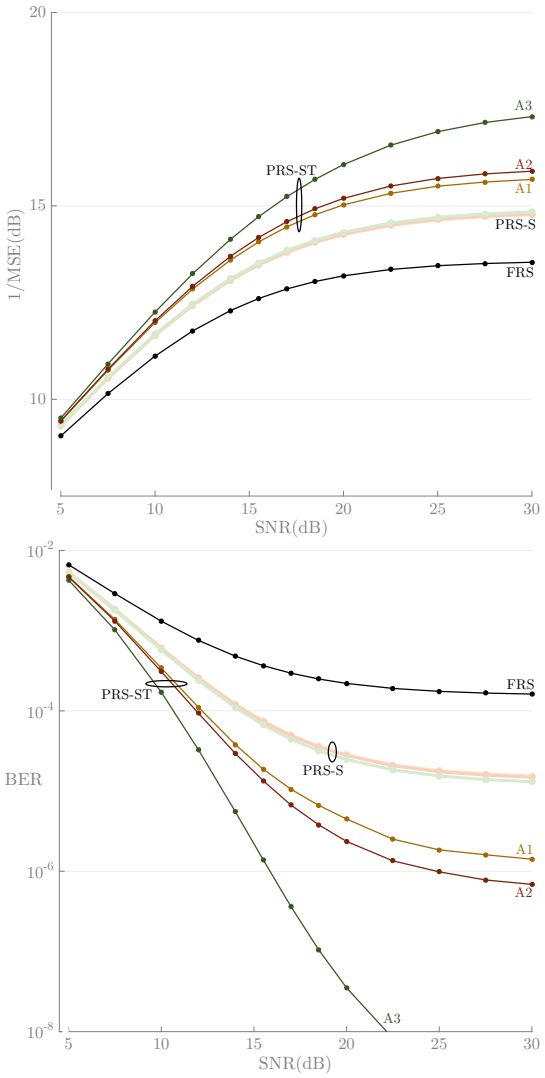


Figure 2. Scenario S1: average MSE (upper plot) and average BER (lower plot) with $\mu = 2$ and a 13-tap $\boldsymbol{W}$. Compared to FRS, PRS drastically improves the performance. In particular, PRS-ST optimized using A3 considerably reduces the BER floor.

BER. For instance, only A3 is able to reduce the error floor below 10^{-8} in S1, whereas its SNR required to reach a BER of 10^{-8} in S2 is approximately 3.0 dB and 2.5 dB lower compared to A1 and A2, respectively; in S3 the gain of A3 compared to A1 and A2 is about 2.2 dB and 1.5 dB, respectively. These results therefore imply not only that A1 and A2 do not yield the global optimum, but also that A3 is to be preferred when performance is essential. Furthermore, the numerical results confirm our expectation that A2 outperforms A1 on average. However, the gain of A2 compared to A1 on average turns out to be quite moderate: a lower error floor is achieved in S1 (factor 2), whereas the SNR required to meet the BER target decreases by less than 0.8 dB in S2 and S3.

When the TIR has only spatial components, i.e., $L_{\mathrm{T}} = 0$, the performance difference between A1, A2, and A3 is almost negligible for all scenarios. Indeed, $\boldsymbol{T}$ contains only $\boldsymbol{T}(0)$, limiting the increments in all algorithms to the set of consecutive row additions. All algorithms are therefore similar, resulting in

Table V
SNR IN DB NEEDED TO REACH AN AVERAGE BER = 10^{-8}

Scenario	FRS	PRS-S			PRS-ST		
		A1	A2	A3	A1	A2	A3
S1	-	-	-	-	-	-	22.17
S2	-	21.44	21.43	21.37	17.40	17.10	14.64
S3	31.69	18.65	18.51	18.40	17.79	17.00	15.53

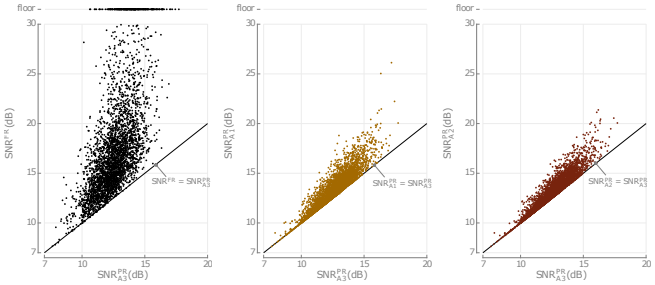


Figure 3. Scenario S2: scatter plot of the points ($\mathrm{SNR}_{\mathrm{A3}}^{\mathrm{PR}}$, $\mathrm{SNR}_{\mathrm{Y}}$), where $\mathrm{SNR}_{\mathrm{Y}}$ equals $\mathrm{SNR}^{\mathrm{FR}}$, $\mathrm{SNR}_{\mathrm{A1}}^{\mathrm{PR}}$ and $\mathrm{SNR}_{\mathrm{A2}}^{\mathrm{PR}}$ in the left, middle, and right plot, respectively. All points above/below the solid straight line represents realizations for which $\mathrm{SNR}_{\mathrm{Y}}$ is larger/smaller than $\mathrm{SNR}_{\mathrm{A3}}^{\mathrm{PR}}$.

