SCALE: A Low-Complexity Distributed Protocol for Spectrum Balancing in Multiuser DSL Networks

John Papandriopoulos, *Member, IEEE*, and Jamie S. Evans, *Member, IEEE*

Abstract—Dynamic spectrum management of digital subscriber lines (DSLs) has the potential to dramatically increase the capacity of the aging last-mile copper access network. This paper takes an important step toward fulfilling this potential through power spectrum balancing. We derive a novel algorithm called SCALE, that provides a significant performance improvement over the existing iterative water-filling (IWF) algorithm in multiuser DSL networks, doing so with comparable low complexity. The algorithm is easily distributed through measurement and limited message passing with the use of a spectrum management center. We outline how overhead can be managed, and show that in the limit of zero message-passing, performance reduces to IWF.

Numerical studies indicate that SCALE converges extremely fast when applied to VDSL, with performance exceeding that of IWF in just a few iterations, and to over 90% of the final rate in under five iterations. We provide a proof to show that SCALE converges to a Karush–Kuhn–Tucker (KKT) point, suggesting that it indeed has the potential to reach true global optimality.

Finally, we return to the IWF problem and derive a new algorithm named SCAWF that is shown to be a very simple way to water-fill, particularly suited to the multiuser context.

Index Terms—Digital subscriber line, dynamic spectrum management, interference channel, iterative water-filling, successive convex optimization.

I. INTRODUCTION

DIGITAL subscriber line (DSL) technology has helped quench our thirst for bandwidth in recent years, extending the life of existing copper twisted-pair networks that now serve over 100 million subscribers around the globe with broadband internet connectivity. While DSL technology has been hugely successful, incumbent telephone operators are increasingly faced with stiff competition from the decreasing cost of optical fiber-fed leased lines, and aggressive cable television companies serving subscribers from much higher bandwidth Hybrid Fiber-Coax (HFC) cable networks. As a

(a) Downstream with remote terminal deployment.

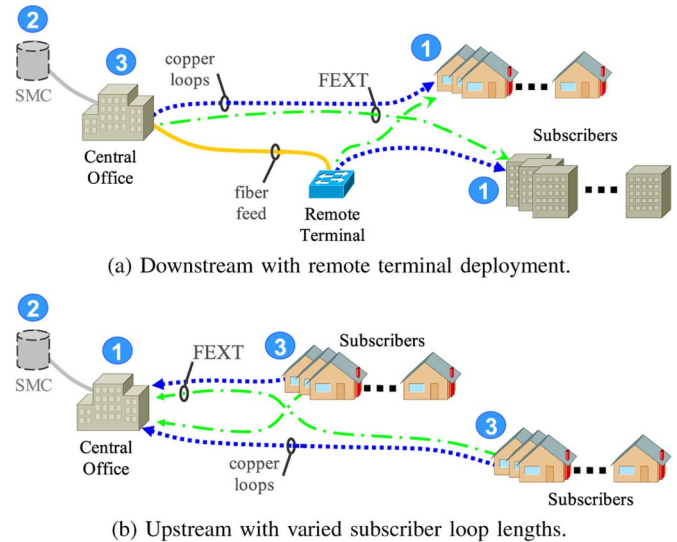(b) Upstream with varied subscriber loop lengths.

Fig. 1. Network topologies where DSM significantly improves performance.

result, telephone companies are in desperate need for increased bit rates on existing DSL lines in an effort to further extend the life of their aging copper plants.

Crosstalk and long loop lengths are the obstacles toward reaching these higher rates. Twisted-pairs are bundled together in groups of 25–100 lines in ducts directed toward subscribers. Lines are sufficiently close such that electromagnetic radiation induces crosstalk coupling between pairs. Near-end crosstalk (NEXT) is caused by transmitters interfering with receivers on the same side of the bundle and is often avoided by using nonoverlapping transmit and receive spectra (FDD) or disjoint time intervals (TDD). Far-end crosstalk (FEXT) is caused by transmitters on opposite sides of the bundle (see Fig. 1). In some cases, this interference can be 10–20 dB larger than the background noise and has been identified as the dominant source of performance degradation in DSL systems [1].

Telephone companies are increasingly shortening the loop using remote terminal (RT) deployments (see Fig. 1(a)), resulting in lower signal attenuation and larger available bandwidths. Unfortunately, this can cause other problems such as the "near–far effect," due to the crosstalk. Common in code-division multiple-access (CDMA) wireless, the near–far effect occurs when a user enjoying a good channel close to the receiver overpowers the received signal of a user further away having a worse channel and where both users transmit at comparable power levels.

Two competing directions to the crosstalk impairment are known: vectored DSL and spectrum balancing. Each falls under the umbrella of dynamic spectrum management (DSM);

see [2] for an overview. Vectoring treats the DSL network as a multiple-input multiple-output (MIMO) system, where modems must coordinate at the signal level to effectively *remove* crosstalk through successive decoding or precoding. In contrast, spectrum balancing involves a much looser level of coordination: much like existing systems, modems employ single-user encoding and decoding (treating interference as noise), however, they may also interact on a more granular time scale to negotiate a spectrum allocation that effectively *avoids* crosstalk as much as possible to improve the overall performance of the network.

This paper is concerned with the balancing of users' power spectral densities (PSDs), explicitly taking crosstalk effects into account. A significant improvement in network capacity is possible by such a judicious allocation of users' power, and especially so in near–far situations as pictured in Fig. 1.

Early work in this area introduced an iterative water-filling (IWF) scheme to balance user PSDs, where each user repeatedly measures the aggregate interference received from all other users, then greedily water-pours their own power allocation without regard for the impact to be had on other users [3]. This process results in a fully distributed and autonomous algorithm having a reasonable computational complexity.

More recent efforts have focused on the underlying optimization problem that spectrum balancing aims to solve. Unfortunately, this optimization (introduced in Section III) is a difficult nonconvex problem. As such, the optimal spectrum balancing (OSB) algorithm [1] makes use of a grid-search to find the optimal power allocation to a predetermined quantum. It suffers from an exponential complexity in the number of users, and so near-optimal iterative spectrum balancing (ISB) algorithms were developed that reduce complexity through a series of line-searches, avoiding the grid-search bottleneck [4], [5]. Appearing simultaneously with the conference version of this work [6] was another variation on OSB making use of a branch-and-bound technique [7]. All of these algorithms are centralized and it is unclear how well-suited they are for practical implementation.

In this paper, we return to the underlying nonconvex spectrum balancing optimization, and show that it is an NP-hard problem. We then apply a novel technique involving a series of convex relaxations to derive an algorithm called SCALE (**S**uccessive **C**onvex **A**pproximation for **L**ow compl**E**xity). We show through numerical simulation that SCALE performs significantly better than IWF, and with comparable complexity.

An important feature of SCALE is that it may be distributed with the help of a spectrum management center (SMC). The resulting method may be viewed as a distributed computation across the DSL network, in contrast to the centralized OSB and ISB schemes. Importantly, we outline how the overhead associated with this approach can be managed, and show that it degrades gracefully to the same performance as that attained by IWF when no inter-user communication facility is available. Finally, we take a fresh look at IWF that leads to the development of a new algorithm called SCAWF (**S**uccessive **C**onvex **A**pproximation for **W**ater-**F**illing). Our development simplifies existing IWF approaches and enjoys a very low implementation complexity.

To summarize, the key contributions of this paper are as follows.

1) We show that the nonconvex spectrum balancing problem, previously thought to be devoid of any convex structure, comprises so-called d.c. (difference of convex) functions. Consequently, we reveal that the problem is inherently NP-hard.

2) A novel algorithm that attempts to solve the nonconvex spectrum balancing problem by exploiting its underlying convexity. Called SCALE (**S**uccessive **C**onvex **A**pprox-imation for **L**ow compl**E**xity), the method involves a sequence of convex relaxations. We prove that the sequence will always converge, to a solution satisfying the necessary Karush–Kuhn–Tucker (KKT) condition for optimality, and with only a mild condition on rate-adaptive users' targets rates, that can typically be relaxed in practice.

3) An associated distributed protocol whereby modems make local decisions based on measurement and limited message-passing with a spectrum management center. This protocol distributes computation across the DSL network, further lowering the computational requirements of the method.

4) We show that SCALE will always perform as well as iterative water-filling. In a realistic very high speed DSL (VDSL) scenario, we provide numerical results to show that SCALE is very nearly globally optimum, significantly outperforming iterative water-filling with only a small handful of iterations.

5) A new family of algorithms for water-filling based on the relaxation technique employed in SCALE. Unlike the well-established water-filling methods that make use of channel sorting or bisection, we propose a purely iterative scheme having low complexity. Consequently, our algorithms are perfect for iterative water-filling: we do not perform a complete (and expensive) water-filling computation at each stage. Instead, each user iterates together with others until the system reaches the simultaneous water-filling condition.

6) A simple and practical approach to discrete bit-loading, where the loading is layered on top of the power allocation produced by the proposed algorithms.

This paper is organized as follows. In Section II, we introduce the system model. Section III introduces the general form of the spectrum balancing problem and reviews known attempts at its solution. Our novel approach is presented in Section IV, where the SCALE algorithm is derived and a distributed protocol for implementation outlined. Techniques for further enhancing its implementation efficiency are also provided.

We revisit IWF in Section V, where direct comparisons with the SCALE protocol unveil new insights into its operation. The results of that section are further shaped in Section VI, where we derive the SCAWF algorithms—a very simple and low-complexity method for water-filling, particularly useful in the IWF context.

A layered approach for discrete loading is presented in Section VII. Numerical performance evaluations follow in Section VIII, where we highlight the speed of convergence and performance improvements that SCALE can provide over

IWF. We also include an investigation into whether the results provided by SCALE are globally optimum. Our conclusions are outlined in Section IX.

II. SYSTEM MODEL

We adopt a standard model for a K user xDSL system employing discrete multitone (DMT) modulation where each user has N tones available that are used to form a set of N intersymbol interference (ISI)-free orthogonal subchannels. We make the usual assumption that users are aligned in frequency such that FEXT coupling occurs on a common tone-by-tone basis.

A fixed band-plan is assumed for simplicity, that partitions each of these tones into separate up- and downstream bands that are the same for all users. While it is known that such a scheme is not optimal [5], partitions are a common way to avoid NEXT. The algorithms developed in this paper are easily extended to include NEXT coupling if required.

