Optimal Multiuser Spectrum Balancing for Digital Subscriber Lines

Raphael Cendrillon, Wei Yu, *Member, IEEE*, Marc Moonen, *Senior Member, IEEE*, Jan Verlinden, and Tom Bostoen

***Abstract*—Crosstalk is a major issue in modern digital subscriber line (DSL) systems such as ADSL and VDSL. Static spectrum management, which is the traditional way of ensuring spectral compatibility, employs spectral masks that can be overly conservative and lead to poor performance. This paper presents a centralized algorithm for optimal spectrum balancing in DSL. The algorithm uses the dual decomposition method to optimize spectra in an efficient and computationally tractable way. The algorithm shows significant performance gains over existing dynamis spectrum management (DSM) techniques, e.g., in one of the cases studied, the proposed centralized algorithm leads to a factor-of-four increase in data rate over the distributed DSM algorithm *iterative waterfilling*.**

***Index Terms*—Digital subscriber line (DSL), dual decomposition, dynamic spectrum management (DSM), interference channel, nonconvex optimization.**

I. Introduction

CROSSTALK is a major issue in modern digital subscriber line (DSL) systems such as ADSL and VDSL. Typically 10–20 dB larger than the background noise, crosstalk is *the* dominant source of performance degradation.

Whilst crosstalk cancellation is a commonly proposed solution [1], [2], in many scenarios, this may not be feasible due to complexity issues or as a result of unbundling. In this case, the effects of crosstalk must be mitigated through spectrum management, where the transmit spectra of all modems are limited in some way to minimize crosstalk.

Static spectrum management is the traditional approach and employs identical spectral masks for all modems. To ensure widespread deployment, these masks are based on worst case scenarios [3]. As a result, they can be overly restrictive and lead to poor performance.

Dynamic spectrum management (DSM), a new paradigm, overcomes this problem by designing the spectra of each modem to match the specific topology of the network [4]. These spectra are adapted based on the direct and crosstalk channels seen by the different modems. They are customized to suit each modem in each particular situation.[1]

A DSM algorithm known as *iterative waterfilling* was recently proposed and demonstrates the spectacular performance gains that are possible [6]. An unanswered question at this point is: how much more can be achieved?

The goal of this paper is to address this question. We consider centralized spectrum management where a spectrum management center (SMC) is responsible for setting the spectra of the modems within the network. This paper will present an algorithm for optimal spectrum balancing in the DSL interference channel. Assuming that all modems employ discrete multitone (DMT) modulation, this algorithm achieves the best possible balance between the rates of the different modems in the network, allowing operation at any point on the rate region boundary.

The algorithm is suitable for direct application when an SMC is available. Note that with centralized algorithms, reoptimization is necessary whenever a line is activated or deactivated in order to ensure optimal performance. This is one disadvantage of centralized algorithms with respect to more distributed algorithms such as iterative waterfilling. Furthermore, centralized spectrum management requires an SMC that is not available in unbundled networks where multiple operators share the same binder. In this case, a distributed algorithm may be preferred. Optimal spectrum balancing can be useful here since it provides an upper bound on performance of all DSM algorithms, both centralized and distributed. Furthermore, the spectra generated by the proposed algorithm provide valuable insight that can be used in distributed algorithm design [7].

One may argue, if centralized control is available (via an SMC), then why not implement full-blown crosstalk cancellation? Although crosstalk cancellation can lead to greater performance gains, it is more complex to implement and is not fea-

Paper approved by S. K. Wilson, the Editor for Multicarrier Modulation of the IEEE Communications Society. Manuscript received April 12, 2004; revised July 22, 2004, and May 11, 2005. This work was carried out in the frame of IUAP P5/22, Dynamical Systems and Control: Computation, Identification and Modeling; and P5/11, Mobile Multimedia Communication Systems and Networks; the Concerted Research Action GOA-MEFISTO-666, Mathematical Engineering for Information and Communication Systems Technology; the IWT SOLIDT Project, Solutions for xDSL Interoperability, Deployment and New Technologies; and FWO Project G.0196.02, Design of Efficient Communication Techniques for Wireless Time-Dispersive Multiuser MIMO Systems; and was supported in part by Alcatel Bell. The work of W. Yu was supported in part by Bell Canada University Laboratories, Communications and Information Technology Ontario (CITO), in part by the Natural Science and Engineering Research Council (NSERC) of Canada, and in part by the Canada Research Chairs Program. This paper was presented in part at the IEEE International Communications Conference (ICC), Paris, France, June 2004.

R. Cendrillon was with the School of Information Technology and Electrical Engineering, University of Queensland, Brisbane, QLD 4072 Australia. He is now with the Marvell Hong Kong Ltd., Mongkok, Hong Kong (e-mail: raphael@cendrillon.org).

W. Yu is with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON Canada M5S 3G4 (e-mail: weiyu@comm.utoronto.ca).

M. Moonen is with the Katholieke Universiteit Leuven, Heverlee 3001, Belgium (e-mail: moonen@esat.kuleuven.be).

J. Verlinden and T. Bostoen are with the DSL Expert Team, Alcatel Bell, Antwerp 2018, Belgium (e-mail: jan.vj.verlinden@alcatel.be; tom.bostoen@alcatel.be).

Digital Object Identifier 10.1109/TCOMM.2006.873096

[1]Recent standardization activities [5] have broadened the scope of DSM to include not only spectra adaptation, but also signal processing algorithms such as crosstalk cancellation and vectoring. This paper uses the term DSM in the narrowly defined sense.

sible when head-end modems are not co-located in the same central office (CO) or remote terminal (RT). Spectrum management, on the other hand, only requires adaptation of the modem transmit spectra. This can be done without any change to the modem hardware currently deployed in the field and is feasible to implement right now. In contrast, crosstalk cancellation uses signal-level coordination, thus requiring an entirely new design of both the DSL access multiplexer (DSLAM) and customer premises (CP) modems.

Under several specific scenarios, crosstalk cancellation is possible even without signal coordination [8]. Whilst performance gains are possible, these techniques are again typically complex. The remainder of this paper assumes that crosstalk cancellation is not performed, and each modem treats crosstalk as additive noise.

The multiuser DSL channel with no signal-level coordination is an example of an interference channel in multiuser information theory. The capacity region and the optimal code design for the interference channel are long-standing open problems in information theory. This paper considers an achievable rate region for the interference channel within the context of currently deployed DSL modems in the field. In this case, interference must be treated as noise, and the optimization of the achievable rate region is reduced to the optimization of the joint spectra amongst all of the users. The solution obtained using the optimal spectrum balancing algorithm proposed in this paper, although not the best possible for the interference channel, is nevertheless optimal within the current capabilities of DSL modems already developed.

The main difficulty in the optimal design of multiuser DSL spectra is the computational complexity associated with the optimization problem. The constrained optimization problem is nonconvex, and a naive exhaustive search leads to an exponential complexity in the number of tones K in the system. In ADSL $K = 256$, whilst in VDSL $K = 4096$. This leads to a computationally intractable problem.

This paper overcomes the exponential complexity in K through the use of a technique called dual decomposition. The computational complexity of the proposed algorithm, although linear in K, is still exponential in the number of users. Nevertheless, it gives a practical way to compute the achievable rate regions for channels with a small number of users. Doing so was not possible prior to this work.

A related work [9] formulated a solution to the optimal balancing problem based on simulated annealing. However, simulated annealing cannot guarantee convergence to the global optimum, and the convergence speed can also be slow. Another paper tries to find the optimal solution in closed form, using convex optimization techniques [10]. Unfortunately, this is only possible when the crosstalk channels are sufficiently weak such that the objective function is convex. This approach is not valid in the general case. In particular, in mixed CO-RT distributions and in VDSL, the crosstalk channels may be stronger than the direct channels. The result is a nonconvex objective function with many local optima. Exploring all local optima requires an exponential complexity in the number of tones K, and the optimization becomes computationally intractable [10]. Other suboptimal solutions, both distributed [6], [11], [12] and centralized [13], have also been proposed.

The remainder of this paper is organized as follows. The system model for a network of interfering DSL modems is formulated in Section II. The problem is then to characterize the achievable rate region and the corresponding transmit spectra. This problem is formulated in Section III, where it is shown that trying to find the solution directly through an exhaustive search is computationally intractable. Section IV shows that the spectrum balancing problem has an equivalent dual problem. This can be decomposed into separate subproblems that are then solved independently on each tone. The resulting algorithm gives an efficient solution to the spectrum balancing problem. Section V compares the performance of the proposed algorithm to existing spectrum balancing techniques. Conclusions are drawn in Section VI.

