KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Lanneer W., Moonen M., Tsiaflakis P., and Maes J. **Low-Complexity Nonlinear Zero-Forcing Precoding Under Per-Line Power Constraints for Improved Downstream G.fast Active-User Peak-Rates** Published in *IEEE Transactions on Communications,* vol. 66, no. 6, pp. 2696-2707, June 2018.
Archived version	Author manuscript: the content is identical to the content of the published paper (except for some additional material in Section V), but without the final typesetting by the publisher
Published version	10.1109/TCOMM.2018.2799613
Journal homepage	http://www.comsoc.org/TC
Author contact	wouter.lanneer@esat.kuleuven.be *+ 32 16 32 79 75*
IR	https://limo.libis.be/primo-explore/fulldisplay?docid=LIRIAS1628280&context=L&vid=Lirias&search_scope=Lirias&tab=default_tab&lang=en_US

(article begins on next page)

Low-Complexity Nonlinear Zero-Forcing Precoding Under Per-Line Power Constraints for Improved Downstream G.fast Active-User Peak-Rates

Wouter Lanneer, *Student Member, IEEE,* Paschalis Tsiaflakis, *Member, IEEE,*
Jochen Maes, *Senior Member, IEEE,* and Marc Moonen, *Fellow, IEEE*

***Abstract*—We consider nonlinear zero-forcing (ZF) precoding design to improve the downstream G.fast peak-rates when only a few users in the cable binder are active. In order to compute the optimal nonlinear ZF precoder under per-line power constraints (PLPCs), we present a novel low-complexity dual decomposition algorithm, in which the key is the use of Lagrange multiplier based virtual precoders to transform the PLPCs into an easier virtual sum-power constraint (SPC), such that the SPC-optimality of the QR decomposition-based precoder may be exploited. We show a reduced computational complexity of this algorithm over the state-of-the-art SVD-block-diagonalization-based dual decomposition algorithm. We present simulations of a 10-line cable binder that demonstrate substantial peak-rate gains over standard QR decomposition-based ZF precoding in DSL, due to the increasingly stronger crosstalk channels in the G.fast frequency range (up to 212 MHz). Furthermore, we show that the proposed algorithm naturally extends to the scenario with multiple lines terminating at the customer premise equipments.**

***Index Terms*—DSL, G.fast, dynamic spectrum management, nonlinear precoding, zero-forcing, per-line power constraints.**

I. Introduction

THE next-generation digital subscriber lines (DSL) access technology called G.fast [1] aims at providing gigabit downstream transmission speeds. Such fiber-like speeds become achievable by transmitting over very short copper lines (below 100 m) and in frequencies up to 212 MHz. Unfortunately, the use of such high frequencies also leads to increasingly stronger levels of crosstalk interference among the lines within a cable binder, making dynamic spectrum management (DSM) techniques for crosstalk precompensation (also called precoding or downstream vectoring) very challenging.

Foremost among the challenges is that the traditional linear zero-forcing (ZF) precoder [2] suffers from large precompensation signals in the G.fast frequency range, which significantly increase the per-line transmit power spectral densities (PSDs). To enforce per-line power constraints, transmit PSD optimization is necessary [3]. The linear ZF precoder may be replaced by the minimum mean squared error (MMSE) precoder which tolerates some residual crosstalk at the benefit of an increased signal-to-noise-ratio (SNR) [4]. An alternative for improving the SNR is nonlinear precoding (NLP). Throughout the paper, NLP is used to refer to multi-user dirty paper coding-like precoding strategies based on sequential interference pre-subtraction. One representative implementation, well-known in the DSL community, is Tomlinson-Harashima precoding (THP) [5].

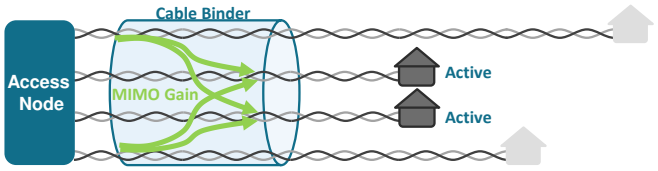


Fig. 1. Supporting improved G.fast active-user peak-rate performance.

On the other hand, these strong crosstalk channels also yield an important opportunity to boost the peak-rates of active users within the cable binder [6]. Whenever some of the users in the binder are inactive, the line drivers of these idle users may still be operated to transmit additional energy into the cable binder (see Fig. 1). In point of fact, the strong crosstalk channels will leak this additional energy into the lines of active users, such that the received signal power or the SNR of the active users increases. In other words, the strong crosstalk channels also provide the cable binder with a so-called multi-user spatial-multiplexing gain or a multi-user multiple-input-multiple-output (MU-MIMO) capacity or array gain, which may thus enable the support of improved active-user peak-rates in G.fast. Naturally, this comes at the cost of increased transmit powers and the inability to turn off some analog functionality of the idle-user line drivers [7].

To exploit this spatial-multiplexing gain within a binder, we study nonlinear ZF precoding (ZF-NLP) design to improve the active-user peak-rates, as a low-complexity alternative for nonlinear MMSE precoding [4]. In the DSL context ZF-NLP design using a sum-power constraint (SPC) as well as per-

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 'The 5th Generation Broadband', Research Project FWO nr. G.0912.13 'Cross-layer optimization with real-time adaptive dynamic spectrum management for fourth generation broadband access networks', the Interuniversity Attractive Poles Programme initiated by the Belgian Science Policy Office: IUAP P7/23 'Belgian network on stochastic modeling analysis design and optimization of communication systems' (BESTCOM) 2012-2017, and the KU Leuven Research Council CoE PFV/10/002 (OPTEC). 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).

P. Tsiaflakis and J. Maes are with the Copper Access and Indoor Team of Nokia Bell Labs, BE 2018 Antwerp, Belgium (e-mail: {paschalis.tsiaflakis, jochen.maes}@nokia-bell-labs.com).

line power constraints (PLPCs)[1] [5], [8]–[15] has most been relying on the QR decomposition (QRD)-based precoder. ZF-NLP design then transforms into pure power allocation with a water-filling solution. This approach can be indeed shown to be optimal for ZF-NLP under a SPC [16]. However, G.fast is subject to PLPCs, for which the QRD-based precoder is not necessarily optimal anymore and may be outperformed by other nonlinear ZF precoders. The optimal nonlinear ZF precoder under PLPCs can be computed by the state-of-the-art algorithm proposed in [17], which basically integrates the SVD-block-diagonalization (SVD-BD)-based method [18], [19] into a dual decomposition approach with an iterative subgradient search.

In this paper, we develop an alternative dual decomposition algorithm for optimal ZF-NLP design under PLPCs. Key in this algorithm is the use of Lagrange multiplier based virtual precoders to transform the PLPCs into an easier virtual SPC, which allows to exploit the SPC-optimality of the QRD-based precoder. We prove that this QRD-based method is equivalent to the state-of-the-art SVD-BD [17] while boasting a significantly lower computational complexity.

To keep the notations clean, most of the paper is limited to the scenario where each user has a single line terminating at its customer premise equipment, corresponding to a so-called multiple-input-single-output broadcast channel (MISO-BC). Nevertheless, we show that our proposed algorithm naturally extends to the scenario where users have multiple lines terminating at their customer premise equipment, corresponding to a so-called MIMO-BC.

This paper is organized as follows. Section II models downstream G.fast transmission. Section III introduces the ZF-NLP design problem and Section IV re-views QRD-based precoding both for PLPCs and for a SPC. Section V presents the state-of-the-art SVD-BD and novel QRD-based dual decomposition algorithm for optimal ZF-NLP design under PLPCs. Section VI discusses the computational complexity. Section VII presents simulation results for a G.fast cable binder. Section VIII discusses the generalization of multiple lines terminating at the customer premise equipment. Finally, Section IX concludes the paper.

Notations: We use lowercase boldface letters to denote vectors and uppercase boldface letters for matrices. We use $\mathbf{I}_A$ as the identity matrix of size A, $(.)^T$ as the transpose, $(.)^H$ as Hermitian transpose, $(.)^*$ as the complex conjugate, $\mathcal{E}\{.\}$ as expectation,$[\mathbf{X}]_{i,j}$ as the i,j-th element of $\mathbf{X}$, $[x]_a^b$ as $\min(b, \max(x, a))$, $\mathrm{Tr}\{.\}$ as trace, $\mathrm{diag}\{\mathbf{x}\}$ as a diagonal matrix with vector $\mathbf{x}$ on the main diagonal, $\mathrm{diag}\{\mathbf{X}\}$ as a diagonal matrix with $\{[\mathbf{X}]_{ii}, \forall i\}$ on the main diagonal, $|\cdot|$ as the absolute value, and $\|\mathbf{x}\|_p$ as the p-norm of $\mathbf{x}$.

II. Downstream G.fast Transmission Model

A. Transmission Model

We consider downstream transmission in a G.fast cable binder consisting of L lines connecting the access node to the customer premise equipments. There are $N(\leq L)$ users in the binder that are active, while all L line drivers are available for transmission. Assuming standard synchronous discrete-multi-tone (DMT) modulation, transmission is modeled independently on each tone (or frequency sub-carrier) $k = [1, \ldots, K]$ as

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

$\mathbf{x}_k \triangleq [x_k^1, \ldots, x_k^L]$ is the transmit vector on tone k, with x_k^l the signal transmitted on line l and tone k. $\mathbf{y}_k \triangleq [y_k^1, \ldots, y_k^N]$ is the receive vector on tone k, with y_k^n the signal received by user n. $\mathbf{z}_k \triangleq [z_k^1, \ldots, z_k^N]$ 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 $N \times L$ channel matrix on tone k. The diagonal elements of $\mathbf{H}_k$ contain the direct channels whilst the off-diagonal elements contain the crosstalk channels.

Although the direct channels of $\mathbf{H}_k$ typically are dominant below 30 MHz (i.e. $|h_k^{n,n}| \gg |h_k^{n,m}|, m \neq n$), recent measurements show that this is not valid anymore for higher frequencies of G.fast where the direct channels may even be weaker than the crosstalk channels [20]. This makes the crosstalk channels very effective for transmitting additional energy to the active users. We assume perfect knowledge of the channel matrices. In practice the channel matrices are estimated and tracked, which requires little additional overhead due to the slow time variations in DSL channels [1].

In G.fast per-line spectral mask constraints are included, in order to not generate too much interference into other technologies, and combined with per-line aggregated transmit power (ATP) constraints [21]. These PLPCs are given as[2]

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

$$\sum_k \mathcal{E}\{|x_k^l|^2\} \leq P^{\mathrm{ATP}}, \quad \forall l. \tag{3}$$

The number of bits that can be loaded on tone k for user n is widely modeled [22] by

$$b_k^n = \left\lfloor \log_2 \left(1 + \frac{\mathrm{SNR}_k^n}{\Gamma} \right) \right\rfloor^{b^{\max}}, \tag{4}$$

where SNR_k^n is the achieved SNR on tone k for user n (which will be specificed in Section III) and Γ denotes the SNR to capacity gap or SNR gap, including a noise margin and a coding gain. The SNR gap achieving a BER of 10^{-7} for uncoded QAM transmission is approximately equal to 9.75 dB [22]. Adding to this a noise margin of 6 dB minus a coding gain of 5.5 dB leads to an overall SNR gap Γ = 10.25 dB. Typically a maximum bitloading (bit cap) $b^{\max}$ is imposed. Finally, the total data rate of user n is $R_n = f_s \sum_k b_k^n$, where f_s is the DMT symbol rate.

