Convex Channel Power Optimization in Nonlinear WDM Systems Using Gaussian Noise Model

Ian Roberts, *Student Member, OSA*, Joseph M. Kahn, *Fellow, IEEE*, and David Boertjes, *Member, OSA*

***Abstract*—Optimization of channel powers to maximize minimum margin or total capacity in WDM systems is studied. Using a Gaussian noise nonlinearity model, the signal-to-noise ratio (SNR) in each channel is expressed as a convex function of the channel powers. Using the SNR expression, convex optimization problems with objectives of maximizing the minimum channel margin or maximizing the fiber capacity minus a coding cap are formulated. Performance gains from software-based power optimization are observed in mesh networks and in point-to-point links having heterogeneous SNR requirements. By contrast, in systems with uniform amplifier noise and modulation formats, the optimized power allocation provides very little improvement over a traditional flat power allocation. In the 14-node NSFNET network, a margin gain of 1.5 dB on average is achieved through power optimization, as compared to a flat power allocation. Margin gains averaging 1.4 dB are found for subsets of this network with three to 13 nodes.**

***Index Terms*—Gaussian noise model, network optimization, nonlinear capacity, optical communications.**

I. INTRODUCTION

THE demand for data communications capacity continues to grow at a rapid pace. As symbol rates increase and fiber spectrum becomes filled, improvement in spectral efficiency becomes key. Advanced modulation formats and coding methods have received significant attention as a route to this goal. Flexibility introduced through adaptive modulation or variable error-control coding allows a single system to provide longer reach or higher data rates tailored to various usage scenarios [1], [2].

Contemporary optical communication systems provide a limited amount of flexibility via modulation format adjustment through a range of BPSK, QPSK, and 16-QAM [3]. This allows for coarse adjustment of the data rate to match the signal-to-noise ratio (SNR) delivered by the channel. Variable-rate coding combined with modulation variation can reduce the granularity of delivered rates [4], [5]. In the limit of fine rate granularity, the relationship between rate and required SNR becomes a continuous function. This continuous relationship can be described conveniently by the information-theoretic channel capacity, but with an SNR gap reflecting the non-ideality of practical coding schemes [6].

The Kerr nonlinearity is the dominant factor limiting the power of signals transmitted through optical fibers. In dispersion-uncompensated fiber systems, the nonlinear phase rotation due to the Kerr effect combines with linear dispersion to produce what is well-modeled as additive Gaussian noise [7]. The Gaussian noise model of nonlinearity provides a convenient, tractable model for intra-channel and inter-channel nonlinear interactions.

The interaction between flexible transceivers, mesh networks, and inter-channel nonlinearity is complex, and few works to date have studied optimization of rate-flexible point-to-point links and networks [8], [9]. In this paper, we formulate convex objectives that enable the performance of such nonlinear networks to be optimized. Convex solution methods are used to guarantee convergence towards the globally optimal solution and provide bounds on sub-optimality of computed results.

The structure of the paper is as follows. In Section II, a discrete version of the Gaussian noise nonlinearity model is developed, allowing the SNR to be expressed as a convex function of the logarithmic channel powers. This SNR expression is then used to obtain a convex optimization problem that maximizes the aggregate communication rate of a point-to-point WDM link. The SNR expression is also used to formulate a convex optimization for maximizing the minimum channel margin. Two methods of solving this optimization are described. In Section III, the optimization problem of maximizing the minimum channel margin is extended to mesh networks. The margins achieved with optimized power allocations are compared with the best flat or one-dimensional optimization. Sections IV and V provide discussion and conclusions, respectively.

II. POINT-TO-POINT POWER OPTIMIZATION

The Gaussian noise nonlinearity model is shown in equation (1) [7]. It expresses the nonlinear noise power spectrum $G_{\mathrm{NL}}(f)$ in terms of the signal power spectrum $G(f)$ at three different frequencies and coefficients calculated from fiber parameters: Attenuation coefficient α, dispersion propagation constant β_2, nonlinear coefficient γ, and span length L_s. The effective span length normalization used in [7] has been incorporated into the ρ coefficient

$$G_{\mathrm{NL}}(f) = \frac{16}{27}\gamma^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G(f_1)G(f_2)G(f_1+f_2-f) \cdot\rho(f_1,f_2,f)\chi(f_1,f_2,f)\partial f_1 \partial f_2 \quad (1)$$

Manuscript received February 01, 2016; revised April 22, 2016; accepted May 10, 2016. Date of publication May 15, 2016; date of current version June 22, 2016. This work was supported by Ciena Corporation, a Stanford Rambus Corporation Graduate Fellowship, and an NSERC scholarship.

I. Roberts and J. M. Kahn are with the E. L. Ginzton Laboratory, Department of Electrical Engineering, Stanford University, Stanford, CA 94305 USA (e-mail: iroberts@stanford.edu; jmk@ee.stanford.edu).

D. Boertjes is with the Ciena Corporation, Nepean, ON K2H 8E9, Canada (e-mail: dboertje@ciena.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2016.2569073

$$\rho(f_1, f_2, f) = \left| \frac{1 - e^{-2\alpha L_s + j4\pi^2 \beta_2 L_s (f_1 - f)(f_2 - f)}}{2\alpha - j4\pi^2 \beta_2 (f_1 - f)(f_2 - f)} \right|^2$$

$$\chi(f_1, f_2, f) = \frac{\sin^2(2N_s \pi^2 (f_1 - f)(f_2 - f)\beta_2 L_s)}{\sin^2(2\pi^2 (f_1 - f)(f_2 - f)\beta_2 L_s)}.$$

If the available spectrum is divided into N discrete channels and it is assumed that there is an identical spectral shape within each channel, the integrals can be separated into a summation with a set of constant coefficients. We optimize the transmitted signal spectrum with a spectral granularity Δf, which can describe the width of a traditional fixed-grid WDM channel or the control granularity of wavelength-selective switches in a flexible-grid network. This paper assumes a power optimization granularity of $\Delta f = 50$ GHz matching a traditional channel grid. Using a finer optimization granularity allows greater optimization freedom, but the calculation time scales nearly cubically with the number of optimization variables, such that a factor-of-eight increase to use a spectral granularity of $\Delta f = 6.25$ GHz increases computation time by a factor of 256.