II. System Model

Assuming that DMT modulation is employed by all modems, transmission can be modeled independently on each tone as

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

The vector $\mathbf{x}_k \triangleq [x_k^1, \cdots, x_k^N]$ contains transmitted signals on tone k. There are N lines in the binder, and x_k^n is the signal transmitted onto line n at tone k. $\mathbf{y}_k$ and $\mathbf{z}_k$ have similar structures. $\mathbf{y}_k$ is the vector of received signals on tone k. $\mathbf{z}_k$ is the vector of additive noise on tone k and contains thermal noise, alien crosstalk, single-carrier modems, and radio frequency interference (RFI). Recall that $1 \leq k \leq K$, where K is the number of tones within the system. We denote the noise power spectral density (PSD) on line n as $\sigma_k^n \triangleq \mathcal{E}\{|z_k^n|^2\}$. $\mathbf{H}_k$ is the $N \times N$ channel transfer matrix on tone k. $h_k^{n,m} \triangleq [\mathbf{H}_k]_{n,m}$ is the channel from TX m to RX n on tone k. The diagonal elements of $\mathbf{H}_k$ contain the direct channels whilst the off-diagonal elements contain the crosstalk channels. We denote the transmit PSD as $s_k^n \triangleq \mathcal{E}\{|x_k^n|^2\}$. For convenience, we denote the vector containing the PSD of user n on all tones as $\mathbf{s}_n \triangleq [s_1^n, \ldots, s_K^n]$. We denote the tone spacing as Δ_f and DMT symbol rate as f_s.

We assume that the modems within the network are synchronized such that each tone operates independently and free from intercarrier interference (ICI). We also assume that each modem employs frequency-division duplexing to separate upstream and downstream transmission.

We denote the maximum bitloading that a modem can support as $b_{\max}$, which lies in the range 8–15 in current standards [14], [15]. It is assumed that each modem treats the interference from other modems as noise. When the number of interfering modems is large, the interference is well approximated by a Gaussian distribution. Under this assumption, the achievable bitloading of user n on tone k is

$$b_k^n \triangleq \min\left(b_{\max}, \log_2\left(1 + \frac{1}{\Gamma}\frac{|h_k^{n,n}|^2 s_k^n}{\sum_{m\neq n}|h_k^{n,m}|^2 s_k^m + \sigma_k^n}\right)\right) \tag{2}$$

where Γ is the SNR gap to capacity and is a function of the desired bit error rate (BER), coding gain, and noise margin [16]. The data rate on line n is thus

$$R_n = f_s \sum_k b_k^n.$$

In practice, the relationship between the received signal-to-interference-plus-noise ratio (SINR) and the bitrate may be more complex and is in fact dependent on the coding scheme employed within the modem. In this paper, (2) will be used for simplicity, however, the algorithms presented here can be applied to any arbitrary function that relates the bitloading to the SINR on each tone.

III. Spectrum Management

A. Spectrum Management Problem

We restrict our attention to the two-user case for ease of explanation. Extensions to more than two users follow naturally from what is presented here. The spectrum management problem for the two-user case is defined as

$$\max_{\mathbf{s}_1,\mathbf{s}_2} R_2 \quad \text{s.t.} \quad R_1 \geq R_1^{\text{target}}. \tag{3}$$

The rate region is a plot of all possible operating points or rate combinations that can be achieved in a multiuser channel. Operating points on the boundary of the region are said to be optimal. These points and their corresponding PSD combinations can be characterized by solving the spectrum management problem (3) for a range of values of R_1^{target}. This is the goal of this paper.

B. Constraints

The optimization (3) is typically subject to a *total power constraint* on each modem

$$\Delta_f \sum_k s_k^n \leq \overline{P}_n, \quad n = 1, 2$$

where $\overline{P}_n$ denotes the total power that modem n can transmit. This arises from limitations on each modem's analog front-end (AFE). For convenience, this is reformulated as

$$\sum_k s_k^n \leq P_n, \quad n = 1, 2 \tag{4}$$

where $P_n \triangleq \overline{P}_n / \Delta_f$. A positivity constraint applies to the transmit spectra

$$s_k^n \geq 0 \quad \forall k, n. \tag{5}$$

Spectral mask constraints may also apply

$$s_k^n \leq s_k^{n,\text{mask}} \quad \forall k, n. \tag{6}$$

C. Continuous Bitloading

Consider the case where the modems can support any possible bitloading. We denote the accuracy with which modems can control their transmit PSD as Δ_s. In current standards, Δ_s is set to 0.5 dBm/Hz [17]. The total power (4) and spectral mask constraints (6) make it possible to upper bound the transmit power on any tone

$$s_k^n \leq s_k^{\text{n,max}}$$

where $s_k^{n,\max} \triangleq \min(P_n, s_k^{n,\text{mask}})$. Combining this with the positivity constraint (5) limits the range of s_k^n to

$$s_k^n \in \{0, \Delta_s, \ldots, s_k^{n,\max}\}. \tag{7}$$

Thus, on tone k, there are $q_k = \prod_n (s_k^{n,\max} \Delta_s^{-1} + 1)$ possible PSD combinations.

D. Discrete Bitloading

Practical DSL modems can only support a fixed set of discrete bitloadings. The search space can thus be reduced to the PSD's corresponding to these discrete bitloadings, thereby reducing complexity without affecting optimality. Define the vectors $\mathbf{b}_k \triangleq [b_k^1, \ldots, b_k^N]^T$ and $\overline{\mathbf{s}}_k \triangleq [s_k^1, \ldots, s_k^N]^T$ which contain the bitloadings and PSDs of all users on tone k, respectively. Provided that $b_k^n \leq b_{\max}, \forall n$, then (2) can be used to find the PSD combination $\overline{\mathbf{s}}_k$ corresponding to a particular bitloading combination $\mathbf{b}_k$, as will now be shown. First, define $\mathbf{A}_k \triangleq \alpha_k^{n,m}$, where

$$\alpha_k^{n,m} \triangleq \begin{cases} 0, & n = m \\ \Gamma \, |h_k^{n,m}|^2, & n \neq m. \end{cases}$$

Also, define $\boldsymbol{\sigma}_k \triangleq \Gamma[\sigma_k^1, \ldots, \sigma_k^N]^T$, $\mathbf{D}_k \triangleq \text{diag}\{|h_k^{1,1}|^2, \ldots, |h_k^{N,N}|^2\}$, and $\Lambda_k \triangleq \text{diag}\{2^{b_k^1} - 1, \ldots, 2^{b_k^N} - 1\}$. Since $b_k^n \leq b_{\max}$, (2) can be rewritten as

$$|h_k^{n,n}|^2 \, s_k^n - \Gamma \left(2^{b_k^n} - 1\right) \sum_{m \neq n} |h_k^{n,m}|^2 \, s_k^m = \Gamma \left(2^{b_k^n} - 1\right) \sigma_k^n \quad \forall n. \tag{8}$$

Note that taking (8) for each n forms a set of n linear equations in $\overline{\mathbf{s}}_k$. These can be written in matrix form as

$$(\mathbf{D}_k - \Lambda_k \mathbf{A}_k)\overline{\mathbf{s}}_k = \Lambda_k \boldsymbol{\sigma}_k.$$

The PSD combination required to support a particular bitloading combination $\mathbf{b}_k$ is then

$$\overline{\mathbf{s}}_k = (\mathbf{D}_k - \Lambda_k \mathbf{A}_k)^{-1} \Lambda_k \boldsymbol{\sigma}_k. \tag{9}$$

In the remainder of this paper, $s_k^n(b_k^1, b_k^2)$ is used to denote the PSD of user n corresponding to the bitloadings b_k^1 and b_k^2, as calculated by (9). Hence, the range of PSD combinations (s_k^1, s_k^2) can be limited to

$$\left(s_k^1, s_k^2\right) \in \left\{ \left(s_k^1\left(b_k^1, b_k^2\right), s_k^2\left(b_k^1, b_k^2\right)\right) | b_k^n \in \{0, \dots b_{\max}\} \quad \forall n \right\}. \tag{10}$$

Thus, on tone k, there are $q_k = (b_{\max} + 1)^2$ possible PSD combinations.

E. Exhaustive Search

At this point, a simplistic algorithm could be proposed to find the optimal PSDs based on an exhaustive search. On tone k, there are q_k possible PSD combinations. Taking all possible PSD levels across all tones results in $\prod_k q_k$ possible PSD combinations. The feasibility of each PSD combination is determined based on any power constraints, as described in Section III-B, and on the target rate constraint for user 1. The PSD combination that maximizes the data rate of user 2 is then selected.

Unfortunately, although this algorithm is simple to implement, its complexity in the discrete bitloading case is $\mathcal{O}((b_{\max}+1)^{2K})$. With $K = 256$ in ADSL and $K = 4096$ in VDSL, this results in a computationally intractable problem. In the continuous bitloading case, the complexity can be even higher.

IV. Optimal Spectrum Balancing

As was shown in Section III-E, an exhaustive search for the optimal PSDs leads to a computationally intractable problem. The reason behind this is as follows. The target rate constraint for the first line and the total power constraint associated with each line couples the power allocation problem across frequency. As such, the PSD combination must be searched in a joint fashion across all tones. This results in an exponential complexity in K and an intractable problem.

The following sections will use dual decomposition to transform this problem into an equivalent one that has a linear complexity in K and is tractable. Since the development has many stages, a brief overview is given here before proceeding with a detailed explanation in Sections IV-A to IV-D.

Section IV-A begins by replacing the original optimization problem (3) with the maximization of a weighted rate sum. With a correctly chosen weight w, maximizing the weighted rate sum implicitly enforces the target rate constraint on user 1.