We consider continuous bit-loading where the achievable loading on tone n, user k is

$$b_k^n(\boldsymbol{P}^n) \triangleq \log\left(1 + \mathrm{SIR}_k^n(\boldsymbol{P}^n)\right) \quad (1)$$

in the units of nats, and where the corresponding signal-to-interference ratio (SIR) is defined as

$$\mathrm{SIR}_k^n(\boldsymbol{P}^n) \triangleq \frac{G_{kk}^n P_k^n}{\sum_{j \neq k} G_{kj}^n P_j^n + \sigma_k^n}. \quad (2)$$

We denote by P_k^n the transmitter power of user k on tone n. For notational convenience, we write $\boldsymbol{P}^n = [P_1^n, P_2^n, \ldots, P_K^n]^T$ as the K-length vector of all transmitter powers on tone n. We will also make use of the notation $\boldsymbol{P}_k = \left[P_k^1, P_k^2, \ldots, P_k^N\right]$ as the N-length PSD vector of user k. The $K \times N$ matrix $\boldsymbol{P}$ is produced by stacking these vectors in the obvious way. This notation makes clear the explicit dependence of the SIR on power. In the sequel, vector- or matrix-inequalities are always element-wise. The gains G_{kj}^n model the channel power transfer on tone n from user j to the receiver of user k. For further notational convenience, we assume the gains G_{kk}^n have been normalized by an appropriate signal-to-noise ratio (SNR)-gap Γ_k^n, that depends on the coding scheme, target probability of error, and noise margin [8].

Each σ_k^n models the received noise power on tone n. We assume the noise powers are constant, modeling receiver thermal noise plus any alien noise injected by other coexisting systems (e.g., HDSL, ISDN, RF noise, etc.).

The achievable rate for user k is then

$$R_k(\boldsymbol{P}) \triangleq \sum_{n=1}^{N} b_k^n(\boldsymbol{P}^n) = \sum_{n=1}^{N} \log\left(1 + \mathrm{SIR}_k^n(\boldsymbol{P}^n)\right) \quad (3)$$

nats per channel use.

We will assume that all K users are coupled through interference to some degree. This is without loss of generality, as any user population can always be partitioned into subgroups of noninterfering users, with each independent subgroup considered separately.

III. THE SPECTRUM BALANCING PROBLEM

The spectrum balancing problem has many forms, categorized by the *rate adaptive* (RA) and *fixed margin* (FM) formulations. The RA problem seeks to maximize the data rate of each user, subject to per-user maximum power constraints. It is inherently a multicriterion optimization problem, where one has the ability to scalarize the rates of each user, forming a weighted sum objective (see below).

The FM problem is concerned with finding a minimal power allocation such that each user has a minimum (or target) data rate that is attained. These target rates must be feasible; that is, there exists a power allocation whereby the maximum power constraint of each user is not violated.

The set of feasible target rates is contained within the so-called rate region. Its boundary corresponds to the Pareto optimal surface of the RA problem, and is often explored by sweeping values of the weights and solving a sequence of such problems. Alternatively, given a set of target rates, one could solve a suitable feasibility problem to determine whether the supplied rates can be supported by the network.

In its most general form, the spectrum balancing problem is written

$$\max_{0 \leq \boldsymbol{P} \leq \boldsymbol{L}} \quad \sum_{k \in \mathrm{RA}} \omega_k R_k(\boldsymbol{P}) - \mathbf{1}_{\mathrm{RA}=\varnothing} \sum_{k \in \mathrm{FM}} \sum_{n=1}^{N} P_k^n \quad (4a)$$

$$\mathrm{s.t.} \quad R_k^{\mathrm{target}} \leq R_k(\boldsymbol{P}), \quad k \in \mathrm{FM} \quad (4b)$$

$$\sum_{n=1}^{N} P_k^n \leq P_k^{\mathrm{max}}, \quad k = 1, \ldots, K \quad (4c)$$

where P_k^{max} is the maximum power constraint of user k and $\mathbf{1}_A$ is an indicator of the event A, taking value 1 when A is true and 0 otherwise. The sets RA and FM denote the index sets of RA and FM users, respectively. By a simple manipulation, the individual RA and FM problems can be recovered as special cases.

Each user $k \in \mathrm{RA}$ is assigned a fixed positive scalarization weight ω_k that allows a tradeoff between the rates allocated to each user. Equivalently, these weights allow the system operator to place differing quality of service (QoS) or importance levels on each user. For example, in accordance with the premium paid by the user for their service.

Every other user $k \in \mathrm{FM}$ is assigned a positive minimum data rate R_k^{target} that must be achieved. These minimum rates are also provided by the system operator and are considered fixed.

The PSD masks $\boldsymbol{L}$ are included for satisfaction of regulatory requirements. They specify the maximum power level L_k^n for user k on tone n. Although they do not add a level of difficulty to the problem, we omit them for brevity in the sequel.[1]

This optimization problem is constructed to maximize the rates of RA users as much as possible while simultaneously ensuring the target rates of FM users are met exactly. For later convenience and to ensure the latter condition, we include the double summation in the objective. Without it, the inequality (4b) should be replaced by an equality.

---

[1] These PSD masks are so-called box-constraints on each power P_k^n. Full treatment is found in [9], [10].

A. Related Work

The individual RA and FM problems have been extensively studied for single-user $K = 1$ systems, for both continuous and discrete bit-loading. Algorithms enjoying $O(N \log N)$ or $O(N)$ complexity are well-established, their solutions usually involving some kind of water-pouring; see [11] and references therein.

For the multiuser case, where $K > 1$, optimization (4) becomes difficult, because of the presence of the achievable rate (3) terms in the objective and constraints. Further insight is gained by rewriting (3) as follows:

$$R_k(\boldsymbol{P}) = \sum_{n=1}^{N} \log\left(\sum_j G_{kj}^n P_j^n + \sigma_k^n\right) - \log\left(\sum_{j \neq k} G_{kj}^n P_j^n + \sigma_k^n\right).$$

Observe that the achievable rate comprises a difference of concave functions in $\boldsymbol{P}$. Optimization problems such as (4) having d.c. structure have been of great interest to the optimization community for the past $30+$ years. Unfortunately these problems are NP-hard [12] and often difficult to solve efficiently for the global optimum.

The IWF approach finds an approximate solution by splitting the problem into K convex subproblems, then iterating over these until convergence. Each subproblem concerns only the powers $\boldsymbol{P}_k$, fixing all other powers $\boldsymbol{P}_{j \neq k}$ and treating their contributions as noise (see Section V). These subproblems are made distributed through SIR measurements. IWF has been shown to converge to a competitive Nash equilibrium [3], and is amenable to practical implementation [13].

A very different approach is made in OSB that attempts to solve an optimization similar to (4) directly [1]. The innovation was to formulate the Lagrangian dual problem. It was then possible to iterate over N separate subproblems for fixed Lagrange multipliers, each subproblem concerning only user powers $\boldsymbol{P}^n$ on tone $n \in [1, N]$. Each subproblem is solved with a brute-force grid-search having $\mathcal{L} = P^{\max}/\Delta_P$ quantized power levels, requiring at least $\mathcal{L}^K$ operations each. An outer loop then updated the Lagrange multipliers via bisection or gradient-based methods.

Although OSB has an *inherent* exponential complexity in the number of users, it has shown significant performance gains are possible over IWF. Recent improvements have been introduced sporting lower complexity, achieved by replacing the grid-search with a sequence of line-searches [4], [5] or a branch-and-bound method [7]. In general, a large number of computations are employed in an attempt to solve each per-tone subproblem completely before moving on to the next: still a large computational hit!

Until now, there has been little in the way of a very low complexity algorithm making use of measurement-based updates devoid of explicit line- or grid-searching. Ideally, such an algorithm would have its computation distributed, with little or no message-passing between modems.

IV. SUCCESSIVE CONVEX APPROXIMATION FOR LOW-COMPLEXITY (SCALE)

Our approach considers a relaxation of the nonconvex problem (4) that avoids the d.c. structure. We will make use of the following lower bound:

$$\alpha \log z + \beta \leq \log(1 + z) \tag{5}$$

that is tight at $z = z_0$ when the approximation constants are chosen as

$$\alpha = \frac{z_0}{1 + z_0} \tag{6a}$$

$$\beta = \log(1 + z_0) - \frac{z_0}{1 + z_0} \log z_0. \tag{6b}$$

The following relaxation to the spectrum balancing problem (4) results:

$$\max_{\boldsymbol{P} \geq 0} \quad \sum_{k \in \mathrm{RA}} \omega_k \check{R}_k(\boldsymbol{P}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) - \mathbf{1}_{\mathrm{RA} = \varnothing} \sum_{k \in \mathrm{FM}} \sum_n P_k^n$$
$$\text{s.t.} \quad R_k^{\mathrm{target}} \leq \check{R}_k(\boldsymbol{P}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k), \quad k \in \mathrm{FM}$$
$$\sum_n P_k^n \leq P_k^{\max}, \quad \forall k \tag{7}$$

where the approximation vectors $\boldsymbol{\alpha}_k = [\alpha_k^1, \ldots, \alpha_k^N]$ and $\boldsymbol{\beta}_k = [\beta_k^1, \ldots, \beta_k^N]$ are fixed for each user k, and we define

$$\check{R}_k(\boldsymbol{P}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) \triangleq \sum_n \alpha_k^n \log\left(\mathrm{SIR}_k^n(\boldsymbol{P}^n)\right) + \beta_k^n \tag{8}$$

as a lower bound on the achievable rate for user k. Although our relaxation still has nonconvex form (since (8) has d.c. structure), the following result gives a way forward.

Lemma 1: The lower bound achievable rate (8) is concavified by the transformation $\widetilde{P}_k^n = \log P_k^n$.

Proof: With $P_k^n = \exp\left(\widetilde{P}_k^n\right)$, the lower bound achievable rate (8) becomes

$$\check{R}_k(e^{\widetilde{\boldsymbol{P}}}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k) =$$
$$\sum_n \alpha_k^n \left[\log G_{kk}^n + \widetilde{P}_k^n - \log\left(\sum_{j \neq k} G_{kj}^n e^{\widetilde{P}_j^n} + \sigma_k^n\right)\right] + \beta_k^n$$

where $\exp(\boldsymbol{x})$ denotes an element-by-element operation on the vector $\boldsymbol{x}$. Concavity follows from the sum of linear and concave terms within the square brackets (where we note that log-sum-exp is convex [14]). $\square$

Lemma 1 ensures that the transformed relaxation

$$\max_{\widetilde{\boldsymbol{P}}} \quad \sum_{k \in \mathrm{RA}} \omega_k \check{R}_k\left(e^{\widetilde{\boldsymbol{P}}}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k\right) - \mathbf{1}_{\mathrm{RA} = \varnothing} \sum_{k \in \mathrm{FM}} \sum_n e^{\widetilde{P}_k^n}$$
$$\text{s.t.} \quad R_k^{\mathrm{target}} \leq \check{R}_k\left(e^{\widetilde{\boldsymbol{P}}}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k\right), \quad k \in \mathrm{FM}$$
$$\sum_n e^{\widetilde{P}_k^n} \leq P_k^{\max}, \quad \forall k \tag{9}$$

is devoid of all d.c. structure. We now have a standard concave maximization problem.