B. Nonlinear Precoding

NLP sequentially encodes the user data signals in order to pre-subtract the crosstalk from previously encoded users without transmit power penalties, which may be implemented

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

[2] Note that $P^{\mathrm{ATP}} \triangleq \overline{P}^{\mathrm{ATP}}/\Delta_f$ with $\overline{P}^{\mathrm{ATP}}$ the maximum ATP and Δ_f the tone spacing.

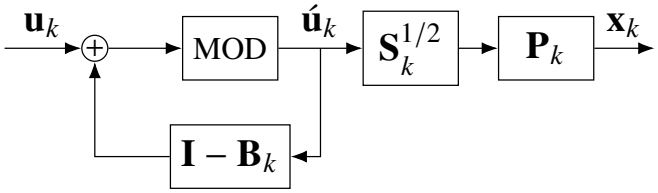


Fig. 2. Tomlinson-Harashima precoding structure on tone k.

in practice with THP [5]. THP (depicted in Fig. 2) applies feedforward and feedback filtering to the user data vector $\mathbf{u}_k \triangleq [u_k^1,\ldots,u_k^N]$, where the user data signals $\{u_k^n\}$ are assumed to be uncorrelated and consist of unit power and zero mean QAM symbols, i.e. $\mathcal{E}\{\mathbf{u}_k\mathbf{u}_k^H\} = \mathbf{I}_N$.

The $N \times N$ feedback precoding matrix $\mathbf{I} - \mathbf{B}_k$ realizes the sequential crosstalk "pre-subtraction". The pre-subtracted signals are prevented from increase in transmit power by applying a nonlinear modulo operation. The intermediate vector $\acute{\mathbf{u}}_k \triangleq [\acute{u}_k^1,\ldots,\acute{u}_k^N]$ is computed sequentially from the first to the last user[3] as

$$\acute{u}_k^n = \mathrm{mod}\left[u_k^n - \sum_{m=1}^{n-1}[\mathbf{B}_k]_{n,m}\acute{u}_k^m\right], \forall n. \tag{5}$$

with $\mathbf{B}_k$ a lower triangular matrix defined as

$$[\mathbf{B}_k]_{n,m} = \begin{cases} \frac{[\mathbf{H}_k\mathbf{P}_k]_{n,m}\sqrt{s_k^m}}{[\mathbf{H}_k\mathbf{P}_k]_{n,n}\sqrt{s_k^n}} & \text{if } m \le n \\ 0 & \text{else.} \end{cases} \tag{6}$$

The transmit vector is then formed by gain scaling the intermediate data vector $\acute{\mathbf{u}}_k$ and multiplication with the $L \times N$ feedforward precoding matrix $\mathbf{P}_k \triangleq \left[\mathbf{p}_k^1,\ldots,\mathbf{p}_k^N\right]$

$$\mathbf{x}_k = \mathbf{P}_k\mathbf{S}_k^{1/2}\acute{\mathbf{u}}_k. \tag{7}$$

The diagonal matrix $\mathbf{S}_k \triangleq \mathrm{diag}\{\mathbf{s}_k\}$ with $\mathbf{s}_k \triangleq [s_k^1,\ldots,s_K^N]$ models the per-user transmit PSDs on tone k. The transmit PSD on line l and tone k is then defined as

$$\mathcal{E}\{|x_k^l|^2\} = \sum_m \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m. \tag{8}$$

At the receiver a second modulo operation is required to make the first modulo operation transparent [5]

$$\begin{aligned}\hat{u}_k^n &= \mathrm{mod}\left[\frac{y_k^n}{[\mathbf{H}_k\mathbf{P}_k]_{n,n}\sqrt{s_k^n}}\right] \\ &\simeq u_k^n + \sum_{m>n}\frac{[\mathbf{H}_k\mathbf{P}_k]_{n,m}\sqrt{s_k^m}}{[\mathbf{H}_k\mathbf{P}_k]_{n,n}\sqrt{s_k^n}}\acute{u}_k^m + \frac{z_k^n}{[\mathbf{H}_k\mathbf{P}_k]_{n,n}\sqrt{s_k^n}}.\end{aligned} \tag{9}$$

Note that in practice the modulo operation at the transmitter results in a precoding loss, since it causes the intermediate signals $\{\acute{u}_k^n\}$ to be (approximately) uniformly distributed [23], such that they have more energy than the QAM signals $\{u_k^n\}$. This loss is most pronounced for low SNR and is maximum 1.25 dB in G.fast [12]. Since it has a low impact on the total data rates, this loss is not considered in our simulations.

[3]We assume without loss of generality the encoding order to be given by the user index, which corresponds to $w_1 \ge w_2 \ge \cdots \ge w_N$ for user weights as introduced in Section III. See [15] for optimizing the user encoding orders.

Furthermore, THP has been shown to be more sensitive to channel estimations errors and other non-idealities than LP, such that less sensitive alternative implementations of THP may be necessary [24].

III. Nonlinear ZF Precoding Design Problem

The ZF condition implies crosstalk interference free transmission for all users. As can be seen from (9), the ZF condition only requires that for each user n the crosstalk interference from users $m > n$ is canceled by the forward precoder $\mathbf{P}_k$, because the crosstalk interference created by users $m < n$ is already canceled by the THP structure. An equivalent condition is then

$$\mathbf{H}_k\mathbf{P}_k = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ * & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & 1 \end{bmatrix}}_{\mathbf{L}_k} \tag{10}$$

where $\mathbf{L}_k$ denotes an $N \times N$ arbitrary lower triangular matrix with unit diagonal elements. When (10) is satisfied, the transmission model (9) reduces to

$$\hat{\mathbf{u}}_k = \mathbf{u}_k + \mathbf{S}_k^{-1/2}\mathbf{z}_k. \tag{11}$$

Since (10) entails some degrees of freedom, the precoder matrix $\mathbf{P}_k$ design is integrated into the ZF-NLP design problem, which is formulated as follows

$$\begin{aligned} \underset{\{\mathbf{P}_k\},\{\mathbf{s}_k\in\mathcal{D}_k\}}{\text{maximize}} \quad & \sum_k\sum_n w_n b_k^n \\ \text{s.t.} \quad & \sum_m \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m \le P_k^{\max}, && \forall k,l \\ & \sum_k\sum_m \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m \le P^{\mathrm{ATP}}, && \forall l \\ & \mathbf{H}_k\mathbf{P}_k = \mathbf{L}_k, && \forall k. \end{aligned} \tag{12}$$

Here w_n is the weight of user n and the bit loading for user n on tone k with ZF precoding is

$$b_k^n = \log_2\left(1 + \Gamma^{-1}\sigma_k^{-1}s_k^n\right). \tag{13}$$

The bit cap $b^{\max}$ is translated here into a PSD cap in (12) where $\mathcal{D}_k \triangleq \{s_k^n | 0 \le s_k^n \le s_k^{\max} = \Gamma\sigma_k\left(2^{b^{\max}}-1\right), \forall n\}$.

IV. QRD-based ZF-NLP

A. Per-Line Power Constraints

In the DSL context [5], [8]–[15], the nonlinear ZF precoder matrix $\mathbf{P}_k$ is most often defined based on the QRD of the conjugated channel matrix $\mathbf{H}_k$ as it was proposed in [5]. Adopting this QRD-based precoder simplifies the ZF-NLP design in (12) which then transforms into a pure transmit PSD optimization problem.

More specifically, this approach uses the following QRD

$$\mathbf{H}_k^H \overset{\mathrm{qr}}{=} \mathbf{Q}_k\mathbf{R}_k, \tag{14}$$

where $\mathbf{Q}_k$ is an $L \times N$ unitary matrix and $\mathbf{R}_k$ is an $N \times N$ upper triangular matrix. Then the feedforward precoding matrix is

Algorithm 1: QRD-based ZF-NLP Under PLPCs

Initialize $\mathbf{P}_k = \mathbf{Q}_k \text{diag}\{\mathbf{R}_k^H\}^{-1}$ using (14), $\forall k$
Initialize $\mathbf{\Omega}_k = \text{diag}\{\theta^1 + \lambda_k^1, \ldots, \theta^L + \lambda_k^L\}, \forall k$
repeat
  Update $\left\{\theta^1, \ldots, \theta^L\right\}$ using subgradient method
  for $k = 1 \cdots K$ **do**
    repeat
      Update $\left\{\lambda_k^1, \ldots, \lambda_k^L\right\}$ using subgradient method
      $\forall n : s_k^n = \left[\frac{w_n}{\sum_{j=1}^L (\theta^j + \lambda_k^j)\left|[\mathbf{P}_k]_{j,n}\right|^2} - \Gamma\sigma_k\right]_0^{s_k^{\max}}$
    until $\lambda_k^l \left|\sum_m \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m - P_k^{\text{mask}}\right| < \epsilon \quad \forall l$
  end
until $\theta^l \left|\sum_k \sum_m \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m - P^{\text{ATP}}\right| < \epsilon \quad \forall l$

set to $\mathbf{P}_k = \mathbf{Q}_k \text{diag}(\mathbf{R}_k^H)^{-1}$. Note that $\mathbf{P}_k$ satisfies the ZF condition in (10). Now, the per-user transmit PSD optimization is a convex problem with strictly feasible constraints meaning that the KKT conditions are sufficient for optimality [25]. Examining these leads to a multi-level water-filling solution for the per-user transmit PSDs

$$s_k^n = \left[w_n \left(\sum_{j=1}^{L} (\theta^j + \lambda_k^j)\left|[\mathbf{P}_k]_{j,n}\right|^2\right)^{-1} - \Gamma\sigma_k\right]_0^{s_k^{\max}} \quad \forall k, n, \quad (15)$$

where $\{\theta^l\}$ and $\{\lambda_k^l\}$ are the non-negative Lagrange multipliers corresponding to the per-line ATP and spectral mask constraints on tone k. The Lagrange multipliers should be chosen such that the KKT complementary conditions are satisfied, i.e.,

$$\theta^l \left(\sum_k \sum_m \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m - P^{\text{ATP}}\right) = 0, \quad \forall l \quad (16)$$

$$\lambda_k^l \left(\sum_m \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m - P_k^{\text{mask}}\right) = 0, \quad \forall k, l. \quad (17)$$

Strategies for searching the optimal Lagrange multipliers are elaborated in Section V. We will refer to this approach as QRD-based ZF-NLP in this paper. A complete algorithmic description is included in Alg. 1.