In Section IV-B, the power-constrained optimization is replaced by an equivalent dual problem. This dual problem consists of an unconstrained optimization of a Lagrangian. In the Lagrangian, the total power constraints are enforced through the use of Lagrangian multipliers which form part of the objective function. When the Lagrangian multipliers are chosen correctly, maximizing the Lagrangian will implicitly enforce the power constraints. The power constraints need not be explicitly enforced, and the problem can be decoupled across frequency.

After this decoupling, the optimization can be solved by maximizing the Lagrangian independently on each tone, an approach which is known as *dual decomposition*. This leads to a complexity which is linear rather than exponential in K, and the problem becomes computationally tractable. This is the main innovation in this paper.

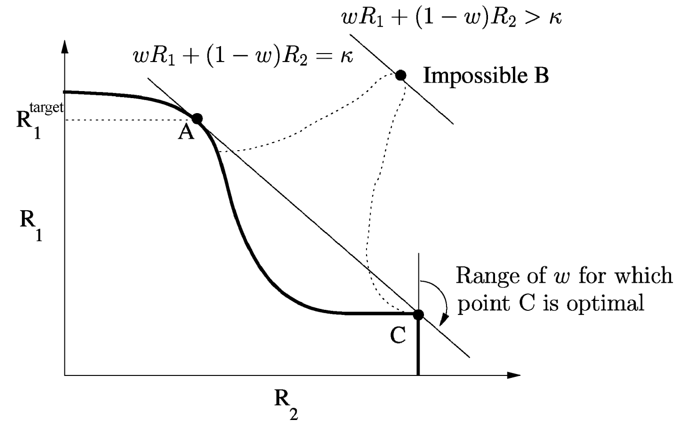


Fig. 1. Optimality of A in the weighted rate sum (11) implies optimality in the spectrum management problem (3).

Section IV-C provides details of the implementation of the algorithm itself, and Section IV-D will show the significant reduction in complexity that it achieves.

The dual-decomposition method is a commonly used approach in convex optimization theory for solving constrained optimization problems through an equivalent unconstrained dual problem. This dual problem can be decomposed into several simpler subproblems. The dual decomposition has been applied in other communication problems with convex objective functions such as joint routing and resource allocation [18] and power allocation in the vector broadcast channel [19]. This study shows that the dual-decomposition method can also be applied to nonconvex optimizations.

A. Weighted Rate Sum

Start by considering the following optimization problem where the objective is to maximize the weighted rate sum

$$\max_{\mathbf{s}_1, \mathbf{s}_2} wR_1 + (1-w)R_2. \tag{11}$$

The following theorem shows that solving this problem is equivalent to solving the original spectrum management problem (3).

Theorem 1: For any $0 \le w \le 1$, there exists at least one R_1^{target} for which the weighted rate-sum optimization (11) is equivalent to the original spectrum management problem (3).

Proof: The proof will be made by illustration. As shown in the rate region in Fig. 1 for any given $0 \le w \le 1$, there will be at least one point which maximizes the weighted rate sum. If there are multiple optimal points, the optimization search will need to explore each point in turn. In this case, there are two points A and C. Consider one of these points $A \triangleq (R_1^a, R_2^a)$. Assume that there exists some other point in the rate region $B \triangleq (R_1^b, R_2^b)$ such that $R_1^b \ge R_1^a$ and $R_2^b > R_2^a$. This would imply that point B leads to a larger weighted rate sum than point A, but this is contradicted by the optimality of A in the weighted rate sum (11). Thus, no such point B can exist. Hence, R_2^a is the highest rate of line 2 that allows a target rate of R_1^a to be achieved on line 1. This implies that point A is optimal in terms

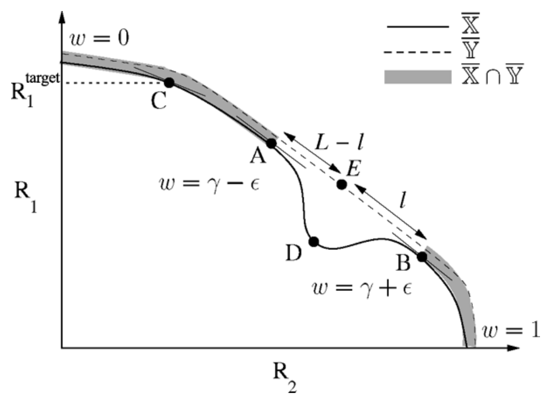


Fig. 2. Operating points in $\mathbb{X}\cap\mathbb{Y}$ can be found through a weighted rate-sum optimization.

of the original spectrum management problem (3) for the target rate $R_1^{\text{target}} = R_1^a$. ■

In Section IV-C, the optimal spectrum balancing algorithm is described, which finds the optimal solution to (11) for any particular w. *Theorem 1* implies that solving (11) is equivalent to solving (3) for some particular R_1^{target}. Thus, the proposed algorithm is guaranteed to always yield an optimal solution to the spectrum management problem (3). The full proof is deferred to *Theorem 3* and the Appendix.

By sweeping through different values of w, a large portion of the rate region can be characterized. Unfortunately, points which lie in the *interior* of the *convex hull* of the rate region, e.g., point D in Fig. 2, cannot be found with the proposed algorithm. The reason for this is that these points are not optimal in terms of a weighted rate sum (11). For example, in Fig. 2, both A and B are superior to point D. This is one of the problems inherent to the use of the weighted rate sum as an optimization metric, however, the weighted rate sum appears to be difficult to avoid, since trying to solve (3) directly leads to an exponential complexity in K and an intractable problem.

Fortunately, all achievable points on the convex hull of the rate region *can* be characterized using a weighted rate sum and hence can be found using optimal spectrum balancing. This statement is formalized in the following theorem.

Theorem 2: For any rate region $\mathbb{X}$, define $\overline{\mathbb{X}}$ as the boundary of $\mathbb{X}$, $\mathbb{Y}$ as the convex hull of $\mathbb{X}$, and $\overline{\mathbb{Y}}$ as the boundary of $\mathbb{Y}$. Consider any operating point $C \triangleq (R_1^c, R_2^c)$ which is achievable $C\in\mathbb{X}$ and on the boundary of the convex hull of the rate region $C\in\overline{\mathbb{Y}}$, as depicted in Fig. 2. There exists some w such that the PSDs at point C can be found through a weighted rate-sum maximization.

Proof: C is on the boundary of the convex set $\mathbb{Y}$. Thus, there exists no point $D \triangleq (R_1^d, R_2^d) \in \mathbb{Y}$ such that $R_1^d > R_1^c$ and $R_2^d > R_2^c$. This implies that for some w, we have

$$wR_1^c + (1-w)R_2^c \geq wR_1^d + (1-w)R_2^d \quad \forall \left(R_1^d, R_2^d\right) \in \mathbb{Y}.$$

Now, since $\mathbb{X}\subset\mathbb{Y}$, we have

$$wR_1^c + (1-w)R_2^c \geq wR_1^d + (1-w)R_2^d \quad \forall \left(R_1^d, R_2^d\right) \in \mathbb{X}.$$

Thus, the point C gives the maximum weighted rate sum of all achievable points within the rate region $\mathbb{X}$ for some particular weight w. Hence, the point C is optimal in the weighted rate sum (11) for that w and can be found through a weighted rate-sum maximization. ■

Corollary 1: For any convex rate region, all optimal operating points on the boundary of the rate region can be found through a weighted rate-sum optimization.

Proof: In a convex rate region, the boundary of the convex hull $\overline{\mathbb{Y}}$ contains the entire boundary of the rate region and $\overline{\mathbb{X}} = \overline{\mathbb{Y}}$. All optimal operating points in terms of the original spectrum management problem (3) lie on the boundary of the rate region. Hence, *Theorem 2* implies that all optimal operating points can be found through a weighted rate-sum optimization. ■

Theorem 2 implies that any achievable operating point on the boundary of the convex hull of the rate region can be found through a weighted rate-sum optimization. If the rate region is close to being convex, then the majority of the optimal operating points can be found. Thankfully, this is the case in DSL channels, as will now be explained.

In the wireline medium, there is some correlation between the channels on neighboring tones. If the channel is sampled finely enough, then neighboring tones will see almost the same channels (both direct and crosstalk).

Imagine that the tone spacing is fine enough such that $h_k^{n,m} \simeq h_{k+l}^{n,m}$, $0 \leq l \leq L-1$. Consider two points in the rate region $A = (R_1^a, R_2^a)$ and $B = (R_1^b, R_2^b)$ and their corresponding PSDs $(s_k^{1,a}, s_k^{2,a})$ and $(s_k^{1,b}, s_k^{2,b})$. It is possible to operate at a point $E = ((l/L)R_1^a + (L-l/L)R_1^b, (l/L)R_2^a + (L-l/L)R_2^b)$ for any $0 \leq l \leq L-1$ as depicted in Fig. 2. This is done by setting the PSDs to $(s_k^{1,a}, s_k^{2,a})$ on tones $k\in\{pL+1,\ldots pL+l\}$ for all integer values of p and to $(s_k^{1,b}, s_k^{2,b})$ on all other tones.