We derive an algorithm to solve this convex relaxation in Section IV-A using gradient methods that are computationally effi-

$$\omega_k \alpha_k^n - P_k^n \left[\lambda_k + \sum_{\substack{j \in \mathrm{RA} \\ j \neq k}} \omega_j \alpha_j^n \frac{\mathrm{SIR}_j^n(\boldsymbol{P})}{P_j^n G_{jj}^n} G_{jk}^n + \sum_{j \in \mathrm{FM}} \mu_j \alpha_j^n \frac{\mathrm{SIR}_j^n(\boldsymbol{P})}{P_j^n G_{jj}^n} G_{jk}^n \right] = 0, \quad k \in \mathrm{RA} \tag{10a}$$

$$\mu_k \alpha_k^n - P_k^n \left[\mathbf{1}_{\mathrm{RA}=\varnothing} + \mu_k + \sum_{j \in \mathrm{RA}} \omega_j \alpha_j^n \frac{\mathrm{SIR}_j^n(\boldsymbol{P})}{P_j^n G_{jj}^n} G_{jk}^n + \sum_{\substack{j \in \mathrm{FM} \\ j \neq k}} \mu_j \alpha_j \frac{\mathrm{SIR}_j^n(\boldsymbol{P})}{P_j^n G_{jj}^n} G_{jk}^n \right] = 0, \quad k \in \mathrm{FM}. \tag{10b}$$

cient, and without the need for a brute-force or heuristic search of any kind. Once a solution is obtained, we may transform back to the P-space with $P_k^n = \exp(\widetilde{P}_k^n)$.

Here we are optimizing a lower bound on the achievable rate: to either maximize the aggregate of all RA users, or when there are none, minimize the aggregate power of all FM users. It then becomes natural to improve these bounds successively, resulting in the following procedure:

1: Initialize all $\boldsymbol{\alpha}_k^{(0)} = \mathbf{1}, \boldsymbol{\beta}_k^{(0)} = \mathbf{0}$ (a high-SIR approximation)

2: Initialize iteration counter $t = 1$

3: **repeat**

4:     *Maximize*: solve subproblem (9) to give solution
    $$\boldsymbol{P}^{(t)} = \exp\left(\widetilde{\boldsymbol{P}}^{(t)}\right)$$

5:     *Tighten*: update elements of $\boldsymbol{\alpha}_k^{(t+1)}, \boldsymbol{\beta}_k^{(t+1)}$ using (6a) at
    $z_0 = \mathrm{SIR}_k^n(\boldsymbol{P}^{(t)})$

6:     Increment t

7: **until** convergence

One caveat results from our choice of initial approximation constants. We require that the set of rate targets $\left\{ R_k^{\mathrm{target}} | k \in \mathrm{FM} \right\}$ are at least "high-SIR feasible." By this we mean that, for the specified rate targets, there exists a feasible solution to the first subproblem that is based on a high-SIR approximation. We later discuss why this requirement can be relaxed in practice.

Lemma 2: All target rate constraints are active at the optimum solution of (9).

Proof: Suppose the feasible solution $\boldsymbol{P}^{(t)}$ is optimal, where at least one user j has a rate target constraint that is inactive.

When there are no RA users, we could reduce the power allocated to user j while still remaining feasible, in turn improving the objective and so our solution was clearly not optimum. Alternatively, when one or more RA users are present, their data rate could be increased, again improving the objective, until the rate target constraint of user j becomes active—the same conclusion is reached. □

Theorem 1: Under the assumption of high-SIR feasible rate targets, the sequence of iterates produces a monotonically increasing objective and always converges. At the converged solution, the lower bound (8) is equal to the actual achievable rate (3).

Proof: The existence of a feasible solution $\boldsymbol{P}^{(1)}$ to the first iteration is assured, by the high-SIR feasible rate targets assumption. For subsequent iterations $t > 1$, the relations

$$R_k^{\mathrm{target}} \stackrel{\mathrm{(a)}}{=} \breve{R}_k \left(\boldsymbol{P}^{(t-1)}; \boldsymbol{\alpha}_k^{(t-1)}, \boldsymbol{\beta}_k^{(t-1)} \right)$$
$$\stackrel{\mathrm{(b)}}{\leq} R_k \left(\boldsymbol{P}^{(t-1)} \right)$$
$$\stackrel{\mathrm{(c)}}{=} \breve{R}_k \left(\boldsymbol{P}^{(t-1)}; \boldsymbol{\alpha}_k^{(t)}, \boldsymbol{\beta}_k^{(t)} \right)$$

hold for each user $k \in FM$ Equality (a) follows from Lemma 2, inequality (b) follows from the definition of the bound (5), while equality (c) is a consequence of the tightening step ("T-step") and (6). Together, these results ensure that iterate $\boldsymbol{P}^{(t-1)}$ is a feasible point of the following subproblem t.

We therefore conclude that the maximization ("M-step") will either improve the objective on the tth iteration, or remain at the same point as the previous $(t-1)$th iterate (since it is feasible).

The T-step can only lead to an improved objective on the next iteration t: for FM users, this means that tightening creates a glut of excess data-rate that is not optimum by Lemma 2 (cf. equality (a) above); while for RA users, tightening may improve the objective value at the feasible point $\boldsymbol{P}^{(t-1)}$, with further potential of improvement during the following M-step. Convergence is brought about when the T-step maps onto itself: exactly the same approximation vectors are selected for the next iteration. Clearly then, performing an M-step would be futile as the solution would also remain the same. Moreover, the lower bound (8) is exactly equal to the actual achievable rate (3) at this point.

Convergence is guaranteed by the monotonically improving objective, bounded above by the global optimum of the original spectrum balancing problem (4), and since each subproblem remains within its feasible region by virtue of our lower bound (8) on the achievable rate. □

One consequence of Theorem 1 is that each subproblem need not be maximized fully; only an improved feasible solution is required. This lends itself toward a distributed T-step: each user need not wait until convergence of subproblem t, each tightens at periodic intervals whenever their constraints are not violated. Moreover, each T-step requires only local information—a simple measurement of the SIR. Our next result deals with a second consequence.

Corollary 1: When the sequence of iterates converges, it does so to a feasible power allocation that satisfies the necessary KKT optimality conditions of the nonconvex spectrum balancing problem (4).

Proof: The KKT conditions for each convex subproblem (9) can be written as (10) (see the top of the page) after reverting back to the original $\boldsymbol{P}$-space, and where $\{\boldsymbol{P}, \boldsymbol{\mu}, \boldsymbol{\lambda}\}$ is a primal-dual triplet satisfying primal feasibility and complementary slackness.

Denote the solution of the final converged subproblem as $\boldsymbol{P}^*$. Then by [15, Proposition 3.3.1], there exists Lagrange multipliers $\{\boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ that, together with $\boldsymbol{P}^*$, satisfy the KKT conditions above as well as feasibility and complementary slackness (regularity is assured by the convexity of the subproblem). Furthermore, Theorem 1 tells us that the lower bound approximation is exact at this point. From the definition of the approximation constants (6), it follows that

$$\alpha_k^{n*} = \frac{SIR_k^n(\boldsymbol{P}^*)}{1 + SIR_k^n(\boldsymbol{P}^*)} \tag{11}$$

for every tone n associated with every user k.

The proof is completed by recognizing that (10) has exactly the same form as the corresponding KKT conditions of (4) when α_k^{n*} is substituted as shown in (11)—and thus the triplet $\{\boldsymbol{P}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ satisfies both sets of KKT conditions. $\square$

Corollary 1 tells us that the sequence of iterates guarantees at least a local solution to the spectrum balancing problem (4). However, it also has the potential to achieve true global optimality. We investigate further in Section VIII-A.

In the sequel, we refer to this procedure as **S**uccessive **C**onvex **A**pproximation for **L**ow compl**E**xity (SCALE). Section IV-B outlines its realization, leading to a *scalable* protocol for spectrum balancing where computation is *distributed* across the network.

A. Subproblem Solution

We begin by deriving an algorithmic solution to subproblem (9) via the dual problem $\min_{\boldsymbol{\mu}, \boldsymbol{\lambda} \geq 0} q_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\lambda})$, where $\boldsymbol{\mu} = \{\mu_k | k \in \mathrm{RA}\}$ and $\boldsymbol{\lambda} = [\lambda_1, \dots, \bar{\lambda}_K]^T$ are Lagrange multipliers. The dual function is given by

$$q_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\lambda}) \triangleq \max_{\widetilde{\boldsymbol{P}}} L_{\mathbb{S}}(\widetilde{\boldsymbol{P}}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \tag{12}$$

and the corresponding Lagrangian is

$$\begin{aligned} L_{\mathbb{S}}(\widetilde{\boldsymbol{P}}, \boldsymbol{\mu}, \boldsymbol{\lambda}) \triangleq & \sum_{k \in \mathrm{RA}} \omega_k \breve{R}_k\left(e^{\widetilde{\boldsymbol{P}}}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k\right) - \mathbf{1}_{\mathrm{RA}=\varnothing} \sum_{k \in \mathrm{FM}} \sum_n e^{\widetilde{P}_k^n} \\ & + \sum_{k \in \mathrm{FM}} \mu_k \breve{R}_k\left(e^{\widetilde{\boldsymbol{P}}}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k\right) - \mu_k R_k^{\mathrm{target}} \\ & - \sum_k \lambda_k \left\{ \sum_n e^{\widetilde{P}_k^n} - P_k^{\max} \right\} \end{aligned} \tag{13}$$

where we recall that the approximation vectors $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ are fixed.