However, for ZF-NLP design under PLPCs the QRD-based precoder matrix is not necessarily optimal. To see this, consider the simple example of one active user ($N = 1$) and a single-carrier channel ($K = 1$) given as $\mathbf{h} = [h_1, \cdots, h_L]$ and with a PLPC P. As can easily be verified, the optimal precoder vector $\mathbf{p}^{\text{opt}}$ and scalar transmit power s^{opt} exhibit the following structure

$$\mathbf{p}^{\text{opt}} = \left[\frac{e^{j\phi_1}}{\|\mathbf{h}\|_1}, \ldots, \frac{e^{j\phi_L}}{\|\mathbf{h}\|_1}\right]^T, \quad \text{with } \phi_l = \angle h_l^* \quad (18)$$

$$s^{\text{opt}} = P\|\mathbf{h}\|_1^2, \quad (19)$$

which satisfies the ZF condition in (10) since $\mathbf{h}\mathbf{p}^{\text{opt}} = 1$. Whereas the QRD-based precoder corresponds to a maximum ratio transmission

$$\mathbf{p}^{\text{QRD}} = \left[\frac{h_1^*}{\|\mathbf{h}\|_2^2}, \ldots, \frac{h_L^*}{\|\mathbf{h}\|_2^2}\right]^T \quad (20)$$

$$s^{\text{QRD}} = P\|\mathbf{h}\|_2^4 \left(\max_m |h_m|^2\right)^{-1}. \quad (21)$$

Although $\mathbf{p}^{\text{opt}}$ is generally outside the row-space of $\mathbf{h}$, its corresponding transmit power will always be larger than that of the QRD-based precoder, as the following inequality demonstrates

$$\frac{s^{\text{QRD}}}{P} = \frac{\|\mathbf{h}\|_2^2 \|\mathbf{h}\|_2^2}{\max_m |h_m|^2} \leq \frac{\|\mathbf{h}\|_1 \|\mathbf{h}\|_1 \max_m |h_m|^2}{\max_m |h_m|^2} = \|\mathbf{h}\|_1^2 = \frac{s^{\text{opt}}}{P} \quad (22)$$

where we used the fact that $\|\mathbf{h}\|_2^2 \leq \|\mathbf{h}\|_1 \|\mathbf{h}\|_\infty$. With (13) s^{opt} will then lead to a higher bit loading than s^{QRD}. This illustrates that in addition to satisfying (10) the optimal precoder matrix $\mathbf{P}_k$ should also result in maximum SNRs. Here the QRD-based precoder generally results in only one active PLPC, whilst the optimal solution has always L active PLPCs.

B. Sum-Power Constraint

As it will be relevant in Section V-B, we also review here the active-user peak-rate maximization problem for ZF-NLP under a single SPC $P^{\text{SPC}} = L \times P^{\text{ATP}}$

$$\begin{aligned} \underset{\{\mathbf{P}_k\}, \{\mathbf{s}_k \in \mathcal{D}_k\}}{\text{maximize}} \quad & \sum_k \sum_n w_n b_k^n \\ \text{s.t.} \quad & \sum_k \sum_m \sum_l \left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m \leq P^{\text{SPC}} \\ & \mathbf{H}_k \mathbf{P}_k = \mathbf{L}_k, \quad \forall k. \end{aligned} \quad (23)$$

Problem (23) is a relaxation and performance upper bound of (12). Here the QRD-based precoder is optimal, in contrast to the case with PLPCs, as stated in the next theorem and proved as follows (similar to the proofs in [16] and [26]).

Theorem 1: The optimal precoder to (23) at tone k is $\mathbf{P}_k = \mathbf{Q}_k \text{diag}(\mathbf{R}_k^H)^{-1}$ where $\mathbf{H}_k^H \overset{\text{qr}}{=} \mathbf{Q}_k \mathbf{R}_k$.

Proof: The general structure for any precoder satisfying (10) is

$$\mathbf{P}_k = \mathbf{H}_k^\dagger \mathbf{L}_k + \mathbf{P}_k^\perp \mathbf{A}_k \quad (24)$$

with $\mathbf{H}_k^\dagger = \mathbf{H}_k^H \left(\mathbf{H}_k \mathbf{H}_k^H\right)^{-1}$ denoting the pseudo-inverse of $\mathbf{H}_k$, and $\mathbf{P}_k^\perp = \mathbf{I}_L - \mathbf{H}_k^\dagger \mathbf{H}_k$ the orthogonal projection matrix onto the null-space of $\mathbf{H}_k$, and where $\mathbf{L}_k$ is a lower triangular matrix with ones on the diagonal, and $\mathbf{A}_k$ is an arbitrary matrix. First, observe that for any set of feasible transmit PSDs $\mathbf{S}_k \succeq \mathbf{0}$

$$\|\mathbf{P}_k \mathbf{S}_k^{1/2}\|_F^2 = \text{Tr}\left\{\mathbf{P}_k \mathbf{S}_k \mathbf{P}_k^H\right\} \quad (25)$$

$$= \text{Tr}\left\{\left(\mathbf{H}_k^\dagger \mathbf{L}_k + \mathbf{P}_k^\perp \mathbf{A}_k\right) \mathbf{S}_k \left(\mathbf{H}_k^\dagger \mathbf{L}_k + \mathbf{P}_k^\perp \mathbf{A}_k\right)^H\right\} \quad (26)$$

$$\geq \text{Tr}\left\{\mathbf{H}_k^\dagger \mathbf{L}_k \mathbf{S}_k \left(\mathbf{H}_k^\dagger \mathbf{L}_k\right)^H\right\} \quad (27)$$

since $\mathbf{P}_k^\perp \mathbf{H}_k^\dagger = \mathbf{0}$ and $\mathbf{P}_k^\perp \mathbf{A}_k \mathbf{S}_k \mathbf{A}_k^H \mathbf{P}_k^\perp \succeq \mathbf{0}$. This means that setting $\mathbf{A}_k = \mathbf{0}$ minimizes the expression in (26).

In addition, using the QRD of $\mathbf{H}_k^H$, $\mathbf{H}_k^\dagger$ can be re-written as $\mathbf{Q}_k(\mathbf{R}_k^H)^{-1}$ such that

$$\mathrm{Tr}\left\{\mathbf{H}_k^\dagger \mathbf{L}_k \mathbf{S}_k \left(\mathbf{H}_k^\dagger \mathbf{L}_k\right)^H\right\} = \mathrm{Tr}\left\{\left(\mathbf{R}_k^H\right)^{-1} \mathbf{L}_k \mathbf{S}_k \mathbf{L}_k^H \left(\mathbf{R}_k\right)^{-1}\right\}. \quad (28)$$

Now since $\left(\mathbf{R}_k^H\right)^{-1}$ is a lower triangular matrix, $\left(\mathbf{R}_k^H\right)^{-1}\mathbf{L}_k$ is also lower triangular and can be decomposed as

$$\left(\mathbf{R}_k^H\right)^{-1}\mathbf{L}_k = \left(\mathbf{I} + \overline{\mathbf{L}}_k\right)\mathbf{D}_k, \quad (29)$$

$$\text{with } \mathbf{D}_k = \mathrm{diag}\{\mathbf{R}_k^H\}^{-1}$$

$$\overline{\mathbf{L}}_k = \left(\mathbf{R}_k^H\right)^{-1}\mathbf{L}_k\left(\mathbf{D}_k\right)^{-1} - \mathbf{I},$$

where $\overline{\mathbf{L}}_k$ is a strictly lower triangular matrix with zeros on the main diagonal. Putting this into (28) leads to

$$\mathrm{Tr}\left\{\left(\mathbf{I} + \overline{\mathbf{L}}_k\right)\mathbf{D}_k\mathbf{S}_k\mathbf{D}_k\left(\mathbf{I} + \overline{\mathbf{L}}_k\right)^H\right\} \geq \mathrm{Tr}\left\{\mathbf{D}_k\mathbf{S}_k\mathbf{D}_k\right\} = \sum_n s_k^n/|r_k^{n,n}|^2, \quad (30)$$

because $\mathrm{Tr}\left\{\overline{\mathbf{L}}_k\mathbf{D}_k\mathbf{S}_k\mathbf{D}_k\right\} = 0$ and $\overline{\mathbf{L}}_k\mathbf{D}_k\mathbf{S}_k\mathbf{D}_k\overline{\mathbf{L}}_k^H \geq \mathbf{0}$. This means that choosing $\overline{\mathbf{L}}_k = \mathbf{0}$, or equivalently $\mathbf{L}_k = \mathbf{R}_k^H\mathrm{diag}\{\mathbf{R}_k^H\}^{-1}$, minimizes the Frobenius norm of $\mathbf{P}_k\mathbf{S}_k^{1/2}$ which indeed leads to the QRD-based precoder, by observing that $\mathbf{P}_k = \mathbf{Q}_k(\mathbf{R}_k^H)^{-1}\mathbf{L}_k = \mathbf{Q}_k\mathrm{diag}\{\mathbf{R}_k^H\}^{-1}$. ■

Consequently, (23) is solved optimally using water-filling, i.e.,

$$s_k^n = \left[w_n\left(\theta \underbrace{\sum_{j=1}^{L}\left|[\mathbf{P}_k]_{j,n}\right|^2}_{1/|r_k^{n,n}|^2}\right)^{-1} - \Gamma\sigma_k\right]_0^{s_k^{\max}} \quad \forall k,n. \quad (31)$$

V. Dual Decomposition for Optimal ZF-NLP Under PLPCs

The basic idea of dual decomposition is to solve problem (12) by maximizing its Lagrangian. The Lagrangian is formed by augmenting the objective function with a weighted sum of the PLPCs, which decouples these constraints across the tones and users. This leads to the following Lagrangian associated with problem (12)

$$\mathcal{L} = \sum_k\left[\sum_n w_n b_k^n - \mathrm{Tr}\{\mathbf{\Omega}_k\mathbf{P}_k\mathbf{S}_k\mathbf{P}_k^H\} + \sum_l \lambda_k^l P_k^{\mathrm{mask}}\right] + \sum_l \theta^l P^{\mathrm{ATP}} \quad (32)$$

where $\mathbf{\Omega}_k = \mathrm{diag}\{\theta^l + \lambda_k^l\}$ is a diagonal matrix containing the non-negative Lagrange multipliers corresponding to the per-line ATP ($\boldsymbol{\theta} \triangleq \left[\theta^1,\ldots,\theta^L\right]$) and spectral mask ($\boldsymbol{\lambda}_k \triangleq \left[\lambda_k^1,\ldots,\lambda_k^L\right]$) constraints on tone k. The corresponding Lagrange dual function is then the constrained maximization of the Lagrangian for a given set of Lagrange multipliers

$$g(\{\mathbf{\Omega}_k\}) = \underset{\{\mathbf{P}_k\},\{\mathbf{s}_k\in\mathcal{D}_k\}}{\text{maximize}} \quad \mathcal{L}(\{\mathbf{\Omega}_k\},\{\mathbf{P}_k\},\{\mathbf{s}_k\}) \quad \text{s.t.} \quad \mathbf{H}_k\mathbf{P}_k = \mathbf{L}_k, \quad \forall k. \quad (33)$$