For example, to operate at a point two thirds between A and B (i.e., on the side closer to A), it is required that $l = 2$ and $L = 3$. Thus, the PSDs are set to $(s_k^{1,a}, s_k^{2,a})$ on tones $k \in \{1,2,4,5,7,8,\ldots,K-1\}$ and to $(s_k^{1,b}, s_k^{2,b})$ on tones $k \in \{3,6,9,\ldots,K\}$. For this to work, the tone spacing must be small enough such that the channel is approximately flat over $L = 3$ neighboring tones. That is, it is necessary that $h_k^{n,m} \simeq h_{k+1}^{n,m} \simeq h_{k+2}^{n,m}, \forall k \in \{1,4,\ldots,K-2\}$.

For large L (small tone spacing), practically any operating point between A and B can be achieved. Thus, for any two points in the rate region, any point between them is also within the rate region. This is the definition of a convex set. As such, the rate region is approximately convex in DMT systems with small tone spacings. This approximation becomes exact as the tone spacing approaches zero. For the remainder of this paper, we assume that the DMT tone spacing is small such that the rate region is convex. This is justified for practical DSL systems for which Δ_f is 4.3125 kHz.

Note that one should not confuse convexity of the rate region with convexity of the objective function (11). In practice, the rate regions are seen to be nearly convex, however, the optimization problem is highly nonconvex, exhibiting many local maxima. For this reason, conventional convex optimization techniques cannot be applied, and an exhaustive search is required on each tone.

B. Dual Decomposition

In the previous section, it was shown that the spectrum management problem (3) can be solved through a weighted rate-sum optimization (11). It was also shown that in DSL, the rate region is approximately convex, allowing almost all optimal operating points to be found. This section will show how the weighted rate-sum optimization can be solved in a computationally tractable way.

The total power constraints (4) can be incorporated into the optimization problem by defining the Lagrangian

$$L \triangleq wR_1 + (1-w)R_2 - \lambda_1 \sum_k s_k^1 - \lambda_2 \sum_k s_k^2. \tag{12}$$

Here, λ_n denotes the Lagrangian multiplier for user n and is chosen such that either the power constraint on user n is tight ($\sum_k s_k^n = P_n$) or $\lambda_n = 0$. The constrained optimization (11) can now be solved via the unconstrained optimization

$$\max_{\mathbf{s}_1,\mathbf{s}_2} L\left(w, \lambda_1, \lambda_2, s_k^1, s_k^2\right). \tag{13}$$

Define the Lagrangian on tone k as

$$L_k \triangleq wb_k^1 + (1-w)b_k^2 - \lambda_1 s_k^1\left(b_k^1, b_k^2\right) - \lambda_2 s_k^2\left(b_k^1, b_k^2\right).$$

Clearly, the Lagrangian (12) can be decomposed into a sum across tones of L_k, i.e.,

$$L = \sum_k L_k.$$

This is known as the dual decomposition. As a result, the optimization can be split into K per-tone optimizations which are coupled only through w, λ_1, and λ_2. This will lead to a linear, rather than exponential, complexity in K and a computationally tractable search.

C. Algorithm

The optimal spectrum balancing algorithm is listed as Algorithm 1. Spectral mask constraints can be incorporated into the optimization by setting L_k to $-\infty$ if $s_k^1 > s_k^{1,\max}$ or $s_k^2 > s_k^{2,\max}$. If discrete bitloading is employed, then the maximization in the function optimize_s is limited to the PSD combinations corresponding to valid bitloading combinations, as described by (10). If continuous bitloading is used, then the maximization is limited to the values of s_k^n supported by the accuracy of the modem's AFE, as described by (7).

The algorithm operates as follows. It is necessary to search through both λ_1 and λ_2 to find values which place sufficient importance on the total power constraint terms within the Lagrangian (12). Varying w makes it possible to map out the optimal points on the convex hull of the rate region.

The algorithm contains three loops: an outer loop that varies w, an intermediate loop that searches for λ_1, and an inner loop that searches for λ_2. Bisection is used in each search.

When searching for λ_n, it is first necessary to find a value of λ_n that ensures that the power constraint of user n is satisfied. This value is stored in $\lambda_n^{\max}$. Note that a larger λ_n places more emphasis on the power constraint of user n in the Lagrangian. As a result, using a larger λ_n will result in a lower total power for user n.

Once $\lambda_n^{\max}$ is found, the algorithm proceeds to bisection. Note that after the algorithm has completed, for each user either $\sum_k s_k^n = P_n$ or $\lambda_n = 0$. Thus, the Lagrangian and the original objective become equivalent. More rigorously, we have the following.

Algorithm 1 Optimal spectrum balancing

Main Function

for $w = 0, \ldots, 1$

$\mathbf{s}_1, \mathbf{s}_2 = \text{optimize\_}\lambda_1(w)$

end

Function $\mathbf{s}_1, \mathbf{s}_2 = \textbf{optimize\_}\lambda_1(w)$

$\lambda_1^{\max} = 1$, $\lambda_1^{\min} = 0$
while $\sum_k s_k^1 > P_1$

$\lambda_1^{\max} = 2\lambda_1^{\max}$
$\mathbf{s}_1, \mathbf{s}_2 = \text{optimize\_}\lambda_2(w, \lambda_1^{\max})$

end
repeat

$\lambda_1 = (\lambda_1^{\max} + \lambda_1^{\min})/2$
$\mathbf{s}_1, \mathbf{s}_2 = \text{optimize\_}\lambda_2(w, \lambda_1)$
if $\sum_k s_k^1 > P_1$, then $\lambda_1^{\min} = \lambda_1$, else $\lambda_1^{\max} = \lambda_1$

until convergence

Function $\mathbf{s}_1, \mathbf{s}_2 = \textbf{optimize\_}\lambda_2(w, \lambda_1)$

$\lambda_2^{\max} = 1$, $\lambda_2^{\min} = 0$
while $\sum_k s_k^2 > P_2$

$\lambda_2^{\max} = 2\lambda_2^{\max}$
$\mathbf{s}_1, \mathbf{s}_2 = \text{optimize\_}s(w, \lambda_1, \lambda_2^{\max})$

end
repeat

$\lambda_2 = (\lambda_2^{\max} + \lambda_2^{\min})/2$
$\mathbf{s}_1, \mathbf{s}_2 = \text{optimize\_}s(w, \lambda_1, \lambda_2)$
if $\sum_k s_k^2 > P_2$, then $\lambda_2^{\min} = \lambda_2$, else $\lambda_2^{\max} = \lambda_2$

until convergence

Function $\mathbf{s}_1, \mathbf{s}_2 = \textbf{optimize\_}s(w, \lambda_1, \lambda_2)$

for $k = 1 \ldots K$

$s_k^1, s_k^2 = \arg\max_{s_k^1, s_k^2} L_k(s_k^1, s_k^2, w, \lambda_1, \lambda_2)$

(solved by exhaustive 2-D search)

end

Theorem 3: For each w, the execution of Algorithm 1 returns a PSD combination that is optimal for the spectrum management problem (3), i.e., for some R_1^{target}, we have

$$\begin{aligned} \mathbf{s}_1, \mathbf{s}_2 = \arg\max_{\mathbf{s}_1,\mathbf{s}_2} & R_2 \\ \text{s.t.} \quad & R_1 \geq R_1^{\text{target}}, \\ & \sum_k s_k^n \leq P_n \quad \forall n \\ & 0 \leq s_k^n \leq s_k^{n,\max} \quad \forall n, k. \end{aligned} \tag{14}$$

Here, R_1^{target} is in fact the rate of user 1 at convergence of the algorithm. Varying w from 0 to 1 allows all optimal operating points which lie on the convex hull of the rate region to be found. If the rate region is convex, then all optimal operating points can be found.

Proof: See the Appendix. ■

Note that the function optimize_s must solve an N-dimensional nonconvex optimization. This requires an exhaustive search, which has an exponential complexity in N. However, since the number of users in a system is typically small, such an exhaustive search is computationally tractable. Through use of the dual decomposition, the complexity of the algorithm is linear in K, and the total optimization is tractable. Compare this with the exhaustive search described in Section III-E, where the optimization was coupled between tones. This coupling led to a KN-dimensional nonconvex optimization, which was computationally intractable.

D. Complexity

This section discusses the complexity of the proposed algorithm and shows that a significant complexity reduction is achieved over the exhaustive search described in Section III-E.

The algorithm begins by finding upper bounds on λ_1 and λ_2, which are stored as $\lambda_1^{\max}$ and $\lambda_2^{\max}$. This is done by simply doubling λ_1 and λ_2 until the power constraints are met. This typically occurs within a few iterations and has a negligible impact on complexity.

The majority of the complexity occurs in the outer loops of the algorithm, where bisection is done on λ_1 and λ_2 such that the power constraints on both users become tight. Assume that an accuracy of ϵ_λ is required in each λ. This will require $\log_2(1/\epsilon_\lambda)$ iterations of optimize_λ_1, which in turn requires $\log_2(1/\epsilon_\lambda)^2$ iterations of optimize_λ_2. Each iteration will result in the function optimize_s being called.