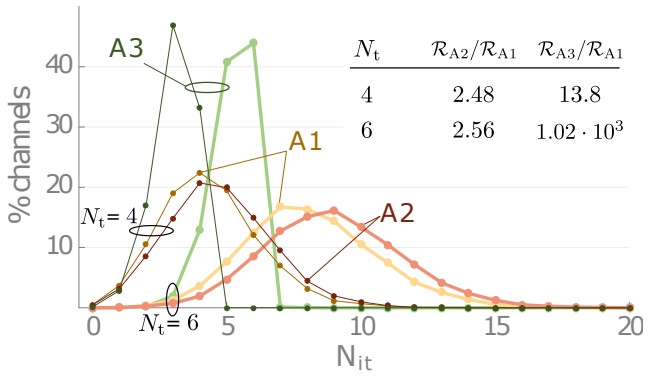


Figure 4. Histogram of number of iterations required for convergence of A1-A3 for a 21-tap feedforward filter with $L_{\mathrm{T}} = 3$ and SNR = 20 dB for both $N_t = N_r = 4$ and $N_t = N_r = 6$. Additionally, the relative runtimes with respect to A1 are also listed for A2 and A3. A small number of high-complexity iterations results in almost instant convergence for algorithm A3, whereas A1 and A2 require more, but significantly less complex iterations.

nearly identical average MSE and BER performance such that the least complex A1 is here the most attractive. For $L_{\mathrm{T}} = 0$, the minimization of (25) is structurally identical to the lattice reduction problem from [18]. Indeed, A1 corresponds to the element-based lattice reduction (ELR) algorithm from [18], while A3 is similar to the improved ELR algorithm.

At low SNR, the matrix $\boldsymbol{G}$ in (25) closely resembles a scaled identity matrix. Consequently, FRS is optimum for infinite noise power, causing the optimized PRS to converge to FRS for low SNR. The performance improvement of PRS is therefore mainly noticeable at high SNR.

As for the mildly selective channel in S3, all PRS configurations outperform FRS, but the difference between PRS-S and PRS-ST is less significant. Hence, the improvement of PRS can mainly be attributed to the spatial components in $\boldsymbol{T}(0)$ and to a lesser extent to the temporal components in $\{\boldsymbol{T}(m) \mid m > 0\}$. As the mildly FS channel generates relatively small temporal ISI, this observation is not unexpected.

The discussion above considers only results that are averaged over 3500 channel realizations. To better understand how FRS and PRS affect the BER performance of an individual realization, $\mathrm{SNR}^{\mathrm{FR}}$, $\mathrm{SNR}_{\mathrm{A1}}^{\mathrm{PR}}$, $\mathrm{SNR}_{\mathrm{A2}}^{\mathrm{PR}}$, and $\mathrm{SNR}_{\mathrm{A3}}^{\mathrm{PR}}$ are defined to denote the SNR required for a *particular* realization to reach a BER of 10^{-8} for FRS and for the three algorithms for PRS, respectively. The scatter plots from Fig. 3 compare $\mathrm{SNR}_{\mathrm{A3}}^{\mathrm{PR}}$ to $\mathrm{SNR}^{\mathrm{FR}}$, $\mathrm{SNR}_{\mathrm{A1}}^{\mathrm{PR}}$, and $\mathrm{SNR}_{\mathrm{A2}}^{\mathrm{PR}}$, for 3500 channel realizations in the case of S2 and $L_{\mathrm{T}} = 3$ (PRS-ST). In these plots, a scatter point is labeled 'floor' when the target BER of 10^{-8} cannot be reached due to an error floor. Immediately evident from the plots is that nearly all scatter points are above the line connecting all points where the SNR on the ordinate equals $\mathrm{SNR}_{\mathrm{A3}}^{\mathrm{PR}}$. Hence, apart from some exceptions, e.g., channels with $\mathrm{SNR}^{\mathrm{FR}} < \mathrm{SNR}_{\mathrm{A3}}^{\mathrm{PR}}$, PRS optimized with A3 requires the lowest SNR to reach the target BER. Additionally, the MSE achieved by A3 has been numerically verified to be the lowest for all realizations (not shown). These observations not only corroborate the superiority of A3 in terms of the MSE, but also illustrate that a smaller MSE does not necessarily guarantee a smaller BER. Next, the largest gain from applying A3 is observed for unfavorable channels, i.e., realizations requiring a rather large $\mathrm{SNR}^{\mathrm{FR}}$, $\mathrm{SNR}_{\mathrm{A1}}^{\mathrm{PR}}$, and $\mathrm{SNR}_{\mathrm{A2}}^{\mathrm{PR}}$, whereas only minor gains are obtained for favorable realizations with smaller values for $\mathrm{SNR}^{\mathrm{FR}}$, $\mathrm{SNR}_{\mathrm{A1}}^{\mathrm{PR}}$, and $\mathrm{SNR}_{\mathrm{A2}}^{\mathrm{PR}}$. Extra simulations confirm that all conclusions apply also to scenarios with different filter lengths and/or base μ.

