Lattice Reduction Aided Precoding Design in Downstream G.fast DSL Networks

Wouter Lanneer, *Member, IEEE,* Carl Nuzman, *Member, IEEE,* Yannick Lefevre, *Member, IEEE,* Paschalis Tsiaflakis, *Member, IEEE,* Werner Coomans, *Member, IEEE,* and Marc Moonen, *Fellow, IEEE*

***Abstract*—As a non-linear precoding alternative to Tomlinson-Harashima precoding (THP), in this paper, so-called lattice reduction aided precoding (LRP) is considered as a crosstalk precompensation technique for downstream transmission in G.fast DSL networks. First, a practically achievable bit-rate expression for LRP is proposed in function of the precoder and integer matrix. The problem then consists of a joint precoder and integer matrix design in order to maximize the weighted sum-rate (WSR) under per-line power constraints. For a fixed integer matrix, zero-forcing (ZF) precoder matrix design simplifies to gain scaling optimization with complex gain scalars, for which a novel successive lower bound maximization method is presented. Additionally, it is established that the achievable ZF-LRP sum-rate is upper bounded by the achievable ZF-THP sum-rate at high SNR. For computing the optimal precoder matrix, on the other hand, an efficient method is developed by leveraging on the equivalence between the WSR maximization and the weighted sum of mean squared error (MSE) minimization, leading to a locally-optimal MMSE-LRP solution. Simulations with a measured G.fast cable binder are provided to compare the proposed LRP schemes with THP schemes.**

I. Introduction

ULTRA-broadband digital subscriber line (DSL) networks like G.fast [2] aim at providing gigabit (i.e. fiber-like) data speeds over very short copper lines (below 100 m) by signaling in high frequencies (up to 212 MHz). The use of such high frequencies leads to increasingly stronger levels of crosstalk interference among the lines within a cable binder [3]. As a result, non-linear precoding (NLP) schemes are currently attracting a lot of attention as an alternative to linear precoding (LP) schemes for far-end crosstalk precompensation in downstream transmission.

NLP schemes include (non-linear) modulo operations at the transmitters and the receivers to transmit equivalent lower power signals. A well-known NLP scheme proposed in DSL is multi-user Tomlinson-Harashima precoding (THP) [4]. However, a disadvantage of THP is the required sequential feedback operation proportional to the number of users (or lines), which in combination with the modulo operations creates a challenging circuit timing problem. In addition, the use of THP schemes may increase the dynamic range at the receiver and is more sensitive to practical non-idealities than LP schemes [5].

In this paper, an alternative NLP scheme is therefore considered, commonly referred to as lattice reduction aided precoding (LRP) [6]–[10]. Although, like THP, LRP uses (non-linear) modulo operations to transmit equivalent lower power signals, it features a complete forward data path, similar to practical LP schemes. This reduces the implementation complexity increase of the NLP scheme and results in a relaxed circuit timing problem compared to THP, especially for scenarios with a large number of users.

A. Main Contributions

In this paper, LRP design is considered for downstream transmission in G.fast DSL networks. Specifically, the LRP scheme precedes the precoder matrix by a (complex) integer matrix, leading to a "reduced" precoder matrix. Using non-linear scalar modulo operations, the integer matrix can be pre-inverted at the transmitter side, such that the receivers are able to detect the desired symbols. A practically achievable bit-rate expression is derived for LRP in function of the precoder matrix and integer matrix, tight at high SNR. The LRP design problem is then formulated as the joint optimization of the precoder matrix and integer matrix in order to maximize the weighted sum-rate (WSR) under per-line power constraints[1].

To tackle this difficult design problem, precoder matrix optimization given a fixed integer matrix is studied. In case of zero-forcing (ZF) precoding, this leads to a gain scaling optimization problem with complex gain scalars (similar to power allocation in linear ZF precoding), for which a novel successive lower bound maximization method is presented. Additionally, it is established that the achievable sum-rate of this ZF-LRP is upper bounded by the achievable sum-rate of ZF-THP at high SNR. This bound is tight when the (output power normalized) reduced precoder matrix is *unitary* and the integer matrix is *unimodular*. To compute the optimal precoder matrix[2], on the other hand, an efficient method is

This research work was carried out at the ESAT Laboratory of KU Leuven in the frame of VLAIO O&O Project nr. HBC.2016.0055 '5GBB' and HBC.2017.1007 'Multi-gigabit Innovations in Access', FWO nr. G.0B1818N 'Real-time adaptive cross-layer dynamic spectrum management for 5th generation broadband copper access networks', Fonds de la Recherche Scientifique - FNRS and the Fonds Wetenschappelijk Onderzoek - Vlaanderen under EOS Project no 30452698 'MUlti-SErvice WIreless NETwork'. This research work is presented in the Ph.D. dissertation of the first author [1]. The scientific responsibility is assumed by its authors.

W. Lanneer and M. Moonen are with the STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, Dept. of Electrical Engineering (ESAT), KU Leuven, BE 3000 Leuven, Belgium (e-mail: {wouter.lanneer, marc.moonen}@esat.kuleuven.be).

C. Nuzman, Y. Lefevre, P. Tsiaflakis, and W. Coomans are with the Copper Access and Indoor Team of Nokia Bell Labs, BE 2018 Antwerp, Belgium and Murray Hill, New Jersey, USA (e-mail: paschalis.tsiaflakis@nokia-bell-labs.com).

[1] Akin to per-antenna power constraints in the wireless field.

[2] Optimal precoding is sometimes also referred to as minimum mean squared error (MMSE) precoding.

developed, by relying on the equivalence between the WSR maximization and the weighted sum of mean square error (MSE) minimization, leading to a locally-optimal MMSE-LRP solution.

Further, a heuristic alternating method is proposed between optimizing the precoder matrix for a fixed integer matrix, and updating the integer matrix based on the lattice reduction of the precoder matrix via the Lenstra-Lenstra-Lovász (LLL) algorithm. Finally, the performance of the proposed LRP schemes is compared with THP schemes, by means of simulations with a measured G.fast cable binder.

B. Related Work

LRP has been introduced for general single-carrier MIMO channels [6]–[9], as a popular low-complexity technique which has been proven to achieve the full diversity order [9]. The original idea of LRP is to precompensate the channel in a suited basis, by decomposing the channel matrix into a reduced one and a unimodular matrix via the LLL algorithm. However, the use of scalar modulo operations (i.e. per-symbol quantization) does not allow for closed-form theoretically achievable bit-rate expressions, such that these prior works typically do not consider the rate maximization problem.

Recently, the rate maximization problem for ZF-LRP has been studied in [10] for downstream G.fast DSL networks. It adopts a closed-form practically achievable bit-rate expression, based on the so-called SNR approximation gap [11]. Subsequently, the authors propose to optimize the bit-rates by introducing a gain scaling with real scalars. The gain scaling problem is approximated in the high SNR regime as a signomial program, which can be locally optimally solved by interior point methods. The approach in this paper differs in that gain scaling with *complex* scalars is considered for ZF-LRP, for which an iterative method is proposed that obtains a local optimum of the *exact* gain scaling problem, and that also the case of MMSE-LRP design is considered. In addition, the combination of lattice reduction and THP in the context of downstream G.fast DSL betworks has been considered in [12]. However, this scheme inherits the sequential feedback operation and complicated hardware implementation of THP.

Closely related to LRP are integer-forcing (IF) precoding schemes [13]–[15], where instead of scalar modulo operations, more involved nested *lattice* encoding and decoding and message precoding is used on blocks of symbols. With appropriate lattice construction and the dimensionality of the blocks going up to infinity, IF precoding leads to closed-form theoretically achievable bit-rate expressions [13]. In [14] an iterative algorithm based on uplink-downlink duality for sum-rate (SR) maximization under a sum-power constraint is proposed for IF precoding systems employing a different shaping lattice for every user. The more practical case of a single shaping lattice for all users is considered in [15]. In particular, a (regularized) ZF precoder structure is proposed with arbitrary scaling by a complex diagonal matrix, referred to as diagonally-scaled exact integer-forcing (DIF). A closed-form expression for the SR optimal scaling and integer matrix under a sum-power constraint is established, for the specific case of two users in the high SNR regime. Different from [15], in this paper more than two users and realistic per-line power constraints are considered, such that optimal closed-form expressions are not available.

C. Organization and Notation

This paper is organized as follows. Section II introduces the LRP system model for downstream G.fast transmission and derives an achievable bit-rate expression. Section III states the general LRP design problem and proposes an alternating optimization approach. Section IV and Section V address the subproblems corresponding to the precoder optimization given a fixed integer matrix for ZF-LRP and MMSE-LRP, respectively. Section VI presents simulation results for a G.fast cable binder. Finally, Section VII concludes the paper.

Lowercase boldface letters are used to denote vectors and uppercase boldface letters are used to denote matrices. $\mathbb{C}^{m\times n}$ and $\mathbb{Z}^{m\times n}[j] \equiv \mathbb{Z}^{m\times n} + j\mathbb{Z}^{m\times n}$ are used to denote the $m \times n$ dimensional complex and complex integer space, respectively. $\mathbf{I}_A$ is used as the identity matrix of size A, $(\cdot)^T$ as the transpose, $(\cdot)^H$ as the Hermitian transpose, $(\cdot)^*$ as the complex conjugate, $\mathcal{E}\{\cdot\}$ as expectation, $\mathrm{Re}(\cdot)$ as real part, $[\mathbf{X}]_{ij}$ as the i,j-th element of $\mathbf{X}$, $[x]_a^b$ as $\min(\max(x,a),b)$, $\mathrm{Tr}\{\cdot\}$ as the trace, $\det(\cdot)$ as matrix determinant, and $|\cdot|$ as scalar absolute value.

II. LRP for Downstream G.fast Transmission

A. Downstream G.fast Transmission Model

Downstream G.fast transmission in a cable binder with N lines or users is considered. Assuming standard synchronous DMT modulation with a sufficiently long cyclic prefix, the transmission is modeled independently on each tone (or frequency sub-carrier) $k = [1,\ldots,K]$ as

$$\mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k + \mathbf{z}_k. \tag{1}$$

$\mathbf{x}_k \triangleq [x_k^1,\ldots,x_k^N]^T$ is the transmit vector on tone k, with x_k^n the signal transmitted on line n. $\mathbf{y}_k \triangleq [y_k^1,\ldots,y_k^N]^T$ is the receive vector on tone k, with y_k^n the signal received by user n. $\mathbf{z}_k \triangleq [z_k^1,\ldots,z_k^N]^T$ is the vector of uncorrelated additive noise signals on tone k, with $\sigma_k \triangleq \mathcal{E}\{|z_k^n|^2\}$ denoting the noise PSD. $\mathbf{H}_k \triangleq [h_k^{n,m}]$ denotes the $N \times N$ channel matrix on tone k. The diagonal elements of $\mathbf{H}_k$ contain the direct channels whereas the off-diagonal elements contain the crosstalk channels. Perfect knowledge of the channel matrices is assumed.