The function optimize_s solves the weighted rate-sum optimization independently on each tone. Since the objective function is nonconvex, the optimization is done exhaustively. This requires $K(b_{\max}+1)^2$ evaluations of L_k in the discrete bitloading case. A similar expression can be written for the continuous bitloading case. Thus, the total complexity of the proposed algorithm is $\mathcal{O}(K(b_{\max}+1)^2 \log_2(1/\epsilon_\lambda)^2)$.

In comparison, solving the problem through an exhaustive search across all tones, as described in Section III-E, requires the evaluation of $(b_{\max}+1)^{2K}$ bitloading combinations. In most cases, this is computationally intractable.

This paper has only shown the algorithm and optimality proof for two user channels. Extensions to more than two users are straightforward and follow naturally from the algorithm and proof presented here. In the general case of N users, there will be a target rate constraint on the first $N-1$ users, and the goal is to maximize the rate of the Nth user. This is equivalent to maximizing a weighted rate sum $\sum_n w_n R_n$, where the weight for the Nth user is arbitrarily defined as $w_N \triangleq 1 - \sum_{n=1}^{N-1} w_n$. To enforce the total power constraints on all users, N Lagrangian multipliers are required $\lambda_1, \ldots, \lambda_N$. The N-user algorithm has a similar form to Algorithm 1, however, it must now sweep through all values of $w_1, \ldots, w_{N-1}$ in *Main Function*.

With N users, bisection must be done on $\lambda_1, \ldots, \lambda_N$, resulting in the function optimize_s being called $\log_2(1/\epsilon_\lambda)^N$ times. In the discrete-bitloading N-user case, the function optimize_s requires $K(b_{\max}+1)^N$ evaluations of L_k. Evaluating L_k requires a weighted rate sum of N users to be calculated, so the total complexity of optimize_s is $\mathcal{O}(KN(b_{\max}+1)^N)$. Taking this all into account, the overall complexity of the proposed algorithm in the N-user case is $\mathcal{O}(KN(b_{\max}+1)^N \log_2(1/\epsilon_\lambda)^N)$. Typically setting ϵ_λ to 1×10^{-10} is sufficient to achieve an accuracy of 1% in the total power constraints [see (4)]. This leads to a complexity

$$V_{\text{OSB}} = \mathcal{O}\left(KN(b_{\max}+1)^N 33^N\right). \tag{15}$$

In comparison, the exhaustive search across all tones in the N-user case requires the evaluation of $(b_{\max}+1)^{KN}$ bitloading combinations. For each bitloading combination, the total rate must be calculated for each user across all tones. Hence, the exhaustive search has a complexity

$$V_{\text{exhaustive}} = \mathcal{O}\left(KN(b_{\max}+1)^{KN}\right).$$

Comparing this with (15) shows that the proposed algorithm leads to a complexity reduction of

$$\Delta V = \mathcal{O}\left((b_{\max}+1)^{(K-1)N} 33^{-N}\right).$$

The first term $(b_{\max}+1)^{(K-1)N}$ can be interpreted as the benefit of replacing the KN-dimensional nonconvex optimization with K separate N-dimensional optimizations. The second term 33^{-N} is the penalty of searching through λ-space.

Typically, $b_{\max} = 14$. In ADSL, $K = 256$, and thus the overall complexity reduction with the proposed algorithm is $\mathcal{O}(10^{298N})$. In VDSL, $K = 4096$, and the overall complexity reduction is even higher at $\mathcal{O}(10^{4815N})$.

Despite the large reduction in complexity that optimal spectrum balancing achieves, at large N, it is still highly complex. Due to changing line conditions and the frequent addition of new users, practical DSM algorithms must be capable of reoptimizing the modem spectra in a matter of minutes. Thus, in practice, it is more interesting to design low-complexity algorithms with near-optimal performance.

There are several options that can be used to reduce the complexity of the proposed algorithm by sacrificing some optimality. An extension of this study has demonstrated that by optimizing the transmit spectra one modem at a time in an iterative fashion, near-optimal performance can be achieved

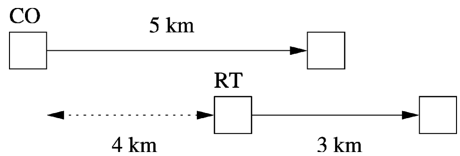


Fig. 3. Downstream ADSL scenario.

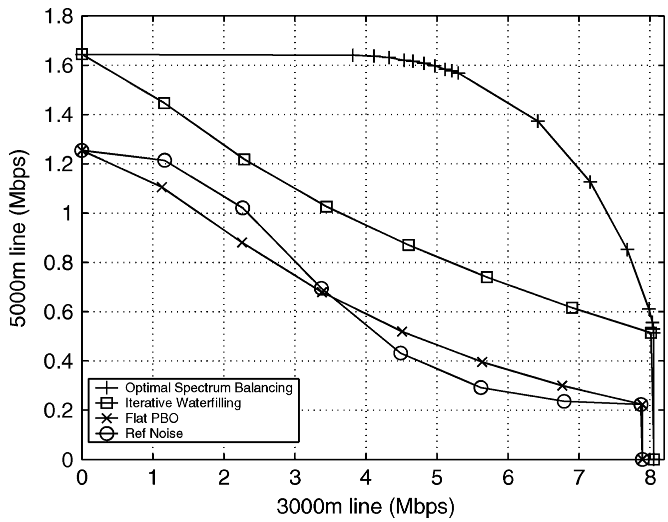


Fig. 4. Rate regions in downstream ADSL.

with a significant reduction in complexity to $\mathcal{O}(KN^2)$ [7]. While this technique helps to reduce complexity considerably, it is important to note that the resulting algorithm is no longer strictly optimal.

V. Performance

This section examines the performance of optimal spectrum balancing when compared with other spectrum management techniques. For all simulations, the line diameter is 0.5 mm (24-AWG). The target symbol error probability is 10^{-7} or less. The coding gain and noise margin are set to 3 and 6 dB, respectively. Continuous bitloading is used and Δ_s is set to 0.1 dBm/Hz. The maximum bitloading is not constrained. As per the DSL standards, the tone spacing Δ_f and DMT symbol rate f_s are set to 4.3125 and 4 kHz, respectively [14], [15].

A. Remote Terminal Distributed Downstream ADSL

Downstream transmission in ADSL was simulated with a 5-km CO distributed line and a 3-km RT distributed line. The RT is located 4 km from the CO, as depicted in Fig. 3.

A maximum transmit power of 20.4 dBm was applied to each modem [14]. A spectral mask was applied to the flat power back-off (PBO) and reference noise method and was set at −40 dBm/Hz [14]. A spectral mask was not applied to iterative waterfilling or optimal spectrum balancing. Background noise included crosstalk from 16 ISDN, four HDSL, and ten conventional ADSL modems which transmit at the spectral mask.

Fig. 4 shows the rate regions corresponding to the various spectrum management algorithms. For comparison, the rate regions with iterative waterfilling, flat PBO, and the reference

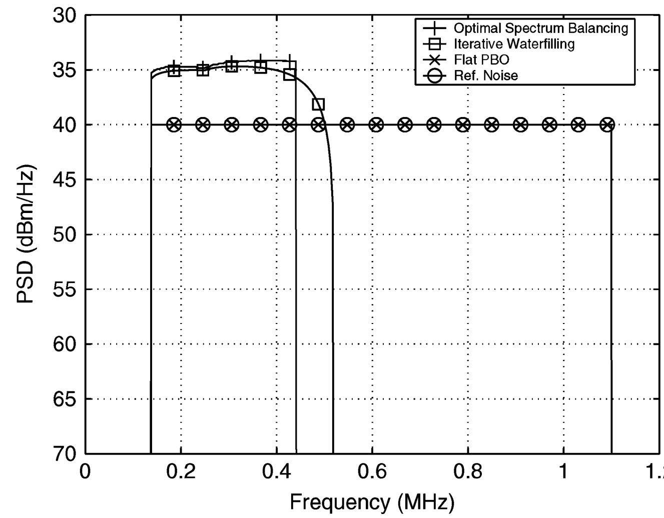


Fig. 5. PSDs on the CO line in downstream ADSL (CO line at 1 Mbps).

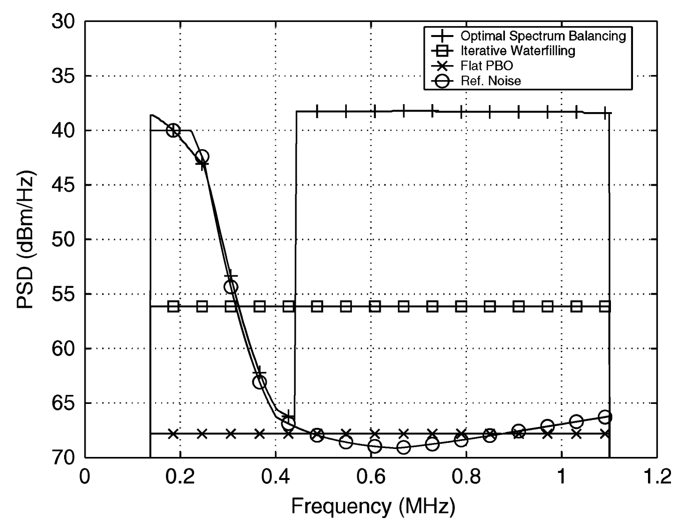


Fig. 6. PSDs on RT line in downstream ADSL (CO line at 1 Mbps).

noise method have been included. In the reference noise method, each modem sets its transmit PSD such that the crosstalk it induces on the victim modem is equal to the background noise seen by that modem, which is the so-called reference noise [12]. With flat PBO, each modem transmits the minimum possible flat PSD required to support its target rate.