We assume that the transmitted signal power is equal to the optical power post-amplification through the entire link, and thus use it as the argument in the Gaussian noise nonlinear model. With a known spectral shape in each channel, we can express the nonlinear power in channel n, NL_n, as a function of the transmitted signal power in each channel, P_n:

$$NL_n = \gamma^2 \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} \sum_{l=-1}^{1} P_{n_1} P_{n_2} P_{n_1+n_2-n+l} D_l(n_1, n_2, n) \tag{2}$$

for $1 \leq n_1 + n_2 - n + l \leq N$. In this equation, $D_l(n_1, n_2, n)$ is an integral that defines a set of system-specific coefficients that determine the strength of the four-wave mixing component that falls on channel n, generated by channels n_1, n_2, and $n_1 + n_2 - n + l$:

$$\begin{aligned} D_l(n_1, n_2, n) = \ & \frac{16}{27} \int_{-\frac{\Delta f}{2}}^{\frac{\Delta f}{2}} \int_{-\frac{\Delta f}{2}}^{\frac{\Delta f}{2}} \int_{-\frac{\Delta f}{2}}^{\frac{\Delta f}{2}} \\ & \rho(\xi_1 + n_1 \Delta f, \xi_2 + n_2 \Delta f, \xi + n \Delta f) \\ & \cdot \chi(\xi_1 + n_1 \Delta f, \xi_2 + n_2 \Delta f, \xi + n \Delta f) g(\xi_1) \\ & \cdot g(\xi_2) g(\xi_1 + \xi_2 - \xi + l\Delta f) R\, g(\xi) \partial\xi_1 \partial\xi_2 \partial\xi. \end{aligned} \tag{3}$$

In (3), $g(f)$ is the spectral shape of each channel, which has been normalized such that $\int g(f) df = 1$. The factor $R\, g(f)$, where R is the symbol rate, describes the receiver matched filtering of the nonlinear noise in moving from power spectral density functions to channel powers. Coefficients for $l = 0$ describe the dominant nonlinear terms, while the coefficients for $l = \pm 1$ describe the corner contributions. This discretization of the Gaussian noise model is similar to that analyzed for XPM and SPM only by Ives *et al.* [8].

This paper assumes $g(f)$ is rectangular with a width of 50 GHz, matching the power optimization granularity. This is intended to represent future communication systems with very small excess bandwidth designed to maximize spectral efficiency.

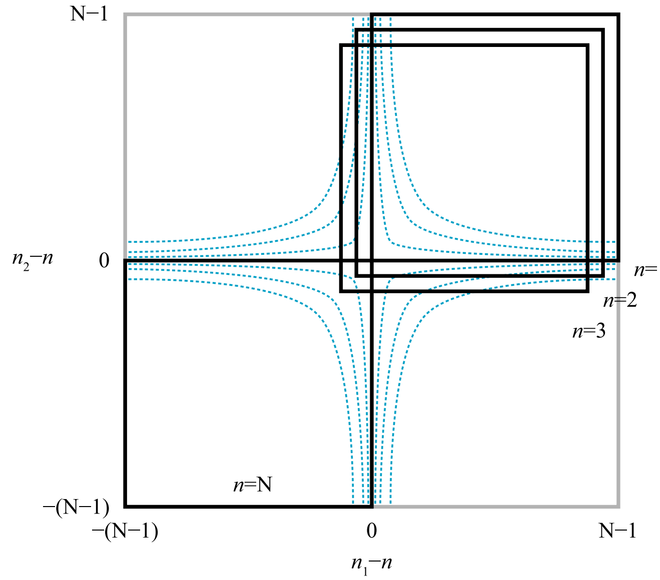

Fig. 1. The $N \times N$ coefficient matrices for each channel described by equation (3) are sub-matrices of a $(2N-1) \times (2N-1)$ matrix.

The χ function in (3) is particularly difficult to integrate numerically due to its periodic sharp peaks. The rigid Cartesian integration bounds of the individual coefficients do not lend themselves to hyperbolic transformation of the integration. An incoherent Gaussian noise model or an epsilon-law approximation to the accumulation of nonlinear noise between spans can replace this term [7].

The D_l coefficient matrices given by (3) have two important properties that facilitate power optimization. The first is that the D_l coefficients are non-negative. The second property is that the ρ and χ coefficients depend only on the magnitude of the product $(f_1 - f)(f_2 - f)$. This means that the three-dimensional collection of coefficients represented by $D_l(n_1, n_2, n)$ is actually a set of sub-matrices of a $(2N-1) \times (2N-1)$ matrix indexed by $n_1 - n$ and $n_2 - n$, as shown in Fig. 1. This overlap in coefficient matrices reduces the number of entries that must be precomputed and stored. The magnitude of the coefficients decreases rapidly for large $(f_1 - f)(f_2 - f)$, following the hyperbolic profile of [7]. This leads to only few percent of the coefficients being significant and the rest being well approximated by zero.

The nonlinear noise power in channel n is conveniently expressed as a function of channel power vector $\boldsymbol{x} \in \mathbf{R}_+^N$:

$$NL_n(\boldsymbol{x}) = \gamma^2 \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{l=-1}^{1} x_i x_j x_{i+j-n+l} D_l(i, j, n) \tag{4}$$

for $1 \leq i + j - n + l \leq N$.

Expression (4) is not an optimization objective, but its properties are important, as it is a part of an expression for the SNR to be introduced shortly. $NL_n(\boldsymbol{x})$ is a convex function of the scalar variable x for the one-dimensional subspace where the powers in all channels are equal, i.e., $\boldsymbol{x} = x[1,\ldots,1]^T$. It is a locally convex function of the vector variable $\boldsymbol{x}$ near this subspace, where the power per channel is approximately equal, but becomes non-convex as the difference in power between adjacent channels increases. Posynomial[1] functions of this form, which are generally not convex, may be transformed into convex functions [10], [11].

As the vector $\boldsymbol{x}$ represents powers that are non-negative, we can express $\boldsymbol{x} = e^{\boldsymbol{y}}$ in terms of a vector $\boldsymbol{y} \in \mathbf{R}^N$. As the D_l coefficients are non-negative, they can be taken as the exponential of a logarithm:

$$\begin{aligned} NL_n(e^{\boldsymbol{y}}) &= \gamma^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=-1}^{1} e^{y_i} e^{y_j} e^{y_{i+j-n+l}} D_l(i,j,n) \\ &= \gamma^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=-1}^{1} e^{y_i+y_j+y_{i+j-n+l}+\log(D_l(i,j,n))}. \end{aligned} \tag{5}$$

In (5), for coefficients with large values of $(f_1 - f)(f_2 - f)$ that have been approximated by zero, the corresponding indices should be excluded from the summation.

