Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 13-100

DMT MIMO IC Rate Maximization in DSL with Per-Transceiver Power Constraints [1]

Rodrigo B. Moraes, Paschalis Tsiaflakis, Jochen Maes and Marc Moonen[2]

Submitted to Elsevier Signal Processing
June 2013

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/sista/rmoraes/reports/13-100.pdf*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SISTA, Kastelpark Arenberg 10, 3001 Leuven, Belgium, Tel. 32/16/32 17 09, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *rodrigo.moraes@esat.kuleuven.ac.be*. This research work was carried out at the ESAT Laboratory of the KU Leuven, in the frame of the KU Leuven Research Council CoE EF/05/006 Optimization in Engineering (OPTEC) and PFV/10/002 (OPTEC); Concerted Research Action GOA-MaNet; IUAP P7/23 (Belgian network on stochastic modeling, analysis, design and optimization of communication systems, BESTCOM, 2012-2017); and research project FWO nr. 6.091213N, 'Cross-layer optimization with real-time adaptive dynamic spectrum management for fourth generation broadband access networks'. The scientific responsibility is assumed by its authors. P. Tsiaflakis is also a post-doctoral fellow funded by the Research Foundation—Flanders (FWO).

DMT MIMO IC Rate Maximization in DSL with Per-Transceiver Power Constraints

Rodrigo B. Moraes[a,*], Paschalis Tsiaflakis[a], Jochen Maes[b], Marc Moonen[a]

[a] *Department of Electrical Engineering (ESAT-SCD), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium*
[b] *Access Research Domain, Alcatel-Lucent Bell Labs, Antwerp, Belgium*

Abstract

This paper deals with the discrete multitone multiple input, multiple output interference channel (DMT MIMO IC) in DSL networks. The scenario consists of a number of users, each with a given number of transceivers, that share the same channel in multiple tones. Our goal is to maximize the weighted rate sum of the users subject to power constraints. A recent paper has treated this problem with per-user power constraints. In this paper we focus on per-transceiver power constraints. We propose two different algorithms. First, we straightforwardly adapt the previously proposed DMT-WMMSE algorithm. Second, we adapt the WMMSE-GDSB, in which we separate the problem in signal and spectrum coordination parts. For the spectrum coordination part, we show that the problem can be solved more efficiently with a change of coordinates: we use a coordinate system consisting of a radius and a direction vector with ℓ^1 norm equal to 1. This can be interpreted as spherical coordinates in taxicab geometry. It is observed that for the radial dimension the problem can be made concave after approximations and it is thus easy to solve. The remaining dimensions are solved iteratively and individually. Simulation results show that the WMMSE-GDSB converges faster.

Keywords: DSL, Dynamic spectrum management, Power control, Optimization, Multiple input, multiple output, Interference channel

1. Introduction

Digital Subscriber Line (DSL) technology has been able to maintain its relevance and continues to be the most widely deployed technology for wireline broadband access communications worldwide. One of the reasons for DSL's enduring success is that the industry and standardization bodies have been able to quickly answer to changing market conditions. As the fiber network expands, DSL operates on shorter lines. The

*Tel.: +32 (0)16 321796; Department of Electrical Engineering (ESAT-SCD), KU Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium

Email addresses: rodrigo.moraes@esat.kuleuven.be (Rodrigo B. Moraes), paschalis.tsiaflakis@esat.kuleuven.be (Paschalis Tsiaflakis), jochen.maes@alcatel-lucent.com (Jochen Maes), marc.moonen@esat.kuleuven.be (Marc Moonen)

answer to that has been evolving DSL standards (e.g. ADSL2+, VDSL2 and G.fast). The main source of performance degradation in DSL networks is multi-user interference (i.e. crosstalk). The answer to crosstalk has been a decade-long continued research effort in dynamic spectrum management (DSM), which is the topic of this paper.

Research on DSM has been often divided into two directions. The first deals with power allocation in each discrete multitone (DMT) sub-channel (or tone) for the interfering users, e.g. [1, 2, 3, 4, 5, 6, 7]. The goal is to assign power levels for each user and tone in the network so that the impact of crosstalk is decreased. The second direction deals with signal coordination (i.e. one or two-sided multiple input, multiple output (MIMO) processing), e.g. [8, 9, 10, 11, 12]. Signal coordination can cancel or even profit from crosstalk. It delivers substantial gains in comparison to spectrum coordination.

In a recent paper [13], the DMT MIMO interference channel (DMT MIMO IC) scenario is suggested as a bridge between these two main directions. In such a scenario, a user is characterized by having multiple transceivers, to which it can apply two-sided MIMO processing—this is for instance the case in networks using common or phantom mode transmission [14, 15]. There are several such users sharing the channel, each with a distinct set of transceivers. Coordination among them should be done both on the signal and on the spectrum levels. More specifically, the goal of the problem is to come up with transmit matrices for all users and tones. These transmit matrices should be set so as to make it easy to detect the desired signal and easy to cancel the undesired ones (signal coordination). The cancelation of undesired signals is not perfect, and hence some interference will remain. To decrease this remaining interference, an appropriate amount of power should be used for every user and tone (spectrum coordination). Every user is also subject to a power constraint (PC), and that complicates the problem significantly.

In [13], two distinct algorithms are proposed for the DMT MIMO IC. Both of them provide noteworthy gains in comparison to, for example, a situation where only spectrum coordination is available. However, these two algorithms consider *per-user* PCs, i.e. the sum of the transmit power of all transmitters is limited. This might not be very realistic in practice. To understand why, consider a situation where one user has three transceivers at its disposal—for example, corresponding to two direct modes and one common mode. Each transceiver is associated with a line driver. Each line driver has a power budget of, say, 100 mW. For the per-user PC case, the constraint is satisfied if the sum of the power consumption of the three line drivers is 300 mW. However, it can be that the individual line drivers use, say, 80, 100 and 120 mW respectively, which would constitute a violation of the budget of the third line driver. In this paper, we consider *per-transceiver* (i.e. per-line driver) PCs. A similar problem occurs in wireless communications. For this situation, we would refer to *per-antenna* PCs.

The optimization problem with per-transceiver PCs is more difficult than the one with per-user PCs as in [13]. There are more variables and more constraints to be satisfied. In this paper, we adapt the two algorithms proposed in [13] to the current situation. First, we derive a new version of the DMT-WMMSE

algorithm. This algorithm solves the signal and spectrum coordination parts of the problem simultaneously. The difference with the algorithm with per-user PCs proposed in [13] is that several Lagrange multipliers have to be found simultaneously. Second, we derive an adaptation of the WMMSE-GDSB algorithm. This algorithm separates the signal and spectrum coordination parts of the problem and solves them sequentially and iteratively. For the spectrum coordination part, we observe that a simple extension of the algorithm presented in [13] fails to produce a functioning algorithm. We propose to optimize the power allocation of all the transceivers of user n on tone k by first doing a change of variables. Consider the vector $\mathbf{p}_n^k$, denoting the allocated power levels of user n on tone k for all transceivers. We re-write this vector in spherical coordinates, i.e. as a function of a radius and a direction vector. We observe that for the radial dimension the problem can be made concave after approximations. It thus becomes easy to solve. To solve for the remaining dimensions more efficiently, we restrict the direction vector to have a fixed ℓ^1 norm. The resulting coordinate system can be interpreted as spherical coordinates in taxicab geometry [16]. This coordinate system allows us to solve for each dimension of the direction vector iteratively and sequentially. Simulation results show good performance with the two proposed algorithms. The WMMSE-GDSB has an advantage in that it is seen to converge much faster than the DMT-WMMSE.

This paper is organized as follows. In Section 2 describes the DMT MIMO IC problem with per-transceiver PCs mathematically. In Section 3 we describe the DMT-WMMSE and in Section 4 we describe the WMMSE-GDSB. Some simulation results are presented in Section 6 and we conclude with Section 6.

We use lower-case boldface letters to denote vectors, upper-case boldface letters for matrices and calligraphic letters for sets (for example, $\mathbf{a}$, $\mathbf{A}$ and $\mathcal{A}$). We use $\mathbf{I}_A$ as the identity matrix of size A, $(\cdot)^{\mathrm{H}}$ as the Hermitian transpose, $\mathrm{E}[\cdot]$ as expectation, $\mathrm{tr}\{\cdot\}$ as trace, $|\cdot|$ as determinant, $\|\cdot\|_2$ as the ℓ^2 norm, $\|\cdot\|_1$ as the ℓ^1 norm and $\mathrm{diag}\{\mathbf{a}\}$ as the matrix with a vector $\mathbf{a}$ in the main diagonal.

2. Problem Statement

We consider a DSL system with N independent users, with user n having A_n transceivers. We also consider discrete multitone (DMT) modulation with K Δ_f-spaced tones. We denote the set of users by $\mathcal{N} = \{1, \ldots, N\}$, the set of transceivers for user n as $\mathcal{A}_n = \{1, \ldots, A_n\}$ and the set of tones by $\mathcal{K} = \{1, \ldots, K\}$. The total number of transceivers in the system is given by $A = \sum_{n \in \mathcal{N}} A_n$. We let $p_{n,(i)}^k$ be the transmit power of transceiver i of user n on tone k and we organize these values in the matrix $\mathbf{P} \in \mathbb{R}^{K \times A}$. A column of $\mathbf{P}$, denoted by $\mathbf{p}_{n,(i)} = \begin{bmatrix} p_{n,(i)}^1 & \cdots & p_{n,(i)}^K \end{bmatrix}^{\mathrm{T}}$, contains the power allocation of transceiver i of user n in all tones. We similarly define the matrix that contains the power allocation of all transceivers of user n in all tones as $\mathbf{P}_n \in \mathbb{R}^{K \times A_n}$, $\mathbf{P}_n = \begin{bmatrix} \mathbf{p}_{n,(1)} & \cdots & \mathbf{p}_{n,(A_n)} \end{bmatrix}$ and the vector with the power allocation of all transceivers of user n in tone k as $\mathbf{p}_n^k$ (the k-row of $\mathbf{P}_n$). Throughout this paper, we focus on a linear design for both transmit and receive matrices and treat interference as noise. All channel gains are considered

perfectly known (not such a tall order in DSL systems). Also, we consider the simplifying assumption of perfect DMT block synchronization between users.[1] Taking that into account, we obtain the received signal for user n on tone k as

$$\mathbf{y}_n^k = \mathbf{H}_{n,n}^k \mathbf{T}_n^k \mathbf{x}_n^k + \sum_{j \neq n} \mathbf{H}_{n,j}^k \mathbf{T}_j^k \mathbf{x}_j^k + \mathbf{z}_n^k. \tag{1}$$

Here $\mathbf{x}_n^k = \begin{bmatrix} x_{n,(1)}^k & \cdots & x_{n,(A_n)}^k \end{bmatrix}^{\mathrm{T}} \in \mathbb{C}^{A_n}$ is the transmit signal vector for user n on tone k; $\mathbf{y}_n^k \in \mathbb{C}^{A_n}$ is the received signal vector for user n on tone k; and $\mathbf{H}_{n,j}^k \in \mathbb{C}^{A_n \times A_j}$, $\mathbf{T}_n^k \in \mathbb{C}^{A_n \times A_n}$ are, respectively, the channel matrix from the transmitter of user j to the receiver of user n on tone k and the transmit matrix for user n on tone k. In (1), we assume $\mathrm{E}\left[\mathbf{x}_n^k (\mathbf{x}_n^k)^{\mathrm{H}}\right] = \mathbf{I}_{A_n} \forall n$, and hence $\mathbf{1}_i^{\mathrm{T}} \mathbf{T}_n^k (\mathbf{T}_n^k)^{\mathrm{H}} \mathbf{1}_i = p_{n,(i)}^k$, where $\mathbf{1}_i$ is a vector with 1 in the ith position and 0 elsewhere. The vector $z_n^k \in \mathbb{C}^{A_n}$ denotes zero mean complex Gaussian noise. Without loss of generality, $\mathbf{z}_n^k$ is assumed to be spatially white with covariance matrix $\mathrm{E}\left[\mathbf{z}_n^k (\mathbf{z}_n^k)^{\mathrm{H}}\right] = \mathbf{I}_{A_n}$. The estimated signal vector for user n on tone k is given by

$$\hat{\mathbf{x}}_n^k = \mathbf{R}_n^k \mathbf{y}_n^k, \tag{2}$$

where $\mathbf{R}_n^k \in \mathbb{C}^{A_n \times A_n}$ is the receive matrix for user n on tone k. The receive matrix used in (2) is the linear MMSE (LMMSE) matrix. For a MIMO IC scenario the LMMSE receiver provides an optimal linear receiver given a set of linear transmit matrices [17, 18]. For a given set of transmit matrices, we have

$$\mathbf{R}_n^k = (\mathbf{T}_n^k)^{\mathrm{H}} (\mathbf{H}_{n,n}^k)^{\mathrm{H}} \left(\mathbf{M}_n^k + \mathbf{H}_{n,n}^k \mathbf{T}_n^k (\mathbf{T}_n^k)^{\mathrm{H}} (\mathbf{H}_{n,n}^k)^{\mathrm{H}} \right)^{-1}, \tag{3}$$

where

$$\mathbf{M}_n^k = \sum_{j \neq n} \mathbf{H}_{n,j}^k \mathbf{T}_j^k (\mathbf{T}_j^k)^{\mathrm{H}} (\mathbf{H}_{n,j}^k)^{\mathrm{H}} + \mathbf{I}_{A_n} \tag{4}$$

is the noise plus interference covariance matrix for user n on tone k. We remark that, although we do not write it explicitly, this matrix may be normalized by a capacity gap Γ.