While the MSE of A3 is numerically verified to be the lowest for all channels, A1 and A2 are only locally optimal. As for their relative performance, results above confirm that A2 outperforms A1 on average, as expected. However, A1 could yield better performance than A2 for a specific channel as a sequence of increments in A1 could enable a large future decrease that was not considered by A2. A scatter plot (not shown) similar to Fig. 3 comparing the MSE of A1 and A2 at SNR = 20 dB supports this claim. For about 92.5% of the channels, the outcome of A1 and A2 is identical, while A2 is superior for 6.5% and A1 outperforms A2 only for <1%.

Above, the optimal row index $l_i^{\star}$ corresponds to the largest reducible $\mathrm{MSE}_i^{(l)}$. However, an alternative selection criterion for $l_i^{\star}$ could be envisaged as well. For instance, the row $l_i^{\star}$ could be chosen such that $\boldsymbol{T}_{\mathrm{inc}}$ induces the maximal reduction in MSE_i. For all algorithms, comparing both selection criteria in terms of the SNR required to attain a target BER of 10^{-8} for 3500 channels reveals that for up to 90% of all channel realizations the performances are identical. Moreover, each criterion is superior in about half of the remaining realizations. Hence, on average, both criteria yield comparable performance. The numerical results also confirm that both criteria require on average nearly the same number of iterations to converge. Contrary to the alternative criterion, the original criterion does not require the computation of $\max_q \delta_{l,q}^{\star}$ for all l in each iteration; therefore, the original criterion is preferred over the alternative as it runs faster with similar performance.

Another key feature of the algorithms is the computational complexity. Fig. 4 therefore presents the histogram (3500 channels) of the number of iterations required for convergence in the case of $L_{\mathrm{T}} = 3$, S2, and SNR = 20 dB for $N_t = N_r = \{4, 6\}$. An extra table lists also the runtimes $\mathcal{R}$ of A2 and A3 relative to the runtime of A1, $\mathcal{R}_{\mathrm{A1}}$. In the case of A1 and A2, convergence is reached for at least 90% of the channels after $2N_t$ iterations, whereas only approximately N_t iterations are needed for A3. However, $\mathcal{R}_{\mathrm{A3}}$ is significantly larger than $\mathcal{R}_{\mathrm{A1}}$ and $\mathcal{R}_{\mathrm{A2}}$, as the complexity of the lattice decoding problem rises exponentially with the product $N_t L_{\mathrm{T}}$, and the complexity of one iteration thus rapidly becomes substantial. Compared to A1, $\mathcal{R}_{\mathrm{A2}}$ is in Fig. 4 approximately 2.5 times as large as $\mathcal{R}_{\mathrm{A1}}$, because A2 requires not only slightly more, but also somewhat more complex iterations than A1. Therefore, algorithms A1 and A2 are to be preferred when either performance is subordinated to complexity or when all algorithms yield similar performance, e.g, at low SNR or for PRS-S. When performance is critical, A3 should be employed.

Above, all simulations are conducted with $\boldsymbol{T}$ initialized as FRS. Additional simulations show, however, that when $\boldsymbol{T}_0$ is initialized with a random feasible TIR, A1 and A2 get trapped in local minima whose performance is often significantly worse than FRS. On the other hand, numerical results suggest that A3 can achieve its excellent performance irrespective of the initialization.

VI. Conclusion and remarks

While the TIR of PRS systems operating on MIMO channels in the literature is mainly restricted to be spatial-only or temporal-only, this paper discusses an equalization scheme with PRS, where a FS MIMO channel is equalized with respect to a general ST-TIR. Importantly, this TIR and these equalizers are jointly optimized according to the MMSE criterion. First, optimal equalization coefficients for a given TIR are derived. Next, three iterative algorithms are described that perform a row-by-row optimization of the TIR by updating the row of $\boldsymbol{T}$ with the largest reducible MSE. The least complex A1 updates in each iteration the entry of the row yielding the largest decrease of the largest reducible MSE. A2 is similar to A1, but considers also the potential of future updates. The most complex A3 updates in each iteration an entire row by solving a lattice decoding problem. Further, this contribution derives the BER expression along with an accurate approximation. The numerical results confirm that optimized PRS substantially outperforms FRS in terms of both the MSE and the BER. These performance improvements are mostly noticeable at high SNR, since PRS converges to FRS when the noise power becomes infinite. Whereas A2 yields on average only a minor enhancement compared to A1, A3 achieves a superior performance, justifying its larger complexity. However, the performance differences between the algorithms are negligibly small for a spatial-only TIR, making the low-complexity A1 more preferable in this situation.

While all channel state information is assumed to be perfect in this study, the design of a algorithm that is robust to delayed and/or noisy channel estimates is also of particular interest. Therefore, our paper [27] discusses the robust and joint design of the ST-TIR and a linear pre-equalizer at the transmitter while channel estimation errors and delays are present.