An affine combination of the vector of variables $\boldsymbol{y}$ is convex, and the exponential of a convex function remains convex, making each summation term in (5) convex [10]. A sum of a set of convex functions is convex, making (5) convex in $\boldsymbol{y} = \log(\boldsymbol{x})$. As the logarithm of each term is affine and the sum of log-convex functions remains log-convex, (5) is also log-convex in $\boldsymbol{y}$.

The SNRs for the channels are given by

$$\mathrm{SNR}_n(\boldsymbol{x}) = \frac{x_n}{\sigma^2{}_n + NL_n(\boldsymbol{x})}, n = 1\ldots N \tag{6}$$

where $\sigma^2{}_n$ is the Gaussian amplifier noise power in channel n. The SNR is quasi-concave in linear power vector $\boldsymbol{x}$ for the region, near the equal-power subspace, where $NL_n(\boldsymbol{x})$ is convex. Where $NL_n(\boldsymbol{x})$ is not convex, the SNR is not quasi-concave and has disjoint sub-level sets [10].

Using the transformation to the logarithmic power variables $\boldsymbol{y}$, a ratio of convex functions is obtained:

$$\mathrm{SNR}_n(e^{\boldsymbol{y}}) = \frac{e^{y_n}}{\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})}. \tag{7}$$

This expression for SNR is neither convex nor concave, but taking a logarithm yields a concave function:

$$\log(\mathrm{SNR}_n(e^{\boldsymbol{y}})) = y_n - \log(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})), \tag{8}$$

as $NL(e^{\boldsymbol{y}})$ is log-convex in $\boldsymbol{y}$, and $\sigma^2{}_n$ is a constant, making $\mathrm{SNR}_n(e^{\boldsymbol{y}})$ log-concave in $\boldsymbol{y}$.

The margin of channel n, M_n, is given by the SNR of each channel divided by the required SNR for the selected modulation format and coding:

$$M_n(\boldsymbol{x}) = \frac{\mathrm{SNR}_n(\boldsymbol{x})}{\mathrm{SNR}_{\mathrm{req},n}}. \tag{9}$$

[1]A sum of monomials with real exponents, but coefficients and domain restricted to positive reals.

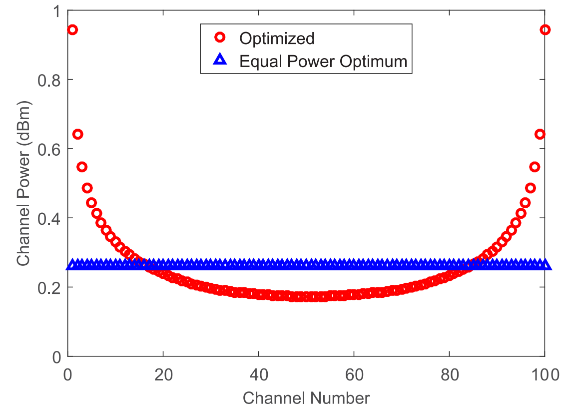

Fig. 2. Capacity-maximizing power allocation and best flat allocation. Simulations consider 40 spans of standard single-mode fiber, each 100 km long, with loss $\alpha = 0.21$ dB/km, dispersion $D = 17$ ps $\cdot$ nm^{-1} km^{-1}, nonlinear coefficient $\gamma = 1.4$ W^{-1} m^{-1}, amplification with flat 4.5-dB noise figure and gain just compensating fiber loss, and a coding gap $\Gamma = 0.794$ (-1 dB).

The log-margin

$$\log(M_n(e^{\boldsymbol{y}})) = y_n - \log(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})) - \log(\mathrm{SNR}_{\mathrm{req},n}) \tag{10}$$

is concave in the logarithmic power vector $\boldsymbol{y}$.

A. Capacity Optimization

The full-spectrum capacity is the sum of the capacities of individual channels. Assuming that a continuous family of codes is available achieving an SNR gap $\Gamma < 1$ from the Shannon capacity limit, the full-spectrum capacity for two complex polarizations is

$$C(\boldsymbol{x}) = 2\sum_{n=1}^{N} \log_2(\Gamma \mathrm{SNR}_n(\boldsymbol{x}) + 1). \tag{11}$$

SNR is log-concave in $e^{\boldsymbol{y}}$, but log-concavity is not preserved under addition like log-convexity [10]. At high SNR when $\log_2(\mathrm{SNR}+1) \approx \log_2(\mathrm{SNR})$, the capacity function is locally concave, allowing the capacity-maximizing power allocation to be found given a reasonable initial power allocation. This leads to a capacity expression

$$C(e^{\boldsymbol{y}}) = 2\sum_{n=1}^{N} \log_2(\Gamma \mathrm{SNR}_n(e^{\boldsymbol{y}}) + 1), \tag{12}$$

which is concave for the region of interest with high SNR.

The capacity-minus-gap-maximizing power allocation may be found by solving the differentiable, unconstrained, convex optimization problem

$$\text{maximize } C(e^{\boldsymbol{y}}). \tag{13}$$

Given an initial power allocation that achieves high SNR, standard ascent methods can be used to rapidly obtain the optimal power allocation and corresponding capacity. A backtracking line search using the gradient was used for the optimization presented in Figs. 2 and 3. Power optimization of a simulated point-to-point system was performed to max-

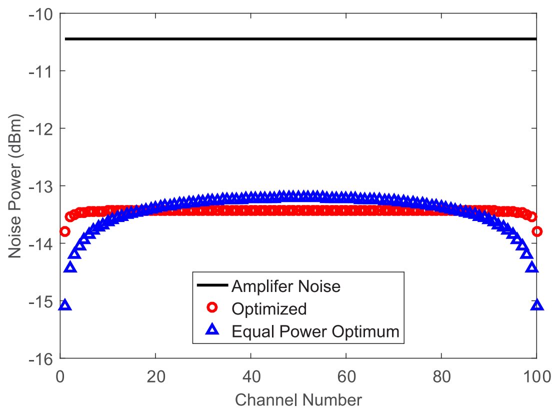

Fig. 3. Amplifier and Gaussian nonlinear noise levels for capacity-maximizing power allocation. Simulation parameters are as in Fig. 2.

imize the total data rate under two scenarios: equal power per-channel, and the true optimum. The best flat capacity can be obtained via a convex optimization by restricting the N-dimensional domain to the one-dimensional subspace where the power is equal in each channel. Simulations consider 40 spans of standard single-mode fiber, each 100 km long, with loss $\alpha = 0.21$ dB/km, dispersion $D = 17$ ps $\cdot$ nm^{-1} km^{-1}, nonlinear coefficient $\gamma = 1.4$ W^{-1} m^{-1}, amplification with flat 4.5-dB noise figure and gain just compensating fiber loss, and a coding gap $\Gamma = 0.794$ (-1 dB).