In G.fast per-line spectral mask and total power constraints are included:

$$\mathcal{E}\{|x_k^l|^2\} \le P_k^{\mathrm{mask}}, \qquad \forall k,l \tag{2}$$

$$\sum_k \mathcal{E}\{|x_k^l|^2\} \le P^{\mathrm{total}}, \qquad \forall l. \tag{3}$$

For ease of representation, the per-tone case is adopted and the tone index is dropped in the remainder of paper. Tone indices are briefly re-introduced in Appendix B in order to address the handling of the total power constraints in (3) algorithm-wise.

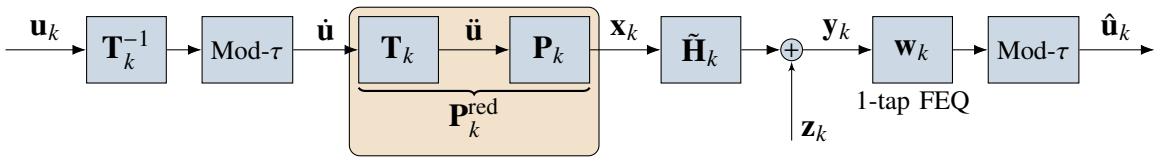


Fig. 1. A block diagram of the LRP scheme on tone k.

B. Transmitter and Receiver Processing

In standard LP, the transmit vector $\mathbf{x}$ is formed by

$$\mathbf{x} = \mathbf{P}\mathbf{u} \tag{4}$$

where $\mathbf{P}$ is the $N \times N$ precoder matrix and $\mathbf{u} \triangleq [u^1, \ldots, u^N]$ is the data vector. The data vector consists of QAM symbols with possibly different constellation sizes for each user, and is assumed to be independently and identically distributed (i.i.d.) with unit power, i.e., $\mathcal{E}\{\mathbf{u}\mathbf{u}^H\} = \mathbf{I}_N$.

The main idea of LRP, as shown in Fig. 1, is then to precede the precoder matrix by a (complex) integer matrix $\mathbf{T}$, leading to a "reduced" precoder matrix

$$\mathbf{P}^{\text{red}} = \mathbf{P}\mathbf{T}. \tag{5}$$

It is required that $\mathbf{T} \in \mathbb{Z}[j]^{N\times N}$ is invertible (i.e. of full-rank). In case $|\det(\mathbf{T})| = 1$, it holds that $\mathbf{T}$ is unimodular and $\mathbf{T}^{-1}$ is also complex integer. The overall goal is to improve the mathematical properties of $\mathbf{P}^{\text{red}}$ over $\mathbf{P}$ (e.g. a reduction of the orthogonality defect and condition number), and hence to achieve a performance gain in terms of a reduced transmit power penalty.

The multiplication with $\mathbf{T}$ has to be compensated, which in downstream transmission, without co-located receivers, may be implemented at the transmitter by multiplying the data vector with $\mathbf{T}^{-1}$, i.e, $\dot{\mathbf{u}} = \mathbf{T}^{-1}\mathbf{u}$. Unfortunately, this may increases the power of $\dot{\mathbf{u}}$ (and thus of $\mathbf{x}$). To deal with this issue, as also used in THP [16], [17], the multiplication is followed by a scalar (component-wise) non-linear modulo operation to bound the value of the transmit signal

$$\dot{\mathbf{u}} = \left[\mathbf{T}^{-1}\mathbf{u}\right] \text{mod-}\tau \tag{6}$$

where the modulo operation is applied to both dimensions of the QAM constellation (corresponding to a complex x)

$$[x] \text{mod-}\tau \triangleq [\text{Re}(x)] \text{mod-}\tau + j[\text{Im}(x)] \text{mod-}\tau \tag{7}$$

with for real y

$$[y] \text{mod-}\tau \triangleq y - \tau\left\lfloor \frac{y + \tau/2}{\tau} \right\rfloor, \qquad y \in \mathbb{R} \tag{8}$$

and $\lfloor\cdot\rfloor$ denoting the floor function. The complex range of mod-τ is denoted by $\Omega_\tau = \{a + jb : a, b \in [-\tau/2, \tau/2)\}$, i.e., $[x] \text{mod-}\tau \in \Omega_\tau$ for all complex x. The base τ of the modulo operation, which is the same for all the users, needs to be chosen such that Ω_τ encloses the QAM constellations of all users. The margin between the boundary $\tau/2$ and any point at the edge of a constellation should be at least d where $2d$ is the minimum distance between adjacent constellation points[3] for square and cross-shaped QAM constellations (i.e. d corresponds to the constellation spacing); otherwise, there will be information loss. However, smaller margins yield larger power savings in $\dot{\mathbf{u}}$. Consequently, a valid choice for τ is

$$\tau = \max_n \tilde{\tau}_n \tag{9}$$

where $\tilde{\tau}_n$ is the (1-dim.) constellation boundary of user n. The values for $\tilde{\tau}_n$ can be derived in function of the constellation size $M = 2^b$ and scaling d (see e.g. [18] for the constellations adopted in G.fast [2]).

The modulo operation in (6) may also be seen as the addition of a scaled complex integer perturbation vector $\mathbf{d} \in \tau\mathbb{Z}^{N\times 1}[j]$, i.e.,

$$\dot{\mathbf{u}} \triangleq \mathbf{T}^{-1}\mathbf{u} + \mathbf{d}. \tag{10}$$

Further, let us introduce the signal vector $\ddot{\mathbf{u}} \triangleq \mathbf{T}\dot{\mathbf{u}} = \mathbf{u} + \mathbf{T}\mathbf{d}$, where $\mathbf{T}\mathbf{d} \in \tau\mathbb{Z}^{N\times 1}[j]$ since $\mathbf{T}$ is a complex integer matrix. Now, the transmit vector $\mathbf{x}$ can be expressed as follows:

$$\mathbf{x} = \mathbf{P}\ddot{\mathbf{u}} = \mathbf{P}\mathbf{T}\dot{\mathbf{u}} = \mathbf{P}(\mathbf{u} + \mathbf{T}\mathbf{d}), \tag{11}$$

which is to be compared with (4) for the LP scheme.

The receiver of user n consists of a per-tone (complex) one-tap frequency domain equalizer (FEQ) followed by a second modulo operation to cancel out the first modulo operation, i.e.,

$$\begin{aligned}
\hat{u}_n &= [w_n y_n] \text{mod-}\tau && (12)\\
&= \Big[\ddot{u}_n + \underbrace{(w_n y_n - \ddot{u}_n)}_{e_n}\Big] \text{mod-}\tau && (13)\\
&= [u_n + [\mathbf{T}\mathbf{d}]_n + e_n] \text{mod-}\tau\\
&= [u_n + e_n] \text{mod-}\tau,
\end{aligned}$$

where e_n is the residual crosstalk interference and noise. From equation (12) it follows that a common base-τ for all users is indeed needed for the modulo operations in (6), instead of using a separate base-τ_n for each user n. In case of the latter, the perturbation term $[\mathbf{T}\mathbf{d}]_n$ at the receiver of user n corresponds to a sum of multiple integers with different scalings $\{\tau_m\}$ which cannot be completely removed by a single mod-τ_n operation.

C. Covariance Matrices of $\dot{\mathbf{u}}$ and $\ddot{\mathbf{u}}$

To determine the transmit power of $\mathbf{x}$, the covariance matrices of the signal vectors $\dot{\mathbf{u}}$ and $\ddot{\mathbf{u}}$ have to be known. In the proposition below, it is shown that when the i.i.d. data vector $\mathbf{u}$ has a uniform distribution over Ω_τ and $\mathbf{T}$ is unimodular, the components of $\dot{\mathbf{u}}$ are pairwise independent and uniformly distributed over Ω_τ. Further, it is shown that this result holds also for QAM-distributed $\mathbf{u}$ and/or non-unimodular $\mathbf{T}$ when including dithering.

[3]For the 2-QAM and 8-QAM constellations specified in G.fast [2] the minimum distance is actually slightly larger and equals $2\sqrt{2}d$.

Proposition 1. If the components of $\mathbf{u}$ are i.i.d. and have a uniform distribution over Ω_τ with unit power and $\mathbf{T}$ is a complex unimodular matrix, then the components of $\dot{\mathbf{u}} = \left[\mathbf{T}^{-1}\mathbf{u}\right] \bmod\text{-}\tau$ are uniformly distributed over Ω_τ with unit power and pairwise independent. Further, $\dot{\mathbf{u}}$ and $\ddot{\mathbf{u}}$ have the following (cross)-covariance matrices:

$$\mathcal{E}\{\dot{\mathbf{u}}\dot{\mathbf{u}}^H\} = \mathbf{I}_N \tag{14}$$

$$\mathcal{E}\{\ddot{\mathbf{u}}\ddot{\mathbf{u}}^H\} = \mathbf{T}\mathbf{T}^H = \begin{bmatrix} * & \mathbf{T}_{12}\mathbf{I}^{c,H}_{N'} \\ \mathbf{I}^{c}_{N'}\mathbf{T}^H_{12} & \mathbf{I}_{N'} \end{bmatrix} \tag{15}$$

$$\mathcal{E}\{\dot{\mathbf{u}}\mathbf{u}^H\} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}^{c,H}_{N'} \end{bmatrix} \tag{16}$$

$$\mathcal{E}\{\ddot{\mathbf{u}}\mathbf{u}^H\} = \mathbf{T}\,\mathcal{E}\{\dot{\mathbf{u}}\mathbf{u}^H\} = \begin{bmatrix} \mathbf{0} & \mathbf{T}_{12}\mathbf{I}^{c,H}_{N'} \\ \mathbf{0} & \mathbf{I}_{N'} \end{bmatrix}. \tag{17}$$

where $\mathbf{I}^c_{N'}$ is a diagonal matrix of size $N' \geq 0$, of which the diagonal entries are complex integers with unit magnitude (i.e. the entries equal either ± 1 or $\pm j$).

Proof: See the proof in Appendix A. ∎

The components of $\mathbf{u}$ being uniformly distributed over Ω_τ is a tight approximation for high SNR channels (with large constellation sizes), which typically dominate the G.fast spectrum. Basically, the conditions in (14)-(17) are tighter as the constellation sizes and the number of users N grow, and if $\mathbf{T}^{-1}$ is a dense matrix with large integer entries. However, for small constellations sizes, the covariance matrix of $\dot{\mathbf{u}}$ (and thus of $\ddot{\mathbf{u}}$) is generally not known.