Appendix A

Proposition 1. *When all entries of the sequence $\{\boldsymbol{c}(k)\}$ are independently and uniformly drawn from the set $C = \{0, \ldots, \sqrt{M} - 1\} + j\{0, \ldots, \sqrt{M} - 1\}$ and $\boldsymbol{b}(k)$ is constructed according to*

$$\boldsymbol{b}(k) = \left[\boldsymbol{T}^{-1}(0) \left(\boldsymbol{c}(k) - \sum_{m=1}^{L_{\mathrm{T}}} \boldsymbol{T}(m) \boldsymbol{b}(k-m) \right) \right]_{\sqrt{M}}, \qquad (49)$$

where $[\cdot]_X$ equals the element-wise modulo-X reduction of the real and the imaginary part, and (i) all entries of the matrices $\boldsymbol{T}^{-1}(0)$ and all $\boldsymbol{T}(m)$ are Gaussian integers, and (ii) $\left|\det(\boldsymbol{T}^{-1}(0)\right| = 1$, then also all entries of the sequence $\{\boldsymbol{b}(k)\}$ are independently and uniformly drawn from C.

Proof: Equation (49) can be interpreted as a finite-state machine with input, state and output at instant k given by $\boldsymbol{c}_k$, $\boldsymbol{s}_k = [\boldsymbol{b}_{k-1}^T, \boldsymbol{b}_{k-2}^T, ..., \boldsymbol{b}_{k-L_t}^T]^T$ and $\boldsymbol{b}_k$, respectively. Next, all components of the initial state $\boldsymbol{s}_0$ are assumed to belong to the set C. As the elements of $\boldsymbol{T}^{-1}(0)$ and $\boldsymbol{T}(m)$ with $m = 1, ..., L_{\mathrm{T}}$ are Gaussian integers, and $\boldsymbol{c}_k \in C^{N_t}$, (49) guarantees that also $\boldsymbol{b}_k \in C^L$ if $\boldsymbol{s}_k \in C^{L_{\mathrm{T}} N_t}$. As $\boldsymbol{s}_0 \in C^{L_{\mathrm{T}} N_t}$ by assumption, we prove by induction that $\boldsymbol{b}_k \in C^{N_t}$ for $k = 0, 1, ..., K$ and $\boldsymbol{s}_k \in C^{L_{\mathrm{T}} N_t}$ for $k = 1, 2, ..., K$.

The joint probability mass function of $\boldsymbol{b}_0, \boldsymbol{b}_1, ..., \boldsymbol{b}_K$ conditioned on $\boldsymbol{s}_0$ can be expressed as

$$\Pr[\boldsymbol{b}_0 = \boldsymbol{\beta}_0, ..., \boldsymbol{b}_K = \boldsymbol{\beta}_K | \boldsymbol{s}_0 = \boldsymbol{\sigma}_0] = \prod_{k=0}^{K} \Pr[\boldsymbol{b}_k = \boldsymbol{\beta}_k | \boldsymbol{s}_k = \boldsymbol{\sigma}_k] \qquad (50)$$

with $\boldsymbol{\beta}_k \in C^{N_t}$ and $\boldsymbol{\sigma}_k \in C^{L_{\mathrm{T}} N_t}$ for $k = 0, ..., K$. When $\boldsymbol{s}_k = \boldsymbol{\sigma}_k$, the only value of $\boldsymbol{c}_k \in C^{N_t}$ giving rise to $\boldsymbol{b}_k = \boldsymbol{\beta}_k$ is $\boldsymbol{c}_k = \boldsymbol{\gamma}_k$, where $\boldsymbol{\gamma}_k = [\boldsymbol{T}(0)\boldsymbol{\beta}_k + [\boldsymbol{T}(1), ..., \boldsymbol{T}(L_t)]\, \boldsymbol{\sigma}_k]_{\sqrt{M}}$. Hence,

$$\Pr[\boldsymbol{b}_k = \boldsymbol{\beta}_k | \boldsymbol{s}_k = \boldsymbol{\sigma}_k] = \Pr[\boldsymbol{c}_k = \boldsymbol{\gamma}_k] = M^{-N_t}, \qquad (51)$$

where the right-most equation results from the statistical properties of $\boldsymbol{c}_k$. Thus, (50) becomes

$$\Pr[\boldsymbol{b}_0 = \boldsymbol{\beta}_0, ..., \boldsymbol{b}_K = \boldsymbol{\beta}_K | \boldsymbol{s}_0 = \boldsymbol{\sigma}_0] = M^{-(K+1)N_t}, \qquad (52)$$

which depends neither on $(\boldsymbol{\beta}_0, ..., \boldsymbol{\beta}_K)$ nor on $\boldsymbol{\sigma}_0$. This indicates that all entries of the vectors $\boldsymbol{b}_0, ..., \boldsymbol{b}_K$ are independently and uniformly drawn from C. ∎

Appendix B

This appendix discusses the detailed derivation of (48), which is an approximation of (47).