Fig. 2 compares the capacity-minus-gap-maximizing power allocation to the best flat power allocation. The powers assigned by the two allocations are very close, except at the extreme edges of the spectrum, where there are fewer interfering neighbors. A capacity of about 28.6 Tb/s is achieved from both optimization methods, with a negligible benefit from the N-dimensional optimization. We conclude that a flat power allocation effectively optimizes the achievable rate for a flat noise spectrum. Fig. 3 shows the amplifier noise power and nonlinear noise power for the optimized and flat power allocations. The optimized allocation has a flat nonlinear noise at the expected level of 3 dB below the amplifier noise power across most channels, but decreasing at the extreme edges of the band. For an isolated channel, the best SNR and hence greatest capacity is achieved when the nonlinear noise power is precisely half that of the amplifier noise power. For the extremal channels in a filled band, a balance must be reached between the powers of those channels and their neighbors, which experience greater nonlinear interference.

B. Minimum Margin Maximization

Achieving reliable communication across all channels is the primary goal of an optical communication system. Reliability is generally quantified through the SNR margin available between the operating conditions and the error correction threshold. This margin allows for system aging, fiber repairs, and transient events. Extra margin ensures that service level agreements are met for the lifetime of the system. The channel with the least margin is the most likely to fail, so maximizing the minimum margin serves to minimize the probability of the most likely failure.

Using the Gaussian noise nonlinearity model, the optimization problem to identify the power allocation that maximizes the minimum margin (9) is:

$$\text{maximize } \min_{n\in[1,N]}(M_n(\boldsymbol{x})). \quad (14)$$

The minimum of a set of convex functions is convex, making this problem log-convex in $\boldsymbol{y}$, following the log-convexity of $M_n(e^{\boldsymbol{y}})$ in $\boldsymbol{y}$. A challenge is presented by the fact that the minimum function is not differentiable, making this objective non-differentiable. We consider two methods to efficiently solve this optimization problem. In concave form, this problem is

$$\text{maximize } \min_{n\in[1,N]}(\log(M_n(e^{\boldsymbol{y}}))). \quad (15)$$

Subgradient methods allow for optimization of non-differentiable objectives, but are not descent methods and are considered slow [12]. The subgradient of the minimum of a set of functions is the gradient of one of those functions that achieves the minimum at the current point. The margin of each channel is predominately affected by nearby channels, making each subgradient step comparatively small. As the margin of each channel must be calculated at each step in order to identify which channel has the minimum margin at the current point, a straight subgradient method can be regarded as computationally infeasible due to the huge number of iterations that would be required.

Using the subgradient method as a starting point, we propose two methods to accelerate the convergence towards the solution. The first method is to solve a different, but similar, problem with a gradient that is non-negligible for a greater portion of the channel variables. The second acceleration is a nonlinear margin correction operation that makes uses properties known about the region surrounding the solution.

A problem similar to the minimum-margin problem is to maximize the sum of the margins of the set A of smallest margins minus an attenuation factor a times the sum of the margins of the set B of largest margins:

$$\text{maximize } \sum_{n\in A}(\log(M_n(e^{\boldsymbol{y}}))) - a\sum_{n\in B}(\log(M_n(e^{\boldsymbol{y}}))). \quad (16)$$

The solution of optimization problem (16) is, in general, not the same as that of the minimum-margin problem. When a goes to zero and $|A|$, the number of elements of the set A, goes to one, this problem turns into the minimum-margin problem (15).

Starting with an instance of (16) and iterating subgradient steps while $|A|$, $|B|$, and a are decreased, transforming problem (16) into (15), provides a significantly faster solution to (15). The subgradient of (16) is the sum of the subgradients for the $|A|$ minimum margin channels. This increases the magnitude of the steps taken, as compared to the single minimum margin objective (15), without over-focusing on a single dimension. As the dominant inter-channel nonlinear component is the XPM interaction with immediately neighboring channels, the single-channel gradient is generally one that raises the current channel power while lowering all other powers, but particularly those of immediate neighbors. If the two lowest-margin channels are ever neighbors, this leads to a slowly decaying oscillation between these two channels as they each raise themselves at the expense of the other. As the subgradient method is not a descent method, and uses predetermined step sizes, this slow oscillation cannot be shortcut by taking the precise step sizes required to raise two neighboring channels just the right amount. Using the subgradient of (16) with $|A| > 1$ helps to prevent simple oscillation patterns from developing. The combination of taking larger steps without overshoot, and avoiding simple oscillation patterns, leads to rapid convergence for problem (16). The solution of (16) is close to that of (15), so by changing the objective from that of (16) to that of (15) during optimization, we can accelerate the initial convergence while still solving the desired problem.

As the Gaussian noise nonlinearity of each channel depends on the power of all others, the optimal solution of the minimum-margin problem has equal margin in all channels. Any margin above this indicates that power can be given up to improve the margin of other channels. At low power levels, the SNR of each channel is linear in the corresponding signal power. As the power of any channel increases, the nonlinear noise component becomes increasingly dominant. At the SNR-maximizing point for any given channel, the derivative of SNR with respect to the channel's power is zero. The minimum-margin optimizing power allocation will never assign channel n a power above that which maximizes SNR_n, holding all other channels constant. From this we can conclude that by assuming a linear relationship between channel power and SNR, we will over-estimate the impact of power on SNR. If we wish to scale SNR_n by at most a factor $\theta > 0$, then scaling P_n by θ will scale SNR_n by some amount less than or equal to θ.

At any point $\boldsymbol{y}$ in the problem domain, channel n has margin $M_n(e^{\boldsymbol{y}})$. Assume channel m has the minimum margin at this point. Decreasing the power level of all other channels $i \neq m$ by a factor smaller than $\frac{M_i(e^{\boldsymbol{y}})}{M_m(e^{\boldsymbol{y}})}$ will always increase the minimum margin. Adding this operation to the optimization process helps to push the subgradient method towards an equal-margin solution. While this operation will alway improve the solution at any point, it decreases the average power of the set of channels. When this margin correction operation is performed frequently, the optimization may converge to a false solution, as the power added by the subgradient steps reaches an equilibrium with the power lost in the correction steps.