Since the Lagrangian (13) is strictly concave in $\widetilde{\boldsymbol{P}}$, the inner maximization (12) has a unique solution. By [15, Proposition 6.1.1], the dual function $q_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\lambda})$ is differentiable everywhere, and we can employ the simultaneous gradient-descent

$$\lambda_k^{(s+1)} = \left[\lambda_k^{(s)} + \epsilon_\lambda \left\{ \sum_n P_k^{n(s)} - P_k^{\max} \right\} \right]^+ \tag{14a}$$

$$\mu_k^{(s+1)} = \left[\mu_k^{(s)} + \epsilon_\mu \left\{ R_k^{\mathrm{target}} - \breve{R}_k\left(\boldsymbol{P}^{(s)}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k\right) \right\} \right]^+ \tag{14b}$$

to solve the outer minimization, where $\epsilon_{\{\mu, \lambda\}}$ are sufficiently small step sizes, $[\cdot]^+ = \max(0, \cdot)$, and s is an iteration number. We denote by $\boldsymbol{P}^{(s)} = \exp(\widetilde{\boldsymbol{P}}^{(s)})$ the maximizer of the Lagrangian with multipliers fixed at iteration s.

Multipliers λ_k and μ_k are conveniently updated by the respective user k using only local information. As with many other Lagrangian dual formulations, the multipliers have a pricing interpretation: as the maximum-power constraint is violated, the price λ_k goes up and *vice versa* until the equilibrium price $\boldsymbol{\lambda}^*$ is reached; and similarly for μ_k with respect to the target rate. At equilibrium, the dual subproblem has been solved.

The dual function (12) is evaluated by finding the stationary point of the Lagrangian (13) with multipliers fixed. This equates to finding a solution to the coupled KKT equations (10). On closer inspection, we may rewrite these with

$$\delta_k = \begin{cases} \omega_k, & k \in \mathrm{RA} \\ \mu_k, & k \in \mathrm{FM} \end{cases} \tag{15}$$

resulting in the unified KKT equation

$$P_k^n = \frac{\delta_k \alpha_k^n}{\mathbf{1}_{\substack{k \in \mathrm{FM} \\ \mathrm{RA}=\varnothing}} + \lambda_k + \sum_{j \neq k} \delta_j \alpha_j^n \frac{\mathrm{SIR}_j^n(\boldsymbol{P})}{P_j^n G_{jj}^n} G_{jk}^n} \tag{16}$$

that has been further rearranged in terms of P_k^n. This is a fixed-point equation, since P_k^n also appears in the denominator of each SIR term, and convenient too, as it forms the basis for our iterative solution of the Lagrange maximization (12) whereby the right-hand side is used to update the power of user k on tone n.

As a power update, (16) admits an elegant interpretation: the gains G_{jk}^n indicate the impact user k has on all other users j, tone n. Power is allocated in such a way so that it *takes other users into account* on a tone-by-tone basis, rather than a selfish allocation as is done in IWF. Further, on violation of the power constraint, the price of power λ_k increases, lowering the allocation to an appropriate level within the budget.

Lemma 3: The fixed-point power update based on (16) always converges to the maximizer of the Lagrangian (13) with multipliers fixed.

Proof: The right-hand side of (16) can be written in the following abstracted form:

$$J_k^n(\boldsymbol{P}) = \frac{A_k^n}{B_k + \sum_{j \neq k} \frac{C_j^n}{D_j^n + \boldsymbol{I}^T \boldsymbol{P}_{-j}^n}} \tag{17}$$

where A_k^n, B_k, C_j^n, D_j^n are nonnegative constants, $\boldsymbol{I}$ is a column vector of all ones, and $\boldsymbol{P}_{-j}^n$ is a $(K-1)$-length column vector equivalent to $\boldsymbol{P}^n$ with the jth element missing.

With reference to the Yates framework [16], we identify $J_k^n(\cdot)$ as the (k, n)th component of the matrix-valued interference function $\boldsymbol{J}_k^n(\cdot)$. Interference functions that are standard enjoy guaranteed convergence of the iteration $\boldsymbol{P}^{(s+1)} = \boldsymbol{J}(\boldsymbol{P}^{(s)})$ to a unique fixed-point (when it exists) from any initial $\boldsymbol{P}^{(0)}$. Updates need not be synchronized.

We are assured a stationary point of the Lagrangian (13) by its strict concavity in $\widetilde{\boldsymbol{P}}$. Therefore, a fixed-point of (17) always exists. We now proceed to prove that (17) satisfies the three properties (positivity, monotonicity, and scalability) required of such functions. Positivity follows from the nonnegativity of each component.

Taking $\boldsymbol{P} = \Psi\boldsymbol{Q}$ with $\Psi \geq 1$, monotonicity follows from

$$J_k^n(\boldsymbol{P}) = \frac{A_k^n}{B_k + \sum_{j \neq k} \frac{C_j^n}{D_j^n + \Psi \boldsymbol{I}^T \boldsymbol{Q}_{-j}^n}}$$

$$\geq \frac{A_k^n}{B_k + \sum_{j \neq k} \frac{C_j^n}{D_j^n + \boldsymbol{I}^T \boldsymbol{Q}_{-j}^n}} = J_k^n(\boldsymbol{Q}).$$

Taking $\Psi > 1$, scalability follows from

$$\Psi J_k^n(\boldsymbol{P}) = \frac{A_k^n}{\frac{1}{\Psi} B_k + \sum_{j \neq k} \frac{1}{\Psi} \frac{C_j^n}{D_j^n + \boldsymbol{I}^T \boldsymbol{P}_{-j}^n}}$$

$$> \frac{A_k^n}{B_k + \sum_{j \neq k} \frac{1}{\Psi} \frac{C_j^n}{D_j^n + \boldsymbol{I}^T \boldsymbol{P}_{-j}^n}}$$

$$> \frac{A_k^n}{B_k + \sum_{j \neq k} \frac{C_j^n}{D_j^n + \Psi \boldsymbol{I}^T \boldsymbol{Q}_{-j}^n}} = J_k^n(\Psi\boldsymbol{Q}).$$

We conclude that (16) has a unique fixed point and the given update will always converge to it. Moreover, the fixed point determines the stationary point of the Lagrangian and therefore maximizes it. $\square$

Each variable (15) is a kind of "QoS parameter"—statically assigned to RA users by the system operator, and dynamically adjusted for FM users—controlled by "market forces" to ensure they are of sufficiently high priority for realization of their target rate. This interpretation explains why SCALE can, in practice, converge to a solution even when the set of rate targets are not "high-SIR feasible," as required by Theorem 1. In this sense, SCALE treats **all users** as if they were RA; their weights δ_k are adjusted by an auxiliary process. Solutions to each intermediate subproblem are thus always feasible whenever the power constraints are met. This is precisely our motivation for including a maximum power constraint for **all users** in our statement of the spectrum balancing problem (4); even though these constraints are not technically required for FM users. Strictly speaking, the (general) feasibility of the rate-targets should be enough to ensure the (then implicit) power constraints are met for associated FM users.

In practice, we do not need to fully maximize the Lagrangian in (12) before updating the multipliers $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$. A single ascent step is sufficient, that equates to one iteration of (16). This observation is the key that enables a low-complexity implementation that is comparable to that of IWF.

B. The SCALE Protocol

The SCALE algorithm just described can be implemented centrally at an SMC, where *a priori* knowledge of all direct

and crosstalk channels is required. This channel state can either be measured (see, for example, [17]) or derived approximately with standard models (e.g., [18]) and knowledge of the loop topology. The computed PSD for each user would then be fed back to its associated transmitting modem, perhaps periodically, to track changes in the network such as modems turning on and off.

This centralized approach requires message-passing between user modems and the SMC for distribution of the computed PSDs and collection of live channel measurements where applicable. Unlike other centralized methods, however, we can take our developments one step further to create a protocol whereby computation of the PSDs is distributed across the network, through a different kind of message-passing that incorporates per-user local measurement. This has the advantage of a reduction in the computational requirements of individual network elements, thereby further enhancing the scalability of the scheme, while also incorporating live channel conditions through the measurement process.

To this end, we rewrite the power update (16) as

$$P_k^{n(s+1)} = \frac{\delta_k \alpha_k^n}{\lambda_k^{(s)} + \mathcal{M}_k^{n(s)}} \tag{18}$$

where $\mathcal{M}_k^{n(s)} \in \mathbb{R}_+$ is a message passed to user k from the SMC, defined as

$$\mathcal{M}_k^{n(s)} = \mathbf{1}_{\substack{k \in \mathrm{FM} \\ \mathrm{RA} = \varnothing}} + \sum_{j \neq k} G_{jk}^n \mathcal{N}_j^{n(s)} \tag{19}$$

and is formed by a weighted sum calculation at the SMC. For this, we require access to estimates of the crosstalk channels G_{jk}^n, obtained through measurement or with crosstalk models and knowledge of the loop topology, in much the same way as the centralized approach described above. Loop makeup information is also useful for determining the 0–1 state of the indicator function value. The terms $\mathcal{N}_j^{n(s)} \in \mathbb{R}_+$ are also messages, from every other user $j \neq k$ on tone n to the SMC

$$\mathcal{N}_j^{n(s)} = \delta_j \alpha_j^n \frac{\mathrm{SIR}_j^n(\boldsymbol{P}^{n(s)})}{G_{jj}^n P_j^{n(s)}} = \frac{\delta_j \alpha_j^n}{\sum_{j \neq k} G_{kj}^n P_j^{n(s)} + \sigma_k^n} \tag{20}$$

and is a local quantity at the receiver of user j: a simple scaled noise *measurement*.

With reference to Fig. 1: we summarize the main steps of the SCALE protocol.
1) Initialization on power-up:
   a) All user PSDs are set to zero: $\boldsymbol{P}_k = \boldsymbol{0}$.
   b) Multipliers are set to zero: $\lambda_k = 0, \forall\, k$ and $\delta_k = 0$, $k \in \mathrm{FM}$.
   c) Approximation vectors set to a "high-SIR approximation": $\boldsymbol{\alpha} = \mathbf{1}, \boldsymbol{\beta} = \mathbf{0}$.
2) The receiver associated with user k performs weighted noise measurements (20) and sends these to the SMC.
3) The SMC computes the weighted sums (19) for all users $k \in [1, K]$ with knowledge of the crosstalk channels, and sends these to the user's corresponding transmitter.
4) The transmitter associated with user k:

a) Updates the target rate multiplier $\delta_k = \mu_k$ according to (14b) if $k \in \mathrm{FM}$.

b) Receives SMC messages (19) and updates their PSD by alternating between the updates (18) and (14a) until convergence.

c) Tightens the approximation vectors $\boldsymbol{\alpha}_k, \boldsymbol{\beta}_k$ according to (6) at the point $z_0 = \mathrm{SIR}_k^n$ obtained by measurement. This tightening step is made periodically after $D \geq 1$ iterations for RA users; and for FM users, whenever the lower bound rate target constraint is active (or within some small ϵ).