First, all pairs of symbols $(c_{\mathrm{ex}}^{(l)}, \hat{c}_{\mathrm{ex}}^{(l)})$ in the extended symbol constellation corresponding to the pair $(\mathrm{c}^{(l)}, \hat{c}^{(l)})$ after the modulo operator are expressed as $(\mathrm{c}^{(l)} + d^{(l)}\sqrt{M}, \hat{c}^{(l)} + \hat{d}^{(l)}\sqrt{M})$ with $(d^{(l)}, \hat{d}^{(l)}) \in \mathbb{Z}[j]^2$. Additionally, all symbols $\boldsymbol{a}(k-m)$ that contribute to $\mathrm{isi}^{(l)}$ are collected in the vector $\boldsymbol{a}^{(l)}$. This appendix employs the notations $p^{(l)}(\boldsymbol{r})$ and $p^{(l)}(\boldsymbol{r}|\boldsymbol{s})$ to denote the probability mass functions $\Pr(\boldsymbol{r}^{(l)} = \boldsymbol{r})$ and $\Pr(\boldsymbol{r}^{(l)} =$

$\boldsymbol{r}|\boldsymbol{s}^{(l)} = \boldsymbol{s})$, where $\boldsymbol{r}^{(l)}$ and $\boldsymbol{s}^{(l)}$ are vectors of discrete random variables. The probability $\Pr\left(\mathrm{c}^{(l)} = c, \hat{c}^{(l)} = \hat{c}\right)$ in (47) is then rewritten as

$$p^{(l)}(c,\hat{c}) = \sum_{(d,\hat{d})\in\mathbb{Z}[j]^2}\sum_{\boldsymbol{a}} p^{(l)}\left(c,\hat{c},d,\hat{d},\boldsymbol{a}\right) \tag{53}$$

$$= \sum_{(d,\hat{d})\in\mathbb{Z}[j]^2}\sum_{\boldsymbol{a}} p^{(l)}\left(\hat{c},\hat{d}|c,d,\boldsymbol{a}\right) p^{(l)}(d|c,\boldsymbol{a})\, p^{(l)}(c,\boldsymbol{a}). \tag{54}$$

Next, the symbols $c_d = c + d\sqrt{M}$ and $\hat{c}_{\hat{d}} = \hat{c} + \hat{d}\sqrt{M}$ in the extended symbol set are defined, allowing to rewrite $p^{(l)}\left(\hat{c},\hat{d}|c,d,\boldsymbol{a}\right)$ as

$$p^{(l)}\left(\hat{c},\hat{d}|c,d,\boldsymbol{a}\right) = \Pr\left(c_d + \mathrm{isi}^{(l)}(\boldsymbol{a}) + n_v^{(l)} \in \mathcal{D}(\hat{c}_{\hat{d}})\right), \tag{55}$$

where $\mathcal{D}(\cdot)$ represents the decision region of a symbol in the extended symbol set, and the notation $\mathrm{isi}^{(l)}(\boldsymbol{a})$ emphasizes the dependence of $\mathrm{isi}^{(l)}$ on the vector $\boldsymbol{a}$. Because the right-hand side of (55) is affected by only the difference $\hat{c} + \hat{d}\sqrt{M} - c - d\sqrt{M} - \mathrm{isi}^{(l)}(\boldsymbol{a})$, (55) reduces to

$$p^{(l)}\left(\hat{c},\hat{d}|c,d,\boldsymbol{a}\right) = \Pr\left(c + \mathrm{isi}^{(l)}(\boldsymbol{a}) + n_v^{(l)} \in \mathcal{D}(\hat{c}_{d_-})\right), \tag{56}$$

where $\hat{c}_{d_-} = \hat{c} + d_-\sqrt{M}$ with $d_- = \hat{d} - d$. After substituting (56) into (53) and replacing the summation over $(d,\hat{d}) \in \mathbb{Z}[j]^2$ by a summation over $(d,d_-) \in \mathbb{Z}[j]^2$, only $p^{(l)}(d|c,\boldsymbol{a})$ in (53) depends on d. As the summation of $p^{(l)}(d|c,\boldsymbol{a})$ over all d amounts to 1, (53) can be written as

$$p^{(l)}(c,\hat{c}) = \sum_{d_-\in\mathbb{Z}[j]}\sum_{\boldsymbol{a}} F(c-\hat{c}_{d_-},\boldsymbol{a},n_v^{(l)})p^{(l)}(c,\boldsymbol{a}), \tag{57}$$

where $F(c-\hat{c}_{d_-},\boldsymbol{a},n_v^{(l)})$ is a shorthand notation for $\Pr\left(c + \mathrm{isi}^{(l)}(\boldsymbol{a}) + n_v^{(l)} \in \mathcal{D}(\hat{c}_{d_-})\right)$. Plugging (57) into (47) then yields

$$\mathrm{BER}^{(l)} = \sum_{\boldsymbol{a}}\sum_{c\in C} G(c,\boldsymbol{a})\Pr(\mathrm{c}^{(l)} = c, \boldsymbol{a}^{(l)} = \boldsymbol{a}), \tag{58}$$

where

$$G(c,\boldsymbol{a}) = \sum_{\hat{c}_{d_-}\in\mathbb{Z}[j]} \frac{N_{\mathrm{bit}}(c,\hat{c}_{d_-})}{\log_2(M)} F(c-\hat{c}_{d_-},\boldsymbol{a},n_v^{(l)}) \tag{59}$$