When $|A| > 1$, the optimization method (16) converges towards a solution with equal margin in all channels not in A. The margin correction operation pushes intermediate optimization states of (16) towards the desired solution for $|A| = 1$ with equal margin in all channels.

The expression for the margin correction used for results in this paper is

$$y_i = y_i + \frac{1}{2}\log\left(\frac{M_{\min}(e^{\boldsymbol{y}})}{M_i(e^{\boldsymbol{y}})}\right). \tag{17}$$

This correction is applied when at least five subgradient steps have been taken since the last correction and the minimum margin is at least two standard deviations below the average margin.

The method described here makes a subgradient-based optimization a feasible way to obtain the solution that maximizes the minimum margin. The challenges of subgradient-based optimization not being a descent method and having only asymptotic convergence remain. With the accelerations presented here, this manifests in the method requiring careful tuning in order to obtain an accurate solution after a given number of iterations. The next section develops a more robust solution method that provides a bound on the sub-optimality of the solution along with the optimized power allocation.

C. Differentiable Minimum Margin Maximization

The minimum-margin-maximization problem can be alternatively expressed as a problem of maximizing the the minimum inverse margin (9):

$$\text{minimize} \max_{n\in[1,N]} (M_n^{-1}(e^{\boldsymbol{y}})). \tag{18}$$

The inverse margin of channel n is a posynomial [10], as the variable $\boldsymbol{x}$ represents power, and is correspondingly a convex function of $\boldsymbol{y}$ when $\boldsymbol{x} = e^{\boldsymbol{y}}$. Expanding the objective of (18) using (9) gives the optimization problem

$$\text{minimize} \max_{n\in[1,N]} \left(\frac{\mathrm{SNR}_{\mathrm{req},n}(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}}))}{e^{y_n}}\right). \tag{19}$$

As the objective is non-negative and a logarithm is monotonic, a logarithm of the objective can be taken while maintaining convexity and forming a problem with the same optimal solution. Taking a logarithm is desired in this scenario due to exponential factors of the optimization variable $\boldsymbol{y}$, which can be linearized with a logarithm, thereby improving the accuracy of the quadratic approximations used in a Newton's method-based optimization [11]. Thus, the optimization problem (19) becomes

$$\begin{aligned} \text{minimize} \quad & \max_{n\in[1,N]} (\log(\mathrm{SNR}_{\mathrm{req},n}) \\ & + \log(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})) - y_n). \end{aligned} \tag{20}$$

The maximum in (20) can be expanded to a set of inequality constraints by introducing a slack variable s. Taking the variable s to the left side of the inequalities leaves a convex inequality-constrained problem in standard form:

$$\begin{aligned} &\text{minimize } s \\ &\text{subject to } \log(\mathrm{SNR}_{\mathrm{req},n}) + \log(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})) \\ &\qquad - y_n - s \leq 0 \quad \forall n \in [1, N]. \end{aligned} \tag{21}$$

In (21), $\log(\mathrm{SNR}_{\mathrm{req},n})$ is a constant, $\log(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}}))$ is convex, and $y_n + s$ is linear, making each constraint convex.

The inequality constraints in (21) can be considered as defining a region where the objective becomes infinite. Any feasible solution will have a finite objective value and will thus fall outside of this region. A barrier function that is infinite when its argument is positive can be used to bring the constraints of (21) into the objective:

$$\mathrm{barrier}(x) = \begin{cases} 0, & x \leq 0 \\ \infty, & \text{otherwise.} \end{cases} \tag{22}$$

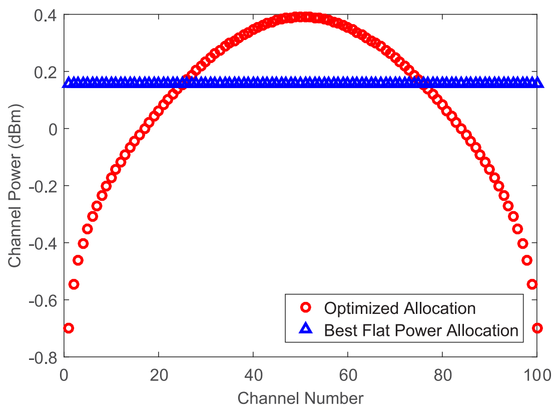

Fig. 4. Optimized power allocation for maximum margin with a homogeneous modulation format and flat amplifier noise. Simulation parameters are as in Fig. 2.

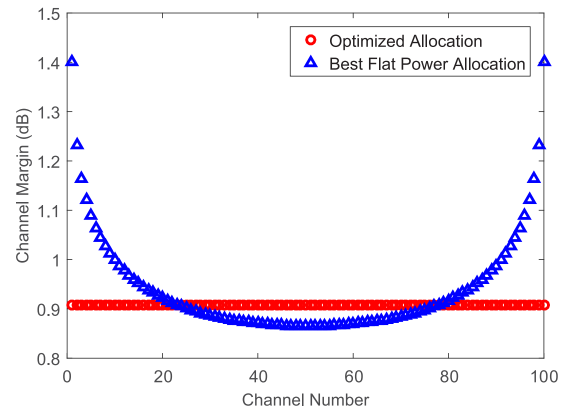

Fig. 5. Channel margins with optimized power allocation for a homogeneous modulation format and flat amplifier noise. Simulation parameters are as in Fig. 2.

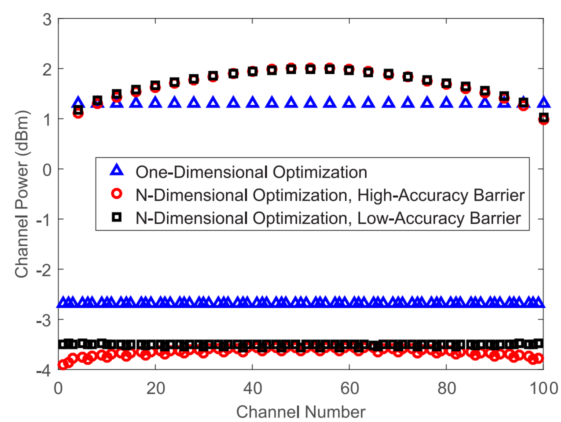

Fig. 6. Power allocation for inhomogeneous required SNRs under various optimization scenarios. Minimum margin optimization with N dimensions assigns higher powers for channels with high SNR requirements and lower powers for channels with lower SNR requirements than a one-dimensional optimization. Simulation parameters are as in Fig. 2.