Note that iterative waterfilling, the reference PSD method, and flat PBO are all distributed algorithms and require no centralized control. In contrast, optimal spectrum balancing is a centralized algorithm requiring knowledge of the direct and crosstalk channel attenuations within the network. Optimal spectrum balancing is suitable for direct application when an SMC is available. In the absence of an SMC, the proposed algorithm is still a useful tool in the design of distributed DSM algorithms, providing both an upper bound on performance and insight into good spectrum management strategy [7].

The PSDs corresponding to a 1-Mbps service on the CO distributed line are depicted in Figs. 5 and 6. The optimal PSD on the RT line decreases with frequency to reflect the increase in crosstalk coupling. This continues until 440 kHz, where the CO

TABLE I
ACHIEVABLE RATES IN DOWNSTREAM ADSL

Scheme	CO Rate	RT Rate
Flat PBO	1.0 Mbps	1.7 Mbps
Ref. Noise	1.0 Mbps	2.3 Mbps
Iterative Waterfilling	1.0 Mbps	3.6 Mbps
Optimal Spectrum Balancing	1.0 Mbps	7.4 Mbps

line becomes inactive due to its low channel SNR above this frequency. Once the CO line becomes inactive, a sudden increase in the optimal PSD on the RT line can be observed.

With the flat PBO algorithm, the RT line must employ a large amount of PBO to protect the CO line. This occurs because, unlike in optimal spectrum balancing, the flat PBO algorithm cannot vary the degree of PBO with frequency.

The iterative waterfilling algorithm gives similar results. Slightly less PBO is required since the CO line PSD has been boosted on the active tones, as shown in Fig. 5. However, the amount of PBO required is still much larger than with optimal spectrum balancing. The iterative waterfilling algorithm does not exploit the fact that crosstalk coupling is low at low frequencies. It also does not exploit the fact that the CO line is inactive above 440 kHz. Both of these facts were used by optimal spectrum balancing to increase the transmit PSD on the RT line at low and high frequencies, thus leading to a large performance gain over iterative waterfilling.

It has been shown that the reference noise method is near optimal when the SINR is high [12]. This is the case in low frequencies. For this reason, the reference noise PSD matches the optimal PSD quite closely at frequencies below 440 kHz.

As shown in Table I, using optimal spectrum balancing instead of iterative waterfilling allows the data rate on the RT distributed line to be increased from 3.6 to 7.4 Mbps whilst still maintaining a 1-Mbps service on the CO distributed line. This corresponds to a gain of over 100%.

B. Near–Far Problem in Upstream VDSL

Upstream VDSL transmission was simulated with 4×600-m lines and 4×1200-m lines with the receivers colocated at a common CO. Each modem had a maximum transmit power of 11.5 dBm available. A spectral mask was applied to the flat PBO, reference noise, and reference PSD method and was set at -60 dBm/Hz [15]. A spectral mask was not applied to iterative waterfilling or optimal spectrum balancing. Alien crosstalk was incorporated into the background noise using ETSI model A [15]. FDD bandplan 998 was used with the frequency bands corresponding to amateur radio frequencies notched off [15].

Fig. 7 shows the rate regions corresponding to various spectrum management algorithms. Included are iterative waterfilling [6], the reference noise method, flat PBO, and the reference PSD method, which is currently adopted in VDSL standards [15].

To give an example of the potential gains of optimal spectrum balancing, we set the target rate to 16 Mbps on the 600-m lines. The resulting achievable rate on the 1200-m lines is listed in Table II for each of the algorithms. As can be seen, using optimal spectrum balancing instead of iterative waterfilling allows the data rate on the 1200-m lines to be increased from 2.6 to 4.5 Mbps, almost doubling the data throughput.

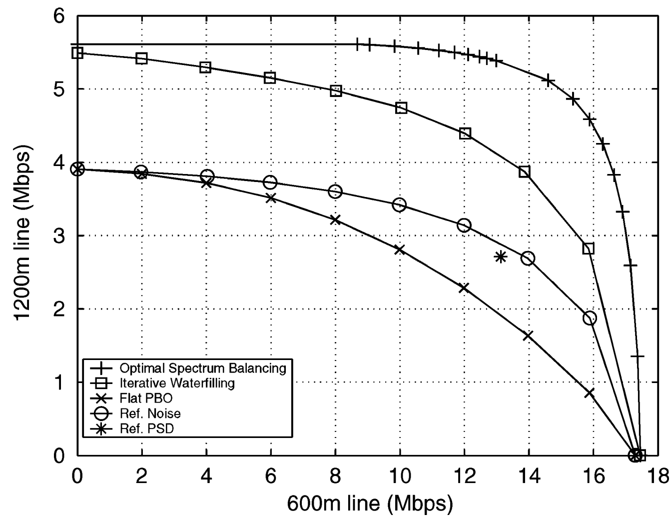


Fig. 7. Rate regions in upstream VDSL.

TABLE II
ACHIEVABLE RATES IN UPSTREAM VDSL

Scheme	600m. Rate	1200m. Rate
Target Rate	16 Mbps	Best Achievable
Ref. PSD	13.1 Mbps	2.7 Mbps
Flat PBO	16.0 Mbps	0.8 Mbps
Ref. Noise	16.0 Mbps	1.7 Mbps
Iterative Waterfilling	16.0 Mbps	2.6 Mbps
Optimal Spectrum Balancing	16.0 Mbps	4.5 Mbps

Note that the optimal rate regions for both the ADSL and VDSL scenarios are convex, as was predicted in Section IV-A.

C. Discrete Bitloading

The same simulations were made with discrete bitloading, with each modem forced to adopt an integer bitloading value. We set the maximum bitloading $b_{\max}$ to 14. All other simulation parameters were the same. In the iterative waterfilling algorithm, the Levin–Campello algorithm was used to ensure integer bitloadings on each tone [16].

In the ADSL scenario, using optimal spectrum balancing instead of iterative waterfilling allowed the data rate on the RT distributed line to be increased from 3.1 to 7.3 Mbps whilst still maintaining a 1-Mbps service on the CO distributed line.

In the VDSL scenario, using optimal spectrum balancing instead of iterative waterfilling allowed the data rate on the 600-m lines to be increased from 3.4 to 13 Mbps while maintaining a 5-Mbps service on the 1200-m lines. This corresponds to a gain of over 280%.

VI. CONCLUSION

This paper presented an optimal algorithm for spectrum balancing in DSL. The algorithm optimizes the spectra of the modems within a network, allowing them to achieve maximal performance and operate on the rate region boundary. The algorithm can operate under a combination of total power and/or spectral mask constraints and can use either continuous or discrete bitloading.

Through the use of a dual decomposition, the inner loop of the algorithm solves the spectrum management problem independently on each tone. The result is a computationally tractable and efficient algorithm. Simulations show that the algorithm yields significant gains over existing spectrum management techniques, e.g., in one of the cases studied, the proposed centralized algorithm leads to a factor-of-four increase in data rate over the distributed DSM algorithm iterative waterfilling.

Optimal spectrum balancing is a centralized algorithm requiring an SMC for direct implementation. In future work, it will be interesting to develop distributed DSM algorithms based on the insight gained from the proposed algorithm. The goal is to find a simple, distributed algorithm which yields near-optimal performance in a broad range of scenarios. Early work in this area is promising [7].

While this paper has focused on the problem of spectrum management in DSL, the algorithm is also applicable to any communication system where interuser interference is a problem. Optimal spectrum balancing could also be applied to broadband cable networks, high-speed Ethernets, or fixed wireless links.

The idea of using dual decomposition to simplify high-dimensional, nonconvex problems can also be applied to many other communications problems. Extensions of the work presented here consider the use of dual decomposition for joint crosstalk canceller—transmit spectra design [20], and for joint bandplan—transmit spectra optimization [21].

APPENDIX
PROOF OF *THEOREM 2*

To prove the optimality of Algorithm 1 as stated in *Theorem 3*, it is first shown that the algorithm converges. It will then be shown that at convergence, maximizing the Lagrangian is equivalent to maximizing the weighted rate sum (11). This implies the optimality of the PSDs generated by the algorithm.

To prove the convergence of Algorithm 1, the convergence of a related routine is first examined. This routine finds the correct value for λ_n, thereby ensuring that the total power constraint on user n, as described by (4), is satisfied. At this value of λ_n, the routine finds the optimal PSD for user n. As will be shown in *Corollaries 2* and *3*, the algorithms optimize_λ_1 and optimize_λ_2 can be seen as special cases of this routine for specific values of the optimization function $f(\mathbf{s}_n)$. In *Lemma 2*, it is proven that this routine converges. This in turn implies the convergence of optimize_λ_1 and optimize_λ_2.