With the LMMSE receiver and assuming Gaussian signaling, the achievable data rate for user n on tone k is given by

$$b_n^k = \log\left| (\mathbf{T}_n^k)^{\mathrm{H}} (\mathbf{H}_{n,n}^k)^{\mathrm{H}} (\mathbf{M}_n^k)^{-1} \mathbf{H}_{n,n}^k \mathbf{T}_n^k + \mathbf{I}_{A_n} \right|. \tag{5}$$

Here $\log(\cdot)$ denotes the natural logarithm. The total data rate of user n in bits per second is given by $r_n = {}^{f_s}/_{\log(2)} \sum_{k \in \mathcal{K}} b_n^k$, where f_s is the symbol rate. Throughout this paper, we ignore the practical constraint of discrete bit loading.

We denote the set of all transmit matrices $\mathbf{T}_n^k$ as $\mathbb{T} = \left\{ \mathbf{T}_n^k \mid n \in \mathcal{N},\ k \in \mathcal{K} \right\}$. The problem we would like to solve is the weighted rate sum (WRS) maximization with per-transceiver PCs, which can be written

[1] For the case when the DMT blocks of different users are offset in relation to each other, inter carrier interference (ICI) arises. ICI complicates the problem significantly. See e.g. [2]

as

$$\begin{aligned} \max_{\mathbb{T}} \quad & \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_n b_n^k \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}} \mathbf{1}_i^{\mathrm{T}} \mathbf{T}_n^k (\mathbf{T}_n^k)^{\mathrm{H}} \mathbf{1}_i \leq P_n^{\max} \ \forall n, i \end{aligned} \tag{6}$$

Here, $P_n^{\max}$ is the maximum transmit power available for transceiver the transceivers of user n (for simplicity, we assume every transceiver has the same power budget) and u_n is the weight assigned to user n. This problem can be written in an equivalent form, where the signal and spectrum coordination parts are more easily distinguished.

$$\begin{aligned} \max_{\mathbb{T},\mathbf{P}} \quad & \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_n b_n^k \\ \text{subject to} \quad & \mathbf{1}_i^{\mathrm{T}} \mathbf{T}_n^k (\mathbf{T}_n^k)^{\mathrm{H}} \mathbf{1}_i = p_{n,(i)}^k \ \forall n, i, k \\ & \sum_{k\in\mathcal{K}} p_{n,(i)}^k \leq P_n^{\max} \ \forall n, i \end{aligned} \tag{7}$$

Note that here, by fixing $\mathbf{P}$, we are left with K independent MIMO IC's—and thus a signal coordination problem. By decomposing $\mathbf{T}_n^k$ as $\mathbf{T}_n^k = \mathrm{diag}\Big\{\Big[\sqrt{p_{n,(1)}^k} \ \cdots \ \sqrt{p_{n,(A_n)}^k}\Big]\Big\}\overline{\mathbf{T}}_n^k$, where $\mathbf{1}_i^{\mathrm{T}}(\overline{\mathbf{T}}_n^k)^{\mathrm{H}}\overline{\mathbf{T}}_n^k\mathbf{1}_i = 1 \ \forall n, i, k$, and fixing $\overline{\mathbf{T}}_n^k \ \forall n, k$, the problem is described as finding power allocation for every user, tone and transceiver. Hence, it is a power coordination problem.

We call (6) and (7) the DMT MIMO IC WRS maximization problem with per-transceiver PCs.

The single tone MIMO IC (i.e. when $K = 1$), either with per-user or per-transceiver PCs, has been the subject of intensive research [19, 20, 21, 22, 23, 24]. Its multitone version, however, has hitherto received much less attention. For example, a recent and quite extensive review of the recent research in the MIMO IC [25] does not treat it explicitly. There is some related work in the wireless communications literature. In [26, 27], sequential beamforming and multitone power allocation are considered, but with simplifying assumptions. In [26], users are restricted to one data stream and in [27] they are restricted to one receive antenna. In [28, 29] the multitone aspect is pretty much passed by because the authors have assumed that power should be equally divided among the tones. Some of the recent work on the single tone MIMO IC can be easily adapted to the multitone situation by just stacking matrices into a block diagonal structure—the first block corresponding to tone 1, the second to tone 2, etc, see, e.g. [19, 13]. However that does not seem to be the case for all single tone solutions. It is difficult to see, for example, how interference alignment algorithms [22, 23] can be easily adapted to the multitone case. Interference alignment algorithms can be used for independent MIMO ICs with fixed PC (e.g. by fixing $\mathbf{P}$ in (7)), but they have no apparent way to perform the power allocation throughout the tones.

It remains a fact that high performance multitone systems benefit considerably by dynamic power allocation through the tones. This is specially true for the DSL case, where the channel is known to be highly frequency selective and the so-called near-far effects are abundant.

In [13] two algorithms are proposed. Both consider the DMT MIMO IC WRS maximization problem with per-user PCs. In the next two sections, we adapt these algorithms to the problem at hand.

3. DMT-WMMSE with per-transceiver power constraints

In [18], the broadcast channel WRS maximization problem is solved through the simpler weighted MMSE (WMMSE) minimization problem. The resulting algorithm is found to provide good performance with reasonable computational cost. Several works have adapted the main idea of [18] to the MIMO IC WRS maximization problem [13, 19, 30]. These works have only treated the per-user PC case. In this section, we show that the same idea can be easily adapted for the per-transceiver PC case.

Since similar derivations have been presented in [13, 19, 30], here we briefly sketch the derivation of the algorithm for the per-transceiver PC case.

We first define the MSE matrix for user n on tone k as

$$\mathbf{E}_n^k = \mathrm{E}\left[(\mathbf{R}_n^k \mathbf{y}_n^k)(\mathbf{R}_n^k \mathbf{y}_n^k)^{\mathrm{H}}\right] \tag{8}$$

$$= \left((\mathbf{T}_n^k)^{\mathrm{H}} (\mathbf{H}_{n,n}^k)^{\mathrm{H}} (\mathbf{M}_n^k)^{-1} \mathbf{H}_{n,n}^k \mathbf{T}_n^k + \mathbf{I}_{A_n}\right)^{-1} \tag{9}$$

and the DMT WMMSE minimization problem as

$$\begin{aligned} \max_{\mathbb{T}} \quad & \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} -\mathrm{tr}\{\mathbf{W}_n^k \mathbf{E}_n^k\} \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}} \mathbf{1}_i^{\mathrm{T}} \mathbf{T}_n^k (\mathbf{T}_n^k)^{\mathrm{H}} \mathbf{1}_i \leq P_n^{\max} \ \forall n, i \end{aligned} \tag{10}$$

Here, $\mathbf{W}_n^k \in \mathbb{C}^{A_n \times A_n}$ is a weighting matrix. Next, we write the Lagrangean of (6) and (10), find the stationary conditions by calculating the gradient in $\mathbf{T}_n^k$ and compare the two. Just like in [13, 19, 30], it is observed that if the weighting matrix is set as

$$\mathbf{W}_n^k = u_n (\mathbf{E}_n^k)^{-1} \tag{11}$$

then a $\mathbb{T}$ that is a stationary point of (10) is also a stationary point of (6). This is why we can solve the WRS maximization problem through the simpler WMMSE minimization problem. To solve the latter, we write

$$\begin{aligned} \max_{\mathbb{T}} \quad & \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} -\mathrm{E}\left[\left(\mathbf{R}_n^k \mathbf{y}_n^k - \mathbf{x}_n^k\right)^{\mathrm{H}} \mathbf{W}_n^k \left(\mathbf{R}_n^k \mathbf{y}_n^k - \mathbf{x}_n^k\right)\right] \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}} \mathbf{1}_i^{\mathrm{T}} \mathbf{T}_n^k (\mathbf{T}_n^k)^{\mathrm{H}} \mathbf{1}_i \leq P_n^{\max} \ \forall n, i \end{aligned} \tag{12}$$

The solution is given by

$$\mathbf{T}_n^k = \left(\sum_{j\in\mathcal{N}} (\mathbf{H}_{j,n}^k)^{\mathrm{H}} (\mathbf{R}_j^k)^{\mathrm{H}} \mathbf{W}_j^k \mathbf{R}_j^k \mathbf{H}_{j,n}^k + \mathrm{diag}\{\boldsymbol{\lambda}_n\}\right)^{-1} (\mathbf{H}_{n,n}^k)^{\mathrm{H}} (\mathbf{R}_n^k)^{\mathrm{H}} \mathbf{W}_n^k \tag{13}$$

Algorithm 1: DMT-WMMSE with per-transceiver PCs

```
Initialize T_n^k;
repeat
    Calculate R_n^k with (3) ∀n, k;
    Calculate W_n^k with (11) ∀n, k;
    for n = 1, ..., N do
        repeat
            Calculate T_n^k with (13) ∀k;
            for i = 1, ..., A_n do
                if Σ_k 1_i^T T_n^k (T_n^k)^H 1_i > P_n^max then
                    increase λ_{n,(i)};
                else
                    decrease λ_{n,(i)};
        until λ_{n,(i)} |Σ_k 1_i^T T_n^k (T_n^k)^H 1_i − P_n^max| < ε ∀i ;
until until convergence ;
```

where $\boldsymbol{\lambda}_n = \begin{bmatrix}\lambda_{n,(1)} & \cdots & \lambda_{n,(A_n)}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{A_n}$ and $\lambda_{n,(i)}$ is the Lagrange multiplier associated to the ith transceiver of user n. Comparing (13) to the equivalent equations in [13, 19, 30], we notice that the difference is that in (13) there are multiple Lagrange multipliers in a diagonal matrix instead of a single Lagrange multiplier times an identity matrix.

We can write a similar algorithm to the one in [13], except that our version includes the search for the Lagrange multiplier vector $\boldsymbol{\lambda}_n$ instead of a scalar Lagrange multiplier. So, as in [13], we first calculate $\mathbf{R}_n^k$ with (3) $\forall n, k$, then $\mathbf{W}_n^k$ with (11) $\forall n, k$ and then $\mathbf{T}_n^k$ with (13) $\forall n, k$. The calculation of $\mathbf{T}_n^k$ for user n should be done inside a loop, wherein (13) is calculated for all tones and the vector $\boldsymbol{\lambda}_n$ is adjusted. This adjustment can be done by a sub-gradient method or by a nested bisection search. For a given transceiver i, the adjustment should aim for $\lambda_{n,(i)}\left|\sum_k \mathbf{1}_i^{\mathrm{T}}\mathbf{T}_n^k(\mathbf{T}_n^k)^{\mathrm{H}}\mathbf{1}_i - P_n^{\max}\right| < \epsilon$, with $\lambda_{n,(i)} \geq 0\ \forall n, i$ and with ϵ being a small positive number. A complete algorithm description is provided in Algorithm 1.

The demonstrations of convergence and of the fact that the DMT-WMMSE algorithm reaches a stationary point provided in [18, 19, 13] are also valid for the current case. The computational cost of the DMT-WMMSE with per-transceiver PCs is $O\big(KN^2 \max_n\{(A_n)^3 \exp(A_n)\}\big)$. The term with $(A_n)^3$ is due to the matrix multiplications and inversion. The term $\exp(A_n)$ is due to the multidimensional search for an appropriate $\boldsymbol{\lambda}_n$.

4. WMMSE-GDSB with per-transceiver power constraints

For this part of the paper, we start from (7). This version of the problem emphasizes more clearly its distinct signal and spectrum coordination parts. The goal in this section is to solve these two parts sequentially and iteratively.

For the signal coordination part, each tone, transceiver and user has a fixed PC and so each tone can be

solved separately. Here we opt for the WMMSE algorithm of Section 3. We detail the implementation later in the paper when we talk about algorithm design. For the remaining of this section, we focus on how to solve the spectrum coordination part.

We solve the spectrum coordination part in an iterative, per-user fashion. For example, for a two user case, we first fix $\mathbf{P}_2$ and solve for $\mathbf{P}_1$, then we fix $\mathbf{P}_1$ and solve for $\mathbf{P}_2$. The optimization problem for a given user n is given by

$$\begin{aligned} \max_{\mathbf{P}_n} \quad & \sum_{j\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_j b_j^k \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}} p_{n,(i)}^k \leq P_n^{\max} \ \forall i \end{aligned} \tag{14}$$

Here, similarly to (4) and (5), we write

$$b_n^k = \log\big|(\overline{\mathbf{T}}_n^k)^{\mathrm{H}}\mathbf{G}_n^k(\mathbf{H}_{n,n}^k)^{\mathrm{H}}(\mathbf{M}_n^k)^{-1}\mathbf{H}_{n,n}^k\mathbf{G}_n^k\overline{\mathbf{T}}_n^k + \mathbf{I}_{A_n}\big|, \tag{15}$$

$$\mathbf{M}_n^k = \sum_{j\neq n}\mathbf{H}_{n,j}^k\mathbf{G}_j^k\overline{\mathbf{T}}_j^k(\overline{\mathbf{T}}_j^k)^{\mathrm{H}}\mathbf{G}_j^k(\mathbf{H}_{n,j}^k)^{\mathrm{H}} + \mathbf{I}_{A_n}. \tag{16}$$

Here $\mathbf{G}_n^k \triangleq \mathrm{diag}\big\{\begin{bmatrix}\sqrt{p_{n,(1)}^k} & \cdots & \sqrt{p_{n,(A_n)}^k}\end{bmatrix}\big\}$. Notice that $\mathbf{G}_n^k = (\mathbf{G}_n^k)^{\mathrm{H}}$.