Nonetheless, for any constellation size, or number of users, or even in case of a non-unimodular $\mathbf{T}$, these results still hold when including the concept of common randomness in the LRP transmission scheme. Proposed in [19], [20], this common randomness is realized by an i.i.d. dither vector $\mathbf{u}^{\text{di}}$ with uniform distribution over Ω_τ, which should be known at both the transmitter and receiver [e.g. through the use of a common seed in a (pseudo)-random number generator]. The dither vector is added to the data vector at the transmitter and subtracted at the receiver, such that (6) and (12) transform into[4]

$$\dot{\mathbf{u}} = \left[\mathbf{T}^{-1}\mathbf{u} + \mathbf{u}^{\text{di}}\right] \bmod\text{-}\tau \tag{18}$$

$$\hat{u}_n = \left[w_n y_n - [\mathbf{T}\,\mathbf{u}^{\text{di}}]_n\right] \bmod\text{-}\tau, \qquad \forall n. \tag{19}$$

Similar to the proof of Proposition 1, it can be shown that the components of $\dot{\mathbf{u}}$ in (18) are uniformly distributed over Ω_τ and pairwise independent, regardless of the distribution of $\mathbf{u}$. A practical consequence of dithering is thus $\dot{\mathbf{u}}$ having an accurate a priori known transmit power even for low constellation sizes (despite the precoding loss, defined later). Additionally, dithering strengthens conditions (16) and (17) to

$$\mathcal{E}\{\dot{\mathbf{u}}\mathbf{u}^H\} = \mathbf{0}_N \qquad \text{and} \qquad \mathcal{E}\{\ddot{\mathbf{u}}\mathbf{u}^H\} = \mathbf{0}_N. \tag{20}$$

D. Achievable Bit-Rate Expression

The key performance metric considered in this paper is the capacity in terms of the maximum achievable bit-rate. Assume that the complex data symbol u_n is uniformly distributed over Ω_τ with unit variance (i.e. $\tau^2 = 6$) and denote the effective noise power as $\varepsilon_n \triangleq \mathcal{E}\{|e_n|^2\}$, where e_n is defined in (13). Then the maximum achievable bit-rate of user n (in bits/channel use) obtained by the LRP scheme can be shown to be lower bounded by:

$$\begin{aligned} I(u_n;\hat{u}_n) &= h(\hat{u}_n) - h(\hat{u}_n|u_n) = 2\log_2(\tau) - h([e_n] \bmod\text{-}\tau\,|u_n) \\ &\geq 2\log_2(\tau) - h([e_n] \bmod\text{-}\tau) \\ &\geq 2\log_2(\tau) - \log_2(\pi e \varepsilon_n) \\ &= \log_2(1/\varepsilon_n) - \log_2(\pi e/6) \end{aligned} \tag{21}$$

where $h(\cdot)$ denotes the differential entropy. The first inequality holds since conditioning never increases entropy, whereas the second inequality holds since $h([e_n] \bmod\text{-}\tau)$ is upper bounded by the differential entropy of a circular symmetric Gaussian random variable with variance ε_n. The constant factor $1/2\log_2(\pi e/6)$ is the well-known shaping loss of 0.255 bits per real dimension at high SNR. Note that IF precoding [13]–[15] is able to recover this shaping loss by using message precoding and nested lattice encoding and decoding on blocks of symbols, together with appropriate lattice construction and the dimensionality of the blocks going up to infinity (instead of using scalar per-symbol modulo operations). In addition, (21) is maximized by using a FEQ that minimizes the effective noise power ε_n, described by the following MMSE equalizer and MMSE expression

$$\begin{aligned} w_n^{\text{opt}} &\triangleq \frac{\mathcal{E}\{y_n^*\ddot{u}_n\}}{\mathcal{E}\{|y_n|^2\}} = \frac{\sum_m \left[\mathbf{HP}\right]^*_{nm}\left[\mathbf{T}\mathbf{T}^H\right]_{nm}}{\sum_m\sum_p \left[\mathbf{HP}\right]_{nm}\left[\mathbf{T}\mathbf{T}^H\right]_{mp}\left[\mathbf{HP}\right]^*_{np} + \sigma} \\ &= \frac{\left[\mathbf{HPTT}^H\right]^*_{nn}}{\left[\mathbf{HPTT}^H\mathbf{P}^H\mathbf{H}^H + \sigma\mathbf{I}_N\right]_{nn}} \end{aligned} \tag{22}$$

$$\begin{aligned} \varepsilon_n^{\text{opt}} &\triangleq \mathcal{E}\left\{\left|\ddot{u}_n - w_n^{\text{opt}} y_n\right|^2\right\} \\ &= \left[\mathbf{T}\mathbf{T}^H\right]_{nn} - \frac{\left|\left[\mathbf{HPTT}^H\right]_{nn}\right|^2}{\left[\mathbf{HPTT}^H\mathbf{P}^H\mathbf{H}^H + \sigma\mathbf{I}_N\right]_{nn}}. \end{aligned} \tag{23}$$

Although (21) provides a theoretical achievability, the SNR approximation gap is adopted for LRP in this paper, which is commonly used in DSL networks [11]. It guarantees practical achievability in the case that QAM constellations, practical coding schemes and noise margins are used. This practically achievable bit-rate is modeled by

$$b_n = \left\lfloor \log_2\left(1 + \frac{\text{SNR}_n(\mathbf{P},\mathbf{T})}{\Gamma}\right)\right\rfloor^{b^{\max}} \tag{24}$$

where $\text{SNR}_n(\mathbf{P},\mathbf{T})$ is the receiver SNR, and Γ is the SNR approximation gap or SNR gap. For uncoded square QAM constellations, the SNR gap is closely approximated to 9.58 dB for the G.fast target bit error rate (BER) of 10^{-7} [21]. When considering trellis coded modulation, a coding gain $\Gamma_c = 5.2$ dB may be subtracted from Γ_{unc} [22]. Additionally, in DSL a 6 dB noise margin is added. A maximum bit loading (bit cap) $b^{\max}$ is also imposed. The total data-rate across all tones k is

$$R_n = f_s \sum_k b_k^n \tag{25}$$

where f_s is the DMT symbol rate.

[4] In case that $\mathbf{T}$ is unimodular, dithering may be equivalently implemented by $\dot{\mathbf{u}} = \left[\mathbf{T}^{-1}(\mathbf{u} + \mathbf{u}^{\text{di}})\right] \bmod\text{-}\tau$ and $\hat{u}_n = \left[w_n y_n - [\mathbf{u}^{\text{di}}]_n\right] \bmod\text{-}\tau$, which admits a slightly easier implementation.

To be able to use (24), it is necessary to neglect the modulo operation at the receiver. This is tight at high SNR values, because in low-BER systems, where $\Pr(e_n > d)$ is very small, it holds for fairly large constellation sizes in (12) that

$$[u_n + e_n] \bmod \tau \approx u_n + e_n, \qquad \forall n. \tag{26}$$

As a result, the LRP scheme can be viewed as a set of N parallel AGWN channels, wherefore the effective noise power ε_n corresponds to a ZF-FEQ assumption[5], i.e.,

$$w_n = \frac{\mathcal{E}\{|\ddot{u}_n|^2\}}{\mathcal{E}\{y_n \ddot{u}_n^*\}} = \frac{\left[\mathbf{T}\mathbf{T}^H\right]_{nn}}{\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\right]_{nn}} \tag{27}$$

and

$$e_n = \frac{\left[\mathbf{T}\mathbf{T}^H\right]_{nn}}{\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\right]_{nn}} \left(\sum_m [\mathbf{H}\mathbf{P}]_{nm} \ddot{u}_m + z_n \right) - \ddot{u}_n, \tag{28}$$

such that $\mathrm{SNR}_n(\mathbf{P}, \mathbf{T}) \triangleq \mathcal{E}\{|u_n|^2\}/\mathcal{E}\{|e_n|^2\}$ is given by

$$\frac{\left|\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\right]_{nn}\right|^2 \Big/ \left[\mathbf{T}\mathbf{T}^H\right]_{nn}}{\left[\mathbf{T}\mathbf{T}^H\right]_{nn} \left(\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\mathbf{H}^H\right]_{nn} + \sigma \right) - \left|\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\right]_{nn}\right|^2}. \tag{29}$$

Note that (29) simplifies to the well-known SNR expression for LP schemes when $\mathbf{T}$ is the identity matrix. Additionally, an important condition for (29) to be valid is that u_n and e_n are uncorrelated. If the components of $\mathbf{u}$ are independently and uniformly distributed (which is an approximation for large constellation sizes) together with $\mathbf{T}$ being unimodular, it can be verified that $\mathcal{E}\{e_n u_n^*\} = 0, \forall n$, based on Proposition 1. Moreover, this condition holds regardless of the distribution of $\mathbf{u}$ or the unimodularity of $\mathbf{T}$ when the use of dithering is assumed.

Two losses are however ignored in (24) arising from the non-linear modulo operations. Fortunately, since these losses are only significant for low-SNR channels (with small constellation sizes), they have a low impact on the total data-rates aggregated over the full G.fast spectrum.

A first loss is the so-called precoding loss resulting from the modulo operation at the transmitter, such that $\dot{u}_n$ is uniformly distributed over Ω_τ, having a slightly larger variance than the QAM distributed u_n (see e.g. [23]). Note that due to the common modulo-base τ in (9) for all users, this power loss is determined by the user with the smallest constellation, in contrast to THP where per-user modulo-bases τ_n are used [18].

A second loss is the neglected modulo loss in (26) entailing in practice a wrap-around effect at the edges of the QAM constellation where received symbols may be shifted to the other side of the constellation, increasing the average number of nearest neighbors. This loss is quantified in [22] for all constellations and channel coding schemes included in G.fast [2]. The modulo operations may be skipped for users corresponding to the last N' rows of $\mathbf{T}$ (except if dithering is used), avoiding any modulo losses.

[5]Since in a true AGWN channel without a modulo operation at the receiver, the use of a MMSE-FEQ over a ZF-FEQ provides no capacity gain.

III. General LRP Design

The general problem statement corresponds to a joint design of the precoder matrix $\mathbf{P}$ and integer matrix $\mathbf{T}$ in order to maximize the achievable WSR under per-line power constraints

$$\begin{aligned} \underset{\mathbf{P}\in\mathbb{C}^{N\times N}, \mathbf{T}\in\mathbb{Z}^{N\times N}[j]}{\text{maximize}} \quad & \sum_n \alpha_n b_n \\ \text{s.t.} \quad & \sum_m |[\mathbf{P}\mathbf{T}]_{lm}|^2 \le P^{\text{mask}}, \qquad \forall l, \\ & \mathrm{Rank}(\mathbf{T}) = N \end{aligned} \tag{30}$$

where α_n denotes the weight of user n and $l \in [1, \cdots, N]$ the line index. The full-rank constraint guarantees that $\mathbf{T}$ is invertible and $|\det(\mathbf{T})| \ge 1$, which is in fact a relaxation of the unimodularity constraint (i.e. $|\det(\mathbf{T})| = 1$). Note that the LRP scheme is able to handle non-unimodular integer $\mathbf{T}$ matrices as well, since the bit-rate expression (24) is valid as long as $\dot{\mathbf{u}}$ is uniformly distributed and independent of e_n (which is always exactly the case if dithering is used).

Since the joint design of $\mathbf{P}$ and $\mathbf{T}$ in (30) is a difficult task, it is appropriate to focus on solving two subproblems instead. The first subproblem amounts to optimizing the precoder matrix $\mathbf{P}$ given a fixed integer matrix $\mathbf{T}$. This subproblem will be addressed in Section IV and V for ZF-LRP and MMSE-LRP, respectively. The second subproblem corresponds to optimizing $\mathbf{T}$ given a fixed $\mathbf{P}$, i.e.,

$$\begin{aligned} \underset{\mathbf{T}\in\mathbb{Z}^{N\times N}[j]}{\text{maximize}} \quad & \sum_n \alpha_n b_n \\ \text{s.t.} \quad & \sum_m |[\mathbf{P}\mathbf{T}]_{lm}|^2 \le P^{\text{mask}}, \qquad \forall l \\ & \mathrm{Rank}(\mathbf{T}) = N. \end{aligned} \tag{31}$$

Unfortunately, due to its integer nature, solving (31) is still a difficult task. As a consequence, considering the large number of tones ($K = 4000$) in the G.fast spectrum, an efficient low-complexity solution of (31) is used, by means of the well-known (complex) LLL algorithm [24], [25] for lattice basis reduction, which produces a unimodular $\mathbf{T}$ matrix. Note that for the particular case of a ZF precoder with flat gain scaling and a sum-power constraint, (31) is equivalent to designing a $\mathbf{T}$ that minimizes the Frobenius norm of $\mathbf{P}\mathbf{T}$, corresponding to a successive minima problem for which an optimal algorithm is provided in [7].