The dual optimization problem is defined as

$$\underset{\{\mathbf{\Omega}_k\}}{\text{minimize}} \quad g(\{\mathbf{\Omega}_k\}) \quad \text{s.t.} \quad \boldsymbol{\theta} \geq \mathbf{0}, \quad \boldsymbol{\lambda}_k \geq \mathbf{0}, \quad \forall k. \quad (34)$$

Strictly speaking, problem (12) is non-convex (i.e. the power constraints are not in a convex form). However, like in [17], one can define the rank-one transmit covariance matrix of user n on tone k as $\mathbf{C}_k^n \triangleq s_k^n\mathbf{p}_k^n\mathbf{p}_k^{n,H}$ with $\mathbf{p}_k^n$ and $\mathbf{h}_k^{n,H}$ being the n-th column of $\mathbf{P}_k$ and n-th row of $\mathbf{H}_k$, respectively, and reformulate problem (12) as follows (and discarding the bit cap constraints[4]):

$$\begin{aligned}
\underset{\{\mathbf{C}_k^n\succeq\mathbf{0}\}}{\text{maximize}} \quad & \sum_k\sum_n w_n\log_2\left(1 + \frac{\mathbf{h}_k^{n,H}\mathbf{C}_k^n\mathbf{h}_k^n}{\Gamma\sigma_k}\right) \\
\text{s.t.} \quad & \sum_m\left[\mathbf{C}_k^m\right]_{l,m} \leq P_k^{\mathrm{mask}}, && \forall k,l, \\
& \sum_k\sum_m\left[\mathbf{C}_k^m\right]_{l,m} \leq P^{\mathrm{ATP}}, && \forall l, \\
& \mathbf{h}_k^{m,H}\mathbf{C}_k^n\mathbf{h}_k^m = 0, && \forall m<n,k.
\end{aligned} \quad (35)$$

Note that problem (35) is in fact a convex relaxation of problem (12), since the non-convex rank-one constraints of $\mathbf{C}_k^n$ are omitted for all users and tones. In addition, problem (35) is convex with strictly feasible constraints, such that strong duality holds in (35). Moreover, (35) always admits an optimal rank-one solution [17], meaning that the relaxation of (12) with (35) is tight (without bit cap constraints). As a result, solving the dual problem (34) can be considered equivalent to solving the dual problem of (35), and therefore dual problem (34) and problem (12) have a zero duality gap as well.

Dual decomposition algorithms iteratively search for the optimal Lagrange multipliers that minimize (34) and enforce the KKT complementary conditions [see (16) and (17)]. In every iteration the Lagrange dual function (33) needs to be solved. A standard search strategy is the subgradient method [27], [28], which updates the Lagrange multipliers as

$$\theta^l = \left[\theta^l + \delta^t\left(\sum_k\sum_m\left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m - P^{\mathrm{ATP}}\right)\Big/P^{\mathrm{ATP}}\right]_0 \quad (36)$$

$$\lambda_k^l = \left[\lambda_k^l + \delta^t\left(\sum_m\left|[\mathbf{P}_k]_{l,m}\right|^2 s_k^m - P_k^{\mathrm{mask}}\right)\Big/P_k^{\mathrm{mask}}\right]_0, \quad (37)$$

where t is the iteration number, and δ^t is the scalar step size at iteration t. Above subgradient update method is guaranteed

[4] In fact, the bit cap constraints are pragmatically enforced in Section V. To enforce the bit cap constraints optimally, it is necessary to include them as linear constraints in function of $\mathbf{C}_k^n$ in problem (35).

Algorithm 2: Dual Decomposition for Optimal ZF-NLP Under PLPCs

1 Initialize $\mathbf{\Omega}_k = \text{diag}\{\theta^1 + \lambda_k^1, \ldots, \theta^L + \lambda_k^L\}, \forall k$
2 **repeat**
3 Update $\{\theta^1, \ldots, \theta^L\}$ using subgradient method
4 **for** $k = 1 \cdots K$ **do**
5 **repeat**
6 Update $\{\lambda_k^1, \ldots, \lambda_k^L\}$ using subgradient method
7 Update $[\mathbf{P}_k, \mathbf{s}_k]$ by solving $\tilde{g}_k(\mathbf{\Omega}_k)$
8 **until** $\lambda_k^l \left| \sum_m \left| [\mathbf{P}_k]_{l,m} \right|^2 s_k^m - P_k^{\text{mask}} \right| < \epsilon \quad \forall l$
9 **end**
10 **until** $\theta^l \left| \sum_k \sum_m \left| [\mathbf{P}_k]_{l,m} \right|^2 s_k^m - P^{\text{ATP}} \right| < \epsilon \quad \forall l$

to converge to the optimal Lagrange multipliers as long as δ^t is chosen sufficiently small [28]. A common choice for the step size sequence δ^t is that it is square summable but not summable, i.e., $\delta^t = \frac{\beta}{t}$ for some positive constant β (if the norm of the subgradient is bounded). Other step size sequence choices include nonsummable diminishing, i.e., $\delta^t = \frac{\beta}{\sqrt{t}}$, and constant $\delta^t = \beta$ [27].

The main difficulty however is in solving the Lagrange dual function (33) for a fixed set of Lagrange multipliers $\{\mathbf{\Omega}_k\}$ every iteration. Each solution requires the computation of the optimal precoder matrices and transmit PSDs as a function of $\{\mathbf{\Omega}_k\}$. Fortunately, maximizing the Lagrangian for a fixed set of Lagrange multipliers can be decoupled across all tones[5]

$$g(\{\mathbf{\Omega_k}\}) = \sum_k \left[\tilde{g}_k(\mathbf{\Omega}_k) + \sum_l \lambda_k^l P_k^{\text{mask}} \right] + \sum_l \theta^l P^{\text{ATP}} \quad (38)$$

with

$$\tilde{g}_k(\mathbf{\Omega}_k) = \underset{\{\mathbf{P}_k\}, \{\mathbf{s}_k \in \mathcal{D}_k\}}{\text{maximize}} \quad \sum_n w_n b_k^n - \text{Tr}\{\mathbf{\Omega}_k \mathbf{P}_k \mathbf{S}_k \mathbf{P}_k^H\}$$
$$\text{s.t.} \quad \mathbf{H}_k \mathbf{P}_k = \mathbf{L}_k. \quad (39)$$

Consequently, we continue with the per-tone Lagrange dual function (39) and drop the tone index for ease of representation (tone indices are re-introduced in Section VIII). A complete algorithmic description of dual decomposition for optimal ZF-NLP under PLPCs is listed in Alg. 2.

To solve (39) (corresponding to line 7 of Alg. 2), we present in the remainder of this section first the state-of-the-art SVD-BD method [17]. Then, we propose a novel optimal method in which the key is the use of Lagrange multiplier based virtual precoders to transform the PLPCs into an easier virtual SPC. This allows to exploit the SPC-optimality of the QRD-based precoder. We show a reduced computational complexity over the SVD-BD in Section VI.

A. SVD-BD-Based Method [17]

The state-of-the-art dual decomposition algorithm for optimal ZF-NLP under PLPCs proposed in [17] is based on the SVD-block-diagonalization and will be referred here to as SVD-BD. Its premise is that the ZF conditions in (10) dictate that the precoder vector $\mathbf{p}_n$ of user n, i.e., the n-th column of $\mathbf{P}$, should lie in the null-space spanned by

$$\mathbf{G}_n \triangleq \begin{bmatrix} \mathbf{h}_1^H \\ \mathbf{h}_2^H \\ \vdots \\ \mathbf{h}_{n-1}^H \end{bmatrix}, \quad (40)$$

with $\mathbf{h}_n^H$ the n-th row vector of $\mathbf{H}$. Therefore, as a first step, the following full SVDs are computed

$$\mathbf{G}_n \overset{\text{SVD}}{=} \mathbf{U}_n \begin{bmatrix} \mathbf{\Sigma}_n & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{V}}_n^H \\ \mathbf{V}_n^H \end{bmatrix}, \forall n \geq 2. \quad (41)$$

In (41), $\tilde{\mathbf{V}}_n$ is an $L \times n-1$ unitary matrix containing the right-singular vectors corresponding to the non-zero singular values and is a row-space basis of $\mathbf{G}_n$. On the other hand, $\mathbf{V}_n$ is an $L \times L - n + 1$ unitary matrix containing the right-singular vectors corresponding to the zero singular values and is a null-space basis of $\mathbf{G}_n$, i.e., $\mathbf{h}_m^H \mathbf{V}_n = \mathbf{0}$ for $m < n$. The first user's $\mathbf{V}_1$ is set to the $L \times L$ identity matrix. To reduce complexity, $\{\mathbf{V}_n\}$ may also be obtained by using a single QRD instead of multiple SVD operations.

The optimal precoding matrix may now be expressed with linear combinations of the null-space bases

$$\mathbf{P} = [\mathbf{V}_1 \hat{\mathbf{p}}_1, \ldots, \mathbf{V}_N \hat{\mathbf{p}}_N], \quad (42)$$

with $\hat{\mathbf{p}}_n$ a vector of length $L - n + 1$. As a result, the ZF condition in (39) weakens to a normalization constraint [corresponding to the ones on the main diagonal in (10)] and the Lagrange dual function is formulated as

$$\tilde{g}(\mathbf{\Omega}) = \underset{\{\hat{\mathbf{p}}_n\}, \{\mathbf{s} \in \mathcal{D}\}}{\text{maximize}} \sum_n \left[w_n \log_2 \left(1 + \frac{s_n}{\Gamma \sigma} \right) - s_n \text{Tr}\{\mathbf{\Omega} \mathbf{V}_n \hat{\mathbf{p}}_n \hat{\mathbf{p}}_n^H \mathbf{V}_n^H\} \right]$$
$$\text{s.t.} \quad \mathbf{h}_n^H \mathbf{V}_n \hat{\mathbf{p}}_n = 1, \quad \forall n. \quad (43)$$

Then defining a new variable $\acute{\mathbf{p}}_n = \left(\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n \right)^{1/2} \hat{\mathbf{p}}_n$ for all users n, (43) may be re-written as

$$\tilde{g}(\mathbf{\Omega}) = \underset{\{\acute{\mathbf{p}}_n\}, \{\mathbf{s} \in \mathcal{D}\}}{\text{maximize}} \sum_n \left[w_n \log_2 \left(1 + \frac{s_n}{\Gamma \sigma} \right) - s_n \text{Tr}\{\acute{\mathbf{p}}_n \acute{\mathbf{p}}_n^H\} \right]$$
$$\text{s.t.} \quad \mathbf{h}_n^H \mathbf{V}_n \left(\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n \right)^{-1/2} \acute{\mathbf{p}}_n = 1, \quad \forall n. \quad (44)$$

Clearly, Lagrange dual function (44) has a trivial solution. Having the minimal Frobenius norm among all vector inverses, the optimal $\acute{\mathbf{p}}_n$ for user n is the pseudo-inverse of $\mathbf{h}_n^H \mathbf{V}_n \left(\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n \right)^{-1/2}$. Then, the transmit PSDs are obtained with the water-filling solution. This results in the following closed-form solution for (39)

$$\mathbf{p}_n^{\text{SVD}} = \mathbf{V}_n \left(\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n \right)^{-1} \mathbf{V}_n^H \mathbf{h}_n / g_n, \quad \forall n \quad (45)$$

$$s_n^{\text{SVD}} = \left[w_n (1/g_n)^{-1} - \Gamma \sigma \right]_0^{s^{\max}}, \quad \forall n \quad (46)$$

with g_n enforcing the normalization constraint

$$g_n = \mathbf{h}_n^H \mathbf{V}_n \left(\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n \right)^{-1} \mathbf{V}_n^H \mathbf{h}_n. \quad (47)$$

[5] The terms $\sum_l \lambda_k^l$ and $\sum_l \theta^l P^{\text{ATP}}$ are independent of $\{\mathbf{P}_k, \mathbf{s}_k\}$ and thus need not to be taken into account in the per-tone maximizations (39).