Consider the Lagrangean of (14),

$$L(\mathbf{P}_n, \boldsymbol{\lambda}_n) = \sum_{j\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_j b_j^k - \sum_{i\in\mathcal{A}_n}\lambda_{n,(i)}\Big(\sum_{k\in\mathcal{K}} p_{n,(i)}^k - P_n^{\max}\Big). \tag{17}$$

Here, $\lambda_{n,(i)}$ is the Lagrange multiplier associated with transceiver i of user n. We also define $\boldsymbol{\lambda}_n = \begin{bmatrix}\lambda_{n,(1)} & \cdots & \lambda_{n,(A_n)}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{A_n}$. Notice that (17) can be decomposed through the tones, i.e. by defining

$$L(\mathbf{p}_n^k, \boldsymbol{\lambda}_n) = \sum_{j\in\mathcal{N}} u_j b_j^k - \sum_{i\in\mathcal{A}_n}\lambda_{n,(i)} p_{n,(i)}^k \tag{18}$$

and rewriting (17) as $L(\mathbf{P}_n, \boldsymbol{\lambda}_n) = \sum_{k\in\mathcal{K}} L(\mathbf{p}_n^k, \boldsymbol{\lambda}_n) + \sum_{i\in\mathcal{A}_n}\lambda_{n,(i)}P_n^{\max}$. By separately solving

$$\max_{\mathbf{p}_n^k} \quad L(\mathbf{p}_n^k, \boldsymbol{\lambda}_n) \tag{19}$$

for each k we also maximize $L(\mathbf{P}_n, \boldsymbol{\lambda}_n)$ in (17). In the rest of this section we will focus on how to solve (14) quickly and accurately. In order to do that we resort to the Lagrangeans in (17) and (18). When using the Lagrangeans, we keep in mind that the search for appropriate Lagrange multipliers is an important part of the problem. It is not difficult to see that (14) is neither concave nor convex in $\mathbf{P}_n$, which makes the optimization challenging.

In the next sub-sections we describe four approaches to solve (14).

First, in Section 4.1 we try to straightforwardly adapt the algorithm described in [13] to the per-transceiver PC case. The idea here is to linearize the non-concave part of $L(\mathbf{p}_n^k, \boldsymbol{\lambda}_n)$, which allows for a

simple solution. We come to the conclusion that such an approach ultimately fails to produce a functioning algorithm.

In Section 4.2 we use an exhaustive search for $\mathbf{p}_n^k$. We observe that, although this can provide good results, computational complexity grows exponentially with A_n, which becomes prohibitive for large A_n.

In Section 4.3, we propose an approach that saves a little on computational complexity. We begin by doing a change of variables: we replace the cartesian vector of powers $\mathbf{p}_n^k$ by a spherical coordinates equivalent. We re-write $\mathbf{p}_n^k$ as $\rho_n^k \mathbf{d}_n^k$, where ρ_n^k is the radius and $\left\|\mathbf{d}_n^k\right\|_2 = 1$ is a direction vector. The important observation is that for the radial dimension the problem can be made concave after approximations. An exhaustive search is still necessary for the other variables, but now only in $A_n - 1$ dimensions. Hence, computational complexity grows exponentially with $A_n - 1$. This still becomes prohibitive for large A_n.

Finally, in Section 4.4 we explain the approach that forms the core of our proposal. We use yet another change of variables. We replace the original cartesian vector with coordinates of the type $\eta_n^k \mathbf{v}_n^k$, with $\left\|\mathbf{v}_n^k\right\|_1 = 1$. This coordinate system can be interpreted as spherical coordinates in taxicab geometry. With this new formulation, it is possible to more easily decouple the optimization through the variables. The strategy is to solve for each variable independently and sequentially while keeping the others fixed. The advantage is that computational cost grows linearly with A_n.

4.1. The limitations of the traditional approach

In [13], the GDSB algorithm is proposed to iteratively solve the spectrum coordination problem with per-user PCs. In such a case, $\mathbf{T}_n^k$ is decomposed as $\mathbf{T}_n^k = p_n^k \overline{\mathbf{T}}_n^k$, with $\mathrm{tr}\{\overline{\mathbf{T}}_n^k (\overline{\mathbf{T}}_n^k)^{\mathrm{H}}\} = 1$, and the variables are $p_n^k,\ k \in \mathcal{K}$. In [13], when solving for user n the Lagrangean is formulated (similar to (18)) and the problem is divided in concave and non-concave parts. The latter is given by $\sum_{j \neq n} b_j^k$. By approximating the non-concave part by its first order Taylor expansion, a concave function is found, which is then solved straightforwardly. Lagrange multipliers have to be adjusted so that the power constraint is respected. The whole procedure repeats until convergence.

In this section, we try to apply the same procedure to the per-transceiver PC case but conclude that it does not work. We depart from the per-tone problem in (18) and re-write it as

$$L(\mathbf{p}_n^k, \boldsymbol{\lambda}_n) = u_n b_n^k + \sum_{j \neq n} u_j b_j^k - \sum_{i \in \mathcal{A}_n} \lambda_{n,(i)} p_{n,(i)}^k. \tag{20}$$

As in [13], we notice that the first and last terms of this equation are increasing with $\mathbf{p}_n^k$. The term with the summation in j is decreasing with $\mathbf{p}_n^k$. We linearize the part with the sum in j by approximating it by its first order Taylor expansion. So, consider

$$b_j^k \approx b_j^k \Big|_{\hat{p}_{n,(i)}^k} + \left(p_{n,(i)}^k - \hat{p}_{n,(i)}^k\right) \frac{\partial b_j^k}{\partial p_{n,(i)}^k} \Bigg|_{\hat{p}_{n,(i)}^k} \tag{21}$$

Here $\hat{p}^k_{n,(i)}$ denotes the power allocation for user n and transceiver i in the previous iteration. To maximize (20) in $p^k_{n,(i)}$, we plug (21) in (20), take the derivative in $p^k_{n,(i)}$ and set it to zero. The derivative of the part that is increasing with $\mathbf{p}^k_n$ is given by

$$\frac{u_n \partial\big(b^k_n - \lambda_{n,(i)} p^k_{n,(i)}\big)}{\partial p^k_{n,(i)}} = -\lambda_{n,(i)} + u_n \mathrm{tr}\big\{\mathbf{E}^k_n (\overline{\mathbf{T}}^k_n)^{\mathrm{H}} \mathrm{diag}\big\{\big[\cdots \;\; 0 \;\; {}^1\!/\!_{\sqrt{p_{(i)}}} \;\; 0 \;\; \cdots\big]\big\} (\mathbf{H}^k_{n,n})^{\mathrm{H}} (\mathbf{M}^k_n)^{-1} \mathbf{H}^k_{n,n} \mathbf{G}^k_n \overline{\mathbf{T}}^k_n\big\}, \tag{22}$$

where $\mathbf{E}^k_n$ is given by (8). Although b^k_n is not a concave function of $p^k_{n,(i)}$, the multiple local maxima and minima can be found by calculating the roots of a polynomial, which does not represent a big problem. We would only have to evaluate roots and pick the best one. The derivative of the part with the sum in j is given by

$$\sum_{j \neq n} \frac{\partial u_j b^k_j}{\partial p^k_{n,(i)}} = \sum_{j \neq n} -u_j \mathrm{tr}\Big\{ (\overline{\mathbf{T}}^k_n)^{\mathrm{H}} \mathbf{G}^k_n (\mathbf{H}^k_{j,n})^{\mathrm{H}} (\mathbf{M}^k_n)^{-1} \mathbf{H}^k_{j,j} \mathbf{T}^k_j \mathbf{E}^k_j (\overline{\mathbf{T}}^k_j)^{\mathrm{H}} (\mathbf{H}^k_{j,j})^{\mathrm{H}} (\mathbf{M}^k_n)^{-1} \mathbf{H}^k_{j,n} \times \mathrm{diag}\big\{\big[\cdots \;\; 0 \;\; {}^1\!/\!_{\sqrt{p_{(i)}}} \;\; 0 \;\; \cdots\big]\big\} \overline{\mathbf{T}}^k_n \Big\}. \tag{23}$$

The problem with this approach is that, for the per-transceiver PC case, the first order Taylor expansion of b^k_j (see (21) and (23)) is a very poor characterization of its behavior. This is so because b^k_j is not a convex, monotonically decreasing function of $\mathbf{p}^k_n$, $\mathbf{p}^k_n \geq 0$. See Fig. 1 for an illustration. In the figure, we calculate b^k_j with (15) for different values of power for an interferer user n that has two transceivers, i.e we calculate b^k_j as a function of $p^k_{n,(1)}$ and $p^k_{n,(2)}$. All matrices involved are randomly chosen and of size two by two. In this figure, the thick line marks the behavior of b^k_j as we fix $p^k_{n,(1)} = 1.5$. This curve is shown more clearly in Fig. 2 along with its first order Taylor expansion around $p^k_{n,(2)} = 0.25$. The Taylor expansion influences the power allocation for the transceiver i of user n by including a price for interference it causes [4, 31]. For the particular point shown in the figure, the Taylor expansion actually represents an incentive. User n has an incentive to increase $p^k_{n,(2)}$, as if b^k_j would increase indefinitely with more interference. Accordingly, note that in (23) the argument of the trace is a matrix with no special properties—it is not necessarily hermitian or positive definite. The result of the trace operation can be a positive or a negative or even a complex number.

If the algorithm finds itself in such a point, i.e. when for a particular tone the prices for loading power become incentives, a very large amount of power is used in this tone. Because of the common PC, all other tones have to decrease power greatly, even if they have favorable channel conditions. Data rate for user n most likely drops. The algorithm does not work.

This is in contrast with the case of the algorithms proposed in [13, 4, 31] that deal with the per-user PC. In these references, b^k_j is a convex and monotonically decreasing function of p^k_n, which means that $\partial u_j b^k_j / \partial p^k_n$ is always real and non-positive, i.e. a true price rather than an incentive. Similar convex approximations proposed to the per-user PC case, i.e. [3] would also not work for the per-transceiver PC case for the same

reason.

4.2. Exhaustive search in $\mathbf{p}_n^k$

One straightforward way of solving (14) is with an exhaustive grid search in $\mathbf{p}_n^k$. This is actually the most direct way to solve (19). The problem is that the grid increases exponentially with the number of transceivers A_n, as does the number of functions evaluations.

For a small number of transceivers, e.g. A_n equals 2 or 3, it is actually possible to perform the exhaustive search. However, such a search is embedded in an algorithm that also adjusts transmit matrices. Each time the transmit matrices are updated, the power allocation needs to be updated as well and the algorithm continues sequentially. This means that, if we choose the exhaustive search to solve the power coordination part, we would have to repeat it tens, maybe hundreds of times. This makes this option not effective at all.

4.3. Changing $\mathbf{p}_n^k$ *to spherical coordinates—exhaustive search for the direction vector*

We can save on computational cost by changing our search slightly. Consider $\mathbf{p}_n^k$ in spherical coordinates.

$$\mathbf{p}_n^k = \rho_n^k \mathbf{d}_n^k, \quad \rho_n^k \geq 0, \quad \left\|\mathbf{d}_n^k\right\|_2 = 1 \tag{24}$$

Here ρ_n^k is the radius, i.e. $\rho_n^k = \left\|\mathbf{p}_n^k\right\|_2$, and $\mathbf{d}_n^k = \begin{bmatrix} d_{n,[1]}^k & \cdots & d_{n,[A_n]}^k \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{A_n}$, $d_{n,[i]}^k \geq 0,\ i \in \mathcal{A}_n$, is a direction. Here we make two remarks. First notice that while in the cartesian vector each element is associated to a transceiver (e.g. $p_{n,(1)}^k$ for transceiver 1), this is not true for the spherical coordinates vector. If, for example, we change the value of ρ_n^k, all transceivers' power change. This is why ρ_n^k has only one subscript and why we add brackets to the subscripts in $d_{n,[i]}^k$. In $d_{n,[i]}^k$, the bracketed subscripts now represent directions, not transceivers. Second, notice that, although the vector $\mathbf{d}_n^k$ is of size A_n, it contains only $A_n - 1$ free variables. For example, for a case with three transceivers, we would have

$$\mathbf{p}_n^k = \rho_n^k \begin{bmatrix} \cos\theta_{n,[2]}^k \cos\theta_{n,[1]}^k \\ \cos\theta_{n,[2]}^k \sin\theta_{n,[1]}^k \\ \sin\theta_{n,[2]}^k \end{bmatrix}, \rho_n^k \geq 0,\ 0 \leq \theta_{n,[1]}^k, \theta_{n,[2]}^k \leq \frac{\pi}{2}. \tag{25}$$

In Appendix A, we write the spherical coordinates vector for the general A_n transceiver case. See Eqs. (A.1), (A.2) and (A.3).

We now define $\mathbf{D}_n^k \triangleq \mathrm{diag}\left\{\begin{bmatrix}\sqrt{d_{n,[1]}^k} & \cdots & \sqrt{d_{n,[A_n]}^k}\end{bmatrix}\right\}$. Notice that $\mathbf{D}_n^k = (\mathbf{D}_n^k)^{\mathrm{H}}$. Next we decompose $\mathbf{T}_n^k$ as $\mathbf{T}_n^k = \sqrt{\rho_n^k}\mathbf{D}_n^k\overline{\mathbf{T}}_n^k$. Here again $\mathbf{1}_i^{\mathrm{T}}\overline{\mathbf{T}}_n^k(\overline{\mathbf{T}}_n^k)^{\mathrm{H}}\mathbf{1}_i = 1\ \forall i$. We re-write (5) and (4) as

$$b_n^k = \log\left|\rho_n^k(\overline{\mathbf{T}}_n^k)^{\mathrm{H}}\mathbf{D}_n^k(\mathbf{H}_{n,n}^k)^{\mathrm{H}}(\mathbf{M}_n^k)^{-1}\mathbf{H}_{n,n}^k\mathbf{D}_n^k\overline{\mathbf{T}}_n^k + \mathbf{I}_{A_n}\right| \tag{26}$$

$$\mathbf{M}_n^k = \sum_{j\neq n}\rho_j^k\mathbf{H}_{n,j}^k\mathbf{D}_j^k\overline{\mathbf{T}}_j^k(\overline{\mathbf{T}}_j^k)^{\mathrm{H}}\mathbf{D}_j^k(\mathbf{H}_{n,j}^k)^{\mathrm{H}} + \mathbf{I}_{A_n} \tag{27}$$

Now we re-write the problem as a function of ρ_n^k and $\mathbf{d}_n^k$, write the Lagrangean and, just as we did in Section 4.1, divide it increasing and decreasing parts. The equivalent of (20) is

$$L(\rho_n^k, \mathbf{d}_n^k, \boldsymbol{\lambda}_n) = u_n b_n^k + \sum_{j \neq n} u_j b_j^k - \sum_{i \in \mathcal{A}_n} \lambda_{n,(i)} \rho_n^k d_{n,[i]}^k. \tag{28}$$

The fundamental advantage of the formulation based on the spherical coordinates is that, if keeping $\mathbf{d}_n^k$ fixed, the first order Taylor expansion of the non-concave part (i.e. $\sum_{j \neq n} u_j b_j^k$) in ρ_n^k provides a good enough approximation of the behavior of the b_j^k's. That is because, for each fixed direction $\mathbf{d}_n^k$, b_j^k is a convex, monotonically decreasing function of ρ_n^k, $\rho_n^k \geq 0$. Additionally, b_n^k is a concave function of ρ_n^k ([32], pg 74).