Alternating between the two subproblems may significantly improve performance. This is motivated by the fact that an integer matrix $\mathbf{T}$ reducing $\mathbf{P}$, is not necessarily a good reducer for $\mathbf{P}'$ which has been optimized for $\mathbf{T}$. However, since the LLL algorithm is merely a suboptimal heuristic for solving (31), such an alternating algorithm does not always feature a monotonically increasing objective value. This is pragmatically handled here by limiting the number of iterations and retaining the $\{\mathbf{P}, \mathbf{T}\}$-pair that achieves the highest WSR. Nevertheless, as will be shown in Section VI, this approach proves to be effective in G.fast DSL networks.

IV. ZF-LRP Design Given A Fixed $\mathbf{T}$

In this section ZF-LRP design is studied given a fixed integer $\mathbf{T}$ matrix. First, a successive lower bound maximization

method for complex gain scaling is proposed, similar to power allocation in linear ZF precoding design. Second, it is established that the achievable ZF-LRP sum-rate is always upper bounded by the achievable ZF-THP sum-rate at high SNR values, leading to a sum-rate gap between both precoding schemes. This sum-rate gap is zero solely for the case that the (output power normalized) reduced ZF precoder is exactly unitary and $\mathbf{T}$ is unimodular.

A. Gain Scaling Optimization

The ZF criterion leads to the following precoder matrix structure[6]

$$\mathbf{P} = \mathbf{H}^{-1}\mathbf{S} \tag{32}$$

where $\mathbf{S} \triangleq \text{diag}\{\mathbf{s}\}$ is a complex diagonal gain scaling matrix with $\mathbf{s} \triangleq [s_1, \ldots, s_N]^T$, offering some degree of freedom that can be exploited to maximize performance. The corresponding gain scaling optimization problem is then formulated as:

$$\underset{\mathbf{s}\in\mathcal{S}}{\text{maximize}} \quad f(\mathbf{s}) \tag{33a}$$

$$\text{s.t.} \quad \sum_m \left|\left[\mathbf{H^{-1}ST}\right]_{lm}\right|^2 \leq P^{\text{mask}}, \qquad \forall l \tag{33b}$$

with

$$f(\mathbf{s}) = \sum_n \alpha_n \log_2\left(1 + \frac{|s_n|^2}{\Gamma\sigma}\right) \tag{34}$$

where $\mathcal{S} \triangleq \left\{\mathbf{s} \in \mathbb{C}^{N\times 1} \middle| |s_n| \leq \sqrt{\Gamma\sigma(2^{b^{\max}}-1)}, \forall n\right\}$ denotes the convex and compact set of gain scalars that satisfy the maximum bit loading constraint, which is translated here into a maximum magnitude constraint.

Fortunately, (33b) is a convex quadratic constraint. Defining $\mathbf{P}^{\text{red}} \triangleq \mathbf{H^{-1}ST}$ and $\tilde{h}_{nm} \triangleq \left[\mathbf{H}^{-1}\right]_{nm}$ and $t_{nm} \triangleq [\mathbf{T}]_{nm}$ yields

$$\begin{aligned}\left|\left[\mathbf{P}^{\text{red}}\right]_{lm}\right|^2 &= \sum_i \sum_j \tilde{h}_{li}\tilde{h}^*_{lj}\, s_i s_j^*\, t_{im}t^*_{jm} \\ &= \sum_i \left|\tilde{h}_{li}s_i t_{im}\right|^2 + \sum_i \sum_{j\neq i} s_i s_j^* \tilde{h}_{li}\tilde{h}^*_{lj} t_{im}t^*_{jm},\end{aligned} \tag{35}$$

such that the transmit power on line l may be re-written as $\sum_m \left|\left[\mathbf{P}^{\text{red}}\right]_{lm}\right|^2 = \mathbf{s}^H\mathbf{A}_l\mathbf{s}$ with $\mathbf{A}_l$ given by

$$[\mathbf{A}_l]_{ji} = \begin{cases} \sum_m \left|\tilde{h}_{li}t_{im}\right|^2 & \text{if } i = j \\ \sum_m \tilde{h}_{li}\tilde{h}^*_{lj}t_{im}t^*_{jm} & \text{if } i \neq j. \end{cases} \tag{36}$$

Since $\sum_m \left|\left[\mathbf{P}^{\text{red}}\right]_{lm}\right|^2$ always produces a positive value, it follows that $\mathbf{s}^H\mathbf{A}_l\mathbf{s} > 0$ for any non-zero $\mathbf{s}$ vector, which means the $\{\mathbf{A}_l\}$-matrices are positive definite by design. Consequently, (33) may be equivalently reformulated as

$$\underset{\mathbf{s}\in\mathcal{S}}{\text{maximize}} \quad f(\mathbf{s}) \tag{37a}$$

$$\text{s.t.} \quad \mathbf{s}^H\mathbf{A}_l\mathbf{s} \leq P^{\text{mask}}, \qquad \forall l. \tag{37b}$$

Unfortunately, the objective function $f(\mathbf{s})$ is non-concave due to the absolute squared gain scalars $|s_n|^2$, making (37) a non-convex problem in general. Observe that when $\mathbf{T} = \mathbf{I}_N$ (i.e. the case of ZF-LP), the $\{\mathbf{A}_l\}$-matrices are diagonal and (37b) is solely in function of $|s_n|^2$. Hence, in this case the $|s_n|^2, \forall n$ may be taken directly as optimization variables such that (33) and (37) reduce to a standard convex power allocation problem with linear power constraints, solvable by means of the well-known water-filling method.

To solve (37) for any $\mathbf{T}$ matrix, it is used that $|s_n - \bar{s}_n|^2 \geq 0$ and thus $|s_n|^2 \geq 2\cdot\text{Re}(\bar{s}_n^* s_n) - |\bar{s}_n|^2$ for any complex scalar $\bar{s}_n$. In combination with $\log_2^+(x)$ being monotonically increasing in x, this leads to the following concave function[7]

$$g(\mathbf{s}|\bar{\mathbf{s}}) = \sum_n \alpha_n \log_2^+\left(1 + \frac{2\cdot\text{Re}(\bar{s}_n^* s_n) - |\bar{s}_n|^2}{\Gamma\sigma}\right) \tag{38}$$

which is a tight lower bound for $f(\mathbf{s})$ at any point $\bar{\mathbf{s}} \in \mathbb{C}^{N\times 1}$:

$$g(\mathbf{s}|\bar{\mathbf{s}}) \leq f(\mathbf{s}) \text{ for any } \mathbf{s} \in \mathbb{C}^{N\times 1} \tag{39}$$

$$g(\bar{\mathbf{s}}|\bar{\mathbf{s}}) = f(\bar{\mathbf{s}}). \tag{40}$$

Hence, based on (38) a successive lower bound maximization method[8] [26] is proposed, which generates a sequence of iterates $\mathbf{s}^{(t)}$ with non-decreasing objective values $f(\mathbf{s}^{(t)})$, i.e.,

$$f\left(\mathbf{s}^{(t)}\right) \geq g\left(\mathbf{s}^{(t)}|\mathbf{s}^{(t-1)}\right) \geq g\left(\mathbf{s}^{(t-1)}|\mathbf{s}^{(t-1)}\right) = f\left(\mathbf{s}^{(t-1)}\right) \quad \text{for all } t = 1, 2, \ldots \tag{41}$$

where $\mathbf{s}^{(t)}$ is the solution of

$$\underset{\mathbf{s}\in\mathcal{S}}{\text{maximize}} \quad g\left(\mathbf{s}|\mathbf{s}^{(t-1)}\right) \tag{42a}$$

$$\text{s.t.} \quad \mathbf{s}^T\mathbf{A}_l\mathbf{s} \leq P^{\text{mask}}, \qquad \forall l. \tag{42b}$$

Although (42) is convex, it is still a generic non-linear program due to the sum-of-log-function (38) as objective. However, (42) may be transformed into an equivalent second-order cone program (SOCP), which is in general more efficiently solved by standard solvers such as e.g. MOSEK [27] of the optimization tool CVX [28]. This transformation is based on re-writing the objective (42a) as a geometric product (or mean)

$$\prod_n \left(\Gamma\sigma + 2\cdot\text{Re}\left(\left(\bar{s}_n^{(t-1)}\right)^* s_n\right) - \left|\bar{s}_n^{(t-1)}\right|^2\right)^{\tilde{\alpha}_n} \tag{43}$$

where the scaling $\tilde{\alpha}_n \triangleq \alpha_n/\max_m(\alpha_m)$ ensures $0 \leq \tilde{\alpha}_n \leq 1, \forall n$, such that $\prod_n (x_n)^{\tilde{\alpha}_n}$ is always concave for positive x_n [29]. In turn, the geometric product can be replaced by a system of SOC constraints [30], which may be implemented in Matlab by using the *geo_mean* function of the CVX tool.

The complete algorithm (summarized in Alg. 1) is guaranteed to converge to a stationary solution as outlined in Theorem 2 below. Further, solving the (transformed) SOCP (42) every outer iteration in Alg. 1 with the interior-point method requires worst-case $O(\sqrt{N}\log(1/\epsilon))$ iterations up to accuracy ϵ, with a per-iteration complexity of approximately $O(N^2\sum_n N + N^3)$, amounting to forming and solving the Newton system [30]. Hence, the total computational complexity of Alg. 1 is $O(I_o N^{4.5}\log(1/\epsilon))$, where I_o is the number of outer iterations. Additionally, the generalization of Alg. 1 to the case with active total per-line power constraints is provided in Appendix B.

[6] This precoder structure is similar to the "diagonally-scaled integer forcing" precoder of [15].

[7] Here $\log_2^+(x)$ denotes the continuous-value extension of $\log_2(x)$ to negative x, with $\log_2^+(x) = -\infty$ if $x \leq 0$ and $\log_2^+(x) = \log_2(x)$ if $x > 0$.

[8] Also known as the Majorization-Minimization (MM) method.

Algorithm 1: A Successive Lower bound Maximization Method for Solving Problem (37)

Find a feasible point $\mathbf{s}^{(0)}$
Set $t = 0$
repeat
  Set $t = t + 1$
  Obtain $\mathbf{s}^{(t)}$ by solving problem (42) with $g(\mathbf{s}|\mathbf{s}^{(t-1)}\})$.
until $f(\mathbf{s}^{(t)}) - f(\mathbf{s}^{(t-1)}) < \epsilon$

Theorem 2. The sequence of iterates $\{\mathbf{s}^{(t)}\}$ generated by Alg. 1 is guaranteed to converge to the set of stationary points of (37), and equivalently of (33).