and the function $N_{\mathrm{bit}}(c,\hat{c})$ is generalized to the extended symbol set, i.e., $N_{\mathrm{bit}}(c,\hat{c}) = N_{\mathrm{bit}}(c,\hat{c}_{d_-})$ for all $d_- \in \mathbb{Z}[j]$. Unfortunately, the summation over all $\hat{c}_{d_-}$ in (59) consists of an infinite number of terms, prohibiting the exact evaluation of $\mathrm{BER}^{(l)}$. However, the properties of the binary reflected Gray mapping impose not only that horizontally and vertically neighboring symbols in the extended symbol set only differ in one bit, but also that diagonally neighboring symbols differ only in two bits, justifying the approximation $N_{\mathrm{bit}}(c,\hat{c}_{d_-}) \approx N_{\mathrm{bit},\approx}(c-\hat{c}_{d_-})$, where

$$N_{\mathrm{bit},\approx}(x) = \begin{cases} 0 & x = 0 \\ 1 & (\mathrm{Re}\,[x] = 0) \vee (\mathrm{Im}\,[x] = 0)\,. \\ 2 & \text{otherwise} \end{cases} \tag{60}$$

The approximation $N_{\mathrm{bit},\approx}(x)$ is visualized in Fig. 5. Interestingly, the number of errors in the decision areas adjacent to $x = 0$ within the dashed square is not altered. For $M = 4$, the approximation is actually an upper bound as only additional errors are introduced. Replacing in (59) $N_{\mathrm{bit}}(c,\hat{c}_{d_-})$ by $N_{\mathrm{bit},\approx}(c-\hat{c}_{d_-})$ yields the function $G_{\approx}(\mathbf{a})$ that no longer depends on c, i.e.,

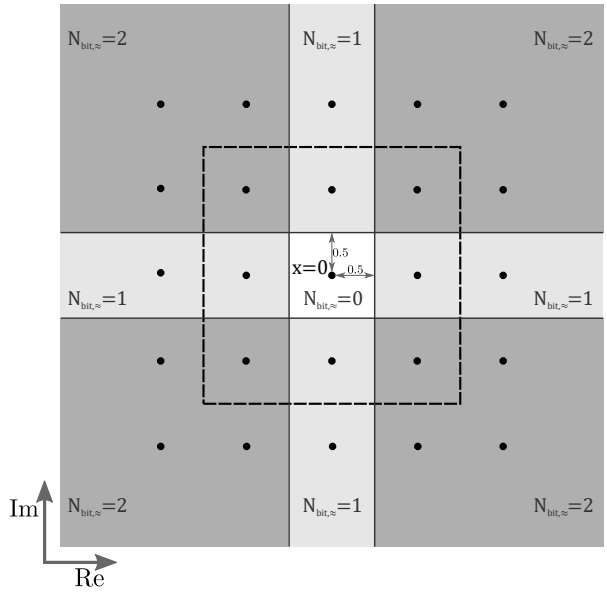


Figure 5. Visualization of the approximation $N_{\mathrm{bit},\approx}(x)$. The white, light gray, and dark gray areas represent the areas for which $N_{\mathrm{bit},\approx} = 0$, $N_{\mathrm{bit},\approx} = 1$ and $N_{\mathrm{bit},\approx} = 2$, respectively.

$$G_{\approx}(\mathbf{a}) = \sum_{x\in\mathbb{Z}[j]} \frac{N_{\mathrm{bit},\approx}(x)}{\log_2(M)} F(x,\mathbf{a},n_v^{(l)}). \tag{61}$$

The infinite summation over x in (61) can then be replaced by the summation over two regions characterized by $N_{\mathrm{bit},\approx}(x) = 1$ and $N_{\mathrm{bit},\approx}(x) = 2$, respectively. This results in

$$G_{\approx}(\mathbf{a}) = Q\left(\frac{0.5 - \mathrm{isi}_R^{(l)}(\boldsymbol{a})}{\sigma_{n_v^{(l)}}}\right) + Q\left(\frac{0.5 + \mathrm{isi}_R^{(l)}(\boldsymbol{a})}{\sigma_{n_v^{(l)}}}\right) + Q\left(\frac{0.5 - \mathrm{isi}_I^{(l)}(\boldsymbol{a})}{\sigma_{n_v^{(l)}}}\right) + Q\left(\frac{0.5 + \mathrm{isi}_I^{(l)}(\boldsymbol{a})}{\sigma_{n_v^{(l)}}}\right), \tag{62}$$

where $Q(\cdot)$ represents the tail distribution function of the standard normal distribution, $\mathrm{isi}_R^{(l)}(\boldsymbol{a}) = \mathrm{Re}\left[\mathrm{isi}^{(l)}(\mathbf{a})\right]$, $\mathrm{isi}_I^{(l)}(\boldsymbol{a}) = \mathrm{Im}\left[\mathrm{isi}^{(l)}(\mathbf{a})\right]$, and $\sigma_{n_v^{(l)}}$ equals the standard deviation of the real part of $n_v^{(l)}$, which is given by $\sigma_w^{(l)} = \sqrt{\frac{(\boldsymbol{W}\boldsymbol{R}_{\mathrm{n}}\boldsymbol{W}^{\boldsymbol{H}})_{l,l}}{8\Delta^2}}$. Substituting $G_{\approx}(\mathbf{a})$ for $G(c,\mathbf{a})$ in (59) yields the approximation of $\mathrm{BER}^{(l)}$. Moreover, the rotational symmetry of the QAM constellation can be exploited by remarking that $\Pr(\boldsymbol{a}) = \Pr(-\boldsymbol{a}) = \Pr(j\boldsymbol{a}) = \Pr(-j\boldsymbol{a})$, simplifying (62) to

$$\mathrm{BER}^{(l)} \approx \frac{1}{\log_2(M)}\mathbb{E}_{\boldsymbol{a}}\left[4Q\left(\frac{0.5 - \mathrm{isi}_R^{(l)}(\boldsymbol{a})}{\sigma_{n_v^{(l)}}}\right)\right]. \tag{63}$$