We envisage steps 2–4 to repeat indefinitely, so that changes in the network can be tracked. These changes may include modems powering down at night, configuration updates of the FM rate targets and/or RA priority weights by the system operator, slow variations in the alien noises σ_k^n, etc.

C. Efficient Lagrange Multiplier Updates

The alternating update of step 4b above is essentially a one-dimensional search for minimum $\lambda_k \geq 0$ such that the PSD allocation remains feasible. We can improve the efficiency of this update tremendously by exploiting the notion underlying complementary slackness: when the maximum power constraint is inactive, the corresponding multiplier is $\lambda_k = 0$, otherwise, it is nonnegative.

We propose the following modified bisection rule: start with $\lambda_k = 0$; if the PSD does not exceed the maximum power constraint then it is minimum and stop. Otherwise, the PSD is infeasible and we require $\lambda_k > 0$ to reduce the PSD back into the feasible set. Search for a coarse upper bound $\overline{\lambda}_k$ by successively trying larger values (every decade, for example) until the PSD becomes feasible. Finally, perform a bisection on the interval $(0, \overline{\lambda}_k)$ to find the value of λ_k that ensures the maximum power constraint is met with equality to within a small tolerance ϵ, as given by

$$P_k^{\max} - \epsilon \leq \sum_n P_k^n \leq P_k^{\max}.$$

The target rate multiplier update of step 4a can also benefit from a more efficient update rule. We will make use of the fact that at equilibrium, the target rate constraint is always met with equality for all FM users. That is, for all $k \in \mathrm{FM}$

$$
\begin{aligned}
R_k^{\mathrm{target}} &\stackrel{(d)}{=} \sum_n \alpha_k^n \log(\mathrm{SIR}_k^n(\boldsymbol{P}^n)) + \beta_k^n \\
&\stackrel{(e)}{=} \sum_n \alpha_k^n \left[\log\left(\frac{\mathrm{SIR}_k^n(\boldsymbol{P}^n)}{\delta_k} \right) + \log \delta_k \right] + \beta_k^n \\
\delta_k &\stackrel{(f)}{=} \exp\left(\frac{R_k^{\mathrm{target}} - \sum_n \left[\alpha_k^n \log\left(\frac{\mathrm{SIR}_k^n(\boldsymbol{P}^n)}{\delta_k} \right) + \beta_k^n \right]}{\sum_n \alpha_k^n} \right)
\end{aligned}
$$
(21)

where (d) follows from (8) and Lemma 2; step (e) multiplies and divides the SIR by the QoS parameter δ_k; and (f) rearranges the expression in terms of δ_k. It does not directly depend on itself, since the $\mathrm{SIR}_k^n(\boldsymbol{P}^n)$ is directly proportional to δ_k, cf. (18) and (2).
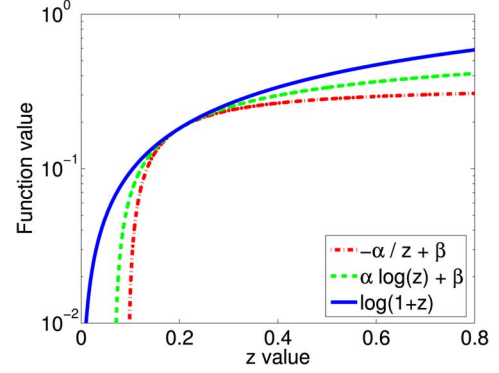

Fig. 2. Two lower bounds (dashed) that lead to convex relaxations, with $z_0 = 0.2$.

These heuristics result in attractive "zero-configuration algorithms," devoid of any step-size selection. They have never failed to produce the desired outcome throughout several numerical studies. In particular, the target rate update (21) was found to significantly speed convergence of FM users to their rate targets in numerical studies, as compared to the update (14b).

D. Other Convex Approximations

Our development hinged on a particular approximation that was used to recast the nonconvex spectrum balancing problem (4) into a convex relaxation. It turns out that our choice is just one of many. For example, the reader can verify that the lower bound

$$-\frac{\alpha}{z} + \beta \leq \log(1 + z) \tag{22}$$

is also suitable, where we recall that $z \geq 0$ is our range of interest.

In general, any lower bound that leads to a convex relaxation is a suitable candidate. Not all bounds, however, are created equal. A tighter bound can hasten convergence to a KKT point of the original nonconvex spectrum balancing problem, since the deviation at each subproblem is relatively smaller as compared to a corresponding subproblem based on a looser bound. Fig. 2 graphically illustrates the tightness of the bound employed in SCALE (5) and the example (22) above.

The bound employed in SCALE (24) is the tightest available for the relaxation to remain convex in the $\widetilde{P}$-space. To see this, we first recast the spectrum balancing problem (4) into the equivalent optimization

$$
\max_{\widetilde{\boldsymbol{P}}, \boldsymbol{t}} \quad \sum_{k \in \mathrm{RA}} \sum_n \omega_k \log(1 + e^{t_k^n}) - \mathbf{1}_{\mathrm{RA} = \varnothing} \sum_{k \in \mathrm{FM}} \sum_n e^{\widetilde{P}_k^n}
$$
$$
\mathrm{s.t.} \quad R_k^{\mathrm{target}} \leq \sum_n \log(1 + e^{t_k^n}), \quad k \in \mathrm{FM}
$$
$$
\sum_n e^{\widetilde{P}_k^n} \leq P_k^{\max}, \quad \forall k
$$
$$
t_k^n \leq \log(\mathrm{SIR}_k^n(\exp(\widetilde{\boldsymbol{P}}^n))), \quad \forall k \tag{23}
$$

where we have introduced the auxiliary variables $\boldsymbol{t} \triangleq \{t_k^n | \forall~k, n\}$. This is the so-called reverse-convex form of

the d.c. spectrum balancing problem: the objective is reverse-convex when the set RA is nonempty; the first constraint set is also reverse-convex. Only the second and third constraint sets are convex, where we recall that $\log(\mathrm{SIR}_k^n)$ is convex in the $\widetilde{P}$-space by Lemma 1.

We could form a convex relaxation to this problem by taking tangential approximations to each $\log\left(1 + e^{t_k^n}\right)$ term using the bound

$$\alpha z + \beta \leq \log(1 + e^z) \tag{24}$$

that is tight at the point $z = z_0$ when the approximation constants are chosen as

$$\alpha = \frac{e^{z_0}}{1 + e^{z_0}} \tag{25a}$$

$$\beta = \log(1 + e^{z_0}) - \frac{e^{z_0}}{1 + e^{z_0}} z_0. \tag{25b}$$

By recognizing that the value of auxiliary variables are by their definition

$$t_k^n = \log\left(\mathrm{SIR}_k^n(\exp(\widetilde{P}))\right) \tag{26}$$

we observe that this tangential approximation (24) is exactly the same as the SCALE lower bound (5) when (26) is substituted into the approximation constants (25).

So in fact, the lower bound employed in SCALE is actually a tangential approximation under a nonlinear transformation. It follows that, in that space, no other bound can be as tight and yet still result in a convex relaxation.

V. A FRESH LOOK AT ITERATIVE WATER-FILLING

A. Rate Adaptive Water-Filling (RA-WF)

Consider the RA-WF subproblem for user $k \in [1, K]$ in the standard IWF procedure, where all PSDs $\boldsymbol{P}_{j \neq k}$ are held fixed and we have choice over only $\boldsymbol{P}_k$. We can apply the same lower bound technique as before, to form the relaxation

$$\max_{\boldsymbol{P}_k \geq 0} \quad \sum_n \delta_k \check{R}_k(\boldsymbol{P}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$$
$$\text{s.t.} \quad \sum_n P_k^n \leq P_k^{\max} \tag{27}$$

where again $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$ are fixed approximation vectors adhering to (6).

This relaxation is maximized by following a similar line of development as outlined in the previous section. That is, we again formulate an appropriate Lagrangian dual problem, this time with a single multiplier λ_k associated with the power constraint of subproblem k. It is straightforward to show that the following solution results:

$$\lambda_k^{(s+1)} = \left[\lambda_k^{(s)} + \epsilon_\lambda \left\{\sum_n P_k^{n(s)} - P_k^{\max}\right\}\right]^+ \tag{28a}$$

$$P_k^{n(s+1)} = \frac{\delta_k \alpha_k^n}{\lambda_k^{(s)}}. \tag{28b}$$

In a similar spirit to the procedure introduced earlier, we can again alternate between maximization and tightening to find a converged solution $\boldsymbol{P}_k^*$. It can be shown that this is exactly the unique global optimum of the kth RA water-filling problem, and follows the proof of Corollary 1 closely, where the KKT conditions are now sufficient for a global optimum due to convexity of the unapproximated RA-WF subproblem.

B. Fixed Margin Water-Filling (FM-WF)

Repetition of the above analysis for the FM problem provides the lower bound relaxation

$$\max_{\boldsymbol{P}_k \geq 0} \quad -\sum_n P_k^n$$
$$\text{s.t.} \quad R_k^{\text{target}} \leq \check{R}_k(\boldsymbol{P}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)$$
$$\sum_n P_k^n \leq P_k^{\max} \tag{29}$$

with rate target multiplier $\delta_k = \mu_k$. The associated dual solution is given by

$$\delta_k^{(s+1)} = \left[\delta_k^{(s)} + \epsilon_\delta \left\{R_k^{\text{target}} - \sum_n \check{R}_k(\boldsymbol{P}^{(s)}; \boldsymbol{\alpha}_k, \boldsymbol{\beta}_k)\right\}\right]^+ \tag{30a}$$

$$\lambda_k^{(s+1)} = \left[\lambda_k^{(s)} + \epsilon_\lambda \left\{\sum_n P_k^{n(s)} - P_k^{\max}\right\}\right]^+ \tag{30b}$$

$$P_k^{n(s+1)} = \frac{\delta_k \alpha_k^n}{1 + \lambda_k^{(s)}} \tag{30c}$$

where again it can be shown that successive maximization and tightening converges to the global optimum $\boldsymbol{P}_k^*$ of the kth FM water-filling problem.

C. Comparison Between IWF and Scale

Let us compare the SCALE algorithm with the solutions to the RA and FM water-filling problems above, described by (28a) and (30a), respectively. We immediately observe that the Lagrange multiplier updates of SCALE (14) are identical (recall that by (15), $\delta_k \triangleq \mu_k$ for FM users). The power updates, (28b) and (30c) are also identical to the SCALE update (18) *when we disregard the impact user k* has on other users.