An ideal barrier is not differentiable, but a differentiable approximation to an ideal barrier is given by

$$-\left(\frac{1}{t}\right)\log(-x). \tag{23}$$

The accuracy of the approximation is controlled by the parameter t such that (23) becomes (22) in the limit $t \to \infty$.

Using (23), problem (21) becomes

$$\text{minimize } s - \left(\frac{1}{t}\right)\sum_{n=1}^{N}\log\left[-\log(\mathrm{SNR}_{\mathrm{req},n})\right.$$
$$\left. -\log({\sigma^2}_n + NL_n(e^{\boldsymbol{y}})) + y_n + s\right], \tag{24}$$

for $t \to \infty$. The objective of (24) is twice differentiable, allowing the use of Newton steps for rapid convergence. For large values of t, the barrier approximation improves, but the Hessian matrix varies rapidly near the inequality boundaries, making the minimization difficult. This type of problem is typically solved for a series of increasing values of t, where the solution from the previous value of t provides the starting point for iterations with the next value of t [10]. The inaccuracy in the optimal minimum inverse margin s^* due to the barrier approximation is bounded by $\frac{N}{t}$ which, combined with the inaccuracy of the Newton optimization, produces a sub-optimality bound for the solution. The Newton steps make use of N, $N \times N$ Hessian matrices of the nonlinearity function to calculate the Hessian of the objective, as shown in Appendix A. If memory becomes a limitation due to a small spectral granularity and correspondingly large N, the Hessian of the nonlinearity for the channels may be found sequentially, at the expense of some redundant computation.

Fig. 4 shows an optimized power allocation when the available spectrum has been divided into 100 channels, each 50 GHz wide, following a typical channel grid. The system parameters simulated are 40 spans of 100-km length with fiber attenuation $\alpha = 0.21$ dB/km, dispersion $D = 17$ ps $\cdot$ nm^{-1} km^{-1}, and nonlinear coefficient $\gamma = 1.4$ W^{-1} m^{-1}. The noise level simulated corresponds to optical amplifiers with a 4.5-dB noise figure and an ideal flat gain profile. The power levels have been optimized to maximize the minimum margin above a 8-dB required SNR across all channels. The increase in minimum margin, shown in Fig. 5, from the optimized power allocation over the best flat power allocation is a minor 0.043 dB. This confirms that a flat power allocation is very close to optimal for a single required SNR for all channels on a homogeneous point-to-point link, as was found in [8].

The flat power allocation can be considered as the one-dimensional approximation to what is in this case a 100-dimensional optimization problem. The SNR, and hence margin, of each channel is very flat near the optimal point, leading the one-dimensional approximation to fall very close to the true optimal value for this homogeneous scenario. A much more useful application of the margin-maximizing optimization is in finding the optimal power allocation in situations where a one-dimensional approximation is inaccurate. Such situations would include those where the noise spectrum is not flat or when the required SNR to be delivered is not equal across all channels.

Fig. 6 shows optimized power allocations, and Fig. 7 shows the margins achieved, for an interleaved combination of two

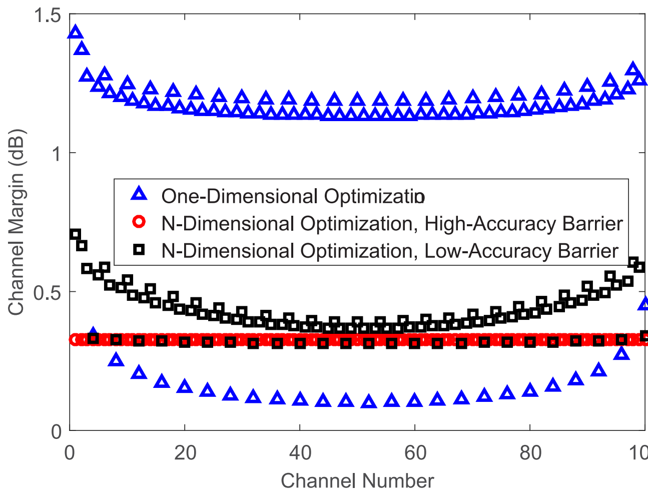

Fig. 7. Channel margins with inhomogeneous required SNRs under various optimization scenarios. Power optimization sacrifices margin in channels with higher margin to maximize the minimum-margin channels resulting in equal margin in all channels. Simulation parameters are as in Fig. 2.

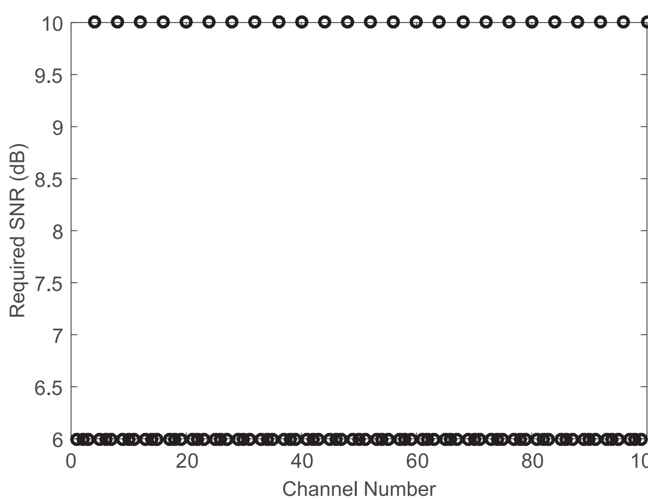

Fig. 8. Required SNR for the inhomogeneous margin optimization scenario of Figs. 6 and 7.

required SNR levels, which is shown in Fig. 8. Such an interleaved set of SNR requirements attempts to minimize the nonlinear interactions of the high power channels. The higher SNR requirements may originate from various sources including: using quantized set of communication rates that fall above and below the highest achievable continuous rate, or from combining different generations of communication hardware. The one-dimensional approximate solution in this scenario maintains a fixed ratio between the powers of the channels equal to the corresponding ratio of their required SNRs. This corresponds to the assumption that worst-case nonlinear noise is shared by all channels. Unlike the homogeneous scenario of Figs. 4 and 5, there is a notable 0.23-dB minimum margin improvement from the full margin optimization, as compared to the one-dimensional method, for the inhomogeneous scenario of Figs. 6–8.