Algorithm 3: SVD-BD-Based Method [17] for Solving Line 7 of Alg. 2

Initialize $\mathbf{V}_k^n$ with (41), $\forall k, n$
for $n = 1 \cdots N$ **do**
  Compute $\mathbf{p}_k^n$ with (45)
  Compute g_k^n with (47)
  $s_k^n = \left[w_n \left(1/g_k^n \right)^{-1} - \Gamma\sigma_k \right]_0^{s_k^{\max}}$
end

An algorithmic description for the SVD-BD-based implementation of line 7 of Alg. 2 is listed in Alg. 3 (including also tone indices for clarity). A practical implementation detail is that the Lagrange multipliers $[\mathbf{\Omega}_k]_{l,l}$ should be lower bounded by a small positive number. Otherwise, when $[\mathbf{\Omega}_k]_{l,l} = 0$ for some l, some of the matrices $\left\{\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n\right\}$ for tone k may become rank-deficient and then the algorithm breaks down[6].

We remark that the dimensions of the null-space bases $\{\mathbf{V}_n\}$ grow large when $L \gg N$ leading to matrix multiplications and inverses in (45) and (47) with large matrices for every user $n > 1$. Therefore a novel low-complexity method for solving (39) is developed in the next sub-section.

B. QRD-Based Method

By transforming the Lagrange dual function (39) into an equivalent version with a virtual SPC, the SPC-optimality of the QRD-based precoder can be exploited, as will be demonstrated. In particular, the transformation adopts a virtual precoding matrix based on the Lagrange multipliers rescaling the channel matrix, similar to [4], [29]. The equivalent transmission model on tone k is then given by (tone indices dropped)

$$\mathbf{y} = \underbrace{\mathbf{H}\mathbf{\Omega}^{-1/2}}_{\tilde{\mathbf{H}}} \tilde{\mathbf{P}}\mathbf{S}^{1/2}\acute{\mathbf{u}} + \mathbf{z}, \tag{48}$$

where $\tilde{\mathbf{P}} \triangleq \mathbf{\Omega}^{1/2}\mathbf{P}$. Now, the transformed Lagrange dual function (39) may be written as

$$\begin{aligned} \tilde{g}(\mathbf{\Omega}) = \quad & \underset{\tilde{\mathbf{P}}, \{\mathbf{s} \in \mathcal{D}\}}{\text{maximize}} \quad \sum_n w_n \log_2 \left(1 + \frac{s_n}{\Gamma\sigma} \right) - \text{Tr}\{\tilde{\mathbf{P}}\mathbf{S}\tilde{\mathbf{P}}^H\} \\ & \text{s.t.} \qquad \tilde{\mathbf{H}}\tilde{\mathbf{P}} = \mathbf{L}. \end{aligned} \tag{49}$$

This transformation hides the Lagrange multipliers in $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{P}}$ such that the transformed Lagrange dual function (49) appears to correspond to ZF precoding under a virtual SPC with its virtual Langrange multiplier equal to one. As a result, for a fixed $\mathbf{\Omega}$, the optimal precoder $\tilde{\mathbf{P}}$ is according to Theorem 1 again $\tilde{\mathbf{P}} = \tilde{\mathbf{Q}}\text{diag}(\tilde{\mathbf{R}}^H)^{-1}$ using the QRD of the scaled channel, i.e., $\tilde{\mathbf{H}}^H \overset{\text{qr}}{=} \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$. Hence, the optimal $\mathbf{P}$ precoder is

$$\mathbf{P}^{\text{QRD}} = \mathbf{\Omega}^{-1/2}\tilde{\mathbf{Q}}\text{diag}(\tilde{\mathbf{R}}^H)^{-1}. \tag{50}$$

[6]Note that optimal ZF-NLP under PLPCs with inactive bit cap constraints on tone k always results in final Lagrange multipliers that are strictly positive for that tone (i.e. it results in $\{[\mathbf{\Omega}_k]_{l,l} > 0, \forall l\}$) [16, Appendix C].

Furthermore, the Lagrange dual function (49) reduces to a per-user transmit PSD optimization with the following water-filling solution

$$s_n^{\text{QRD}} = \left[w_n \left(1/|\tilde{r}_{n,n}|^2 \right)^{-1} - \Gamma\sigma \right]_0^{s^{\max}} \quad \forall n. \tag{51}$$

The complete algorithmic description of the QRD-based implementation of line 7 of Alg. 2 is listed in Alg. 4. Also here the $[\mathbf{\Omega}]_{l,l} = 0$ should be lower bounded by a small positive number to prevent rank-deficiency in $\tilde{\mathbf{H}}$ and a break down of the algorithm every subgradient iteration.

Clearly, by re-working its closed-form formula (50)

$$\mathbf{P}^{\text{QRD}} = \mathbf{\Omega}^{-1}\mathbf{H}^H(\tilde{\mathbf{R}})^{-1}\text{diag}(\tilde{\mathbf{R}}^H)^{-1}, \tag{52}$$

we observe that the optimal nonlinear ZF precoder matrix is generally outside the row-space of $\mathbf{H}$. Another observation is that the optimal nonlinear ZF precoder generally consists of non-orthogonal columns. Only when $\mathbf{\Omega} = \omega\mathbf{I}_L$ (corresponding to a SPC) the precoder $\mathbf{P}$ is in the row-space of $\mathbf{H}$ and consists of orthogonal columns, such that it reduces to the standard QRD-based precoder of Section IV-A. The fact that these additional degrees of freedom are not being used by the standard QRD-based precoder results in suboptimality for the PLPC case.

Since the QRD-based precoder strucure $\mathbf{P}^{\text{QRD}}(\mathbf{S}^{\text{QRD}})^{1/2}$ can be shown to be the unique solution to Lagrange dual function (49) [and thus also to (39)] for any $\mathbf{\Omega} \succ \mathbf{0}$ using Theorem 1, it should be equivalent to the SVD-BD-based solution structure. The next theorem states that this is indeed the case.

Theorem 2: The solutions to the Lagrange dual function (39) for any given $\mathbf{\Omega} \succ \mathbf{0}$ obtained by the SVD-BD (by Alg. 3) and the QRD-based method (by Alg. 4) are identical at each tone.

Proof: For the first encoded user $n = 1$, the proof is trivial since the null-space basis $\mathbf{V}_1 = \mathbf{I}_L$, which leads to

$$\mathbf{p}_1^{\text{SVD}} = \mathbf{\Omega}^{-1}\mathbf{h}_1 \left(\mathbf{h}_1^H \mathbf{\Omega}^{-1} \mathbf{h}_1 \right)^{-1} = \mathbf{\Omega}^{-1/2}\tilde{\mathbf{q}}_1/\tilde{r}_{1,1}^* = \mathbf{p}_1^{\text{QRD}}. \tag{53}$$

For users $n > 1$, we first incorporate a virtual precoding matrix based on the Lagrange multipliers into the null-space bases $\{\mathbf{V}_n\}$ that rescales the channel vectors, similar as in the QRD-based method. We define $\tilde{\mathbf{h}}_n^H = \mathbf{h}_n^H \mathbf{\Omega}^{-1/2}$, $\tilde{\mathbf{V}}_n = \mathbf{\Omega}^{1/2}\mathbf{V}_n$, and $\tilde{\mathbf{p}}_n^{\text{SVD}} = g_n \mathbf{\Omega}^{1/2}\mathbf{p}_n^{\text{SVD}}$ for user n. Hence, $\tilde{\mathbf{p}}_n^{\text{SVD}}$ can be formulated as

$$\tilde{\mathbf{p}}_n^{\text{SVD}} = \underbrace{\tilde{\mathbf{V}}_n \left(\tilde{\mathbf{V}}_n^H \tilde{\mathbf{V}}_n \right)^{-1} \tilde{\mathbf{V}}_n^H}_{\text{orth. projector}} \tilde{\mathbf{h}}_n. \tag{54}$$

Important here is that (54) may be interpreted as the orthogonal projection of $\tilde{\mathbf{h}}_n$ onto the null-space of $\tilde{\mathbf{G}}_n \triangleq [\tilde{\mathbf{h}}_1^H, \ldots, \tilde{\mathbf{h}}_{n-1}^H]^T$, due to the fact that $\tilde{\mathbf{V}}_n$ is a non-orthogonal null-space basis of $\tilde{\mathbf{G}}_n$.