This is a significant result: SCALE degrades to IWF when message passing is not available, or not desired. More importantly, it motivates the use of *reduced* communication to form a hybrid SCALE–IWF scheme whereby no communication is made on tones enjoying little or no FEXT (i.e., those at low frequencies), and making full use of neighboring line conditions on tones heavily affected by FEXT to improve performance beyond IWF. As the level of such communication reduces to zero, SCALE degrades gracefully to the same performance as what IWF would provide. Put simply, SCALE will always perform at least as well as IWF. In practice, SCALE can outperform IWF considerably.

VI. THE SCAWF ALGORITHM: AN IMPROVED IWF

In this section, we develop novel water-filling methods that are purely iterative in nature and therefore admit low

complexity implementation, unlike traditional solutions that make use of channel sorting or bisection. We name these algorithms **S**uccessive **C**onvex **A**pproximation for **W**ater-**F**illing (SCAWF), as they are based on the relaxed water-filling problems of the previous section. Our new methods stem from closed-form solutions to the relaxations (27) and (29), allowing simultaneous evaluation of the maximization and tightening steps.

A. Rate Adaptive SCAWF (RA-SCAWF)

We first consider the RA-WF relaxation (27). Substituting the closed-form power update (28a) directly into the multiplier update (28b) reveals the very simple form of the dual problem

$$\lambda_k^{(s+1)} = \left[\lambda_k^{(s)} + \epsilon_\lambda \left\{ \sum_n \frac{\delta_k \alpha_k^n}{\lambda_k^{(s)}} - P_k^{\max} \right\} \right]^+.$$

This is a beautiful example of Lagrange duality: the N-dimensional constrained optimization (27) has been transformed into a one-dimensional search over multiplier λ_k.

Going further, it is easy to see that the dual solution occurs at the equilibrium point

$$\lambda_k^* = \frac{1}{P_k^{\max}} \sum_n \delta_k \alpha_k^n \tag{31}$$

where the associated primal solution is given by

$$P_k^{n*} = P_k^{\max} \frac{\delta_k \alpha_k^n}{\sum_m \delta_k \alpha_k^m} \tag{32}$$

and follows by substitution of (31) into (28b). This is a somewhat surprising closed-form solution to the relaxed optimization problem (27). Combined with the tightening operation (6) around the point $z_0 = \mathrm{SIR}_k^n(\boldsymbol{P})$, we obtain the RA-SCAWF algorithm

$$P_k^n = P_k^{\max} \frac{\frac{\mathrm{SIR}_k^n}{1+\mathrm{SIR}_k^n}}{\sum_{m=1}^{N} \frac{\mathrm{SIR}_k^m}{1+\mathrm{SIR}_k^m}} \tag{33}$$

where the denominator sum is common to the allocation of all tones, and needs to be calculated only once (consequently, numerators are calculated for free). This algorithm is a particularly attractive alternative to the current IWF procedure where a conventional water-filling solution is computed for every user at every iteration. Such water-filling computations often require a sorting step of the channel gains or the use of complex data structures. Instead, the SCAWF algorithm computes the partial solution (33) that adapts a user's PSD and converges together with all other users to the simultaneous multiuser water-filling solution.

The SCAWF algorithm is also extremely simple: the SIR on each tone is periodically measured and (33) computed to form a new power allocation that is immediately updated. No channel sorting or complex data structures are required.

B. Fixed Margin SCAWF (FM-SCAWF)

We now develop a similar algorithm for the FM-WF relaxation (29). Substituting the closed-form power update (30c) into (30a) gives

$$\delta_k^{(s+1)} = \left[\delta_k^{(s)} + \epsilon_\delta R_k^{\mathrm{target}} \right.$$
$$\left. - \epsilon_\delta \sum_n \left\{ \alpha_k^n \log \left(\frac{\delta_k \alpha_k^n}{(1+\lambda_k)\widetilde{\sigma}_k^n} \right) + \beta_k^n \right\} \right]^+$$

where $\widetilde{\sigma}_k^n$ is the thermal noise plus (fixed) interference experienced by the receiver of user k on tone n. The equilibrium condition is given by

$$R_k^{\mathrm{target}} = \sum_n \alpha_k^n \log \left(\frac{\delta_k \alpha_k^n}{(1+\lambda_k)\widetilde{\sigma}_k^n} \right) + \beta_k^n$$
$$= \log \left(\frac{\delta_k}{1+\lambda_k} \right) \sum_n \alpha_k^n + \sum_n \left[\alpha_k^n \log \left(\frac{\alpha_k^n}{\widetilde{\sigma}_k^n} \right) + \beta_k^n \right]$$
$$\frac{\delta_k}{1+\lambda_k} = \exp \left(\frac{R_k^{\mathrm{target}} - \sum_n [\alpha_k^n \log (\alpha_k^n/\widetilde{\sigma}_k^n) + \beta_k^n]}{\sum_n \alpha_k^n} \right).$$

The right-hand side is a constant, resembling (21). Substituting it into (30c) results in the closed-form power update

$$P_k^{n*} = \alpha_k^n \exp \left(\frac{R_k^{\mathrm{target}} - \sum_m [\alpha_k^m \log (\alpha_k^m/\widetilde{\sigma}_k^m) + \beta_k^m]}{\sum_m \alpha_k^m} \right). \tag{34}$$

While it may seem surprising that this update is independent of the multiplier λ_k, the result is indeed sensible. The preceding development supposed existence of the equilibrium condition. In other words, we had assumed that the rate target R_k^{target} was feasible. It then follows that the maximum power constraint becomes redundant (reminiscent of our discussion at the end of Section IV-A), and so the actual power allocation should not depend on λ_k.

This maximum power constraint is still important nonetheless. Without it, we would require "high-SIR feasible" rate targets, as discussed in Section IV-A. We therefore propose that the PSD allocated according to (34) is paired back whenever it violates the maximum power constraint. We elect a simple projection that scales all components of the PSD equally, resulting in the power update

$$P_k^{n(s+1)} = \begin{cases} Q_k^{n(s)}, & \text{when } \left\| \boldsymbol{Q}_k^{(s)} \right\|_1 \leq P_k^{\max} \\ P_k^{\max} \frac{Q_k^{n(s)}}{\left\| \boldsymbol{Q}_k^{(s)} \right\|_1}, & \text{otherwise} \end{cases} \tag{35}$$

where $\boldsymbol{Q}_k = [Q_k^1, \ldots, Q_k^N]$ is an N-length PSD with components given by (34), and we note that the total power required by the update is given by $\|\boldsymbol{Q}_k\|_1$.

This update solves the relaxation (29) in closed form when it is feasible. Otherwise, the relaxation does not have a solution

and we provide a compromise: a PSD that satisfies the maximum power constraint with equality, having the "same shape" of the PSD that would have been allocated without any maximum power restriction. Combined with an alternating tightening operation (6) around the point $z_0 = \text{SIR}_k^n(\boldsymbol{P})$, we obtain the FM-SCAWF algorithm that solves the RA water-filling problem at convergence.

This algorithm, much like its close relative RA-SCAWF, is an attractive alternative to existing methods for water-filling under a rate constraint, and especially so in an IWF procedure for similar reasons already discussed.

VII. DISCRETE BIT-LOADING

All of our developments thus far have hinged on the assumption of continuous bit-loading. That is, once the PSD $\boldsymbol{P}_k$ for user k is found with the SCALE or SCAWF algorithm, the bit-loading on each tone is computed with (1).

It may be desirable to limit the bit-loading to a discrete set in practice. We propose this by "layering" the PSD optimization and bit-loading operations: all user PSDs are first optimized with the SCALE or SCAWF algorithms as presented. With PSDs fixed, the associated continuous bit-loadings are mapped to the desired discrete set.

For example, we may be interested in the discrete set of integers $\widehat{b}_k^n \in [0, b_{\max}^n]$, where $b_{\max}^n$ is a maximum integer bit-loading on tone n, and a bit-loading of zero indicates that the associated tone will not be used for data communication. In this case, the continuous-to-discrete mapping may be realized with

$$\widehat{b}_k^n = \min \left\{ b_{\max}^n, \left\lfloor \log_2 \left(1 + \text{SIR}_k^n \right) \right\rfloor \right\}$$

where $\lfloor \cdot \rfloor$ is the floor operator. The SIR_k^n can be obtained through a noise measurement and knowledge of the PSD $\boldsymbol{P}_k$. We always truncate—not round—to ensure that the loading associated with a particular SIR is supported.

This layered approach is typically suboptimal: slightly more power will be expended in general than what is optimally required, unless by chance, the continuous bit-loading aligns with the desired discrete set. Conceptually, we may think of this layering as being similar to the well-known practice of "gain scaling": a perturbation of the spacings between constellation points by multiplying each complex symbol by a small positive factor [19]. Although gain scaling is typically used to equalize the bit-error rate (BER) on each tone, we can think of it as the gap in power between the optimal PSD based on discrete loading and the continuous result. On the upside, the resulting tonal BERs may be improved since we may be supporting a bit-loading with a slightly higher SIR_k^n than what is really required.

We do not advocate that each PSD be reduced accordingly (for example, with true gain scaling to counteract the gap), as the corresponding interference between users would also change. This interference is an important message-passing mechanism, albeit an implicit one, achieved through measurement.

Despite these caveats, this layered approach provides a simple and practical way in which to operate with discrete bit-loadings.
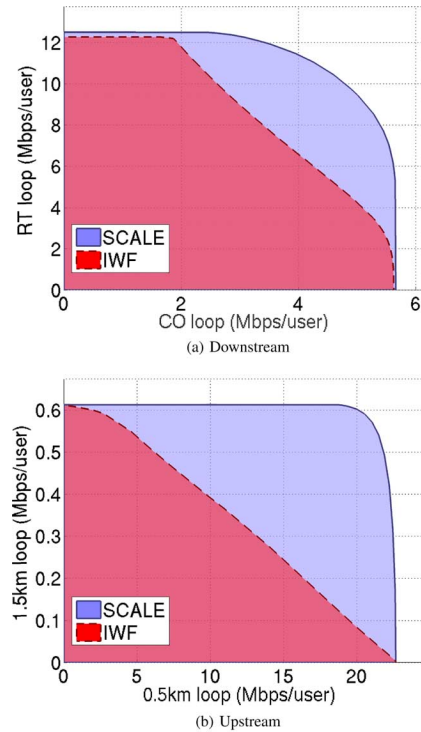

Fig. 3. SCALE can significantly enlarge the VDSL rate region compared to IWF in both down- and upstream topologies.