First, define the objective function

$$G(\mathbf{s}_n, \lambda_n) \triangleq f(\mathbf{s}_n) - \lambda_n \sum_k s_k^n. \tag{16}$$

Denote the optimal power allocation for a given λ_n as

$$\mathbf{s}_n(\lambda_n) \triangleq \arg\max_{\mathbf{s}_n} G(\mathbf{s}_n, \lambda_n)$$

with $s_k^n(\lambda_n) \triangleq [\mathbf{s}_n(\lambda_n)]_k$. The routine for user n is then given as follows.

Routine for user n

$\lambda_n^{\max} = 1, \lambda_n^{\min} = 0$
while $\sum_k s_k^n > P_n$
    $\lambda_n^{\max} = 2\lambda_n^{\max}$
    $\mathbf{s}_n = \arg\max_{\mathbf{s}_n} f(\mathbf{s}_n) - \lambda_n^{\max} \sum_k s_k^n$
end
repeat
    $\lambda_n = (\lambda_n^{\max} + \lambda_n^{\min})/2$
    $\mathbf{s}_n = \arg\max_{\mathbf{s}_n} f(\mathbf{s}_n) - \lambda_n \sum_k s_k^n$
    if $\sum_k s_k^n > P_n$, then $\lambda_n^{\min} = \lambda_n$, else $\lambda_n^{\max} = \lambda_n$
until convergence

The following lemma is used to prove the convergence of this routine.

Lemma 1: Fix n. $\sum_k s_k^n(\lambda_n)$ is monotonic decreasing in λ_n. Furthermore, $\lim_{\lambda_n \to \infty} \sum_k s_k^n(\lambda_n) = 0$.

Proof: Consider two Lagrangian multipliers λ_n^a and λ_n^b and their corresponding optimal PSDs $\mathbf{s}_n^a \triangleq \mathbf{s}_n(\lambda_n^a)$ and $\mathbf{s}_n^b \triangleq \mathbf{s}_n(\lambda_n^b)$. Denote the elements of these PSDs as $s_k^{n,a}$ and $s_k^{n,b}$, respectively. Let

$$\lambda_n^b \geq \lambda_n^a. \tag{17}$$

Define

$$A \triangleq f(\mathbf{s}_n^a) - \lambda_n^a \sum_k s_k^{n,a}$$
$$B \triangleq f(\mathbf{s}_n^b) - \lambda_n^a \sum_k s_k^{n,b}$$
$$C \triangleq f(\mathbf{s}_n^b) - \lambda_n^b \sum_k s_k^{n,b}$$
$$D \triangleq f(\mathbf{s}_n^a) - \lambda_n^b \sum_k s_k^{n,a}.$$

Now $G(\mathbf{s}_n^a, \lambda_n^a) \geq G(\mathbf{s}_n^b, \lambda_n^a)$ by the optimality of $\mathbf{s}_n^a$ in $G(\mathbf{s}_n, \lambda_n^a)$. Hence, $A \geq B$. Similarly, the optimality of $\mathbf{s}_n^b$ in $G(\mathbf{s}_n, \lambda_n^b)$ implies $C \geq D$. Furthermore, (17) implies $B > C$. Now, $A \geq B \geq C \geq D$ implies $A - D \geq B - C$. Hence

$$(\lambda_n^b - \lambda_n^a) \sum_k s_k^{n,a} \geq (\lambda_n^b - \lambda_n^a) \sum_k s_k^{n,b}$$

which implies

$$\sum_k s_k^{n,a} \geq \sum_k s_k^{n,b}. \tag{18}$$

Thus, a larger λ_n leads to a smaller $\sum_k s_k^n$. This implies that $\sum_k s_k^n$ is monotonically decreasing in λ_n.

The second part of the lemma will now be proven. From (16), it can be shown that for large λ_n, $G(\mathbf{s}_n, \lambda_n) \simeq -\lambda_n \sum_k s_k^n$ with the approximation becoming exact as $\lambda_n \to \infty$. Hence, $\lim_{\lambda_n \to \infty} \mathbf{s}_n(\lambda_n) = \mathbf{0}_K$, where $\mathbf{0}_K$ is the length K vector with all elements equal to zero. ■

Lemma 2: In DMT systems with small tone spacing, the *Routine for user n* converges. At convergence, we have

$$\mathbf{s}_n = \arg\max_{\mathbf{s}_n} f(\mathbf{s}_n) \text{ s.t.} \sum_k s_k^n \le P_n.$$

Proof: Before establishing convergence, we first show that $\sum_k s_k^n(\lambda_n)$ is a continuous function of λ_n. This is true for DMT systems with small tone spacing and with a large number of DMT tones. In such a system, there would always be one or more tones that would change their bit and power allocation in response to any small change in λ_n. Thus, the total transmit power of the optimal power allocation $\sum_k s_k^n(\lambda_n)$ is continuous in λ_n.

Now, we are ready to prove convergence of the routine. The routine consists of two stages: a preamble that determines $\lambda_n^{\max}$ and the actual routine itself. The preamble clearly converges since, from *Lemma 1*, $\sum_k s_k^n(\lambda_n) \to 0$ as $\lambda_n \to \infty$.

The convergence of the main part of the *Routine for user n* can be shown as follows: $\lambda_n^{\max} - \lambda_n^{\min}$ decreases by half in each iteration. Thus, λ_n converges to a fixed value. Let us now consider two cases, depending on whether $\sum_k s_k^n(\lambda_n^{\min}) > P_n$ or not.

Suppose that $\sum_k s_k^n(\lambda_n^{\min}) > P_n$ at $\lambda_n^{\min} = 0$, then, since the preamble ensures that $\sum_k s_k^n(\lambda_n^{\max}) \le P_n$, throughout the algorithm, it is always the case that $\sum_k s_k^n(\lambda_n^{\min}) > P_n$ and $\sum_k s_k^n(\lambda_n^{\max}) \le P_n$. Since $\lambda_n^{\max} \ge \lambda_n \ge \lambda_n^{\min}$, $\lambda_n^{\min}$ and $\lambda_n^{\max}$ converge to a fixed value, and since $\sum_k s_k^n(\lambda_n)$ is monotonic in λ_n, this implies that $\sum_k s_k^n(\lambda_n)$ must converge to P_n. On the other hand, suppose that $\sum_k s_k^n(\lambda_n^{\min}) \le P_n$ at $\lambda_n^{\min} = 0$. Then, λ_n will converge to zero.

Hence, the algorithm will converge, and at convergence, either $\lambda_n = 0$ or $\sum_k s_k^n(\lambda_n) = P_n$. Thus, at convergence, we have

$$G(\mathbf{s}_n, \lambda_n) = f(\mathbf{s}_n) - \lambda_n P_n$$

which is simply $f(\mathbf{s}_n)$ modified by a term that is independent of $\mathbf{s}_n$. In the routine, we have

$$\begin{aligned}\mathbf{s}_n &= \arg\max_{\mathbf{s}_n} G(\mathbf{s}_n, \lambda_n)\\ &= \arg\max_{\mathbf{s}_n} f(\mathbf{s}_n) \text{ s.t.} \sum_k s_k^n \le P_n.\end{aligned}$$

To see this, clearly $\mathbf{s}_n$ satisfies the constraint. Further, if there is some other feasible $\mathbf{s}'_n$ that does better than $\mathbf{s}_n$ for the objective function $f(\mathbf{s}_n)$, then $\mathbf{s}'_n$ should do better than $\mathbf{s}_n$ for the objective $G(\mathbf{s}_n, \lambda_n)$ also. This is contradicted by the optimality of $\mathbf{s}_n$ in $G(\mathbf{s}_n, \lambda_n)$. Hence, $\mathbf{s}_n$ must be optimal in $f(\mathbf{s}_n)$. ■

Lemma 3: The function optimize_s yields PSDs $\mathbf{s}_1$ and $\mathbf{s}_2$ which maximize the Lagrangian.