As $N_{\mathrm{bit},\approx}(x)$ maintains the number of bit errors in the decision areas adjacent to the decision area of the symbol c, the approximation (63) is expected to be very accurate when $\mathrm{isi}^{(l)}(\boldsymbol{a})$ and $\sigma_{n_v^{(l)}}$ are small compared to 1, e.g, in the case of powerful equalization and large SNR.

ACKNOWLEDGEMENT

Part of this research has been supported by the Research Foundation Flanders (FWO) under Grant no. EOS-30458693. Jelle Bailleul is supported by an UGent-BOF PhD scholarship.

REFERENCES

[1] G. Forney, "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Trans. Inform. Theory*, vol. 18, no. 3, pp. 363–378, May 1972.

[2] W. van Etten, "Maximum Likelihood Receiver for Multiple Channel Transmission Systems," *IEEE Trans. Commun.*, vol. 24, no. 2, pp. 276–283, Feb. 1976.

[3] N. Al-Dhahir and A. H. Sayed, "The Finite-Length Multi-Input Multi-Output MMSE-DFE," *IEEE Trans. Signal Process.*, vol. 48, no. 10, pp. 2921–2936, Oct. 2000.

[4] J. Bailleul, L. Jacobs, P. Manfredi *et al.*, "MIMO Time-Domain Equalization for High-Speed Continuous Transmission Under Channel Variability," *IEEE Trans. Commun.*, vol. 66, no. 8, pp. 3394–3406, Aug. 2018.

[5] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics Letters*, vol. 7, no. 5, pp. 138–139, Mar. 1971.

[6] H. Harashima and H. Miyakawa, "Matched-transmission technique for channels with intersymbol interference," *IEEE Trans. Commun.*, vol. 20, no. 4, pp. 774–780, Aug. 1972.

[7] O. Simeone, Y. Bar-Ness, and U. Spagnolini, "Linear and nonlinear preequalization/equalization for mimo systems with long-term channel state information at the transmitter," *IEEE Trans. Wireless Commun.*, vol. 3, no. 2, pp. 373–378, Mar. 2004.

[8] M. Joham, D. A. Schmidt, J. Brehmer *et al.*, "Finite-Length MMSE Tomlinson-Harashima Precoding for Frequency Selective Vector Channels," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3073–3088, Jun. 2007.

[9] P. Kabal and S. Pasupathy, "Partial-Response Signaling," *IEEE Trans. Commun.*, vol. 23, no. 9, pp. 921–934, Sep. 1975.

[10] L. Jacobs, J. Bailleul, P. Manfredi *et al.*, "On Partial Response Signaling for MIMO Equalization on Multi-Gbit/s Electrical Interconnects," in *26th European Signal Processing Conference(EUSIPCO 2018)*, Sep. 2018.

[11] N. Eiselt, D. Muench, A. Dochhan *et al.*, "Performance Comparison of 112-Gb/s DMT, Nyquist PAM4, and Partial-Response PAM4 for Future 5G Ethernet-Based Fronthaul Architecture," *Journal of Lightwave Technology*, vol. 36, no. 10, pp. 1807–1814, May 2018.

[12] R. F. H. Fischer, "Lattice-Reduction-Aided Equalization and Generalized Partial-Response Signaling for Point-to-Point Transmission over Flat-Fading MIMO Channels," in *4th Int. Symp. on Turbo Codes Related Topics; 6th Int. ITG-Conference on Source and Channel Coding*, Apr. 2006, pp. 1–6.

[13] R. F. H. Fischer and C. Siegl, "Lattice-reduction-aided equalisation for transmission over intersymbol-interference channels," *Electronics Letters*, vol. 41, no. 17, pp. 969–970, Aug. 2005.

[14] ——, "On the Relation between Lattice-Reduction-Aided Equalization and Partial-Response Signaling," in *2006 Int. Zurich Seminar on Communications*, 2006, pp. 34–37.

[15] R. F. Fischer, C. Windpassinger, C. Stierstorfer *et al.*, "Lattice-Reduction-Aided MMSE Equalization and the Successive Estimation of Correlated Data," *AEU-Intern. Journal of Electronics and Communications*, vol. 65, no. 8, pp. 688 – 693, 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1434841111000069

[16] X. Ma and W. Zhang, "Performance Analysis for MIMO Systems with Lattice-Reduction Aided Linear Equalization," *IEEE Trans. Commun.*, vol. 56, no. 2, pp. 309–318, Feb. 2008.