For a point-to-point communication system, the enhanced subgradient method and the differentiable barrier method solve the minimum margin problem with similar speed. The solution provided by the enhanced subgradient method always provides equal margin to all channels. The barrier method supports tuning of the barrier accuracy depending upon the largest value of t used. In an inhomogeneous point-to-point link, certain channels with higher SNR requirements will generally limit the minimum margin. Channels far away from the limiting channels will have a negligible nonlinear impact on the minimum margin. Terminating with more slack in the barrier function approximation will trade off potentially large amounts of margin in these channels against small decreases in the minimum margin. This is illustrated in Figs. 6 and 7, which include a third optimized power allocation making use of a low-accuracy barrier function. The high-accuracy barrier has a relative sub-optimality bound $\frac{N}{t}$ of $2^{-22} \approx 2.4 \times 10^{-7}$, while the low-accuracy barrier has a relative sub-optimality bound of $2^{-10} \approx 10^{-3}$. The lower-accuracy barrier reduces the minimum margin by 0.01 dB, but achieves gains on non-limiting channels.

The dual variables of the barrier constraints provide a measure of the relative tightness of each of the inequality constraints. The dual variables can be evaluated during the optimization process following (25) for each channel n [10]:

$$\lambda_n = -\frac{1}{t(\log(\mathrm{SNR}_{\mathrm{req},n}) + \log({\sigma^2}_n + NL_n(e^{\boldsymbol{y}})) - y_n - s)}. \tag{25}$$

The barrier function dual variables serve to identify which channels would increase in margin should the barrier parameter t be relaxed. This information can be useful for derivative optimization problems that solve multiple instances of the margin maximization problem. One such example is the relaxed minimum-margin problem where the ordering of SNR requirements is not fixed.

III. Mesh Power Optimization

Isolated point-to-point links are often filled by a homogeneous set of channels using a common modulation and coding scheme. In mesh networks, by contrast, different channels generally traverse different distances, accumulating different amounts of noise, and may use different modulation or coding schemes. Mesh networks may also experience wavelength fragmentation, and partial utilization, such that different channels have different numbers of interfering neighbors. A homogeneous network supports a near-optimal one-dimensional optimization, as shown in Section II-C. A one-dimension-per-section optimization is unable to accommodate the inhomogeneities commonly found in mesh networks. The optimal minimum-margin power allocation takes advantage of network and wavelength inhomogeneity and is able to provide the greatest margin increase in mesh scenarios.

Wavelength-selective switches situated at the nodes of a mesh can provide wavelength-dependent attenuation. The combination of wavelength-selective switches and adjustable amplification allows flexible power control for each wavelength on each section of the network. We thus use independent signal power variables for each channel on each section of the network. Despite the independent channel powers for each network section, a mesh should generally not be considered as a set of point-to-point links. Channels that traverse multiple sections of the network couple the optimization objectives associated with separate sections.

Achieving reliable communication between a pair of transceivers through a mesh network using a particular modulation and coding scheme imposes SNR requirements on the spectrum employed over the network sections traversed. We assume that there is a set $\mathcal{S}$ of network sections, and the number of networks sections is $|\mathcal{S}|$. To populate the network, we assume that there exists a set $\mathcal{R}$ of non-overlapping requirements, each with assigned spectra and set of traversed network sections.

For channel n, given requirement $r \in \mathcal{R}$ that makes use of this channel, the SNR seen by this requirement over its corresponding set $\mathcal{S}^{(r)}$ of network sections is

$$\mathrm{SNR}_n^{(r)} = \frac{1}{\sum_{s \in \mathcal{S}^{(r)}} \frac{1}{\mathrm{SNR}_s}}. \tag{26}$$

The margin seen by requirement $r \in \mathcal{R}$ on channel n is

$$M_n^{(r)} = \frac{\mathrm{SNR}_n^{(r)}}{\mathrm{SNR}_{\mathrm{req}}^{(r)}}. \tag{27}$$

Let $\boldsymbol{y}^s$ be a vector of logarithmic powers for channels on network section s. Let $\boldsymbol{y}$ be a vector of logarithmic power variables that is the concatenation of the vectors $\boldsymbol{y}^s$ for each $s \in \mathcal{S}$:

$$\boldsymbol{y} = \left[\boldsymbol{y}^{1^T} \quad \cdots \quad \boldsymbol{y}^{|\mathcal{S}|^T} \right]^T. \tag{28}$$

Making use of the vector of variables logarithmic power variables $\boldsymbol{y}$ for each channel on each section of the mesh network, the margin for each requirement on the network has a log-concave expression:

$$\log\left(M_n^{(r)}\right) = \log\left(\frac{1}{\mathrm{SNR}_{req}^{(r)}}\right) - \log\left(\sum_{s \in \mathcal{S}^{(r)}} \frac{\sigma_n^{s\,2} + NL_n^s\left(e^{\boldsymbol{y}^s}\right)}{e^{y_n^s}}\right). \tag{29}$$

In general the coefficients used for each section of the mesh will vary due to differing fiber lengths and types. Using the epsilon-law approximation [7] for χ in the coefficient integration can improve computational efficiency for a mesh network with consistent span lengths, as it may enable coefficient reuse with appropriate scaling for the number of spans per section.

The two solution methods described in Section II for optimizing minimum margin in a point-to-point link can both be extended for optimizing a mesh network. The number of variables and margin inequalities increase with the number of network sections. The enhanced subgradient method's performance scales better as the network size increases, if the optimization parameters are tuned correctly, but this tuning process is laborious and there are no useful bounds on the sub-optimality of a solution. A further possible drawback of the barrier method with Newton steps, as compared to the subgradient-based method, is the memory needed for an array of Hessian matrices for each channel for each section of the network.