Next we focus on (28), we calculate the first order Taylor expansion of the non-concave part of $L(\rho_n^k, \mathbf{d}_n^k, \boldsymbol{\lambda}_n)$ (i.e. $\sum_{j \neq n} b_j^k$) and we write the stationary condition as a function of ρ_n^k for fixed $\mathbf{d}_n^k$. We obtain

$$\frac{\partial L(\rho_n^k, \mathbf{d}_n^k, \boldsymbol{\lambda}_n)}{\partial \rho_n^k} = u_n \mathrm{tr}\Big\{ \big(\rho_n^k \mathbf{S}_n^k(\mathbf{D}_n^k) + \mathbf{I}_{A_n}\big)^{-1} \mathbf{S}_n^k(\mathbf{D}_n^k) \Big\} - \tau_n^k(\mathbf{D}_n^k) - \boldsymbol{\lambda}_n^{\mathrm{T}} \mathbf{d}_n^k = 0. \tag{29}$$

Here

$$\mathbf{S}_n^k(\mathbf{D}_n^k) = (\overline{\mathbf{T}}_n^k)^{\mathrm{H}} \mathbf{D}_n^k (\mathbf{H}_{n,n}^k)^{\mathrm{H}} (\mathbf{M}_n^k)^{-1} \mathbf{H}_{n,n}^k \mathbf{D}_n^k \overline{\mathbf{T}}_n^k. \tag{30}$$

The variable $\tau_n^k(\mathbf{D}_n^k)$ is obtained after the linearization with the Taylor expansion and is given by

$$\tau_n^k(\mathbf{D}_n^k) \triangleq -\sum_{j \neq n} u_j \frac{\partial b_j^k}{\partial p_n^k} = \sum_{j \neq n} u_j \mathrm{tr}\Big\{ (\overline{\mathbf{T}}_n^k)^{\mathrm{H}} \mathbf{D}_n^k (\mathbf{H}_{j,n}^k)^{\mathrm{H}} (\mathbf{M}_j^k)^{-1} \mathbf{H}_{j,j}^k \mathbf{T}_j^k \mathbf{E}_j^k (\mathbf{T}_j^k)^{\mathrm{H}} (\mathbf{H}_{j,j}^k)^{\mathrm{H}} (\mathbf{M}_j^k)^{-1} \mathbf{H}_{j,n}^k \mathbf{D}_n^k \overline{\mathbf{T}}_n^k \Big\}. \tag{31}$$

Here $\mathbf{E}_n^k$ is given by (8). Note that, in contrast to (29), the argument of the trace operator is now a hermitian positive semidefinite matrix. Its trace will always be non-negative and, as a consequence, $-\tau_n^k(\mathbf{d}_n^k)$ acts as a price for power loaded for user n. For each fixed $\mathbf{d}_n^k$, we obtain an easy, concave problem after the linearization of the non-concave parts.

Eq. (29) can be solved by finding the roots of a polynomial of degree A_n. It can be shown that there is at most one non-negative root [13], which implies that the only root of interest is the rightmost one. The problem in ρ_n^k is in essence the problem with the per-user PCs in [13].

If solving for ρ_n^k is easy, the same cannot be said for the angles $\theta_{n,[i]}^k$. The $\theta_{n,[i]}^k$'s are coupled among themselves and with ρ_n^k. The way to solve the full problem is to do an exhaustive search in the direction vector $\mathbf{d}_n^k$. The advantage is that, since there are $A_n - 1$ free variables in $\mathbf{d}_n^k$, the exhaustive search would be in a grid with one dimension less in comparison to the cartesian exhaustive search of Section 4.2. For each fixed $\mathbf{d}_n^k$, solving for ρ_n^k is easy. We can then develop an algorithm where all possible directions are searched for exhaustively, for example by sampling $[0, \pi/2]$ with Q points and building a $(A_n - 1)$-dimensional grid with such points. For a given point in the grid, we have values for $\theta_{n,[i]}^k$, $i = 1, \ldots, A_n - 1$, we calculate $\mathbf{d}_n^k$ and solve for ρ_n^k. Lagrange multipliers have to be searched for in an outer loop so that the power budgets

are respected. We have tested such algorithm and it works well. However, it only works for small A_n. The grid search grows exponentially with $A_n - 1$, which quickly becomes unfeasible. The fact of the matter is that such an algorithm would still be very limited. The shortcomings of the cartesian vector exhaustive search described in Section 4.2 are but slightly mitigated.

4.4. Changing $\mathbf{p}_n^k$ to spherical coordinates in taxicab geometry—iterative search for the direction vector

Consider that the vector $\mathbf{p}_n^k$ is re-written as

$$\mathbf{p}_n^k = \eta_n^k \mathbf{v}_n^k, \quad \eta_n^k \geq 0, \quad \left\|\mathbf{v}_n^k\right\|_1 = 1 \tag{32}$$

where, just as before, $\mathbf{v}_n^k = \begin{bmatrix} v_{n,[1]}^k & \cdots & v_{n,[A_n]}^k \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{A_n}$, where $v_{n,[i]}^k \geq 0, i \in \mathcal{A}_n$, points to a direction and η_n^k is the radius. Because of the ℓ^1 norm constraint in $\mathbf{v}_n^k$, $\eta_n^k \mathbf{v}_n^k$ describes a sphere of radius η_n^k in taxicab geometry [16]. Hence we refer to this system of coordinates as spherical coordinates in taxicab geometry.

In order to make the exposition clearer and easier, for the remaining of this section we focus on an example where we want to find $\mathbf{P}_n$ for a user with $A_n = 3$. We will recover the general case in the section about algorithm design. The extension is straightforward. For this case with $A_n = 3$, we write $\mathbf{p}_n^k$ as

$$\mathbf{p}_n^k = \eta_n^k \mathbf{v}_n^k = \eta_n^k \begin{bmatrix} (1 - \phi_{n,[2]}^k)(1 - \phi_{n,[1]}^k) \\ (1 - \phi_{n,[2]}^k)\phi_{n,[1]}^k \\ \phi_{n,[2]}^k \end{bmatrix} \tag{33}$$

Here $0 \leq \phi_{n,[2]}^k \leq 1$ and $0 \leq \phi_{n,[1]}^k \leq 1$ define a point in the sphere of radius 1 in taxicab geometry. Given a $\mathbf{p}_n^k$ vector it is straightforward to obtain η_n^k, $\phi_{n,[1]}^k$ and $\phi_{n,[2]}^k$ and write the equivalent formulation (33). In Appendix A, we write the spherical coordinates in taxicab geometry for the general case in (A.5)-(A.7).

In Section 4.3, we remarked that the spherical coordinates (in Euclidean geometry) representation of $\mathbf{p}_n^k$ is a bit unusual in the sense that the variables are not directly related to users' powers. This still applies to the spherical coordinates in taxicab geometry. However, for the latter it is easier to control the share of power that is allocated to each transceiver and to thus optimize the variables separately. Because of the fixed ℓ^1 norm, changing either $\phi_{n,[1]}^k$ or $\phi_{n,[2]}^k$ does not change the total per-user power, i.e.

$$\sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{A}_n} \eta_n^k v_{n,[i]}^k = \sum_{k \in \mathcal{K}} \eta_n^k \left\|\mathbf{v}_n^k\right\|_1 = \sum_{k \in \mathcal{K}} \eta_n^k. \tag{34}$$

This is clearly not true for the (Euclidean) spherical coordinates (see Section 4.3). What we do next is to solve for the radius η_n^k with a per-user PC and then solve for each $\phi_{n,[i]}^k$ separately and sequentially. By adjusting prices for the angles in the form of Lagrange multipliers, we can satisfy all per-transceiver PCs.

We proceed by writing the Lagrangean as a function of η_n^k, $\phi_{n,[1]}^k$ and $\phi_{n,[2]}^k$. Before that, we notice that

the A_n per-transceiver PCs of the form $\sum_{k\in\mathcal{K}} \eta_n^k v_{n,[i]}^k \leq P_n^{\max}$, $i = 1, \ldots, A_n$ can be written in the form

$$\sum_{k\in\mathcal{K}} \eta_n^k \leq A_n P_n^{\max} \tag{35}$$

$$\sum_{k\in\mathcal{K}} \eta_n^k v_{n,[i]}^k \leq P_n^{\max},\ i = 2, \ldots, A_n \tag{36}$$

Here, we consider, without loss of generality, that the transceivers are ordered with ascending loaded powers, i.e. they are ordered with ascending $\sum_{k\in\mathcal{K}} \eta_n^k v_{n,[i]}^k$. With (35) and (36) in hands, we write the Lagrangean as

$$L(\boldsymbol{\eta}_n, \boldsymbol{\phi}_n, \boldsymbol{\lambda}_n) = \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_n b_n^k - \lambda_{n,[1]}\Big(\sum_{k\in\mathcal{K}} \eta_n^k - A_n P_n^{\max}\Big) - \sum_{\substack{i\in\mathcal{A}_n \\ i\neq 1}} \lambda_{n,[i]}\Big(\sum_{k\in\mathcal{K}} \eta_n^k v_{n,[i]}^k - P_n^{\max}\Big). \tag{37}$$

Here we define $\boldsymbol{\eta}_n = \begin{bmatrix}\eta_n^1 & \cdots & \eta_n^K\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^K$ and $\boldsymbol{\phi_n} = \begin{bmatrix}\boldsymbol{\phi}_{n,[1]} & \cdots & \boldsymbol{\phi}_{n,[A_n]}\end{bmatrix} \in \mathbb{R}^{K\times A_n}$, where $\boldsymbol{\phi}_{n,[i]}$ is defined analogously to $\boldsymbol{\eta}_n$; and $\boldsymbol{\lambda}_n = \begin{bmatrix}\lambda_{n,[1]} & \cdots & \lambda_{n,[A_n]}\end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{A_n}$, $\lambda_{n,[i]} \geq 0$, $i \in \mathcal{A}_n$.

To solve the whole problem, we separate the Lagrangean in two parts, one of them relating to the radius η_n^k and the other to the $\phi_{n,[i]}^k$'s, i.e.

$$L(\boldsymbol{\eta}_n, \lambda_{n,[1]}) = \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_n b_n^k - \lambda_{n,[1]}\Big(\sum_{k\in\mathcal{K}} \eta_n^k - A_n P_n^{\max}\Big) - \sum_{\substack{i\in\mathcal{A}_n \\ i\neq 1}} \lambda_{n,[i]}\Big(\sum_{k\in\mathcal{K}} \eta_n^k v_{n,[i]}^k - P_n^{\max}\Big) \tag{38}$$

$$L(\boldsymbol{\phi}_n, \lambda_{n,[2]}, \lambda_{n,[3]}) = \sum_{n\in\mathcal{N}}\sum_{k\in\mathcal{K}} u_n b_n^k - \sum_{\substack{i\in\mathcal{A}_n \\ i\neq 1}} \lambda_{n,[i]}\Big(\sum_{k\in\mathcal{K}} \eta_n^k v_{n,[i]}^k - P_n^{\max}\Big) \tag{39}$$

In (38), the constraint we use is per-user instead of per-transceiver. The maximization of (38) can be solved per tone. After approximating its convex part by the first order Taylor expansion, the stationary condition of this problem is given by

$$\frac{\partial L(\eta_n^k, \lambda_{n,[1]})}{\partial \eta_n^k} = u_n \mathrm{tr}\Big\{\big(\eta_n^k \mathbf{S}_n^k(\mathbf{V}_n^k) + \mathbf{I}_{A_n}\big)^{-1}\mathbf{S}_n^k(\mathbf{V}_n^k)\Big\} - \tau_n^k(\mathbf{V}_n^k) - \lambda_{n,[1]} - \sum_{\substack{i\in\mathcal{A}_n \\ i\neq 1}} \lambda_{n,[i]} v_{n,[i]}^k = 0, \tag{40}$$

where $\mathbf{S}_n^k(\mathbf{V}_n^k)$ is given by (30), $\tau_n^k(\mathbf{V}_n^k)$ is given by (31) and $\mathbf{V}_n^k = \mathrm{diag}\big\{\big[\sqrt{v_{n,[1]}^k} \quad \cdots \quad \sqrt{v_{n,[A_n]}^k}\big]\big\}$. This problem is the same as that already solved in [13] and in Section 4.3. To solve for $\boldsymbol{\eta}_n$, we solve (40) for every tone and search in an outer loop for the appropriate Lagrange multiplier $\lambda_{n,[1]}$.