Proof: It is already established that $g(\mathbf{s}|\bar{\mathbf{s}})$ is a continuous tight lower bound of the objective $f(\mathbf{s})$ in (39) and (40); and that the sequence of generated objective values $f(\mathbf{s}^{(t)})$ is monotonically non-decreasing in (41). In addition, it can be verified that $g(\mathbf{s}|\bar{\mathbf{s}})$ has the same first-order derivative as the original objective $f(\mathbf{s})$ at the point where the lower bound is tight (i.e in point $\mathbf{s} = \bar{\mathbf{s}}$). Hence, in combination with the iterates $\{\mathbf{s}^{(t)}\}$ lying in a closed and bounded set, convergence of $\{\mathbf{s}^{(t)}\}$ can be established to the set of stationary points of (37) [26]. It is remarked that the results in [26] are also valid for the case with complex variables. ■

B. Achievable Sum-Rate Gap with ZF-THP

Recall that ZF-LRP uses the reduced precoder matrix $\mathbf{P}^{\text{red}} = \mathbf{H}^{-1}\mathbf{S}\mathbf{T}$ where $\mathbf{T}$ is a (possibly) non-unimodular integer matrix with $|\det(\mathbf{T})| \geq 1$. To satisfy the per-line power constraint P, the scaling matrix $\mathbf{S}$ has to be chosen such that the row norms of $\mathbf{P}^{\text{red}}$ are less than unity (assuming that $\mathbf{S}$ is normalized by $\sqrt{P}$). Consequently, Hadamard's inequality yields $|\det(\mathbf{P}^{\text{red}})| \leq 1$. In turn, this yields $|\det(\mathbf{S})| = \left|\det(\mathbf{P}^{\text{red}})\right| \cdot |\det(\mathbf{H})| \cdot |\det(\mathbf{T})|^{-1} \leq |\det(\mathbf{H})|$. The achievable sum-rate of ZF-LRP may hence be approximated as follows:

$$\begin{aligned} R_{\text{LRP}} &= \sum_n \log_2\left(1 + \frac{|s_n|^2 P}{\Gamma\sigma}\right) \\ &= \log_2\left(\prod_n \left(1/|s_n|^2 + P/\Gamma\sigma\right) \prod_m |s_m|^2\right) \quad (44a) \\ &\approx N \log_2\left(\frac{P}{\Gamma\sigma}\right) + 2\log_2|\det(\mathbf{S})|. \quad (44b) \end{aligned}$$

where the approximation in (44b) is valid at high SNR values with $|s_n|^2 \gg \Gamma\sigma/P$.

ZF-THP [4], on the other hand, is based on the QR decomposition of the conjugate transpose of the channel matrix, i.e., $\mathbf{H}^H \overset{\text{qr}}{=} \mathbf{Q}\mathbf{R}$. The implementation then combines a feedforward (orthogonal) precoder matrix $\mathbf{P} = \mathbf{Q}\tilde{\mathbf{S}}$, with $\tilde{\mathbf{S}}$ a diagonal scaling matrix, that has to be designed, together with a sequential feedback operation. Likewise, the scaling matrix $\tilde{\mathbf{S}}$ has to be chosen such that the row norms of the feedforward precoder matrix $\mathbf{P}$ are less than unity (assuming that $\tilde{\mathbf{S}}$ is normalized by $\sqrt{P}$). This yields by Hadamard's inequality that $|\det(\mathbf{P})| = \left|\det(\tilde{\mathbf{S}})\right| = \prod_n |\tilde{s}_n| \leq 1$. In addition, it holds that $|\det(\mathbf{H})| = |\det(\mathbf{R}^H)| = \prod_n |r_{nn}|$ where r_{nn} are the diagonal values of $\mathbf{R}^H$. This leads to the following achievable sum-rate for ZF-THP [4]:

$$\begin{aligned} R_{\text{THP}} &= \sum_n \log_2\left(1 + \frac{|r_{nn}|^2|\tilde{s}_n|^2 P}{\Gamma\sigma}\right) \\ &= \log_2\left(\prod_n \left(1/r_{nn}|^2|\tilde{s}_n|^2 + P/\Gamma\sigma\right) \prod_m |r_{mm}|^2|\tilde{s}_m|^2\right) \quad (45a) \\ &\overset{(i)}{\approx} N \log_2\left(\frac{P}{\Gamma\sigma}\right) + 2\log_2|\det(\mathbf{H})| + 2\log_2\left|\det(\tilde{\mathbf{S}})\right| \\ &\overset{(ii)}{=} N \log_2\left(\frac{P}{\Gamma\sigma}\right) + 2\log_2|\det(\mathbf{H})|, \quad (45b) \end{aligned}$$

where the approximation in (i) is valid at high SNR values with $|r_{nn}|^2|\tilde{s}_n|^2 \gg \Gamma\sigma/P$, and (ii) is due to the fact that flat gain scaling (i.e. $|\tilde{s}_n| = 1, \forall n$) is optimal for ZF-THP at high SNR values. However, for low/moderate SNR values, some spread in the $1/|r_{nn}|^2|\tilde{s}_n|^2 + P/\Gamma\sigma$ factors may counter the loss of $\prod_n |\tilde{s}_n|$ being smaller than unity.

Subtracting (44b) from (45b) leads thus to the achievable sum-rate gap ΔR between the ZF-THP and ZF-LRP, approximated at high SNR values by

$$\Delta R \approx 2\log_2 \frac{|\det(\mathbf{H})|}{|\det(\mathbf{S})|} = 2\log_2|\det(\mathbf{T})| - 2\log_2\left|\det(\mathbf{P}^{\text{red}})\right| \geq 0. \quad (46)$$

In order for this bound to be tight: (1) the (output power normalized) reduced precoder matrix $\mathbf{P}^{\text{red}}$ has to have orthogonal rows with unit norm by Hadamard's inequality, corresponding to a *unitary* matrix (such that $|\det(\mathbf{P}^{\text{red}})| = 1$); and (2) the $\mathbf{T}$ matrix has to be *unimodular* (such that $|\det(\mathbf{T})| = 1$). A unimodular $\mathbf{T}$ matrix leads to $\prod_n |s_n| \leq \prod_n |r_{nn}|$ being tight when $\mathbf{P}^{\text{red}}$ is unitary.

By consequence, if $\mathbf{P}^{\text{red}}$ is not exactly orthogonal, some gain can be expected from lattice reducing the ZF precoder $\mathbf{H}^{-1}\mathbf{S}$, in an attempt to further improve the orthogonality of $\mathbf{P}^{\text{red}}$. This motivates the heuristic method proposed in Section III that alternates between lattice reduction of $\mathbf{H}^{-1}\mathbf{S}$ and gain scaling optimization of $\mathbf{S}$ given $\mathbf{T}$. In addition, if instead of e.g. the LLL algorithm which results in $|\det(\mathbf{T})| = 1$, a non-unimodular lattice reduction algorithm is used (see e.g. [7]), the loss from $|\det(\mathbf{T})| > 1$ should be compensated by the extra orthogonality gain in $|\det(\mathbf{P}^{\text{red}})|$.

Finally, it is pointed out that these conclusions are in line with an information-theoretic point of view, since THP can be viewed as practical implementation of "dirty paper coding", which is the capacity-achieving transmission scheme for the broadcast channel [31].

V. MMSE-LRP Design Given A Fixed $\mathbf{T}$

Instead of considering a ZF precoder matrix with complex gain scaling optimization, the goal in this section is to optimally design the entire precoder matrix given a fixed integer $\mathbf{T}$

matrix. The MMSE-LRP design is formulated as the following WSR maximization problem:

$$\underset{\mathbf{P}}{\text{maximize}} \quad \sum_n \alpha_n \log_2\left(1+\frac{\text{SNR}_n(\mathbf{P},\mathbf{T})}{\Gamma}\right) \tag{47a}$$

$$\text{s.t.} \quad \left[\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\right]_{ll} \le P^{\text{mask}}, \qquad \forall l \tag{47b}$$

where $\text{SNR}_n(\mathbf{P},\mathbf{T})$ is defined in (29) and (47b) is a convex quadratic constraint. Unfortunately, (47) is a difficult problem, with a non-convex objective function even when $\mathbf{T}=\mathbf{I}_N$ (i.e. the case of linear MMSE precoding).

A feasible approach to solve (47) is to approximate the objective by using the theoretically achievable bit-rate expression defined in (21). Dropping the constant shaping loss in (21), it is actually equivalent to the achievable bit-rate of IF precoding schemes [14], [15] in the broadcast channel. Assuming optimal MMSE-FEQs [see (22) and (23)], the achievable WSR is

$$\begin{aligned} C &= \sum_n \alpha_n \log_2\left(\frac{1}{\varepsilon_n^{\text{opt}}}\right) \\ &= \sum_n \alpha_n \log_2\left(\frac{1}{\left[\mathbf{T}\mathbf{T}^H\right]_{nn}}+\text{SNR}_n(\mathbf{P},\mathbf{T})\right), \end{aligned} \tag{48}$$

leading to the following WSR maximization problem:

$$\underset{\mathbf{P}}{\text{maximize}} \quad C$$

$$\text{s.t.} \quad \left[\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\right]_{ll} \le P^{\text{mask}}, \qquad \forall l. \tag{49}$$

Note that for high SNR values problem (47) and (49) are equivalent, meaning that the optimum precoder $\mathbf{P}$ will be identical for both problems. Further, the maximum bit cap constraints are omitted in (47) and (49) because these are non-convex in general for MMSE precoding.[9]

Nevertheless the approximate problem (49) remains non-convex, its main advantage is that it may be reformulated as an equivalent WMMSE problem and then solved by a block coordinate descend method leading eventually to a stationary point of (49). The equivalence between WSR maximization and WMMSE minimization has been established first for LP in the MIMO broadcast channel [32] and later extended to the MIMO interference channel in [33]. In addition, it can be readily extended to problem (49) for IF precoding in the broadcast channel as well, as outlined below:

Proposition 3. Define $\boldsymbol{\mu} \triangleq [\mu_1,\ldots,\mu_N]$ with μ_n the MSE weight for user n, and $\mathbf{w} \triangleq [w_1,\ldots,w_N]$ with w_n the equalizer tap for user n, and ε_n as the corresponding MSE for user n given by

$$\begin{aligned} \varepsilon_n \triangleq \mathcal{E}\{|\ddot{u}_n - w_n y_n|^2\} &= \left[\mathbf{T}\mathbf{T}^H\right]_{nn} - w_n^*\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\right]_{nn}^* \\ &\quad - w_n\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\right]_{nn} + w_n w_n^*\left[\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\mathbf{H}^H + \sigma\mathbf{I}_N\right]_{nn}. \end{aligned} \tag{50}$$

Then the following problem

$$\underset{\boldsymbol{\mu},\mathbf{w},\mathbf{P}}{\text{minimize}} \quad \sum_n \alpha_n(\mu_n\varepsilon_n - \log(\mu_n) - 1) \tag{51a}$$

$$\text{s.t.} \quad \left[\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\right]_{ll} \le P^{\text{mask}}, \qquad \forall l \tag{51b}$$

[9] In G.fast, bit cap constraints are typically only active for low frequency tones with diagonally dominant channels. Hence, for these tones, the ZF precoder with gain scaling optimization is near-optimal.