VIII. PERFORMANCE

In this section, we compare the performance of the SCALE and SCAWF algorithms to IWF under continuous bit-loading. Our evaluations consider VDSL over 26-AWG (0.4-mm) lines. A coding gain of 3 dB, with a 6-dB noise margin is assumed, giving an SNR gap $\Gamma_k^n = 12.8$ dB for an error probability of 10^{-7} [8]. Each modem has maximum transmission power 11.5 dBm, and can transmit in *both* 1U and 2U upstream bands (regional-specific band; former plan 998) [18, Table 1] with amateur RF bands notched off [18, Table 17]. No other spectral masks are enforced. A DMT symbol rate of 4 kHz is assumed, with tone spacing of 4.3125 kHz. Users are subject to -140-dBm/Hz background noise and alien noises corresponding to European Telecommunications Standards Institute (ETSI) models XA.$\{L, N\}T.\{A, D\}$ as appropriate [18, Tables 21–22]. The cross-gains G_{ij}^n are calculated according to [20] without FSAN combination of FEXT sources, and using standard FEXT models [8].

Our simulations consider $K = 8$ users, split into two equal groups of four users each. The downstream topology of Fig. 1(a) has a central office (CO)-based group placed at 3 km and an RT deployment at a distance of 2 km, with the second RT-based user-group placed 2 km further along. The upstream topology of Fig. 1(b) has one group placed 0.5 km from the CO, the other at 1.5 km.

Due to the inherent symmetry in the channel models [8], the resulting rates for users having equal loop lengths end up the same. Fig. 3 then shows the rate region between two users, one from each user group. The performance improvement of SCALE is clearly significant, where the rate region is almost doubled over IWF in the upstream direction.
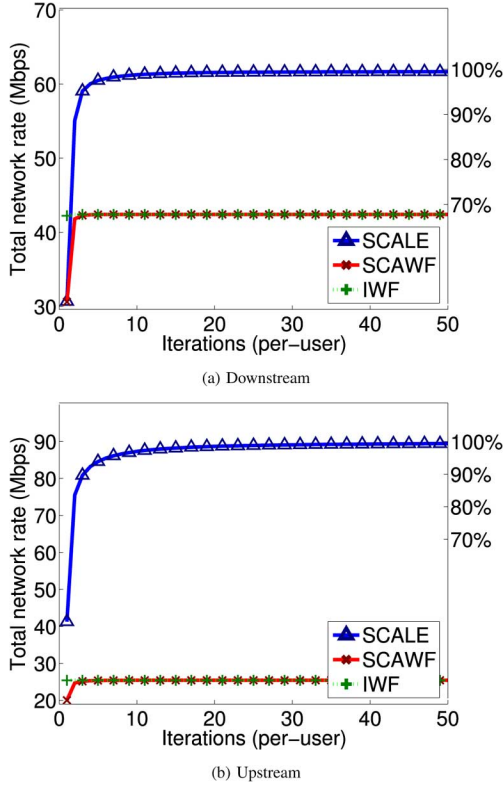
(a) Downstream

(b) Upstream

Fig. 4. Convergence and performance comparison: SCALE has improved system rate with speedy convergence to rates well above IWF and SCAWF.

The SCALE region is produced by sweeping the weights δ_k and solving a complete RA problem for each tuple. A similar goal is achieved in IWF by pairing back the maximum power budget of each user k by a factor $\Psi_k \leq 1$. That is, each user may use a maximum total power of $\Psi_k P_k^{\max}$, rather than the full power budget $P_k^{\max}$. Selection of the scaling factors Ψ_k was a challenging process, as minute changes resulted in large differences of the final allocated rates.

We now compare the convergence properties of each algorithm, having selected specific weightings that correspond to particular points within the rate region.

For the downstream, our selection corresponds to a 4-Mbps/user service on CO-based loops. Fig. 4(a) shows the convergence of each scheme, where iterates are shown after all users have updated their PSD. The IWF algorithm converges within two iterations. While the SCAWF algorithm takes an additional iteration to converge to the same result, it requires significantly less computation as outlined in Section VI. The SCALE algorithm also converges very quickly, and in just two iterations, far exceeds the final performance of IWF. These gains stem from an almost disjoint frequency-division separation of the near and far user groups, negotiated automatically by a power allocation that takes other users into account on a tone-by-tone basis. These PSDs are depicted in Fig. 5. In contrast, the IWF scheme overlaps the spectrum of each user group due to its selfish nature.

On the upstream, we select weightings that correspond to a 500-kbps/user service on 1.5-km loops. Fig. 4(b) shows the corresponding iterations. Convergence rates similar to the downstream direction are observed, where SCALE outperforms IWF after a single iteration. Comparing the associated PSDs in Fig. 5,

we clearly see that SCALE allocates power over a much larger bandwidth, employing tones that IWF neglects.

A. Global Optimality

In this section, we revisit the topic of global optimality. Recall that at convergence, SCALE guarantees at least a local optimum solution: a point that satisfies the necessary KKT optimality condition. Nonetheless, this solution may or may not be globally optimum, due to the nonconvex nature of the problem. We therefore resort to a numerical technique to test the previous two examples for global optimality. Our basis is the following general sufficiency condition, listed in [15] and recast below in terms of a *maximization* problem (not necessarily convex).

Proposition 1 (General Sufficiency Condition): Consider the problem

$$
\begin{aligned}
\max \quad & f(x) \\
\text{s.t.} \quad & x \in X, \quad g_j(x) \leq 0, \quad j = 1, \dots, r
\end{aligned}
$$

where f and g_j are real-valued functions on $\mathbb{R}^n$ and X is a given subset of $\mathbb{R}^n$. Let x^* be a feasible vector which together with a vector $\mu^* = [\mu_1^*, \dots, \mu_r^*]$, satisfies

$$
\begin{aligned}
\mu_j^* &\geq 0, \quad j = 1, \dots, r \\
\mu_j^* &= 0, \quad \forall j \notin A(x^*)
\end{aligned}
$$

and maximizes the Lagrangian function $L(x, \mu^*)$ over $x \in X$

$$
x^* = \arg \max_{x \in X} L(x, \mu^*). \tag{36}
$$

Then x^* is a global maximum of the problem.

Consider the converged primal–dual triplet $\{\boldsymbol{P}^*, \boldsymbol{\lambda}^*, \boldsymbol{\delta}^*\}$ resulting from the SCALE algorithm. By Corollary 1, it satisfies the feasibility and multiplier conditions of Proposition 1 with respect to the spectrum balancing problem (4). All that remains is evaluation of condition (36). The associated Lagrangian is given by

$$
L(\boldsymbol{P}, \boldsymbol{\lambda}, \boldsymbol{\delta}) \triangleq \sum_n L^n(\boldsymbol{P}^n, \boldsymbol{\lambda}, \boldsymbol{\delta})
$$
$$
+ \sum_k \lambda_k P_k^{\max} - \sum_{k \in \text{FM}} \delta_k R_k^{\text{target}} \tag{37}
$$

where we have made use of the unifying multiplier (15) for notional simplicity and

$$
L^n(\boldsymbol{P}^n, \boldsymbol{\lambda}, \boldsymbol{\delta}) \triangleq \sum_k \delta_k \log \left(1 + \text{SIR}_k^n(\boldsymbol{P}^n)\right) - P_k^n \left(\lambda_k + \mathbf{1}_{\substack{k \in \text{FM} \\ \text{RA}=\varnothing}}\right)
$$

represents a partial Lagrangian involving only the user powers on a particular tone n.

We exploit this separability to simplify the maximization (36) from dimension $K \times N$ to a sequence of N smaller maximizations, each involving only K dimensions. This leads to a vast computational saving, since typically $N \gg K$. It follows that when the condition

$$
\boldsymbol{P}^{n*} = \arg \max_{\left\{0 \leq P_k^n \leq P_k^{\max} \mid k \in [1, K]\right\}} L^n(\boldsymbol{P}^n, \boldsymbol{\lambda}^*, \boldsymbol{\delta}^*) \tag{38}
$$

holds for all tones $n \in [1, N]$, the SCALE power allocation $\boldsymbol{P}^*$ is globally optimal.
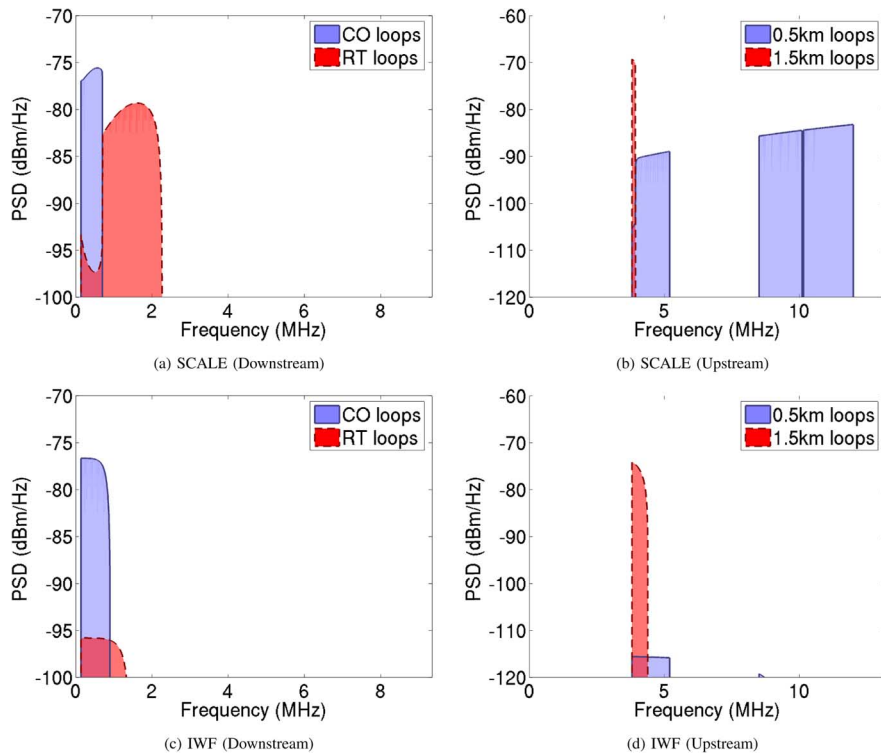
Fig. 5. Power spectrum density allocation: SCALE spreads power over a wider bandwidth, with less overlapping spectrum between near and far users.