After, the solution for $\boldsymbol{\eta}_n$, we obtain $\sum_{k\in\mathcal{K}} \eta_n^k \leq A_n P_n^{\max}$. However, at this point it can be that the per-transceiver PCs are not satisfied. As in the example given in the introduction of this paper, it can be that, if $P_n^{\max} = 100$ mW, the individual transceivers use, respectively, 80, 100 and 120 mW, which violates the power constraint of the third transceiver (as mentioned, the transceivers are in ascending order of allocated powers). If at least one of the transceivers surpasses its power constraint, then we have to solve for the $\phi_{n,[i]}^k$'s.

To satisfy the per-transceiver PCs's, we solve first for $\boldsymbol{\phi}_{n,[2]}$ and then $\boldsymbol{\phi}_{n,[1]}$ for fixed $\lambda_{n,[2]}$ and $\lambda_{n,[3]}$. For $\phi^k_{n,[2]}$, we write

$$L(\phi^k_{n,[2]}, \lambda_{n,[2]}, \lambda_{n,[3]}) = \sum_{j\in\mathcal{N}} u_j b^k_j - \lambda_{n,[2]}\eta^k_n(1-\phi^k_{n,[2]})\phi^k_{n,[1]} - \lambda_{n,[3]}\eta^k_n\phi^k_{n,[2]}. \tag{41}$$

We solve (41) for each tone independently with an exhaustive search. Notice that, for every tone, $\eta^k_n(-\lambda_{n,[3]}+\lambda_{n,[2]}\phi^k_{n,[1]})$ works as a price for $\phi^k_{n,[2]}$. The greater this price is, the smaller $\phi^k_{n,[2]}$ tends to be. With smaller values for $\phi^k_{n,[2]}$, there is less power for transceiver 3 and more for transceiver 1 and 2 (see (33)).

Next, we solve for $\boldsymbol{\phi}_{n,[1]}$ (still for fixed $\lambda_{n,[2]}$ and $\lambda_{n,[3]}$). For $\phi^k_{n,[1]}$, we write

$$L(\phi^k_{n,[1]}, \lambda_{n,[2]}) = \sum_{j\in\mathcal{N}} u_j b^k_j - \lambda_{n,[2]}\eta^k_n(1-\phi^k_{n,[2]})\phi^k_{n,[1]} \tag{42}$$

We solve (42) for each tone independently with an exhaustive search. Notice that, for every tone, $-\lambda_{n,[2]}\eta^k_n(1-\phi^k_{n,2})$ works as a price for $\phi^k_{n,[1]}$. The greater this price is, the smaller $\phi^k_{n,[1]}$ tends to be. With smaller values for $\phi^k_{n,[1]}$, there is less power for transceiver 2 and more for transceiver 1 (see (33)).

Because we order the transceivers is ascending order of allocated powers, it is natural that $\lambda_{n,[2]}$ and $\lambda_{n,[3]}$ are non-negative—transceivers with excessive power should have a price. They have to be non-negative because they also act as prices for the solution of the radius (see (40)). It is also natural that at convergence $(-\lambda_{n,[3]}+\lambda_{n,[2]}\phi^k_{n,[1]}) \geq 0\ \forall k$, i.e. the third transceiver has a price instead of an incentive. By adjusting the Lagrange multipliers we can apply correct prices to both $\phi^k_{n,[2]}$ and $\phi^k_{n,[1]}$ such that the PCs for transceivers 2 and 3 are satisfied. A joint search for the Lagrange multipliers $\lambda_{n,[2]}$ and $\lambda_{n,[3]}$ has to be performed. After appropriate values are found, the constraint for transceiver 1 is automatically satisfied.

4.4.1. Algorithm

The pseudo-code of the proposed method for the general case is shown in Algorithm 2. In line 2, we sample the ϕ-space, $\phi \in [0,1]$. The set of points is represented as $\mathcal{Q}$. Note here that $\mathcal{Q}$ does not grow exponentially with the number of transceivers. We refer to an element of $\mathcal{Q}$ as $\phi^{\{q\}}$. In line 3, we order the transceivers of every user in ascending order of channel quality, i.e. from bad channels to good channels. This is important to have the ascending order of allocated powers after the solution for the radius. In the DSL case, a rule of thumb is to order the transceivers in descending order of line lengths. In a more general case, we can solve for the radius once (with lines 7-10 in the algorithm) and order the transceivers with ascending powers.

Lines 4 to 13 contain the steps for the optimization of $\mathbf{P}_n$. In Lines 7 to 10 we solve for the radius η^k_n—the Lagrange multiplier is found with a bisection search. After the solution for $\boldsymbol{\eta}_n$, we obtain $\sum_{k\in\mathcal{K}}\eta^k_n \leq A_n P^{\max}_n$.

In lines 11-13 we solve for the $\boldsymbol{\phi}_{n,[i]}$'s, one at a time and for fixed $\lambda_{n,[2]},\ldots,\lambda_{n,[A_n]}$. These Lagrange

Algorithm 2: WMMSE-GDSB with per-transceiver PCs

1 Initialize $\mathbf{T}_n^k$, $\boldsymbol{\lambda}_n$;
2 $\mathcal{Q} \leftarrow$ discretized ϕ-space with $\phi \in [0,1]$;
3 For each user, order transceivers with ascending channel quality;
4 **repeat**
5 Calculate $\tau_n^k(\mathbf{E}_n^k)$ with (31) and $\mathbf{S}_n^k(\mathbf{E}_n^k)$, $n \in \mathcal{N}$, $k \in \mathcal{K}$;
6 **for** $n = 1, \ldots, N$ **do**
7 **repeat**
8 Find η_n^k with (40), $k \in \mathcal{K}$;
9 Adjust $\lambda_{n,[1]}$;
10 **until** $\lambda_{n,[1]} \big| \sum_k \eta_n^k - A_n P_n^{\max} \big| < \epsilon$;
11 **for** $i = A_n - 1, \ldots, 1$ **do**
12 Calculate $L(\phi_{n,[i]}^k, \lambda_{n,[2]}, \ldots, \lambda_{n,[A_n]})$ with $\phi_{n,[i]}^k = \phi^{\{q\}}$, $k \in \mathcal{K}$, $q \in \mathcal{Q}$;
13 $\phi_{n,[i]}^k = \operatorname{argmax}_{\phi^{\{q\}} \in \mathcal{Q}} L(\phi_{n,[i]}^k, \lambda_{n,[2]}, \ldots, \lambda_{n,[A_n]})$, $k \in \mathcal{K}$;
14 $\mathbf{p}_n^k = \eta_n^k \mathbf{v}_n^k, n \in \mathcal{N}, k \in \mathcal{K}$;
15 Calculate $\mathbf{R}_n^k$ with (3), $n \in \mathcal{N}$, $k \in \mathcal{K}$;
16 Calculate $\mathbf{W}_n^k$ with (11), $n \in \mathcal{N}$, $k \in \mathcal{K}$;
17 **for** $n = 1, \ldots, N$ **do**
18 **for** $k = 1, \ldots, K$ **do**
19 **repeat**
20 Calculate $\mathbf{T}_n^k$ with (13);
21 Adjust $\boldsymbol{\mu}_n$;
22 **until** $\big| \mathbf{1}_i^{\mathrm{T}} \mathbf{T}_n^k (\mathbf{T}_n^k)^{\mathrm{H}} \mathbf{1}_i - p_{n,(i)}^k \big| < \epsilon \ \forall i$;
23 **for** $n = 1, \ldots, N$ **do**
24 **for** $i = 2, \ldots, A_n$ **do**
25 $\lambda_{n,[i]} = \max\big[\lambda_{n,[i]} + \gamma\big(\sum_{k \in \mathcal{K}} \eta_n^k v_{n,[i]}^k - P_n^{\max}\big), 0\big]$;
26 **until** *until convergence* ;

multipliers are adjusted at the very end of the algorithm. In line 12, we calculate

$$L(\phi_{n,[i]}^k, \lambda_{n,[2]}, \ldots, \lambda_{n,[A_n]}) = \sum_{n \in \mathcal{N}} u_n b_n^k - \sum_{\substack{i \in \mathcal{A}_n \\ i \neq 1}} \lambda_{n,[i]} \eta_n^k v_{n,[i]}^k \tag{43}$$

for all tones and all elements in $\mathcal{Q}$. Here, the per-tone data rates are calculated as a function of η_n^k (calculated before) and $\phi_{n,[i]}^k = \phi^{\{q\}}$. Accordingly, we calculate (43) with (remember that $\mathbf{V}_n^k = \operatorname{diag}\big\{\big[\sqrt{v_{n,[1]}^k} \ \cdots \ \sqrt{v_{n,[A_n]}^k}\big]\big\}$)

$$b_n^k = \log\big|\eta_n^k (\overline{\mathbf{T}}_n^k)^{\mathrm{H}} \mathbf{V}_n^k (\mathbf{H}_{n,n}^k)^{\mathrm{H}} (\mathbf{M}_n^k)^{-1} \mathbf{H}_{n,n}^k \mathbf{V}_n^k \overline{\mathbf{T}}_n^k + \mathbf{I}_{A_n}\big| \tag{44}$$

where $\mathbf{M}_n^k$ is given by (4), and, for $j \neq n$,

$$b_j^k = \log\big|(\mathbf{T}_j^k)^{\mathrm{H}} (\mathbf{H}_{j,j}^k)^{\mathrm{H}} (\mathbf{M}_j^k)^{-1} \mathbf{H}_{j,j}^k \mathbf{T}_j^k + \mathbf{I}_{A_j}\big| \tag{45}$$

$$\mathbf{M}_j^k = \sum_{\substack{m \neq j \\ m \neq n}} \mathbf{H}_{j,m}^k \mathbf{T}_m^k (\mathbf{T}_m^k)^{\mathrm{H}} (\mathbf{H}_{j,m}^k)^{\mathrm{H}} + \eta_n^k \mathbf{H}_{j,n}^k \mathbf{V}_n^k \overline{\mathbf{T}}_n^k (\overline{\mathbf{T}}_n^k)^{\mathrm{H}} \mathbf{V}_n^k (\mathbf{H}_{j,n}^k)^{\mathrm{H}} + \mathbf{I}_{A_j}. \tag{46}$$

After power allocation, in line 14 we transform back to cartesian coordinates and we continue with the

signal coordination in lines 15-22. As in [13], we first calculate $\mathbf{R}_n^k$ and then $\mathbf{W}_n^k$ for all users and tones. In the loop on lines 19-22 we calculate $\mathbf{T}_n^k$ and search for adequate Lagrange multipliers (here denoted by $\boldsymbol{\mu}_n$).

At the end of the first iteration, the PC power constraints are most likely not satisfied. If at least one transceiver surpasses its PC, we have to adjust its Lagrange multiplier. This is done in lines 23-25. Here γ is a predefined step size.

We repeat the whole process until convergence.

4.4.2. Computational Complexity

The computational cost of the WWMSE-GDSB with per-transceiver PCs is estimated as follows. For the solution for $\boldsymbol{\eta}_n$ for one user n, the algorithm calculates $\tau_n^k(\mathbf{E}_n^k)$, which implies a cost of the order of $KN(A_n)^3$. For the solution for one of the $\boldsymbol{\phi}_{n,[i]}$, say $i = 1$, for user n, the cost is proportional to $KNQ(A_n)^3$ (the term $(A_n)^3$ is due to the matrix multiplications and inversions). This should be repeated for all i, which implies a cost proportional to $KNQ(A_n)^3(A_n - 1)$. The whole process should repeat for all users, so the computational cost of the spectrum coordination part of Algorithm 1 (i.e. lines 4-17) is $O(KN^2 \max\{(A_n)^3\} + KN^2Q \max_n\{(A_n)^3(A_n - 1)\})$.

Computational cost of the signal coordination part is $O(KN^2 \max\{(A_n)^3 \exp(A_n)\})$.

4.5. Convergence

It is very difficult to show convergence of the whole procedure, but some partial results provide some guarantee. For example, with the same line of thinking as [13], we can show that the update for the radius η_n^k converges to a unique fixed point if we consider $\phi_{n,[i]}^k$ and $\overline{\mathbf{T}}_n^k$ fixed for $i \in \mathcal{A}_n,\ n \in \mathcal{N},\ k \in \mathcal{K}$, $A_n = 2\ \forall n$, a high SNR approximation and if

$$(N-1) \max_{\substack{j,n,k \\ j \neq k}} \left\{ \frac{\mathrm{tr}\{\mathrm{adj}\{\overline{\mathbf{H}}_{j,n}^k\}\overline{\mathbf{H}}_{n,n}^k\}}{2|\overline{\mathbf{H}}_{n,n}^k|} \right\} \leq 1. \tag{47}$$

Here $\overline{\mathbf{H}}_{n,j}^k = \mathbf{H}_{n,j}^k \mathbf{E}_j^k \overline{\mathbf{T}}_j^k (\overline{\mathbf{T}}_j^k)^{\mathrm{H}} \mathbf{E}_j^k (\mathbf{H}_{n,j}^k)^{\mathrm{H}}$.

The following proposition treats the convergence of the $\phi_{n,[i]}^k$'s.

Proposition 1. Consider a two user case where η_n^k, $\overline{\mathbf{T}}_n^k$ are fixed $\forall k$ and $A_n = 2\ \forall n$. The update for the angles $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ converges to a critical point in a finite number of steps.

We leave the proof of this proposition to Appendix B.

The signal coordination part is guaranteed to converge if $\mathbf{P}$ is fixed. See, e.g. [18, 19].

To summarize our results about convergence: For a two user case where each user has two transceivers, consider the three variables of the problem as $\overline{\mathbf{T}}_n^k$, η_n^k and ϕ_n^k. We have partial convergence results showing the convergence of each one when the other two are held fixed. For the general case, we always observe that the whole procedure converges—some results are shown in the next section. We performed hundreds of experiments, and this was always the case.