Algorithm 2: WSR Maximization for MMSE-LRP

Find a feasible precoder $\mathbf{P}$
repeat
  Update the FEQs $\{w_n\}_{n=1}^N$ and MSE $\{e_n\}_{n=1}^N$ according to (22) and (50), respectively
  Update the MSE weights $\{\mu_n\}_{n=1}^N$ using (52)
  repeat
    Set/update$\{\lambda_n\}_{n=1}^N$ using the subgradient method
    Update the MMSE precoder $\mathbf{P}$ with (54)
  until $\lambda_n\left|\left[\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\right]_{nn} - P^{\text{mask}}\right| < \epsilon \quad \forall n$
until convergence

is equivalent to the WSR maximization problem (49), such that the optimum precoder matrix $\mathbf{P}$ for (49) and (51) will be identical. This equivalence can be shown by checking the first-order conditions for $\boldsymbol{\mu}$ and $\mathbf{w}$ [see (52)], substituting them in (51) and making some straightforward simplifications.

The essence of the equivalent WMMSE problem (51) is that the objective function is convex in each of the optimization variables when the other variables are fixed. This allows the use of a block coordinate descent method by successively updating the variables to solve (51) [33]. For fixed $\mathbf{P}$, the first order conditions of w_n and μ_n lead to the following update expressions for w_n and μ_n:

$$w_n^{\text{opt}} = \text{r.h.s. of (22)} \qquad \text{and} \qquad \mu_n^{\text{opt}} = \varepsilon_n^{-1}, \qquad \forall n. \tag{52}$$

For fixed $\boldsymbol{\mu}$ and $\mathbf{w}$ on the other hand, (51) reduces to the following convex problem:

$$\begin{aligned} \underset{\mathbf{P}}{\text{minimize}} \quad & \text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\mathbf{H}^H\mathbf{W}^H\} \\ & \qquad -2\text{Re}\left(\text{Tr}\{\mathbf{D}\mathbf{W}\mathbf{H}\mathbf{P}\mathbf{T}\mathbf{T}^H\}\right) \end{aligned} \tag{53a}$$

$$\text{s.t.} \quad \left[\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\right]_{ll} \le P^{\text{mask}}, \qquad \forall l, \tag{53b}$$

where $\mathbf{W} = \text{diag}\{\mathbf{w}\}$ and $\mathbf{D} = \text{diag}\{\{\mu_n\alpha_n\}_{n=1}^N\}$. As strong duality holds in problem (53)[10], Lagrange dual decomposition can be used in order to obtain an optimal closed-form solution [33], i.e.,

$$\mathbf{P}^{\text{opt}} = \left(\mathbf{H}^H\mathbf{W}^H\mathbf{D}\mathbf{W}\mathbf{H} + \text{diag}\{\boldsymbol{\lambda}\}\right)^{-1}\mathbf{H}^H\mathbf{W}^H\mathbf{D}, \tag{54}$$

where $\boldsymbol{\lambda} \triangleq [\lambda_1,\cdots,\lambda_N]$ are the Lagrange multipliers, which should be chosen such that the per-line power constraints (53b) are either tight or inactive. These Lagrange multipliers may be found using e.g. standard subgradient search with

$$\lambda_n = \left[\lambda_n + \delta\left(\left[\mathbf{P}\mathbf{T}\mathbf{T}^H\mathbf{P}^H\right]_{nn} - P^{\text{mask}}\right)\right]^+, \qquad \forall n, \tag{55}$$

and δ being a pre-defined step size, or by using the ellipsoid method [34].

The block coordinate descent method is summarized in Alg. 2. When $\mathbf{T}$ is the identity matrix (i.e. the case of LP), problem (47) and (49) become equivalent (for zero SNR gap),

[10] Problem (53) is convex and Slater's condition is satisfied, e.g. choose any feasible precoder matrix.

and Alg. 2 reduces to the original WMMSE algorithm for the broadcast channel. In addition, Alg. 2 may be generalized to the case with active total per-line power constraints, see (3), by re-introducing the tone indices and including additional Lagrange multipliers $\boldsymbol{\theta} \triangleq [\theta_1, \cdots, \theta_N]$ in (54). Further details are omitted for brevity.

VI. G.fast Cable Binder Simulation

In this section, a cable binder is simulated consisting of 10 lines of 80 m for the downstream G.fast 212 MHz profile. The channel matrices have been obtained by measurements. The observed crosstalk levels in this particular cable binder are rather high compared to other reported G.fast measurements. Following the G.fast recommendation [2], the ATP constraints are set to 8 dBm while the per-tone PSD spectral masks are obtained from [35] ranging from −65 dBm/Hz to −79 dBm/Hz. The total SNR gap Γ is 10.37 dB and the tone spacing Δ_f is 51.75 kHz. The noise PSD is assumed to be −140 dBm/Hz. The symbol rate is 48 kHz and the bitcap is 14. Only sum-rate optimization with $\{\alpha_n = 1\}_{n=1}^{N}$ is considered. In these simulations, the ATP constraints are observed to be always inactive, due to the per-tone spectral mask and maximum bit loading constraints. For this cable binder, various NLP schemes are evaluated:

- *ZF-LRP-RNS*: represents the baseline LRP scheme with ZF precoding and a flat row norm scaling (RNS). The precoder matrix is hence $\mathbf{P} = \mathbf{H}^{-1}s$, where $\mathbf{T} = \text{LLL}(\mathbf{H}^{-1})$ is obtained by the lattice reduction of $\mathbf{H}^{-1}$ via the LLL algorithm and s is set to the square-root of

$$\left[\frac{P^{\text{mask}}}{\max_l \sum_m \left|\left[\mathbf{H}^{-1}\mathbf{T}\right]_{lm}\right|^2}\right]^{\Gamma\sigma(2^{b^{\max}}-1)}. \tag{56}$$

- *ZF-LRP (Alg. 1)*: uses $\mathbf{P} = \mathbf{H}^{-1}\mathbf{S}$ with $\mathbf{T} = \text{LLL}(\mathbf{H}^{-1})$ and the scaling matrix $\mathbf{S}$ obtained by means of Alg. 1.
- *ZF-LRP* [10]: is the ZF-LRP approach proposed in [10], by means of Alg. 1 with a *real* gain scaling matrix $\mathbf{S}$ and a high SNR approximation, and $\mathbf{T} = \text{LLL}(\mathbf{H}^{-1})$.
- *ZF-LRP-AO*: is the final ZF-LRP scheme, as proposed in Section III, which uses alternating optimization (AO) between gain scaling optimization of $\mathbf{S}$ for fixed $\mathbf{T}$ (using Alg. 1) and lattice reduction of $\mathbf{T} = \text{LLL}(\mathbf{H}^{-1}\mathbf{S})$.
- *MMSE-LRP*: uses the optimized MMSE precoder matrix obtained by means of Alg. 2 given $\mathbf{T} = \text{LLL}(\mathbf{H}^{-1})$.
- *MMSE-LRP-AO*: alternates between MMSE precoder matrix optimization and lattice reduction.
- *ZF-THP Max-Min*: corresponds to α-fair QRD-based THP with joint power allocation and user ordering [36]. A large α is used here (i.e. min-rate maximization)
- *ZF-THP Max-Sum*: is α-fair QRD-based THP with $\alpha = 0$ (i.e. sum-rate maximization).
- *MMSE-THP*: is provided by the BC-DSB-NLP [37] together with the same user ordering of ZF-THP Max-Sum.

Foremost among the numerical results, summarized in Table I and Fig. 2, is that LRP consistently is outperformed by THP, both in terms of performance and fairness (i.e. in terms of multi-tone sum-rate and min-rate, respectively).

TABLE I
PERFORMANCE COMPARISON BETWEEN VARIOUS NLP SCHEMES FOR THE G.FAST 212 MHZ PROFILE IN A 10×80 M CABLE BINDER.

	Precoding scheme	Mean-Rate		Min-Rate	
		[Mbps]	[%]	[Mbps]	[%]
LRP	ZF-LRP-RNS	1213	86.5	1213	89.0
	ZF-LRP [10]	1263	90.0	1227	90.0
	ZF-LRP (Alg. 1)	1281	91.4	1249	91.6
	ZF-LRP-AO	1319	94.1	1260	92.4
	MMSE-LRP	1287	91.8	1255	92.1
	MMSE-LRP-AO	1330	94.9	1248	91.6
THP	ZF-THP Max-Min	1369	97.6	1364	100
	ZF-THP Max-Sum	1397	99.6	1213	88.9
	MMSE-THP	1402	100	1200	87.9

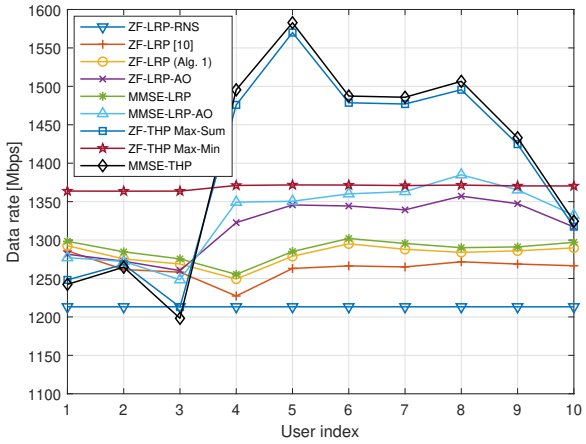


Fig. 2. User data-rates of various NLP schemes in a 10-user cable binder with the G.fast 212 MHz profile.