For the purposes of evaluating this per-tone nonconvex maximization, we employ an exhaustive grid search, reminiscent of the OSB method [1]. We employ a nonuniform grid spacing, in contrast to OSB, with 150 points in each dimension. This exploits the fact that the per-user maximum power $P_k^{\max}$ is typically spread over a large number of tones and therefore the power allocated to an individual will likely be small. In turn, this allows us to achieve high accuracy with reasonable computation. We further exploit the symmetry in the channel models [8], where each user within each group ends up with exactly the same power allocation and multiplier values, to further reduce the search space. This latter technique is not uncommon [21]. Fig. 6 shows the per-tone Lagrangian difference

$$L_\Delta^n = L^n(\boldsymbol{P}^{n*}, \boldsymbol{\lambda}^*, \boldsymbol{\delta}^*) - \max_{\left\{0 \le P_k^n \le P_k^{\max} | k \in [1,K]\right\}} L^n(\boldsymbol{P}^n, \boldsymbol{\lambda}^*, \boldsymbol{\delta}^*)$$

where the grid-search is employed to compute the maximization over $\boldsymbol{P}^n$. Assuming sufficient accuracy in the exhaustive search, small nonnegative differences imply that condition (38) holds and the associated power allocation $\boldsymbol{P}^n$ on tone n is optimal for the supplied multiplier values. Negative values indicate otherwise. We make use of the SCALE solution $\{\boldsymbol{P}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\}$ from the previous two examples, taken at the 50th iterate.

Fig. 6(a) illustrates the differences for the downstream example, where only 10 out of 1602 tones (0.6%) resulted in a negative difference. Interestingly, these tones were located at the crossover point between CO- and RT-based loops, observed in Fig. 5(a) at around 500 kHz. The corresponding powers resulting from the grid-search were "dimensionally opposite" to those provided by SCALE in this case. i.e., power allocated to the RT-based user group should have been allocated to the CO-based group and *vice versa*. Otherwise, whenever $L_\Delta^n \ge 0$,
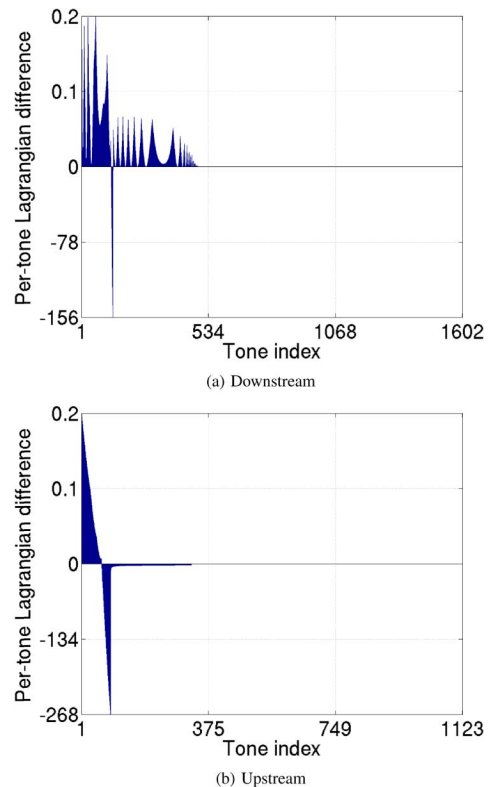

Fig. 6. Difference of per-tone Lagrangians, between SCALE and a grid-based search.

the powers obtained by grid search were very nearly $\boldsymbol{P}^{n*}$. This gives us confidence that our grid search was of sufficient accuracy and that condition (38) was met on all such tones.

The upstream example is illustrated in Fig. 6(b), where only 27 tones out of 1123 (2.4%) did not satisfy condition (38). These violations were similarly located at the crossover point between the long and short loops, shown at around 4 MHz in Fig. 5(b).

These minute differences give us confidence that the SCALE solution is very nearly globally optimum in this case. This is a very significant result: with a mere 50 iterations, SCALE has nearly achieved a globally optimum solution on an NP-hard problem having approximately 8000–12 000 variables—very large scale indeed. Although not shown here, additional SCALE iterations were found to further reduce the number of tonal violations of condition (38). The associated total-network data rate remained virtually the same in this case, giving further weight to our claims.

IX. CONCLUSION

Two novel algorithms for spectrum balancing in multiuser DSL networks have been introduced, each enjoying a low-complexity implementation. SCALE, the first algorithm, explicitly accounts for the "damage" a user's power allocation has on other users, resulting in higher achievable rates than the selfish competitive-optimal rates resulting from simultaneous water-filling. Through measurement and limited message passing, SCALE is easily distributed with the help of a SMC. Message-passing overhead can be arbitrarily traded for performance, and in the limit of zero overhead, SCALE can perform no worse than IWF.

Convergence to a solution satisfying the necessary KKT optimality condition was proven under a mild assumption on RA users that can be typically relaxed in practice. Numerical studies indicate that SCALE can converge to rates far exceeding that of IWF with just two to three iterations, and to within 90% of the final rate in under five iterates. These results were further shown to be very nearly global optimum, where less than 3% of the 1000+ subchannels employed were observed to fail a general sufficiency condition for global optimality.

The second algorithm, SCAWF, was shown to be an extremely simple way to water-pour that is particularly well-suited to iterative multiuser water-filling problems. Unlike existing methods, it does not employ expensive channel sorting or bisection operations.

While the focus of this paper was on continuous bit-loading, we recognize the importance of discrete bit-loading and have outlined a layered approach that paves the way for a simple and practical method based on the results herein. Any further improvements in this area would require a very different approach—that of solving a discrete optimization problem—and remains an attractive area of future work.

REFERENCES

[1] R. Cendrillon, W. Yu, M. Moonen, J. Verlinden, and T. Bostoen, "Optimal multi-user spectrum management for digital subscriber lines," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 922–933, May 2006.

[2] K. B. Song, S. T. Chung, G. Ginis, and J. M. Cioffi, "Dynamic spectrum management for next-generation DSL systems," *IEEE Commun. Mag.*, vol. 40, no. 10, pp. 101–109, 2002.

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

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

[5] R. Lui and W. Yu, "Low-complexity near-optimal spectrum balancing for digital subscriber lines," in *IEEE Int. Conf. Communications*, Seoul, Korea, 2005, vol. 3, pp. 1947–1951.

[6] J. Papandriopoulos and J. S. Evans, "Low-complexity distributed algorithms for spectrum balancing in multi-user DSL networks," in *IEEE Int. Conf. Communications*, Istanbul, Turkey, Jun. 2006, pp. 3270–3275.

[7] Y. X. S. Panigrahi and T. Le-Ngoc, "A concave minimization approach to dynamic spectrum management for digital subscriber lines," in *IEEE Int. Conf. Communications*, Istanbul, Turkey, Jun. 2006, pp. 84–89.

[8] T. Starr, J. M. Cioffi, and P. J. Silverman, *Understanding Digital Subscriber Line Technology*. Englewood Cliffs, NJ: Prentice-Hall, 1999.

[9] J. Papandriopoulos and J. S. Evans, "Method for Distributed Spectrum Management of Digital Communication Systems," U.S. application US11/433,025, Apr. 2006.

[10] J. Papandriopoulos and J. S. Evans, "Method for Distributed Spectrum Management of Digital Communication Systems," Australia Patent application 2006202136, Apr. 2006.

[11] J. Campello, "Practical bit loading for DMT," in *IEEE Int. Conf. Communications*, Vancouver, BC, Canada, Jun. 1999, vol. 2, pp. 801–805.

[12] R. Horst and H. Tuy, *Global Optimization: Deterministic Approaches*, 2nd ed. Berlin, Germany: Springer-Verlag, 1993.

[13] E. V. den Bogaert, T. Bostoen, J. V. Elsen, R. Cendrillon, and M. Moonen, "DSM in practice: Iterative water-filling implemented on ADSL modems," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, Montreal, QC, Canada, May 2004, vol. 5, pp. 337–340.

[14] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[15] D. P. Bertsekas, *Nonlinear Programming*. Nashua, NH: Athena Scientific, 1995.

[16] R. D. Yates, "A framework for uplink power control in cellular radio systems," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1341–1347, Sep. 1995.

[17] C. Zeng, C. Aldana, A. A. Salvekar, and J. M. Cioffi, "Crosstalk identification in xDSL systems," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 8, pp. 1488–1496, Aug. 2001.

[18] *ETSI TS 101 270-1 v1.3.1 Standard. Part 1: Functional Requirements*, 2003-07.

[19] *ETSI TS 101 270-2 v1.2.1 Standard. Part 2: Tranceiver Specification*, 2003-07.

[20] R. Baldemair, M. Horvat, and T. Nordstöm, "Proposed method on crosstalk calculations in a distributed environment," in *Proc. ETSI STC TM6 Plenary Meeting*, Sophia-Antipolis, France, 2003.

[21] R. Cendrillon, private communication Sep./Oct. 2006.

John Papandriopoulos (S'01–M'07) received the combined B.E. degree in communications engineering and the B.Appl.Sci. degree in computer science from the Royal Melbourne Institute of Technology (RMIT University), Melbourne, Australia, in 2001 and the Ph.D. degree in electrical engineering from the University of Melbourne, in 2006.

His research interests are in nonconvex optimization, DSL and CDMA resource allocation and cross-layer design. He enjoys delving into practical engineering implementations, and has spent some time with Telstra, Agilent Technologies, and the 3G Mobile R&D Division of NEC Australia. He is currently with a startup Adaptive Spectrum and Signal Alignment Inc. in the San Francisco Bay Area.

Dr. Papandriopoulos was awarded the J. N. McNicol Prize (University Medal) upon graduating from RMIT University in 2002, and the Chancellor's Prize for excellence in his Ph.D. dissertation from the University of Melbourne in 2007. In 2005, he received a Victoria Fellowship from the Victoria State Government of Australia.

Jamie S. Evans (S'93–M'98) received the B.S. degree in physics and the B.E. degree in computer engineering from the University of Newcastle, Newcastle, Australia, in 1992 and 1993, respectively, and the M.S. and Ph.D. degrees from the University of Melbourne, Melbourne, Australia, in 1996 and 1998, respectively, both in electrical engineering.

From March 1998 to June 1999, he was a Visiting Researcher with the Department of Electrical Engineering and Computer Science, University of California, Berkeley. He returned to Australia to take up a position as Lecturer at The University of Sydney, Sydney, Australia, where he stayed until July 2001. Since that time, he has been with the Department of Electrical and Electronic Engineering, University of Melbourne, where he is now an Associate Professor and Reader. His research interests are in communications theory, information theory, and statistical signal processing with current focus on wireless communications networks.

Dr. Evans was the recipient of the University Medal of the University of Newcastle and the Chancellor's Prize for Excellence for his Ph.D. dissertation.