5. Simulation results

5.1. ADSL downstream

We simulate a system with cables of 0.5 mm (AWG 24) and an SNR gap of 12.8 dB. We consider the near-far downstream ADSL scenario depicted in Fig. 3, with two users, one with three and one with two transceivers (all direct modes). A good solution is one in which user 2 spares user 1 from excessive interference. To dramatize the effect of the per-transceiver PCs, we consider the case where some lines have a 'detour', i.e. they start and end on the same node but their line lengths are not the same. This effect may represent practical unevenness between the pairs due to different patching applied to the different pairs, and due to differences in the quality and integrity of the pairs and connectors. Referring to Fig. 3, we define $\mathbf{l}_n = \begin{bmatrix} l_{n,[1]} & \cdots & l_{n,[A_n]} \end{bmatrix}$ and set $\mathbf{l}_1 = \begin{bmatrix} 4.8 & 4.5 & 4.2 \end{bmatrix}$ km and $\mathbf{l}_2 = \begin{bmatrix} 3.3 & 3 \end{bmatrix}$ km. We also use $d_2 = 3$ km. Each transceiver has a total power budget of 20.4 dBm (equal to 109.64 mW). For each line, noise model ANSI A is adopted.

In Fig. 4 we depict the rate regions of the WMMSE-GDSB with per-user PCs (i.e. the algorithm in [13]), the DMT-WMMSE with per-transceiver PCs (i.e. Algorithm 1 in this paper) and the WMMSE-GDSB with per-transceiver PCs (i.e. Algorithm 2 in this paper). Not surprisingly, the algorithm with the per-user constraints does better through all the rate region. The algorithms with per-transceiver PCs are more limited in their solution space, and this has an impact on the performance. However, as we mentioned in Section 1, the per-transceiver PCs are in many situations more realistic.

Both algorithms that consider the per-transceiver PCs perform equally well. When they exist, the small differences in performance are due to a badly chosen starting point. However, in terms of convergence speed the advantage is clearly with the WMMSE-GDSB (a fact that was already mentioned in [13]). In Fig. 5 we illustrate the weighted rate sum (WRS) after each iteration of both algorithms. As we can see, the DMT-WMMSE converges much slower. It needs approximately 800 iterations to converge, while the WMMSE-GDSB needs less than 70. Because of its faster convergence, we observe that, although more complex, the WMMSE-GDSB takes most of the time less time to finish in comparison to the DMT-WMMSE. One iteration of the former takes approximately 10 seconds (we use $Q = 100$) while one iteration of the latter takes approximately 1 seconds. The faster convergence of the WMMSE-GDSB also means advantages in terms of how fast the algorithm can respond to changing channel conditions, specially if the algorithms are implemented in a distributed way with message exchanges.

In Fig. 6 and 7 we illustrate the final PSD's of the WMMSE-GDSB with per-user PCs and with per-transceiver PCs. We observe that, for user 1, the former allocates excessive power on the third transceiver—it is the one connected to the shortest line. The first transceiver, with the longest line, receives little power. This makes the transceivers uneven. For the result in the figure, we obtain a total power of approximately 98.15, 110.79 and 119.98 mW, respectively, for the transceivers of user 1. That constitutes a breach of the

PCs for second and third transceivers. For user 2, these numbers are 104.62 and 114.67 mW, which violates the PC of the second transceiver. The achieved data rates are approximately $r_1 = 8.0$ Mb/s and $r_2 = 13.74$ Mb/s. For the situation in Fig. 7, all transceivers of both users satisfy the PC. The achieved data rates for the WMMSE-GDSB per-Tx PC are approximately $r_1 = 8$ Mb/s and $r_2 = 13.69$ Mb/s.

5.2. VDSL upstream

We also simulate a VDSL upstream scenario. Cables of 0.4 mm (AWG 26) are used and an SNR gap of 9.45 dB is considered. Each user has a total power budget of 11.5 dBm. For each line, noise model ETSI A is adopted with a background noise level of -140 dBm/Hz. We use the FDD 998 frequency bandplan over POTS up to 12 MHz.

We consider three users, the first two with 3 transceivers and the third with two. In this experiment, we want to assess the effect of the 'detours' in the solutions for both the per-user and the per-transceiver PC cases. We define, just as for the ADSL scenario, $\mathbf{l}_1 = \begin{bmatrix}1.2+\beta & 1.2 & 1.2-\beta\end{bmatrix}$ km, $\mathbf{l}_2 = \begin{bmatrix}0.8+\beta & 0.8 & 0.8-\beta\end{bmatrix}$ km and $\mathbf{l}_3 = \begin{bmatrix}0.6+\beta & 0.6-\beta\end{bmatrix}$, where $\beta \geq 0$ introduces some imbalance between the DSL lines. We simulate the system for 4 distinct values of β, ranging from 0 to 75 m, and calculate the solutions of the WMMSE-GDSB with per-user PCs (i.e. the algorithm in [13]) and with per-transceiver PCs (i.e. the algorithm discussed in Section 4 in this paper). We depict the respective WRS as a function of β in Fig. 8.

For perfectly even lines (i.e. $\beta = 0$), both algorithms return the same result. As β increases, so does the disparity between the lines. The algorithm with per-user PCs is able to load more power on the shorter lines, thus increasing its WRS, while the algorithm with the per-transceiver PCs cannot explore this effect.

6. Conclusion

In this paper, we have focused on the DMT MIMO IC WRS maximization problem with per-transceiver PCs. Two solutions were proposed. The first builds on the recent work showing the relation between the WRS maximization problem and the MMSE minimization problem. The algorithm, called DMT-WMMSE with per-transceiver PCs, solves the signal and spectrum coordination parts of the problem simultaneously. The second proposal, called WMMSE-GDSB with per-transceiver PCs, separates these two parts. For the spectrum coordination part, we observe that the problem can be solved more efficiently with a change of variables. By using spherical coordinates in taxicab geometry (i.e. in the form of $\eta_n^k \mathbf{v}_n^k$, with $\left\|\mathbf{v}_n^k\right\|_1 = 1$), we obtain for the radius η_n^k a concave problem. For the angles, it is shown that they can be solved independently.

Both algorithms achieve good results. It is observed, however, that the WMMSE-GDSB with per transceiver PCs converges much faster.

Appendix A. General formulas for $\boldsymbol{\rho_n^k}$, $\boldsymbol{\theta_{n,[i]}^k}$ and $\boldsymbol{\eta_n^k}$, $\boldsymbol{\phi_{n,[i]}^k}$

To re-write $\mathbf{p}_n^k \in \mathbb{R}^{A_n}$ as (24) for the general case, we have $\mathbf{p}_n^k = \rho_n^k \mathbf{d}_n^k$, where

$$d_{n,[1]}^k = \prod_{j=1}^{A_n-1} \cos\theta_{n,[j]}^k \tag{A.1}$$

$$d_{n,[i]}^k = \prod_{j=i}^{A_n-1} \cos\theta_{n,[j]}^k \sin\theta_{n,[i-1]}^k, \quad i = 2,\ldots,A_n-1 \tag{A.2}$$

$$d_{n,[A_n]}^k = \sin\theta_{n,[A_n-1]}^k \tag{A.3}$$

Here $0 \le \theta_{n,[i]}^k \le \pi/2,\ \forall i$. To find $\theta_{n,[i]}^k$ from $\mathbf{p}_n^k$, we write

$$\theta_{n,[i]}^k = \arctan\left(\frac{p_{n,(i+1)}^k}{\sqrt{(p_{n,1}^k)^2 + \cdots + (p_{n,i}^k)^2}}\right), \quad i = 1,\ \ldots,\ A_n - 1. \tag{A.4}$$

The radius is given by $\rho_n^k = \left\|\mathbf{p}_n^k\right\|_2$.

To re-write $\mathbf{p}_n^k$ as (32) for the general case, we have $\mathbf{p}_n^k = \eta_n^k \mathbf{v}_n^k$, where

$$v_{n,[1]}^k = \prod_{j=1}^{A_n-1} (1 - \phi_{n,[j]}^k) \tag{A.5}$$

$$v_{n,[i]}^k = \prod_{j=i}^{A_n-1} (1 - \phi_{n,[j]}^k)\phi_{n,[i-1]^k}, \quad i = 2,\ldots,A_n-1 \tag{A.6}$$

$$v_{n,[A_n]}^k = \phi_{n,[A_n-1]}^k \tag{A.7}$$

Here $0 \le \phi_{n,[i]}^k \le 1\ \forall i$. To re-write $\mathbf{p}_n^k$ as (32), we calculate $\eta_n^k = \sum_{i\in\mathcal{A}_n} p_{n,(i)}^k$. To find $\phi_{n,[i]}^k$, we first define $\mathbf{v}_n^k = \mathbf{p}_n^k/\eta_n^k$ then straightforwardly calculate each $\phi_{n,[i]}^k$.

Appendix B. Proof of Proposition 1

Since $A_n = 2$ for all users, we can drop one of the subscripts for ϕ_n^k. We will refer to $\boldsymbol{\phi}_1 = [\phi_1^1 \ \cdots \ \phi_1^K]^{\mathrm{T}}$ and $\boldsymbol{\phi}_2 = [\phi_2^1 \ \cdots \ \phi_2^K]^{\mathrm{T}}$. For simplicity, we assume that $\sum_{k\in\mathcal{K}} \eta_n^k = 2P,\ n = 1,2$ and that the per-transceiver PCs are given by P. The optimization to be solved in $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ is given by

$$\begin{aligned} \max_{\boldsymbol{\phi}_1,\ \boldsymbol{\phi}_2} \quad & \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} u_n b_n^k \\ \text{subject to} \quad & \sum_{k\in\mathcal{K}} \eta_1^k \phi_1^k \le P, \quad 0 \le \phi_1^k \le 1 \\ & \sum_{k\in\mathcal{K}} \eta_2^k \phi_2^k \le P, \quad 0 \le \phi_2^k \le 1 \end{aligned} \tag{B.1}$$

As explained in Section 4.4, we solve this problem with a block coordinate descent method. These constraints are convex and the feasible set of the problem is now the cartesian product $\mathcal{B}_1 \times \mathcal{B}_2$, where

$\mathcal{B}_n = \{\boldsymbol{\phi}_n \mid \sum_{k\in\mathcal{K}} \eta_n^k \phi_n^k \leq P,\ \phi_n^k \in [0,1]\}$. We solve sequentially, i.e. $n = 1$, $n = 2$ and repeat until convergence, the following problem:

$$\boldsymbol{\phi}_n = \text{argmax}_{\boldsymbol{\phi}_n \in \mathcal{B}_n} \sum_{n\in\mathcal{N}} \sum_{k\in\mathcal{K}} u_n b_n^k. \tag{B.2}$$

Because of the dual decomposition and the exhaustive search, each step returns the optimal solution. Since b_n^k is continuously differentiable and the feasible set of the problem is the cartesian product of non-empty, convex subsets, the block coordinate descent with two blocks is guaranteed to converge to a critical point [33].

Acknowledgment

This research work was carried out at the ESAT Laboratory of the KU Leuven, in the frame of the KU Leuven Research Council CoE EF/05/006 Optimization in Engineering (OPTEC) and PFV/10/002 (OPTEC); Concerted Research Action GOA-MaNet; IUAP P7/23 (Belgian network on stochastic modeling, analysis, design and optimization of communication systems, BESTCOM, 2012-2017); and research project FWO nr. 6.091213N, 'Cross-layer optimization with real-time adaptive dynamic spectrum management for fourth generation broadband access networks'. The scientific responsibility is assumed by its authors.

P. Tsiaflakis is also a postdoctoral fellow funded by the Research Foundation—Flanders (FWO).

References

[1] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, T. Bostoen, Optimal multiuser spectrum balancing for digital subscriber lines, IEEE Trans. Commun. 54 (5) (2006) 922–933.

[2] R. Cendrillon, J. Huang, M. Chiang, M. Moonen, Autonomous spectrum balancing for digital subscriber lines, IEEE Trans. Signal Process. 55 (8) (2007) 4241–4257.

[3] J. Papandriopoulos, J. S. Evans, SCALE: a low-complexity distributed protocol for spectrum balancing in multiuser DSL networks, IEEE Trans. Inf. Theory 55 (8) (2009) 3711–3724.

[4] P. Tsiaflakis, M. Diehl, M. Moonen, Distributed spectrum management algorithms for multiuser DSL networks, IEEE Trans. Signal Process. 56 (10) (2008) 4825–4843.

[5] W. Yu, G. Ginis, J. M. Cioffi, Distributed multiuser power control for digital subscriber lines, IEEE J. Sel. Areas of Commun. 20 (5) (2002) 1105–1115.

[6] R. B. Moraes, B. Dortschy, A. Klautau, J. R. i Riu, Semiblind spectrum balancing for DSL, IEEE Trans. Signal Process. 58 (7) (2010) 3717–3727.

[7] M. Wolkerstorfer, J. Jaldén, T. Nordström, Column generation for discrete-rate multi-user and multi-carrier power control, IEEE Trans. Commun. 60 (9) (2012) 2712–2722.

[8] R. Cendrillon, G. Ginis, E. Van den Bogaert, M. Moonen, A near-optimal linear crosstalk canceler for upstream VDSL, IEEE Trans. Signal Process. 54 (8) (2006) 3136–3146.

[9] R. Cendrillon, G. Ginis, E. Van den Bogaert, M. Moonen, A near-optimal linear crosstalk precoder for downstream VDSL, IEEE Trans. Commun. 55 (5) (2007) 860–863.

[10] G. Ginis, J. M. Cioffi, Vectored transmission for digital subscriber line systems, IEEE J. Sel. Areas Commun. 20 (5) (2002) 1085–1104.

[11] B. Lee, J. M. Cioffi, S. Jagannathan, M. Mohseni, Gigabit DSL, IEEE Trans. Commun. 55 (9) (2007) 1689–1692.