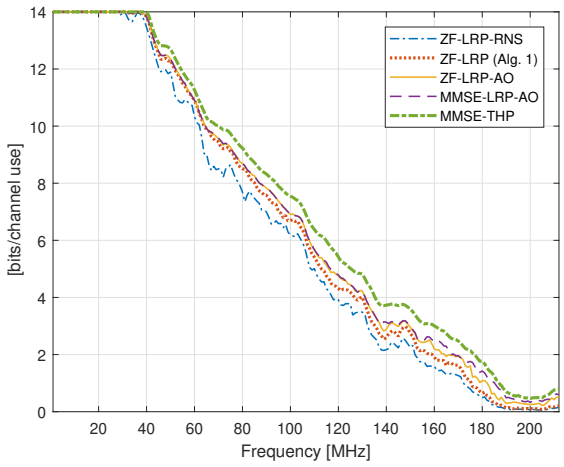
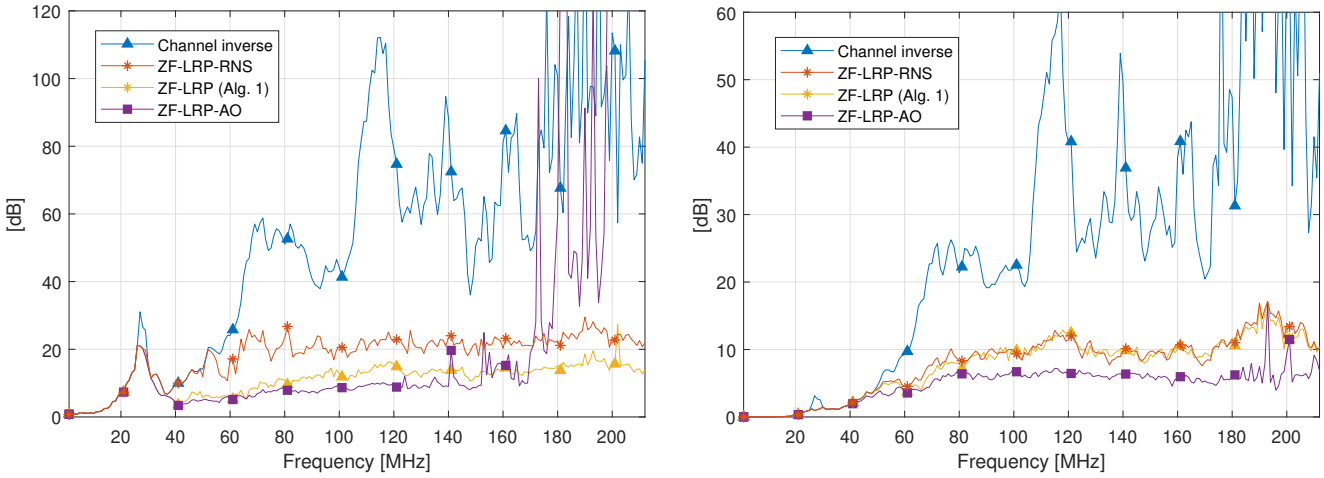


Fig. 3. Mean-user bit-rates across the frequency range in a 10-user cable binder with the G.fast 212 MHz profile. The values are smoothened by averaging over 1 MHz frequency bins.

This is expected from the results in Section IV-B, where the achievable ZF-LRP sum-rate is shown to be upper bounded at high SNR by the achievable ZF-THP sum-rate. Additionally,



(a) Factor ψ (b) Othogonality Defect

Fig. 4. Factor ψ and the orthogonality effect across the frequency range. Channel inverse corresponds to setting $\mathbf{T}$ to the identity matrix together with RNS. The values are smoothened by averaging over 1 MHz frequency bins.

ZF-THP in combination with joint fair power allocation and user ordering [36] is able to achieve a higher minimum user rate than LRP. However, as mentioned in Section I, the main disadvantage of THP is the required sequential feedback operation proportional to the number of users, which complicates hardware implementation. In addition, LRP typically has a natural (more or less) fair user rate distribution, while THP without fairness aware user encoding ordering optimization typically results in a large user rate spread. As expected, the proposed ZF-LRP (Alg. 1) with complex scaling achieves a higher sum-rate than ZF-LRP [10] with real scaling, and significantly outperforms ZF-LRP-RNS. It is stressed that although the total bit-rate gains summed over all frequency tones might be rather small, the relative bit-rate gains on an individual tone can be quite significant (see Fig. 3).

The factor $\psi \triangleq \prod_n |r_{nn}|/|\tilde{s}_n| \geq 1$ together with the so-called orthogonality defect[11] [38] of the ZF precoders across the frequency range is shown in Fig. 4. The former determines the sum-rate gap (46) with ZF-THP at high SNR, whereas the latter is a measure for matrix orthogonality. Fig. 4 shows that the lattice reduction of $\mathbf{T} = \text{LLL}(\mathbf{H}^{-1})$ substantially reduces the orthogonality defect of the channel inverse, leading to a significant sum-rate-gain for ZF-LRP-RNS. Additional scaling of the ZF precoder by optimizing $\mathbf{S}$ (i.e. ZF-LRP) then further reduces factor ψ, without significantly improving the orthogonality defect. It thus seems that gain scaling optimization primarily leads to the inequality $\prod_n |\tilde{s}_n| \leq \prod_n |r_{nn}|$ being as tight as possible given a certain $\mathbf{T}$, which maximizes the sum-rate at high SNR. Further improving the orthogonality of the reduced and scaled ZF precoder (and hence the sum-rate at high SNR) is the goal of ZF-LRP-AO, which alternates between gain scaling optimization of $\mathbf{S}$ and lattice reduction of $\mathbf{T} = \text{LLL}(\mathbf{H}^{-1}\mathbf{S})$. It results in a further reduction of both the factor ψ and the orthogonality defect, which illustrates the potential of joint gain scaling and lattice reduction design (i.e. to jointly optimize $\mathbf{S}$ and $\mathbf{T}$). Notice that some users may be allocated low or zero $|s_n|$ magnitudes at high frequency tones with low SNR, which blows up factor ψ for ZF-LRP-AO.

A final remark is that generally the use of NLP schemes decreases the benefit of using MMSE over ZF precoding. The sum-rate gain between MMSE and ZF precoding for THP and LRP is smaller than 1%, while the gain between MMSE-LP and ZF-LP has been reported to be about 10% for the same binder [37].

[11]The orthogonality defect of the ZF precoder $\tilde{\mathbf{P}} \triangleq \mathbf{H}^{-1}\mathbf{S}\mathbf{T}$ with $\tilde{\mathbf{P}} = [\mathbf{p}_1, \ldots, \mathbf{p}_N]$ is computed by $\prod_i \|\mathbf{p}_i\|/\sqrt{|\det(\tilde{\mathbf{P}}^H\tilde{\mathbf{P}})|}$, and with any zero columns of $\tilde{\mathbf{P}}$ removed.

VII. Conclusion

In this paper, the potential of LRP for downstream transmission in G.fast DSL networks has been studied, as an NLP alternative for THP with a reduced implementation complexity. A practically achievable bit-rate expression has been proposed for LRP, in order to cast the design problem as a WSR maximization in function of a precoder matrix and an integer matrix. For a fixed integer matrix, ZF precoding optimization reduces to a gain scaling optimization, for which a successive lower bound maximization method has been presented. Additionally, it has been established that the achievable ZF-LRP sum-rate is upper bounded by the achievable ZF-THP sum-rate at high SNR. For MMSE precoding optimization an efficient method has been proposed as well, by relying on the equivalence between the WSR maximization and the WMMSE minimization. The LRP design methods may be improved by alternating between precoder matrix optimization for a fixed integer matrix, and updating the integer matrix based on the lattice reduction of the precoder matrix. Simulations of a measured G.fast cable binder have been provided to compare the performance of the proposed LRP schemes with THP schemes.

Appendix A
Proof of Proposition 1

The following two lemmas are considered first (as a generalization of [19, Lemma 1]):

Lemma 4. If $\{X_1,\ldots,X_N\}$ are independent variables uniformly distributed over Ω_τ and $[a_1,\ldots,a_N]$ are non-zero complex integers with $N \geq 2$, then $Z = [\sum_n a_n X_n] \,\text{mod-}\tau$ is uniformly distributed over Ω_τ, and pairwise independent of $\{X_1,\ldots,X_N\}$, and mutually independent of any subset of $\{X_1,\ldots,X_N\}$ with at most $N-1$ variables.

Proof: First, the property $\text{mod}(a+b) = \text{mod}(\text{mod}(a) + \text{mod}(b))$ is used to observe that:

$$\left[\sum_n a_n X_n\right] \text{mod-}\tau = \left[\sum_n \left[a_n X_n\right] \text{mod-}\tau\right] \text{mod-}\tau \tag{57}$$

with $\overline{X}_n \triangleq \left[a_n X_n\right] \text{mod-}\tau$ being uniformly distributed over Ω_τ, since a_n is complex integer. It is seen that $\{\overline{X}_1,\ldots,\overline{X}_N\}$ are still independent variables, and moreover, that $\overline{X}_n$ is mutually independent of $\{X_1,\ldots,X_{n-1},X_{n+1},\ldots,X_N\}$, due to the fact that (measurable) functions of independent variables are independent variables themselves (see e.g. [39, Th. 2.1.6]). Now, for any $n \in [1,\ldots,N]$

$$Z = \left[\sum_n \overline{X}_n\right] \text{mod-}\tau = \left[A_n + \overline{X}_n\right] \text{mod-}\tau \tag{58}$$

is considered, where $A_n \triangleq \left[\sum_{m\neq n} \overline{X}_m\right] \text{mod-}\tau$ forms a random variable in Ω_τ. Since A_n is independent of $\overline{X}_n$ [39, Th. 2.1.6], $\left[a + \overline{X}_n\right] \text{mod-}\tau$ is uniformly distributed over Ω_τ for any possible value $A_n = a \in \Omega_\tau$. Since this holds for any value a, it holds for any combination of values as well, meaning that Z is uniformly distributed over Ω_τ. Further, it is seen that the probability density of Z is constant over Ω_τ for any value A_n or $\overline{X}_n$ (i.e. knowing both A_n and $\overline{X}_n$ fixes Z, but knowing only one of them gives no information about Z). As a result, Z is pairwise independent of A_n and $\overline{X}_n$, and consequently, mutually independent of $\{X_1,\ldots,X_{n-1},X_{n+1},\ldots,X_N\}$. Therefore, repeating this for every n, Z can be shown to be mutually independent of any subset of $\{X_1,\ldots,X_N\}$ with at most $N-1$ variables. This also means that Z is pairwise independent of $\{X_1,\ldots,X_N\}$. ∎

Lemma 5. If $\{X_1,\ldots,X_N\}$ are independent variables uniformly distributed over Ω_τ, and $\mathbf{a} \triangleq [a_1,\ldots,a_N]$ and $\mathbf{b} \triangleq [b_1,\ldots,b_N]$ are non-zero complex integer vectors and linearly independent, then $Z_1 = [\sum_n a_n X_n] \,\text{mod-}\tau$ and $Z_2 = [\sum_n b_n X_n] \,\text{mod-}\tau$ are independent.

Proof: The set of non-zero elements in $\mathbf{a}$ and $\mathbf{b}$ is denoted by $\mathcal{A}$ and $\mathcal{B}$, respectively. When $\mathcal{A}$ and $\mathcal{B}$ are disjoint, Z_1 and Z_2 are independent since they are functions of independent variables [39, Th. 2.1.6]. For the case of full overlap (i.e. $\mathcal{A} = \mathcal{B}$), Z_2 can always be written w.l.o.g. as:

$$Z_2 = \left[cZ_1 + Z_r\right] \text{mod-}\tau \tag{59}$$

with $Z_r \triangleq \left[\sum_{m\in\mathcal{A}} d_m X_m\right] \text{mod-}\tau$ a non-zero uniform variable in Ω_τ, c a non-zero complex integer, and $b_m = ca_m + d_m, \forall m \in \mathcal{A}$. Moreover, since $\mathbf{a}$ and $\mathbf{b}$ are linearly independent (i.e. they are not an integer multiple of each other), it is always possible to choose c and $\{d_m\}$ such that at least one element of $\{d_m\}$ is zero. Hence, this means that Z_r is a function of a subset of $\{X_m\}_{m\in\mathcal{A}}$ with at most $|\mathcal{A}|-1$ variables, leading to Z_r being independent of Z_1 (since Z_1 is mutually independent of any $|\mathcal{A}|-1$ subset of $\{X_n\}_{n\in\mathcal{A}}$). Thus, as a result of Lemma 4, Z_1 and Z_2 are independent. In case of only partial overlap, it can be shown similarly that Z_2 and Z_1 are always independent. ∎

Any unimodular matrix $\mathbf{T}$, obtained by the lattice reduction algorithm, can be without loss of generality shown to be structured as follows:

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{11} & \mathbf{T}_{12} \\ \mathbf{0} & \mathbf{I}^c_{N'} \end{bmatrix}, \qquad \mathbf{T}^{-1} = \begin{bmatrix} \mathbf{T}_{11}^{-1} & -\mathbf{T}_{11}^{-1}\mathbf{T}_{12}\mathbf{I}^{c,H}_{N'} \\ \mathbf{0} & \mathbf{I}^{c,H}_{N'} \end{bmatrix} \tag{60}$$

where $\mathbf{I}^c_{N'}$ is a diagonal matrix of size $N' \geq 0$, of which the diagonal entries are complex integers with unit magnitude (i.e. the entries equal either ± 1 or $\pm j$). Note that one could include permutation matrices in the LRP scheme of Fig. 1 to enforce this particular structure, and that if a row of a unimodular matrix with determinant ± 1 has a single entry, this entry must have unit magnitude, since scalar multiplication of a row by a constant c multiplies the determinant by c.