At the same time, when considering the QRD of $\tilde{\mathbf{H}}^H$ as a Gram-Schmidt procedure, $\tilde{\mathbf{q}}_n$ can be seen as the orthogonal projection of $\tilde{\mathbf{h}}_n$ onto the complement of $[\tilde{\mathbf{q}}_1, \ldots, \tilde{\mathbf{q}}_{n-1}]$ with $\tilde{r}_{n,n}$ providing the corresponding projection length. Since $[\tilde{\mathbf{q}}_1, \ldots, \tilde{\mathbf{q}}_{n-1}]$ is an orthonormal basis for $[\tilde{\mathbf{h}}_1, \ldots, \tilde{\mathbf{h}}_{n-1}]$, its orthogonal complement is in fact the null-space of $\tilde{\mathbf{G}}_n$, i.e.,

Algorithm 4: QRD-Based Method for Solving Line 7 of Alg. 2

$\tilde{\mathbf{H}}_k = \mathbf{H}_k \mathbf{\Omega}_k^{-1/2}$
$\tilde{\mathbf{H}}_k^H \overset{\text{qr}}{=} \tilde{\mathbf{Q}}_k \tilde{\mathbf{R}}_k$
$\tilde{\mathbf{P}}_k = \tilde{\mathbf{Q}}_k \text{diag}(\tilde{\mathbf{R}}_k^H)^{-1}$
$\mathbf{P}_k = \mathbf{\Omega}_k^{-1/2} \tilde{\mathbf{P}}_k$
for $n = 1 \cdots N$ **do**
  $s_k^n = \left[w_n \left(1/\left|\tilde{r}_k^{n,n}\right|^2 \right)^{-1} - \Gamma \sigma_k \right]_0^{s_k^{\max}}$
end

$\tilde{\mathbf{h}}_m^H \tilde{\mathbf{q}}_n = 0, \forall m < n$. Then by uniqueness of orthogonal projection onto the same sub-space, we conclude that $\tilde{\mathbf{p}}_n^{\text{SVD}} = \tilde{\mathbf{q}}_n \tilde{r}_{n,n}$. Consequently, it can be verified that

$$g_n = \tilde{\mathbf{h}}_n^H \tilde{\mathbf{p}}_n^{\text{SVD}} = \tilde{\mathbf{h}}_n^H \tilde{\mathbf{q}}_n \tilde{r}_{n,n} = |\tilde{r}_{n,n}|^2.$$

This shows that $\mathbf{p}_n^{\text{SVD}} = \mathbf{p}_n{}^{\text{QRD}}$ and hence $s_n^{\text{SVD}} = s_n{}^{\text{QRD}}$ for all users n. ∎

VI. Computational complexity

In this section, we analyze the computational complexity reduction of the QRD-based dual decomposition algorithm over the state-of-the-art SVD-BD [17] for optimal ZF-NLP under PLPCs. As both dual decomposition algorithms require the same number of subgradient iterations, we first focus on the complexity per iteration of Alg. 3 and Alg. 4.

Computational complexity is often approximated by the number of floating point operations (flops). A real addition, and multiplication or division are counted as one flop [30]. A complex addition and multiplication add up to two and six flops, respectively. A real square root is approximated as one flop. We use the following flop counts for specific matrix operations, with $\mathbf{X}$ denoting an $m \times n$ complex valued matrix ($m \geq n$):

- Multiplication of $\mathbf{X}^H\mathbf{X}$ requires $nm(n+1)/2$ complex multiplications and $nm(n+1)/2 - n(n+1)/2$ complex additions; hence, the flop count is approximated by $4nm(n+1)$.
- Solving a real linear system $\mathbf{Ax} = \mathbf{b}$ using Gaussian elimination with $\mathbf{A}$ an $n \times n$ matrix and $\mathbf{x}$ and $\mathbf{b}$ an n-vector takes $2n^3/3$ flops [30]. Treating every operation as a complex multiplication, solving a complex linear system takes $6 \times 2n^3/3$ flops.
- Computing the QRD of $\mathbf{X}$ using a complex Householder factorization algorithm takes $8n^2(m - n/3)$ flops [30].

Since in each iteration of the Lagrange multiplier $\mathbf{\Omega}$ search, the precoder matrix $\mathbf{P}$ computation dominates the computational complexity in both algorithms, we compare the flop count needed for this operation.

1) SVD-BD-Based Method (Alg. 3) [17]: For the n-th user ($n \geq 2$), following computations are significant[7]:

- The square root of the real diagonal matrix $\mathbf{\Omega}$ and multiplication with $\mathbf{V}_n$ takes $L + 2L(L - n + 1)$ flops.
- Computing $\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n$ involves a matrix multiplication taking $4L(L - n + 1)(L - n + 2)$ flops.
- $\dot{\mathbf{p}}_n = (\mathbf{V}_n^H \mathbf{\Omega} \mathbf{V}_n)^{-1} \mathbf{V}_n^H \mathbf{h}_n^H$ is computed by solving a linear system of equations taking $6 \times 2(L - n + 1)^3/3$ flops.
- Computing $\mathbf{V}_n \dot{\mathbf{p}}_n$ takes an additional $8L(L - n + 1)$ flops.

Thus the total flop count per iteration for the SVD-BD is

$$\psi_{\text{SVD}} \approx \sum_{n=2}^{N} \Big\{ 4L(L-n+1)^2 + 6 \times 2(L-n+1)^3/3 + 10L(L-n+1) \Big\} + L.$$

2) QRD-Based Method (Alg. 4): The flop count per iteration of the QRD-based dual decomposition algorithm corresponds to computing to a single QRD of $\tilde{\mathbf{H}}^H$, the inverse of the real diagonal matrix $\mathbf{\Omega}$ and $\text{diag}(\tilde{\mathbf{R}}^H)$, the square root of $\mathbf{\Omega}^{-1}$, and three complex matrix by real diagonal matrix multiplications, i.e.,

$$\psi_{\text{QRD}} \approx 8N^2(L - N/3) + 6LN + 2L + N.$$

Leveraging on the SPC-optimal QRD solution thus clearly results in a significant computational complexity reduction per iteration, which is illustrated in Fig. 3a for increasing L with $N = L$. If $N \ll L$, the complexity reduction increases dramatically (see Fig. 3b), since for this case the dimensions of the null-space bases $\mathbf{V}_n$ get increasingly larger.

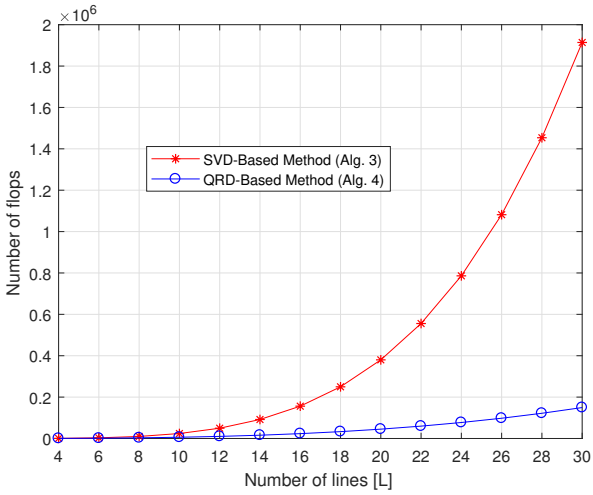
Next, we present simulation results on the convergence behavior of various step size sequences by the subgradient method in Alg. 2. Since we observed in our G.fast simulations (see Section VII for details) that the ATP constraints are always inactive due to the per-tone spectral mask and bit cap constraints, $\boldsymbol{\theta}$ is set to zero. Fig. 4 then illustrates the search for $\boldsymbol{\lambda}_k$ at tone $k = 2000$, with $L = 10$ and $N = 10$, and using $\{\lambda_k^l = 10^8\}$ as the Lagrange multiplier starting point. As can be seen in the figure, adequate tuning of β leads to acceptable solution accuracies after a reasonable number of iterations (20-100). We remark that for higher frequency tones (>175 MHz) with high direct channel attenuation or few active users ($N \ll L$) sometimes slightly smaller initial step sizes are needed for proper convergence.

Finally, the computational complexity of the QRD-based dual decomposition algorithm is also compared with standard QRD-based ZF-NLP (Alg. 1 in Section IV-A) and MMSE-NLP (by means of the BC-DSB-NLP algorithm [4]). The comparison is summarized in Table I and also accounts for the number of tones K. Here I_1 denotes the number of subgradient iterations needed for convergence. Standard QRD-based ZF-NLP only needs to compute the QRDs of the complex conjugated channel matrices once at initialization, and can be re-used for all subgradient iterations. The BC-DSB-NLP algorithm with Jacobi updates has a complexity of $O(I_2 KNL^2)$ per subgradient iteration, where I_2 is the number of inner iterations and is relatively small compared to I_1 [4].

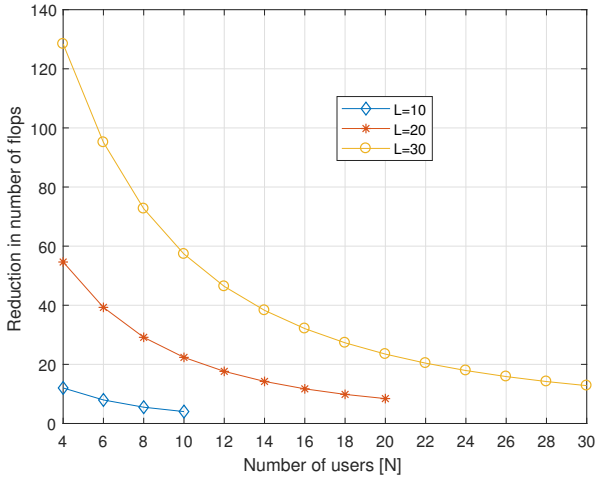
VII. G.fast Cable Binder Simulation

In this section, we simulate a cable binder consisting out of 10 lines of 80 m for the downstream G.fast 212 MHz profile.

[7] Also, the null-space bases $\{\mathbf{V}_n\}$ must be calculated. However this only needs to be done once at initialization, and can be re-used for all subgradient iterations. Hence this has a minimal impact on complexity.



(a) The number of flops versus the number of lines L with $N = L$.



(b) The reduction in number of flops (i.e. $\frac{\psi_{\text{SVD}}}{\psi_{\text{QRD}}}$) versus the number of users N with $L = \{10, 20, 30\}$.

Fig. 3. The QRD-based method (Alg. 4) has a significantly lower computational complexity per iteration than the SVD-BD-based method [17] (Alg. 3), illustrated here for $N = L$ (a) and $N < L$ (b).

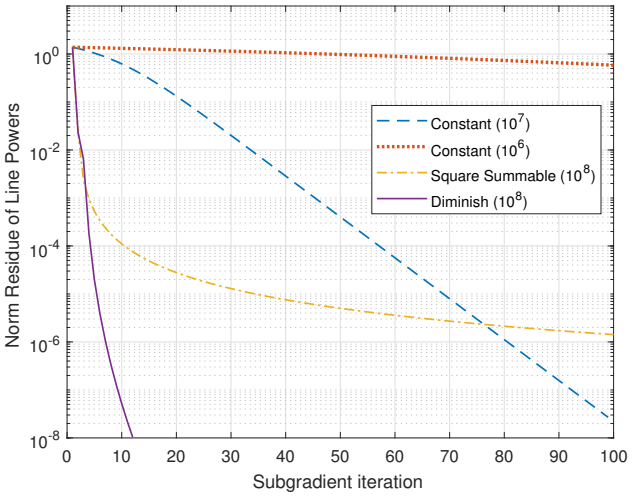


Fig. 4. Comparison between various step size sequences for the subgradient method by Alg. 2 for a ($L = 10$) G.fast cable binder with $N = L$ at tone $k = 2000$. Constant(β) refers to the update rule $\delta^t = \beta$, Square Summable(β) refers to the update rule $\delta^t = \frac{\beta}{t}$. Diminish($\beta$) refers to the update rule $\delta^t = \frac{\beta}{\sqrt{t}}$. Here $\sqrt{\sum_l |P_k^{\text{mask}} - \sum_m [\mathbf{P}_k]_{l,m} s_k^m)|^2 / P_k^{\text{mask}}}$ versus the iteration number is shown.