[12] P. Tsiaflakis, J. Vangorp, J. verlinden, M. Moonen, Multiple access channel optimal spectrum balancing for upstream DSL transmission, IEEE Commun. Lett. 11 (4) (2007) 398–400.

[13] R. B. Moraes, P. Tsiaflakis, J. Maes, M. Moonen, DMT MIMO IC rate maximization in DSL with combined signal and spectrum coordination, IEEE Trans. Signal Process. 61 (7) (2013) 1756–1769.

[14] T. Magesacher, P. Ödling, P. O. Börjesson, T. Nordström, Exploiting the common-mode signal in xDSL, in: European Signal Process. Conf., Vienna, Austria, 2004.

[15] J. M. Cioffi, S. Jagannathan, M. Mohseni, G. Ginis, CuPON: The copper alternative to PON 100 Gb/s DSL networks, IEEE Commun. Mag. 45 (6) (2007) 132–139.

[16] E. F. Krause, Taxicab geometry : an adventure in non-Euclidean geometry, New York (N.Y.): Dover, 1986.

[17] F. Negro, S. P. Shenoy, I. Ghauri, D. T. M. Slock, On the MIMO interference channel, in: Inf. Theory and Appl. Workshop, San Diego, USA, 2010.

[18] S. S. Christensen, R. Agarwal, E. de Carvalho, J. M. Cioffi, Weighted sum-rate maximization using weighted MMSE for MIMO-BC beamforming design, IEEE Trans. Wireless Commun. 7 (12) (2008) 4792–4799.

[19] Q. Shi, M. Razaviyayn, Z.-Q. Luo, C. He, An iteratively weighted MMSE approach to distributed sum-utility maximization for a MIMO interfering broadcast channel, IEEE Trans. Signal Process. 59 (9) (2011) 4331–4340.

[20] S. J. Kim, G. B. Giannakis, Optimal resource allocation for MIMO ad-hoc cognitive radio networks, in: Allerton Conf. Comm., Control and Comput., Urbana-Champaign, USA, 2008.

[21] Z. K. M. Ho, D. Gesbert, Balancing egoism and altruism on the interference channel: The mimo case, in: IEEE Int. Conf. Commun., Cape Town, South Africa, 2010.

[22] V. R. Cadambe, S. A. Jafar, Interference alignment and degrees of freedom of the K-user interference channel, IEEE Trans. Inf. Theory 54 (8) (2008) 3425–3441.

[23] K. Gomadam, V. R. Cadambe, S. A. Jafar, Approaching the capacity of wireless networks through distributed interference alignment, in: IEEE Global Telecomm. Conf., New Orleans, USA, 2008.

[24] J. Starr, O. El Ayach, R. W. Heath, Jr., Interference alignment with per-antenna power constraints, in: IEEE Int. Symp. Inf. Theory, S. Petersburg, Russia, 2011.

[25] M. Hong, Z.-Q. Luo, Signal processing and optimal resource allocation for the interference channel, E-Reference Signal Processing, to appear in.
URL `http://arxiv.org/abs/1206.5144v1`

[26] W. Yu, T. Kwon, C. Shin, Multicell coordination via joint scheduling, beamforming and power spectrum adaptation, in: IEEE INFOCOM, Shanghai, China, 2011.

[27] L. Venturino, N. Prasad, X. Wang, Coordinated linear beamforming in downlink multi-cell wireless networks, IEEE Trans. Wireless Commun. 9 (4) (2010) 1451–1461.

[28] H. Bölcskei, D. Gesbert, A. J. Paulraj, On the capacity of OFDM-based spatial multiplexing systems, IEEE Trans. Commun. 50 (2) (2002) 225–234.

[29] O. El Ayach, S. W. Peters, R. W. Heath, Jr., The feasibility of interference alignment over measured MIMO-OFDM channels, IEEE Trans. Veh. Technol. 59 (9) (2010) 4309–4321.

[30] J. Shin, J. Moon, , J. Ahn, Weighted-sum-rate-maximizing linear transceiver filters for the K-user MIMO interference channel, in: Global Telecomm. Conf., Houston, USA, 2011.

[31] W. Yu, Multiuser water-filling in the presence of crosstalk, in: Inf. Theory and Appl. Workshop, San Diego, USA, 2007.

[32] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[33] L. Grippo, M. Sciandrone, On the convergence of the block nonlinear Gauss-Seidel method under convex constraints, Operations Research Letters 23 (3) (2000) 127–136.

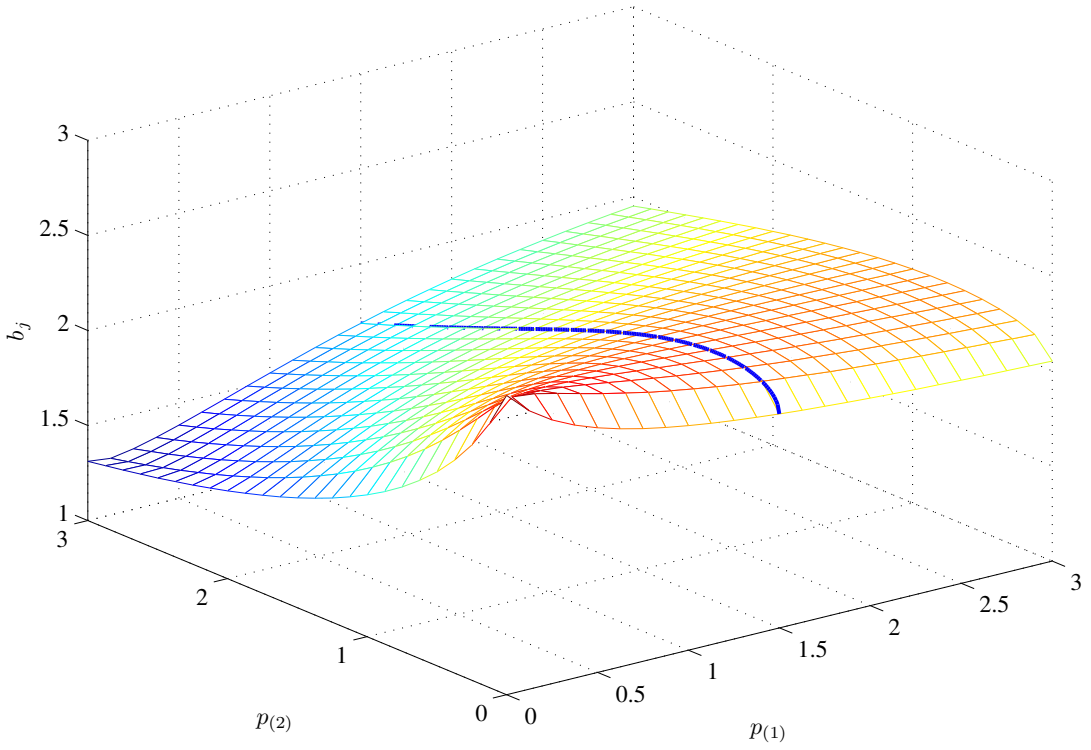

Figure 1: Illustration of b_j as a function of interferer powers $p_{(1)}$ and $p_{(2)}$ (we suppress the subscripts and superscripts for simplicity). All matrices involved, including direct channel, crosstalk channel and transmit matrices, are randomly chosen. The thick line illustrates the behavior of the function for fixed $p_{(1)}$.

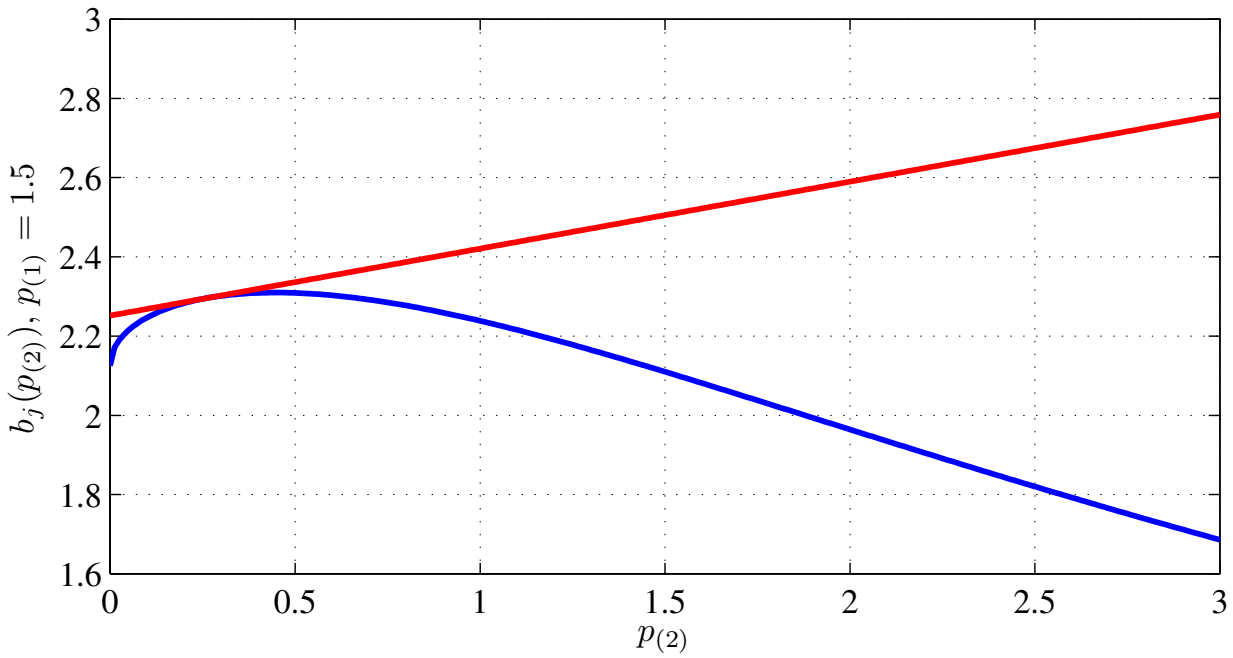

Figure 2: Detail of b_j as a function of $p_{(2)}$ for fixed $p_{(1)} = 1.5$ (we suppress the subscripts and superscripts for simplicity). A first order Taylor expansion around point $p_{(2)} = 0.25$ shows that such approximation is a poor characterization of the behavior of the original function.

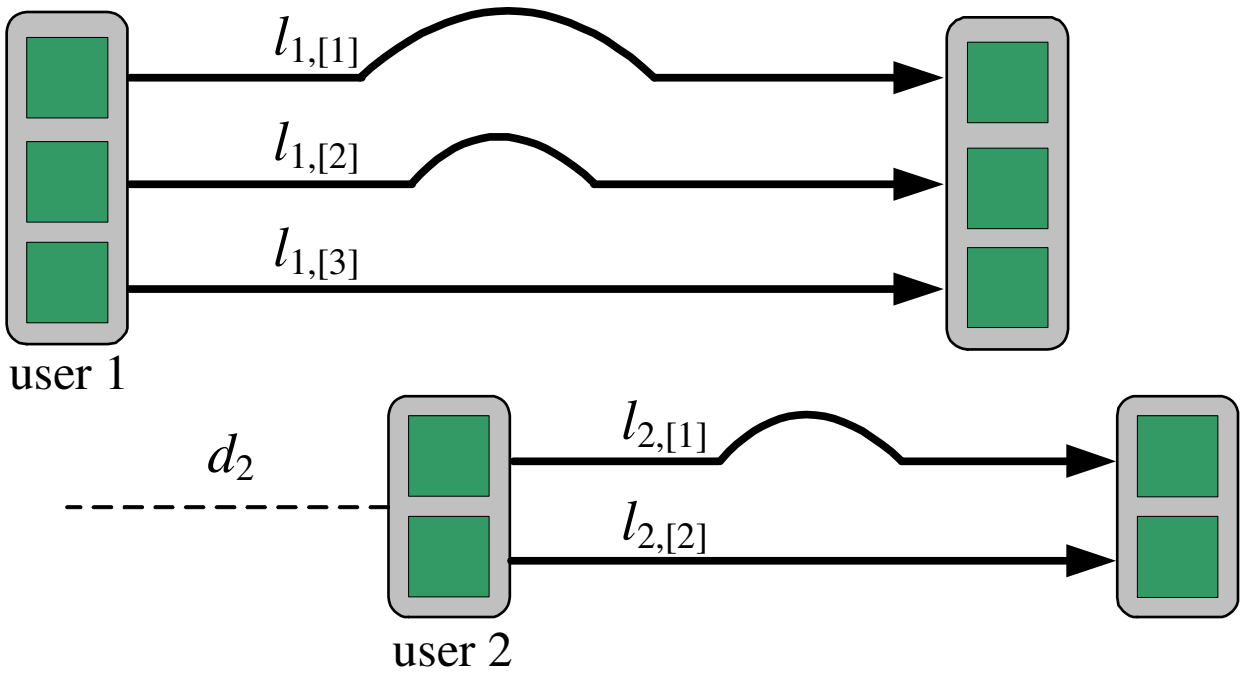

Figure 3: Near-far downstream ADSL scenario.