[17] D. Wubben, R. Bohnke, V. Kuhn *et al.*, "Near-Maximum-Likelihood Detection of MIMO Systems using MMSE-Based Lattice- Reduction," in *2004 IEEE Int. Conf. on Communications (IEEE Cat. No.04CH37577)*, vol. 2, Jun. 2004, pp. 798–802 Vol.2.

[18] Q. Zhou and X. Ma, "Improved Element-Based Lattice Reduction Algorithms for Wireless Communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4414–4421, Sep. 2013.

[19] A. K. Lenstra, H. W. Lenstra, and L. Lovasz, "Factoring Polynomials with Rational Coefficients," *MATH. ANN*, vol. 261, pp. 515–534, 1982.

[20] J. Park, J. Chun, and F. T. Luk, "Lattice Reduction Aided MMSE Decision Feedback Equalizers," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 436–441, Jan. 2011.

[21] M. Niroomand and M. Derakhtian, "A Low Complexity Diversity Achieving Decoder Based on a Two-Stage Lattice Reduction in Frequency-Selective MIMO Channels," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2465–2477, Apr. 2017.

[22] I. Berenguer, J. Adeane, I. J. Wassell *et al.*, "Lattice-Reduction-Aided Receivers for MIMO-OFDM in Spatial Multiplexing Systems," in *2004 IEEE 15th Int. Symp. on Personal, Indoor and Mobile Radio Communications (IEEE Cat. No.04TH8754)*, vol. 2, Sep. 2004, pp. 1517–1521 Vol.2.

[23] A. H. Mehana and A. Nosratinia, "Performance of Linear Receivers in Frequency-Selective MIMO Channels," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2697–2705, Jun. 2013.

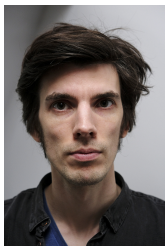
[24] E. Agrell, T. Eriksson, A. Vardy *et al.*, "Closest Point Search in Lattices," *IEEE Trans. Inform. Theory*, vol. 48, no. 8, pp. 2201–2214, Aug. 2002.

[25] A. Schrijver, *Theory of Linear and Integer Programming*. John Wiley and Sons, Jun. 1998.

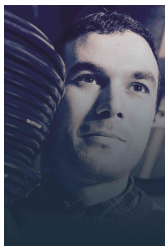
[26] M. Guenach, L. Jacobs, B. Kozicki *et al.*, "Performance analysis of pre-equalized multilevel partial response modulation for high-speed electrical interconnects," *Computers & Electrical Engineering*, vol. 58, pp. 30 – 48, 2017.

[27] J. Bailleul, L. Jacobs, M. Guenach *et al.*, "Robust Spatio-Temporal Partial-Response Signaling over a Frequency-Selective Fading MIMO Channel with Imperfect CSI," in *30st Annual IEEE Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Sep. 2019, pp. 1–7.

Jelle Bailleul received the Master's degree in electrical engineering from Ghent University in 2015. He is now pursuing a Ph.D degree at the Telecommunications and Information Processing Department, Ghent University, Belgium. His research interests include signal processing, modulation and coding and statistical communication theory.



Lennert Jacobs received the M.S. degree and the Ph.D. degree in electrical engineering from Ghent University, Ghent, Belgium, in 2006 and 2012, respectively. He is currently serving as a post-doctoral researcher in the Department of Telecommunications and Information Processing at Ghent University. His research interests are in fading channels, MIMO techniques, signal processing, and modulation and coding for digital communications, as well as in automotive radar and sensor fusion.



Mamoun Guenach (SM'11) is a research scientist with Imec Leuven since 2019. He graduated in 1995 (EMI, Morocco), and received a Ph.D. degree in 2002 (UCL, Belgium). He served as a post-doctoral researcher at Ghent University (2002-2006) where, since 2015, he is a part-time visiting professor. He was a member of technical staff of Nokia Bell Labs (2006-2019) where he was mainly working on high-speed wired communication technologies. He published several conference and journal papers, patents, and standardization contributions, and is currently associate editor of IEEE Transactions on Communications. His main research interests include coding, modulation, equalization, cell-free massive MIMO both for high-speed wired and wireless communication technologies.

Marc Moeneclaey (M'93, SM'99, F'02) received the diploma of electrical engineering and the Ph.D. degree in electrical engineering from Ghent University, Gent, Belgium, in 1978 and 1983, respectively. He is Professor at the Department of Telecommunications and Information Processing (TELIN), Gent University. His main research interests are in statistical communication theory, carrier and symbol synchronization, bandwidth-efficient modulation and coding, spread-spectrum, satellite and mobile communication. He is the author of more than 500 scientific papers in international journals and conference proceedings. Together with Prof. H. Meyr (RWTH Aachen) and Dr. S. Fechtel (Siemens AG), he coauthors the book Digital communication receivers - Synchronization, channel estimation, and signal processing. (J. Wiley, 1998). He is co-recipient of the Mannesmann Innovations Prize 2000. During the period 1992-1994, was Editor for Synchronization, for the IEEE Transactions on Communications. He served as co-guest editor for special issues of the Wireless Personal Communications Journal (on Equalization and Synchronization in Wireless Communications) and the IEEE Journal on Selected Areas in Communications (on Signal Synchronization in Digital Transmission Systems) in 1998 and 2001, respectively.