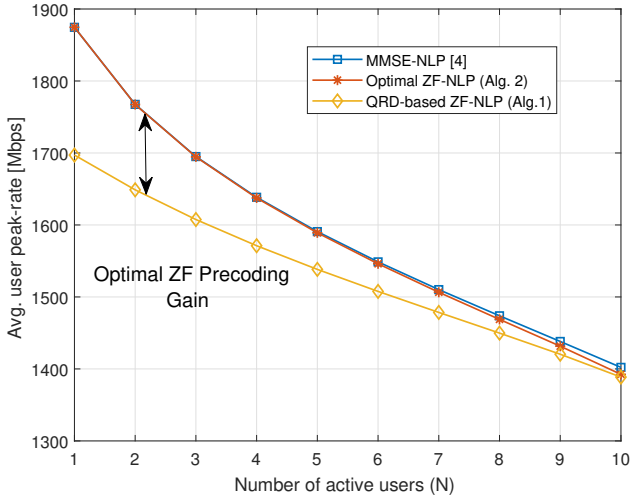


Fig. 5. The average user peak-rate is plotted versus the number of active users N in a $L = 10$ G.fast cable binder.

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON

Algorithm	Computational complexity
QRD-based ZF-NLP (Alg. 1)	$O(KN^2L + I_1KN)$
QRD-based dual decomposition (Alg. 4)	$O(I_1KN^2L)$
BC-DSB-NLP with Jacobi update [4]	$O(I_1I_2KNL^2)$

The channel matrices have been obtained by measurements. We remark that the observed crosstalk levels in this particular cable binder are rather high compared to other G.fast measurements. Considering the latest recommended G.fast parameters [1], the ATP constraints are set to 8 dBm while the per-tone PSD spectral masks are obtained from [21] ranging from −65 dBm/Hz to −79 dBm/Hz. The capacity gap Γ is set to 10.25 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. We consider unweighted sum-rate optimization with $\{w_n = 1 | \forall n\}$.

For this cable binder, we evaluate various NLP schemes when varying the number of active users N and considering all $\binom{L}{N}$ combinations (see Fig. 5). Included schemes are optimal ZF-NLP (corresponding to the QRD and SVD-BD-based implementation of Alg. 2 in Section V), standard QRD-based ZF-NLP (Alg. 1 in Section IV-A), and MMSE-NLP (by means of the BC-DSB-NLP algorithm [4] where we set the

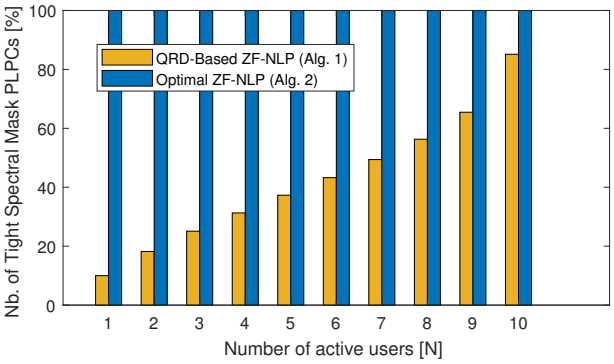


Fig. 6. Optimal ZF-NLP is able to transmit at full spectral mask power on all lines and tones, in contrast to QRD-based ZF-NLP.

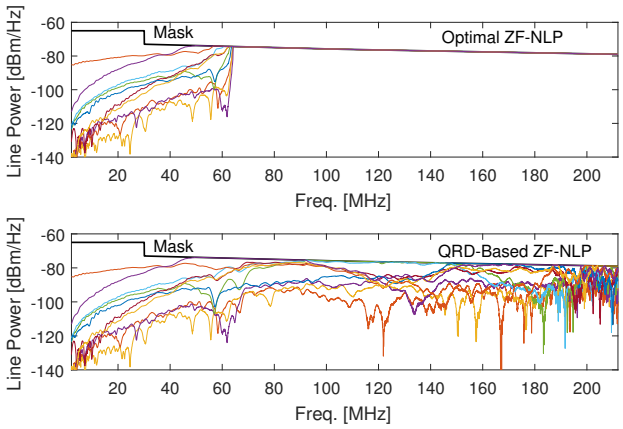


Fig. 7. The per-line transmit powers in the cable binder with $N = 1$ active user for optimal ZF-NLP (top) and QRD-based ZF-NLP (bottom).

weights of inactive users to zero). All three precoding schemes always use all available lines for transmission regardless the number of active users.

Clearly, the results show that the peak-rate performance may be substantially boosted over standard QRD-based ZF-NLP. In particular when only a few users in the binder are active, with up to 175 Mbps improvement for one active user. Remarkably, when all users are active (i.e. $N = L$) the gain of optimal ZF-NLP over QRD-based ZF-NLP is negligible. Furthermore, MMSE-NLP yields only very small gains over optimal ZF-NLP.

These optimal ZF precoding gains are explained by the additional degrees of freedom in the ZF-NLP design problem under PLPCs that are left unexploited by the standard QRD-based precoder. Naturally, the largest precoding gains occur for the case with $N \ll L$ leading to large channel matrix null-spaces to be used. Additionally, the columns of the optimal nonlinear ZF precoder are non-orthogonal, unlike the columns of the standard QRD-based precoder. This is the very reason why even for a square channel (i.e. $N = L$), when all channel matrix null-spaces are empty, there remains a ZF precoding gain, although very small.

Point of fact, these additional degrees of freedom enable optimal ZF-NLP to utilize the full transmit power on all lines, in contrast to QRD-based ZF-NLP. We illustrate this in Fig. 6 where the number of active spectral mask PLPCs[8] versus the number of active users N is shown. Interestingly, when only one user is active, QRD-based ZF-NLP transmits only on one out of ten lines at full spectral mask power. This is explained by the QRD-based precoder being now a single column with its entries having in general different magnitudes, see the 'simple example' in Section IV-B, and is illustrated in Fig. 7.

VIII. Generalization to Bonded Transmission

Until now we have considered only a single line terminating at the customer premise equipment of each user. Here we consider $A_n \geq 1$ lines terminating at the customer premise equipment of user n, which is referred to as bonded transmission in DSL [31] and corresponds to a so-called MIMO-BC scenario. The received signal for user n on tone k is now a vector

$$\hat{\mathbf{u}}_k^n = \mathbf{G}_k^n \mathbf{H}_k^n \left(\mathbf{P}_k^n \left(\mathbf{S}_k^n \right)^{1/2} \mathbf{u}_k^n + \sum_{m>n} \mathbf{P}_k^m \left(\mathbf{S}_k^m \right)^{1/2} \mathbf{u}_k^m \right) + \mathbf{G}_k^n \mathbf{z}_k^n, \tag{55}$$

where the $A_n \times L$ matrix $\mathbf{H}_k^n$ represents the channel gains from L transmitters to A_n receivers for user n at tone k. $\mathbf{P}_k^n$ is the $L \times A_n$ precoder matrix, $\mathbf{G}_k^n$ is the $A_n \times A_n$ receiver matrix, $\mathbf{S}_k^n$ is the $A_n \times A_n$ diagonal gain scaling matrix which contains the transmit PSDs, and $\mathbf{u}_k^n$ and $\mathbf{z}_k^n$ are the A_n data and noise signal vector respectively for user n and tone k.

Similar to (10), the ZF condition for the MIMO-BC scenario with NLP on tone k is defined as

$$\mathbf{G}_k \mathbf{H}_k \mathbf{P}_k = \underbrace{\begin{bmatrix} \mathbf{I}_{A_1} & 0 & \cdots & 0 \\ * & \mathbf{I}_{A_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \cdots & \mathbf{I}_{A_N} \end{bmatrix}}_{\overline{\mathbf{L}}_k} \tag{56}$$

where $\mathbf{G}_k = \text{blockdiag}\{\mathbf{G}_k^1, \cdots, \mathbf{G}_k^N\}$ and $\overline{\mathbf{L}}_k$ is an arbitrary $\sum_n A_n \times \sum_n A_n$ lower triangular block matrix with identity matrices on the main diagonal.

These identity constraints are without loss of optimality since we consider linear separation of the per-line data streams and independent decoding at each terminated line. Moreover, two-sided intra-user signal coordination boasts a degree of freedom that allows to rotate the precoder and receiver matrices of all users n on tone k such that the intra-user data streams are decorrelated.

As a result, the Lagrange dual function of the ZF-NLP design problem for given Lagrange multipliers $\{\mathbf{\Omega}_k\}$ can be formulated similar to (39) for the MIMO-BC scenario as (and

[8] We observed in our simulations for this particular G.fast cable binder that the 8 dBm ATP constraints are always inactive due to the per-tone spectral mask and bit cap constraints. Furthermore, in Fig. 6, we disregarded all low frequencies where the spectral mask PLPCs are inactive due to the maximum bit cap constraint.

dropping the tone index again)

$$\tilde{g}(\mathbf{\Omega}) = \underset{\mathbf{G},\mathbf{P},\mathbf{s}\in\mathcal{D}}{\text{maximize}} \quad \sum_{n,i} \log_2\left(1+\frac{s_n^i}{\Gamma\sigma}\right) - \text{Tr}\{\mathbf{\Omega P S P}\}$$
$$\text{s.t.} \qquad \mathbf{GHP} = \overline{\mathbf{L}}, \tag{57}$$

with s_n^i the ith diagonal element of $\mathbf{S}_n$.

To solve (57), we now exploit the SPC-optimality of the MIMO-BC version of the QRD-based precoder proposed in [32]. First, the QRD of the scaled channel $\tilde{\mathbf{H}} = \mathbf{H\Omega}^{-1/2}$ is computed, i.e., $\tilde{\mathbf{H}}^H \overset{\text{qr}}{=} \tilde{\mathbf{Q}}\tilde{\mathbf{R}}$. Then, the optimal precoder matrix $\tilde{\mathbf{P}} = \left[\tilde{\mathbf{P}}_1,\ldots,\tilde{\mathbf{P}}_N\right]$ has the following structure

$$\tilde{\mathbf{P}} = \tilde{\mathbf{Q}}\mathbf{W} = \left[\tilde{\mathbf{Q}}_1,\ldots,\tilde{\mathbf{Q}}_N\right]\mathbf{W}, \tag{58}$$

where $\mathbf{W} = \text{blockdiag}\{\mathbf{W}_1,\cdots,\mathbf{W}_N\}$, and $\mathbf{W}_n$ is computed using the following SVD

$$\tilde{\mathbf{H}}_n\tilde{\mathbf{Q}}_n = \left[\tilde{\mathbf{R}}^H\right]_n \overset{\text{SVD}}{=} \tilde{\mathbf{U}}_n\tilde{\mathbf{\Sigma}}_n\tilde{\mathbf{V}}_n^H, \tag{59}$$

with $\left[\tilde{\mathbf{R}}^H\right]_n$ denoting the $A_n \times A_n$ nth submatrix on the main diagonal of $\tilde{\mathbf{R}}^H$. This leads to choosing $\mathbf{W}_n = \tilde{\mathbf{V}}_n\tilde{\mathbf{\Sigma}}_n^{-1}$ and $\mathbf{G}_n = \tilde{\mathbf{U}}_n^H$ which satisfies (56). The transmit PSDs are then obtained by the water-filling solution over the singular values $\{\tilde{\sigma}_n^i\}$ with $\tilde{\sigma}_n^i$ the ith diagonal element of $\tilde{\mathbf{\Sigma}}_n$

$$s_n^i = \left[w_n\left(1/\left[\tilde{\sigma}_n^i\right]^2\right)^{-1} - \Gamma\sigma\right]_0^{s^{\max}} \quad \forall n,i. \tag{60}$$

The proposed approach for ZF precoding under PLPCs in the MIMO-BC scenario leads to a significant complexity reduction over the state-of-the-art SVD-BD [17], as already is demonstrated for $A_n = 1$ for all users n in Section VI. Furthermore, equivalence between both approaches may be shown by extending the proof in Section V-B which is omitted here for brevity.