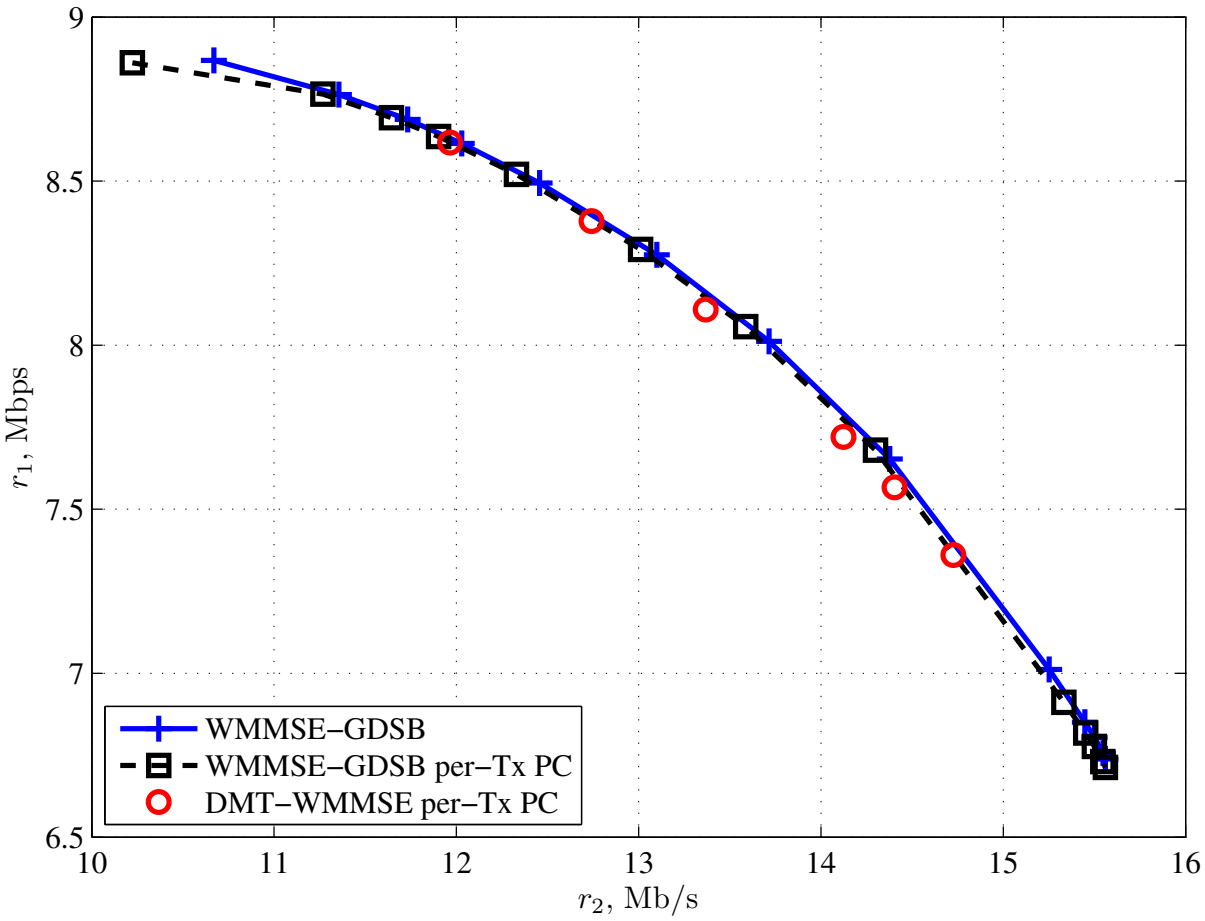

Figure 4: Rate region for the near-far ADSL scenario. We set $\mathbf{l}_1 = [4.8 \;\; 4.5 \;\; 4.2]$ km, $\mathbf{l}_2 = [3.3 \;\; 3]$ km and $d_2 = 3$ km

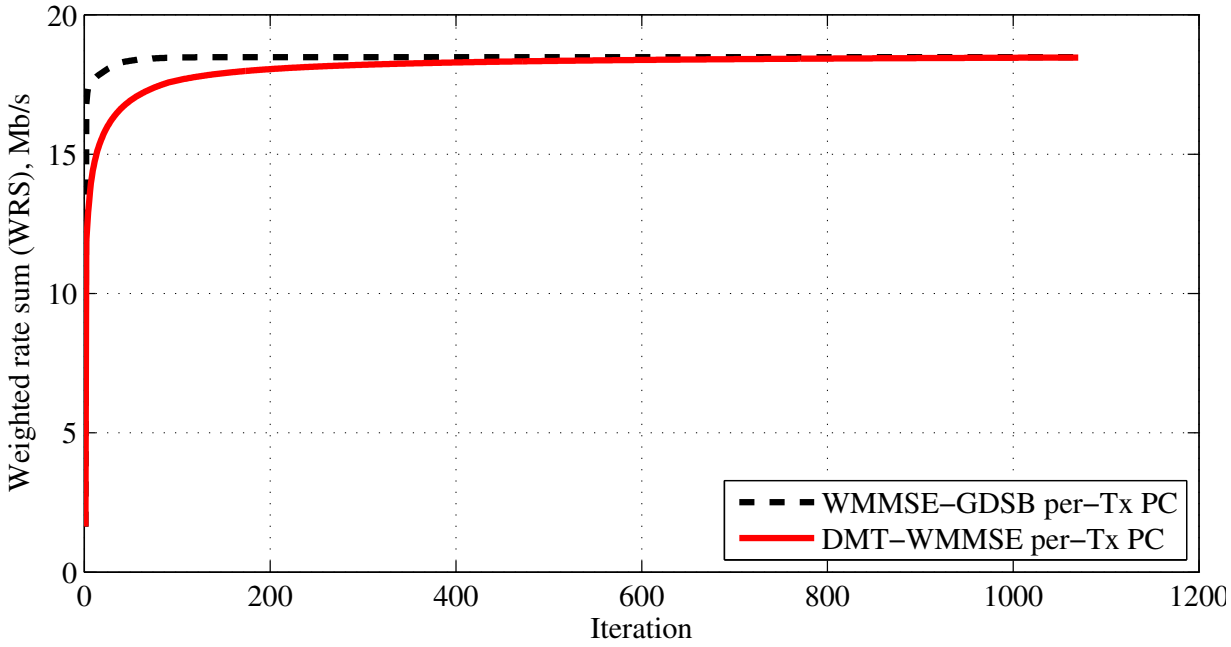

Figure 5: Convergence speed for the WMMSE-GDSB per-Tx PC and the DMT-WMMSE per-Tx PC.

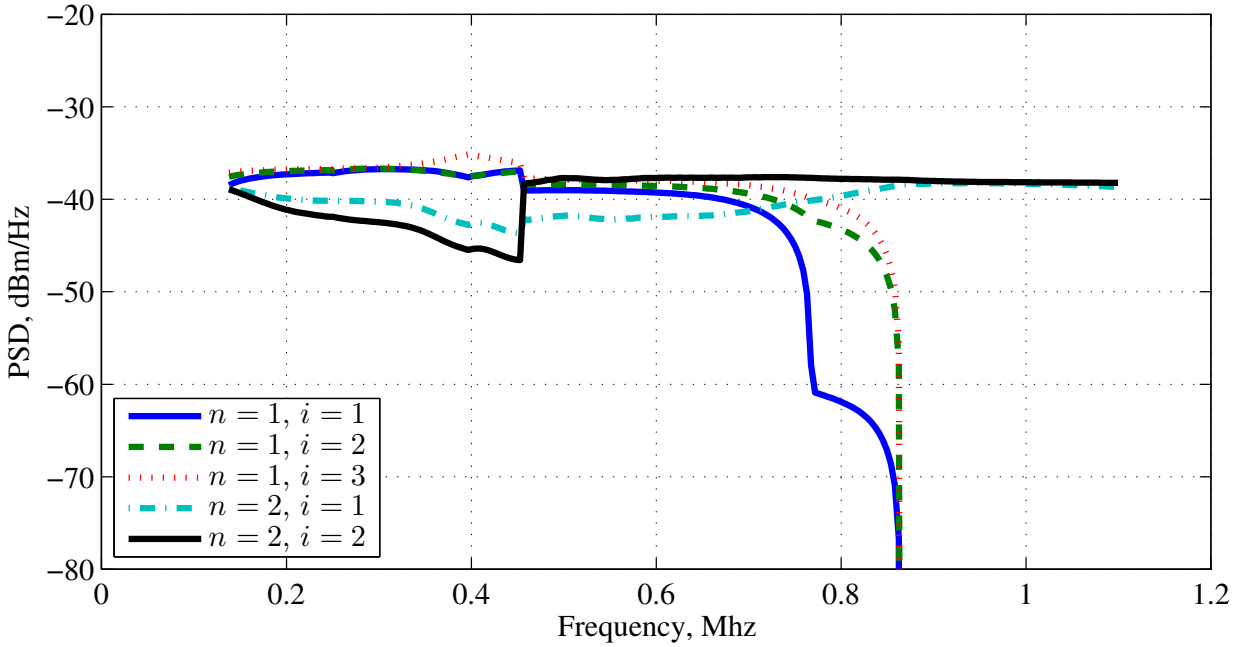

Figure 6: PSD's for the WMMSE-GDSB with per-user PCs.

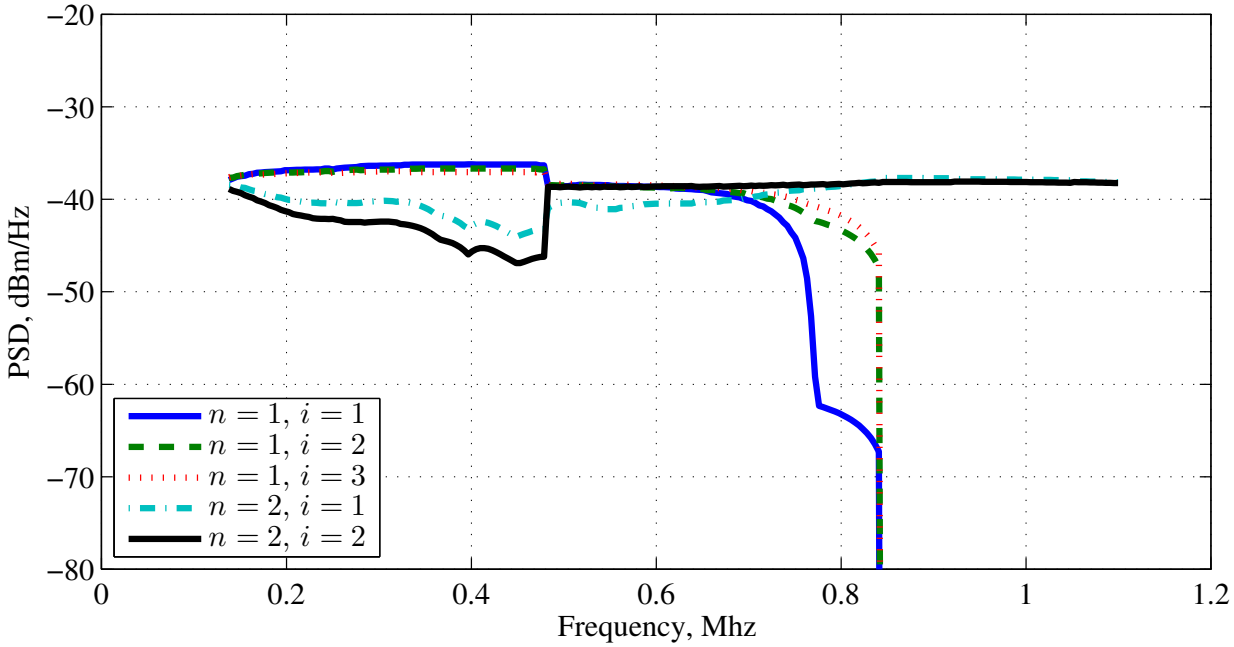

Figure 7: PSD's for the WMMSE-GDSB per-Tx PC.

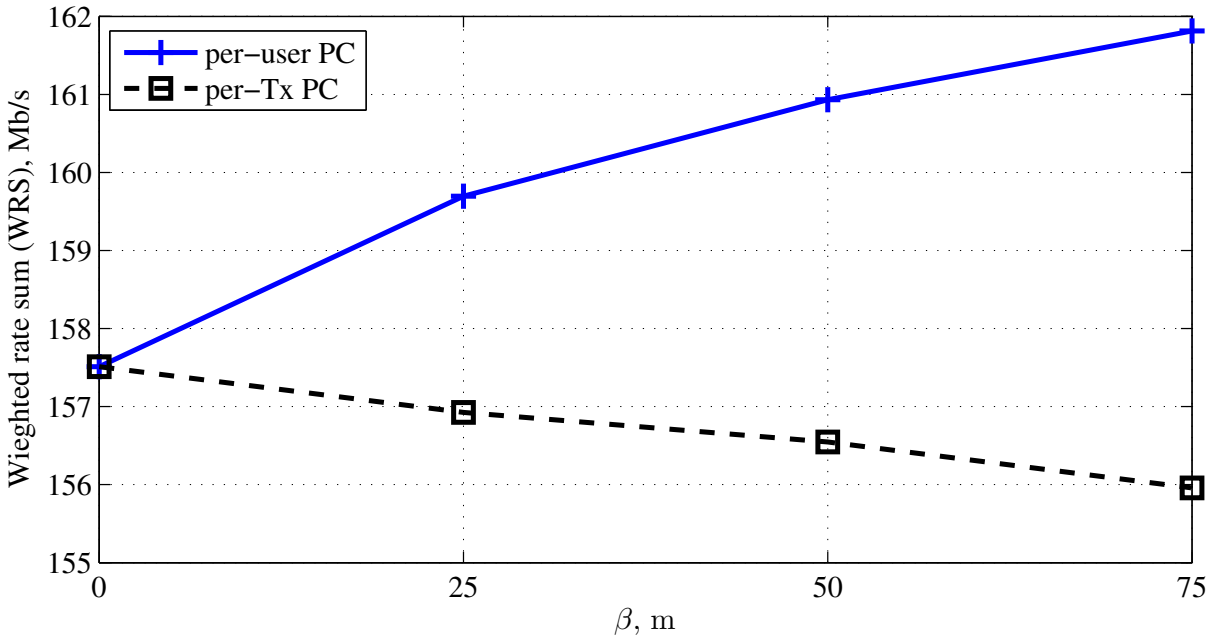

Figure 8: Results for the WMMSE-GDSB with per-user and per-transceiver PCs as a function of β.