Proof: From function optimize_s, we have

$$s_k^1, s_k^2 = \arg\max_{s_k^1, s_k^2} L_k\left(w, \lambda_1, \lambda_2, s_k^1, s_k^2\right).$$

Since $L = \sum_k L_k$, and since the optimization of the Lagrangian is unconstrained,[2] this implies

$$\mathbf{s}_1, \mathbf{s}_2 = \arg\max_{\mathbf{s}_1, \mathbf{s}_2} L(w, \lambda_1, \lambda_2, \mathbf{s}_1, \mathbf{s}_2).$$

■

Corollary 2: The function optimize_λ_2 converges. At convergence, we have

$$\begin{aligned}\mathbf{s}_2 = \arg\max_{\mathbf{s}_1, \mathbf{s}_2} &\, wR_1 + (1-w)R_2 - \lambda_1 \sum_k s_k^1,\\ \text{s.t.} &\sum s_k^2 \le P_2.\end{aligned} \tag{19}$$

Proof: Let $n^k = 2$ and

$$f(\mathbf{s}_2) \triangleq \max_{\mathbf{s}_1} wR_1 + (1-w)R_2 - \lambda_1 \sum_k s_k^1.$$

Lemma 3 implies that optimize_λ_1 and the routine are equivalent. Hence, *Lemma 2* implies optimize_λ_2 converges, and that at convergence, (19) is satisfied. ■

Corollary 3: The function optimize_λ_1 converges. At convergence, we have

$$\begin{aligned}\mathbf{s}_1 = \arg\max_{\mathbf{s}_1, \mathbf{s}_2} &\, wR_1 + (1-w)R_2,\\ \text{s.t.} &\sum s_k^1 \le P_1, \ \sum_k s_k^2 \le P_2.\end{aligned} \tag{20}$$

Proof: Let $n = 1^k$ and

$$\begin{aligned}f(\mathbf{s}_1) \triangleq \max_{\mathbf{s}_2} &\, wR_1 + (1-w)R_2,\\ \text{s.t.} &\sum_k s_k^2 \le P_2.\end{aligned}$$

Then, *Lemma 2* and *Corollary 2* imply that optimize_λ_1 converges and that, at convergence, (20) is satisfied. ■

From *Theorem 1*, for any particular w, there exists some R_1^{target} for which the weighted rate-sum optimization (20) is equivalent to the original spectrum management problem (14). Hence, for any particular w, the weighted rate-sum optimization leads to an optimal operating point.

Corollary 3 implies that for each value of w in Algorithm 1, the PSD combination returned by the algorithm maximizes a weighted-rate sum. Hence, the PSD combination is also an optimal solution to (14). Furthermore, *Theorem 2* states that by varying w from 0 to 1, it is possible to map out all achievable operating points on the boundary of the convex hull of the rate region.

References

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

[2]Recall that the constraints are incorporated into the objective function and need not be explicitly enforced.

[2] R. Cendrillon, M. Moonen, G. Ginis, K. Van Acker, T. Bostoen, and P. Vandaele, "Partial crosstalk cancellation for upstream VDSL," *EURASIP J. Appl. Signal Process.*, vol. 10, pp. 1433–1448, Oct. 2004.

[3] *Spectrum Management for Loop Transmission Systems*, ANSI Std. T1.417, 2003.

[4] J. Cioffi, "Dynamic spectrum management," in *DSL Advances*. Upper Saddle River, NJ: Prentice-Hall, 2002, ch. 11.

[5] *Dynamic Spectrum Management*, ANSI Draft Std. T1E1.4/2003-018, 2004, Rev. 15.

[6] W. Yu, G. Ginis, and J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 5, pp. 1105–1115, Jun. 2002.

[7] R. Cendrillon and M. Moonen, "Iterative spectrum balancing for digital subscriber lines," in *Proc. IEEE Int. Conf. Commun.*, Seoul, Korea, May 2005, vol. 3, pp. 1937–1941.

[8] K. Cheong, W. Choi, and J. Cioffi, "Multiuser soft interference canceler via iterative decoding for DSL applications," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 2, pp. 363–371, Feb. 2002.

[9] G. Cherubini, "Optimum upstream power back-off and multiuser detection for VDSL," in *Proc. IEEE Global Telecommun. Conf.*, 2001, pp. 375–380.

[10] J. Lee, R. Sonalkar, and J. Cioffi, "A multi-user power control algorithm for digital subscriber lines," *IEEE Commun. Lett.*, vol. 9, no. 3, pp. 193–195, Mar. 2005.

[11] K. Jacobsen, "Methods of upstream power backoff on very high-speed digital subscriber lines," *IEEE Commun. Mag.*, pp. 210–216, Mar. 2001.

[12] F. Sjoberg, S. Wilson, and M. Isaksson, "Power back-off in the upstream of VDSL," in *Proc. Radiovetenskaplig Koferens (RVK'99)*, Karlskrona, Sweden, Jun. 1999, vol. 1, pp. 436–440.

[13] J. Lee, V. Sonalkar, and J. Cioffi, "A multi-user rate and power control algorithm for VDSL," in *Proc. IEEE Global Telecommun. Conf.*, 2003, pp. 1264–1268.

[14] *Asymmetrical Digital Subscriber Line Transceivers 2 (ADSL2)*, ITU Std. G.992.2, 2002.

[15] *Very High Speed Digital Subscriber Line (VDSL); Functional Requirements*, ETSI Std. TS 101 270-1, 2003, Rev. V.1.3.1.

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

[17] *Physical Layer Management for Digital Subscriber Line (DSL) Transcievers*, ITU Std. G.997.1, 2003.

[18] L. Xiao, M. Johansson, and S. Boyd, "Simultaneous routing and resource allocation via dual decomposition," *IEEE Trans. Commun.*, vol. 52, no. 7, pp. 1136–1144, Jul. 2004.

[19] W. Yu, "A dual decomposition approach to the sum power Gaussian vector multiple access channel sum capacity problem," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 754–759, Feb. 2006.

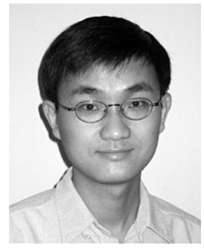
[20] V. Chan and W. Yu, "Joint multiuser detection and optimal spectrum balancing for digital subscriber lines," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 3, Mar. 2005, pp. 333–336.

[21] W. Yu, R. Lui, and R. Cendrillon, "Dual optimization methods for multiuser OFDM systems," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, 2004, pp. 225–229.

Raphael Cendrillon was born in Melbourne, Australia, in 1978. He received the electrical engineering degree (Honors First Class) from the University of Queensland, Brisbane, Australia, in 1999, and the Ph.D. degree in electrical engineering (*summa cum laude*) from the Katholieke Universiteit Leuven (KU Leuven), Heverlee, Belgium, in 2004.

His research focuses on the application of multiuser communication theory to xDSL. In 2002, he was a Visiting Scholar with the Information Systems Laboratory, Stanford University, Stanford, CA. In 2005, he was a Postdoctoral Research Fellow with the University of Queensland. During this time, he was also a Visiting Research Fellow with Princeton University, Princeton, NJ. He is currently a Senior DSP Engineer with Marvel Hong Kong Ltd., Hong Kong.

Dr. Cendrillon was awarded the Alcatel Bell Scientific Prize in 2004, the IEEE Travel Grants in 2003, 2004, and 2005, the K.U. Leuven Bursary for Advanced Foreign Scholars in 2004, and the UniQuest Trailblazer Prize for Commercialization in 2005.



Wei Yu (M'00) received the B.A.Sc. degree in computer engineering and mathematics from the University of Waterloo, Waterloo, ON, Canada, in 1997, and M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1998 and 2002, respectively.

Since 2002, he has been an Assistant Professor with the Electrical and Computer Engineering Department, University of Toronto, Toronto, ON, Canada, where he also holds a Canada Research Chair. His main research interests include multiuser information theory, coding, optimization, wireless communications, and broadband access networks.

Dr. Yu is currently an Associate Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and a Guest Editor for a Special Issue on Nonlinear Optimization in Communications of the IEEE JOURNAL ON SELECTED AREAS OF COMMUNICATIONS.

Marc Moonen (M'94–SM'04) received the electrical engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium, in 1986 and 1990, respectively.

Since 2004, he has been a Full Professor with the Electrical Engineering Department, KU Leuven, where he is currently heading a research team of 16 Ph.D. candidates and postdocs, working in the area of signal processing for digital communications, wireless communications, DSL, and audio signal processing. He has been Editor-in-Chief for the *EURASIP Journal of Applied Signal Processing* since 2003, and is a member of the editorial boards of *Integration, VLSI Journal,* and the *EURASIP Journal on Wireless Communications and Networking*

Dr. Moonen was the recipient of the 1994 KU Leuven Research Council Award and the 1997 Alcatel Bell (Belgium) Award (with P. Vandaele) and was a 1997 Laureate of the Belgium Royal Academy of Science. He was chairman of the IEEE Benelux Signal Processing Chapter from 1997 to 2002 and has been an AdCom Member of the European Association for Signal, Speech, and Image Processing (EURASIP) since 2000. He is a member of the editorial board of the *IEEE Signal Processing Magazine*. He was on the editorial board of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: ANALOG AND DIGITAL SIGNAL PROCESSING from 2002 to 2003.

Jan Verlinden received the electrical engineering degree from the Katholieke Universiteit Leuven (KU Leuven), Leuven, Belgium, in 2000.

He is currently a member of the DSL Expert Team of the Access Network Division, Alcatel, Belgium. He joined the Research and Innovation Division of Alcatel in September 2000, where he focussed on echo cancellation techniques. Since 2002, he has focused on dynamic spectrum management (DSM). As such, he participated in the VDSL Olympics by introducing DSM into the VDSL prototype. Within the DSL Experts Team, he is currently studying emerging DSL physical layer technologies. He also contributes to ANSI NIPP-NAI standardization.

Tom Bostoen received the M.S. degree in physical engineering from Ghent University, Ghent, Belgium, in 1998.

Since 1998, he has been with the Research and Innovation Department, Alcatel, Antwerp, Belgium. His research interests include time-domain reflectometry, multiuser transmission theory, interference mitigation, and their applications to digital subscriber line technology.