This shows that $\dot{u}_n$ equals u_n up to a possible phase shift for components $n \in [N-N'+1,\ldots,N]$. Additionally, the first $N-N'$ rows in $\mathbf{T}$ and $\mathbf{T}^{-1}$ have at least two non-zero integer entries. Therefore, the variables $\{\dot{u}_1,\ldots,\dot{u}_{N-N'}\}$ are pairwise independent and uniformly distributed over Ω_τ, and pairwise independent of $\{u_1,\ldots,u_N\}$ based on Lemma 4 and 5 and [39, Th. 2.1.6]. From these results the conditions (14)-(17) follow readily.

Appendix B
Generalization of Alg. 1 to the case with active total per-line power constraints

Alg. 1 can be generalized to the case with active total per-line power constraints [see (3)], which then couple the K convex subproblems in (42) at iteration t. This is formulated as follows by re-introducing the tone indices k:

$$\begin{aligned} \underset{\{\mathbf{s}_k\in\mathcal{S}_k\}}{\text{maximize}} \quad & \sum_k g_k\left(\mathbf{s}_k|\mathbf{s}_k^{(t-1)}\right) \\ \text{s.t.} \quad & \mathbf{s}_k^T\mathbf{A}_{lk}\mathbf{s}_k \leq P_k^{\text{mask}}, \qquad \forall l,k \\ & \sum_k \mathbf{s}_k^T\mathbf{A}_{lk}\mathbf{s}_k \leq P^{\text{total}}, \qquad \forall l. \end{aligned} \tag{61}$$

Since strong duality holds in (61), it may be decoupled across the tones by using Lagrange dual decomposition. Then, for every fixed set of Lagrange multipliers $\boldsymbol{\theta} \triangleq [\theta_1,\cdots,\theta_N]$, the independent per-tone slave problem for tone k is given by

$$\begin{aligned} h_k(\boldsymbol{\theta}) = \underset{\mathbf{s}_k\in\mathcal{S}}{\text{maximize}} \quad & g_k\left(\mathbf{s}_k|\mathbf{s}_k^{(t-1)}\right) - \sum_{l=1}^N \theta_n\left(\mathbf{s}_k^T\mathbf{A}_{lk}\mathbf{s}_k - P^{\text{total}}\right) \\ \text{s.t.} \quad & \mathbf{s}_k^T\mathbf{A}_{lk}\mathbf{s}_k \leq P_k^{\text{mask}}, \qquad \forall l, \end{aligned} \tag{62}$$

which is again a small-scale convex problem that may be solved by e.g. CVX. The master dual problem of (61) is

$$\underset{\boldsymbol{\theta}\succeq\mathbf{0}}{\text{minimize}} \quad \sum_k h_k(\boldsymbol{\theta}), \tag{63}$$

where the optimal $\boldsymbol{\theta}$ are found with e.g. a standard subgradient search

$$\theta_l = \left[\theta_l + \delta\left(\sum_k \mathbf{s}_k^T \mathbf{A}_{lk}\mathbf{s}_k - P^{\text{total}}\right)\right]^+, \quad \forall l, \tag{64}$$

which is guaranteed to converge if the step size δ is chosen sufficiently small [34].

REFERENCES

[1] W. Lanneer, "Vectoring Design Optimization in Ultra-Broadband DSL Networks," Ph.D. dissertation, KU Leuven, Feb. 2019.

[2] *Fast Access to Subscriber Terminals (FAST) - Physical Layer Specification*, Recommendation ITU-T G.9701 - Amendment 3, Apr. 2017.

[3] R. Strobel, R. Stolle, and W. Utschick, "Wideband modeling of twisted-pair cables for MIMO applications," in *IEEE Global Commun. Conf. (GLOBECOM)*, 2013, pp. 2828–2833.

[4] G. Ginis and J. Cioffi, "A multi-user precoding scheme achieving crosstalk cancellation with application to DSL systems," in *Conf. Record 34th Signals, Systems and Computers*, vol. 2, Oct. 2000, pp. 1627–1631 vol.2.

[5] J. Maes, C. Nuzman, and P. Tsiaflakis, "Sensitivity of nonlinear precoding to imperfect channel state information in G.fast," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sept. 2016.

[6] C. Windpassinger, R. F. H. Fischer, and J. B. Huber, "Lattice-reduction-aided broadcast precoding," *IEEE Tran. Commun.*, vol. 52, no. 12, pp. 2057–2060, Dec. 2004.

[7] S. Stern and R. F. H. Fischer, "Advanced factorization strategies for lattice-reduction-aided preequalization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2016, pp. 1471–1475.

[8] C. Windpassinger and R. F. H. Fischer, "Low-complexity near-maximum-likelihood detection and precoding for MIMO systems using lattice reduction," in *Proc. IEEE Info. Theory Workshop*, Mar. 2003, pp. 345–348.

[9] M. Taherzadeh, A. Mobasher, and A. K. Khandani, "Communication Over MIMO Broadcast Channels Using Lattice-Basis Reduction," *IEEE Tran. Inf. Theory*, vol. 53, no. 12, pp. 4567–4582, Dec. 2007.

[10] M. Hekrdla, A. Matera, U. Spagnolini, and W. Wang, "Per-line Power Controlled Lattice-Reduction Aided Zero-Forcing Precoding for G.fast Downstream," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[11] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1999.

[12] Y. Zhang, R. Zhang, A. F. A. Rawi, and L. Hanzo, "Approximate perturbation aided lattice encoding (apple) for g.fast and beyond," *IEEE Access*, vol. 6, pp. 53 438–53 451, 2018.

[13] S. Hong and G. Caire, "Compute-and-Forward Strategies for Cooperative Distributed Antenna Systems," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5227–5243, Sep. 2013.

[14] W. He, B. Nazer, and S. Shamai, "Uplink-downlink duality for integer-forcing: Effective SINRs and iterative optimization," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun.*, Jun. 2014, pp. 474–478.

[15] D. Silva, G. Pivaro, G. Fraidenraich, and B. Aazhang, "On Integer-Forcing Precoding for the Gaussian MIMO Broadcast Channel," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4476–4488, Jul. 2017.

[16] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electron. Lett.*, vol. 7, no. 5, pp. 138–139, March 1971.

[17] 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.

[18] J. Neckebroeck, "Error correction, precoding and bitloading algorithms in high-speed access networks," *Ph.D. dissertation, UGent*, 2016. [Online]. Available: https://biblio.ugent.be/publication/8205721/file/8205722.pdf

[19] U. Erez and R. Zamir, "Achieving 1/2 log (1+SNR) on the AWGN channel with lattice encoding and decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[20] U. Erez, S. Shamai, and R. Zamir, "Capacity and lattice strategies for canceling known interference," *IEEE Tran. Inf. Theory*, vol. 51, no. 11, pp. 3820–3833, Nov. 2005.

[21] S. T. Chung and A. J. Goldsmith, "Degrees of freedom in adaptive modulation: a unified view," *IEEE Trans. Commun.*, vol. 49, no. 9, pp. 1561–1571, Sep. 2001.

[22] R. Strobel, A. Barthelme, and W. Utschick, "Implementation aspects of nonlinear precoding for G.fast - coding and legacy receivers," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2017, pp. 111–115.

[23] J. Neckebroek *et al.*, "Novel bitloading algorithms for coded G.fast DSL transmission with linear and nonlinear precoding," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2015, pp. 945–951.

[24] A. K. Lenstra, H. W. Lenstra, and L. Lovász, "Factoring polynomials with rational coefficients," *Mathematische Annalen*, vol. 261, no. 4, pp. 515–534, Dec 1982. [Online]. Available: https://doi.org/10.1007/BF01457454

[25] Y. H. Gan, C. Ling, and W. H. Mow, "Complex lattice reduction algorithm for low-complexity full-diversity mimo detection," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2701–2710, July 2009.

[26] M. Razaviyayn, M. Hong, and Z. Luo, "A Unified Convergence Analysis of Block Successive Minimization Methods for Nonsmooth Optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013. [Online]. Available: https://doi.org/10.1137/120891009

[27] M. ApS, *The MOSEK optimization toolbox for MATLAB manual. Version 8.1.*, 2017. [Online]. Available: http://docs.mosek.com/8.1/toolbox/index.html

[28] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.

[30] M. S. Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of Second-order Cone programming," *LINEAR ALGEBRA AND ITS APPLICATIONS*, vol. 284, pp. 193–228, 1998.

[31] S. Vishwanath, N. Jindal, and A. Goldsmith, "Duality, achievable rates, and sum-rate capacity of Gaussian MIMO broadcast channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2658–2668, Oct. 2003.

[32] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 4792–4799, Dec. 2008.

[33] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively Weighted MMSE Approach to Distributed Sum-Utility Maximization for a MIMO Interfering Broadcast Channel," *IEEE Trans. Signal Process.*, vol. 59, pp. 4331–4340, 2011.

[34] W. Yu and R. Lui, "Dual methods for nonconvex spectrum optimization of multicarrier systems," *IEEE Trans. Commun.*, vol. 54, no. 7, pp. 1310–1322, July 2006.

[35] *Fast Access to Subscriber Terminals (FAST) - Power Spectral Density Specification*, Recommendation ITU-T G.9700, April 2014.

[36] W. Lanneer, P. Tsiaflakis, J. Maes, and M. Moonen, "α-Fair Dynamic Spectrum Management for QRD-Based Precoding with User Encoding Ordering in Downstream G.fast Transmission," *IEEE Trans. Commun.*, pp. 1–1, 2018.

[37] W. Lanneer, P. Tsiaflakis, J. Maes, and M. Moonen, "Linear and Nonlinear Precoding Based Dynamic Spectrum Management for Downstream Vectored G.fast Transmission," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1247–1259, Mar. 2017.

[38] D. Wubben, D. Seethaler, J. Jalden, and G. Matz, "Lattice reduction," *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 70–91, May. 2011.

[39] R. Durrett, *Probability: Theory and Examples, 4th edition*. Cambridge University Press, 2010.