We simulated the same G.fast cable binder of Section VII for bonded transmission by re-using it as an $N = 5$ user scenario with $A_n = 2$ lines per user. Fig. 8 includes the average user peak-rates versus the number of active users of the optimal ZF-NLP, the QRD-based ZF-NLP, and MMSE-NLP [4] algorithm for the MIMO-BC and BC (the latter which only considers tranmitter signal coordination). Likewise for bonded transmission, a substantial optimal ZF precoding gain is observed. However, as can be expected from [4], the performance gain of bonded transmission (in the sense of the additional receiver signal coordination) is quite small for NLP, and decreases for a decreasing number of active users. Notice that larger gains for the MIMO-BC in G.fast are expected when linear precoding is employed [4].

IX. Conclusion

We have developed a novel optimal low-complexity dual decomposition algorithm for the ZF-NLP design problem under PLPCs to maximize the active-user peak-rates. Key in this algorithm is the transformation of the Lagrange dual function with PLPCs into an easier virtual SPC, such that the SPC-optimality of the QRD-based precoder may be exploited. We have demonstrated that a significant computational complexity reduction is achieved over the state-of-the-art SVD-BD [17].

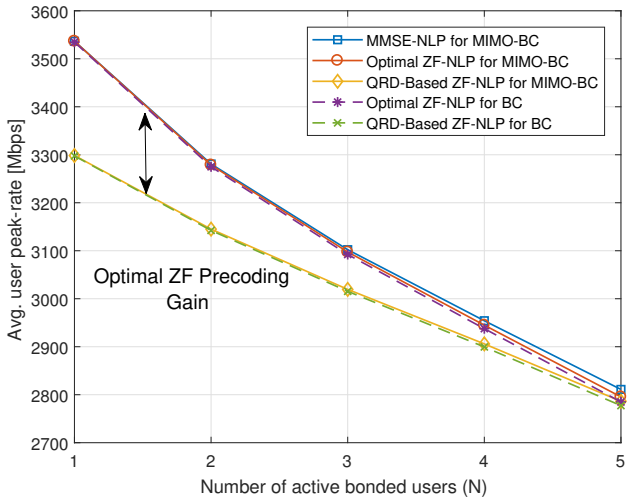


Fig. 8. The average user peak-rate is plotted vs. the number of active users N with bonded transmission $\{A_n = 2, \forall n\}$ in a $L = 10$ G.fast cable binder.

Simulations of a G.fast cable binder have revealed substantial peak-rate gains when transmitting on all lines over QRD-based ZF-NLP. Furthermore, we have shown that the proposed algorithm naturally extends to the MIMO-BC scenario.

References

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

[2] R. Cendrillon, G. Ginis, E. Van den Bogaert, and M. Moonen, "A Near-Optimal Linear Crosstalk Precoder for Downstream VDSL," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 860–863, May 2007.

[3] M. Guenach, C. Nuzman, P. Tsiaflakis, and J. Maes, "Power optimization in vectored and non-vectored G.fast transmission," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2014, pp. 2229–2233.

[4] 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.

[5] G. Ginis and J. M. Cioffi, "Vectored transmission for digital subscriber line systems," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1085–1104, June 2002.

[6] Y. Huang, T. Magesacher, E. Medeiros, C. Lu, P.-E. Eriksson, and P. Ödling, "Rate-Boosting Using Strong Crosstalk in Next Generation Wireline Systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2015, pp. 1–6.

[7] J. Maes and C. Nuzman, "Energy efficient discontinuous operation in vectored G.fast," in *Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 3854–3858.

[8] R. Cendrillon, "Multi-user signal and spectra coordination for digital subscriber lines," Ph.D. dissertation, KU Leuven, Dec. 2014.

[9] A. Forouzan, M. Moonen, J. Maes, and M. Guenach, "Joint Level 2 and 3 Dynamic Spectrum Management for Downstream DSL," *IEEE Trans. Commun.*, vol. 60, no. 10, pp. 3111–3122, Oct. 2012.

[10] F. C. B. F. Müller *et al.*, "Optimizing power normalization for G.fast linear precoder by linear programming," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 4160–4165.

[11] R. Strobel, A. Barthelme, and W. Utschick, "Zero-Forcing and MMSE Precoding for G.fast," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[12] 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.

[13] M. Hekrdla, A. Matera, W. Wang, D. Wei, and U. Spagnolini, "Ordered Tomlinson-Harashima Precoding in G.fast Downstream," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[14] 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.

[15] W. Lanneer, P. Tsiaflakis, J. Maes, and M. Moonen, "Joint alpha-fairness based DSM and user encoding ordering for zero-forcing nonlinear precoding in G.fast downstream transmission," in *Proc. IEEE Inter. Conf. Acoustics, Speech and Signal Process. (ICASSP)*, Mar. 2017, pp. 3654–3658.

[16] L.-N. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Beamformer Designs for MISO Broadcast Channels with Zero-Forcing Dirty Paper Coding," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, pp. 1173–1185, March 2013.

[17] R. Zhang, "Cooperative multi-cell block diagonalization with per-base-station power constraints," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1435–1445, Dec. 2010.

[18] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 461–471, Feb. 2004.

[19] L.-U. Choi and R. D. Murch, "A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach," *IEEE Trans. Wireless Commun.*, vol. 3, no. 1, pp. 20–24, Jan. 2004.

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

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

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

[23] C. Windpassinger, R. F. H. Fischer, T. Vencel, and J. B. Huber, "Precoding in multiantenna and multiuser communications," *IEEE Trans. Wireless Commun.*, vol. 3, no. 4, pp. 1305–1316, Jul. 2004.

[24] 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.

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

[26] A. Wiesel, Y. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sept 2008.

[27] 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.

[28] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1995. [Online]. Available: https://books.google.be/books?id=QeweAQAAIAAJ

[29] V. Le Nir, M. Moonen, J. Verlinden, and M. Guenach, "Optimal power allocation for downstream xDSL with per-modem total power constraints: Broadcast Channel Optimal Spectrum Balancing (BC-OSB)," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 690–697, Feb. 2009.

[30] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[31] M. Peeters and S. Vanhastel, "The Copper Phantom," *OSP Mag.*, Dec. 2010. [Online]. Available: https://www.isemag.com/2017/01/the-copper-phantom/

[32] L. N. Tran, M. Juntti, M. Bengtsson, and B. Ottersten, "Weighted sum rate maximization for MIMO broadcast channels using dirty paper coding and zero-forcing methods," *IEEE Trans. Commun.*, vol. 61, no. 6, pp. 2362–2373, June 2013.



Wouter Lanneer (S'15) received the M.Sc. degree in electrical engineering from the KU Leuven, Leuven, Belgium, in 2014, where he is currently pursuing the Ph.D. degree under the supervision of Prof. M. Moonen at the Electrical Engineering Department of KU Leuven. He is currently involved in joint projects with KU Leuven and Nokia Bell Labs, Antwerp, Belgium. His research concerns about signal processing and optimization for digital communication systems with an emphasis on DSL wireline access networks.



Paschalis Tsiaflakis (S'06-M'09) received the M.Sc. and Ph.D. degree in electrical engineering from the KU Leuven, Leuven, Belgium, in 2004 and 2009, respectively. From 2010 to 2013, he was a Post-Doctoral Researcher, during which he supervised four Ph.D. students. During his academic period, he conducted research with Princeton University in 2007, UCLA in 2010, Université catholique de Louvain in 2011, and Tsinghua University in 2013. Since 2013, he has been a Research and Innovation Engineer at Nokia Bell Labs, where he drives innovation into next generation communication systems. He has authored over 60 papers in internationally peer-reviewed journals and conferences and has filed 20 patents. He was a recipient of the Nokia Innovation Award in 2017, the Nokia Spotlight Recognition Award in 2016, the Belgian Young ICT Personality Award in 2010, two IEEE ICC best paper awards in 2013, the Best Multimedia Master Thesis Prize Award in 2001. He was a top-12 finalist for the European ERCIM Cor Baayen Award 2010.



Jochen Maes (SM'11) received the M.Sc. degree in physics and the Ph.D. degree in science from KU Leuven, Leuven, Belgium. He joined Bell Labs in 2006, where he continuously shifts the limits of broadband. He heads the Fixed Networks Group, Nokia Bell Labs, involved in innovations in optical and copper access networks. He contributes to several standardization bodies. His work received the Broadband Infovision Awards (2010, 2014) and the Bell Labs President's Award (2011). He has been a Bell Labs Fellow since 2016.



Marc Moonen (M'94-SM'06-F'07) is a Full Professor at the Electrical Engineering Department of KU Leuven, Leuven, Belgium, where he is heading a research team working in the area of numerical algorithms and signal processing for digital communications, DSL and audio signal processing.

He received the 1994 KU Leuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), the 2004 Alcatel Bell (Belgium) Award (with Raphael Cendrillon), and was a 1997 Laureate of the Belgium Royal Academy of Science. He received journal best paper awards from the IEEE Transactions on Signal Processing (with Geert Leus and with Daniele Giacobello) and from Elsevier Signal Processing (with Simon Doclo).

He was the chairman of the IEEE Benelux Signal Processing Chapter from 1998 to 2002, a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications, and President of EURASIP (European Association for Signal Processing), from 2007 to 2008 and from 2011 to 2012.

He has served as an Editor-in-Chief for the EURASIP Journal on Applied Signal Processing from 2003 to 2005, an Area Editor for Feature Articles in IEEE Signal Processing Magazine from 2012 to 2014, and has been a member of the editorial board of Signal Processing, IEEE Transactions on Circuits and Systems II, IEEE Signal Processing Magazine, Integration-the VLSI Journal, EURASIP Journal on Wireless Communications and Networking and EURASIP Journal on Advances in Signal Processing.