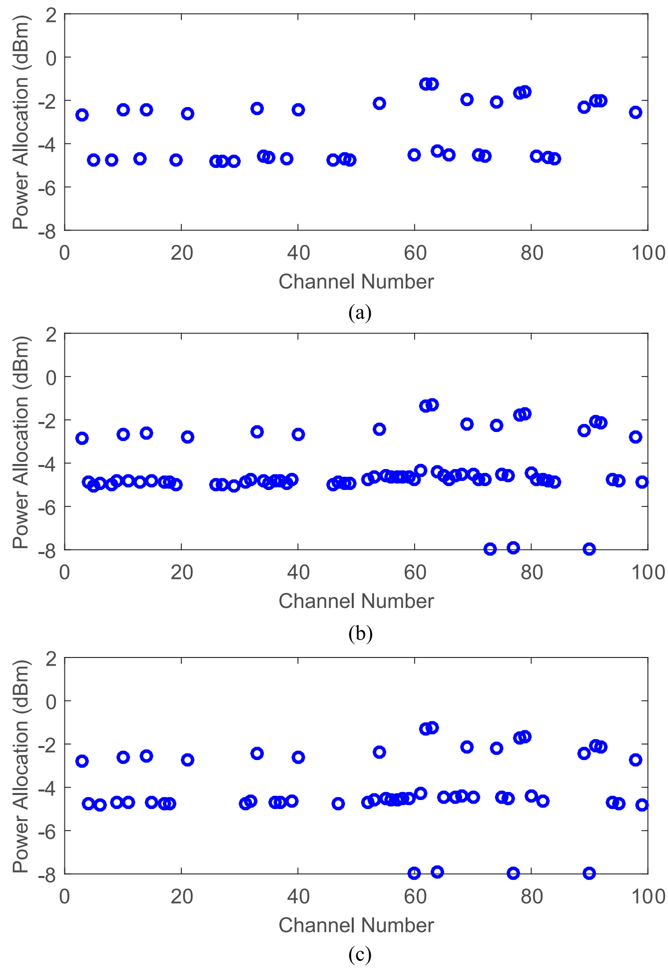

Fig. 9. Margin-maximizing power allocations for randomly assigned paths in three-section mesh network. (a) Section I. (b) Section II. (c) Section III. The mesh network uses 10 spans per section. Other parameters are as in Fig. 2.

Fig. 10. Network layout for three-section linear network used for the optimization reported in Fig. 9 with link lengths in km.

In a mesh network, the presence of an unused channel on a section in the network can lead to a singular or nearly singular Hessian matrix for the overall objective function of the barrier method. Such a matrix prevents the calculation of the optimal step direction, as it cannot be inverted. This can be avoided by adding in dummy requirements across the unused channels with SNR targets significantly below unity, or by indexing and removing the variables corresponding to unused channels.

Fig. 9 shows the optimized power allocation for the three-section linear network illustrated in Fig. 10. This trivial network is used to illustrate a mesh network optimization at an interpretable scale. The demands have been randomly populated to pass over one to three of the network sections, each comprising 20 spans of fiber. The same required SNR has been assumed for all paths, leading to the three distinct tiers of power levels. The limiting channels for maximizing the margin are those that pass

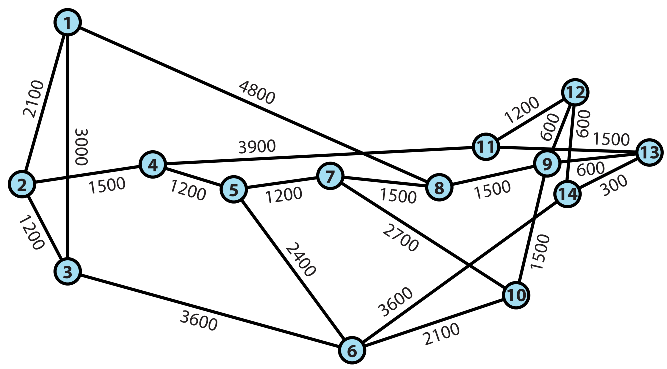

Fig. 11. The NSFNET network used as a reference network in simulations, with link lengths in km [13].

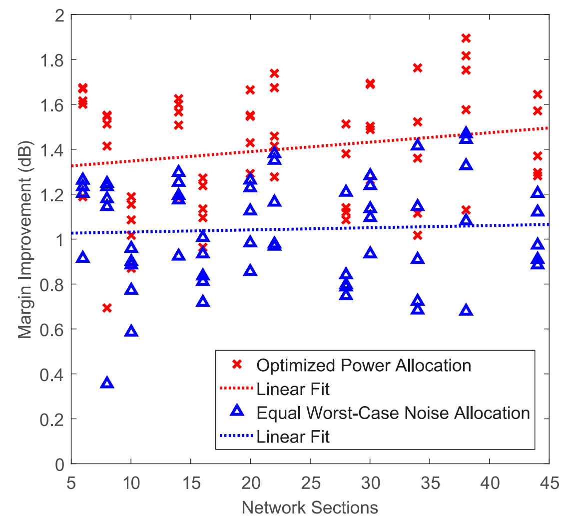

Fig. 12. Margin improvement from power optimization compared to optimized flat power allocation for each network section in subsets of the NSFNET network of Fig. 11. The equal worst-case noise allocation is the $|\mathcal{S}|$-dimensional optimization that assumes the worst channel's nonlinear noise for all wavelengths, where $|\mathcal{S}|$ is the number of fiber sections. Network subsets are formed using node ranges from 1–3 up to 1–14 with separate links for each direction along each network path. Networks are filled with randomly generated uniform traffic demands, with identical modulation formats, that were routed with congestion routing [13]. Parameters other than path lengths are as in Fig. 2.

through all three sections of the network, and particularly those with neighbors that also traverse all network sections. These channels receive the highest power assignment and experience noticeable inter-channel nonlinearity when they are neighbors with other limiting channels. The channels that pass through all of the network sections also exhibit the most variability in optimal power level depending upon the number and power of their neighbors.

The variable path lengths in mesh networks lead to variations in the delivered SNR. Even with a fixed modulation and coding scheme leading to a fixed SNR requirement across all channels, as is used for the scenario of Fig. 9, the variation in delivered SNR makes margin gains from power optimization possible. The minimum margin of this three-section network was increased by 1.0 dB over the best flat power allocation and by 0.30 dB over a three-dimensional power optimization that assumes the worst-case nonlinear noise for all channels for each section.

The NSFNET continental network shown in Fig. 11 was selected for use as a reference network to test the benefits of optimized power allocations [13]. Any potential scaling of the margin benefit available from optimizing the power allocation was detected by optimizing over a series of subsets of the network by including only nodes selected from the range 1–3 up to the range 1–14. The networks were filled with randomly generated traffic demands with a uniform demand distribution between every node pair. All demands were satisfied using communication channels with identical modulation formats and coding. Routing and wavelength assignments were performed using a shortest-path computation across all N wavelengths with the length metric augmented by a linear estimation of the expected inter-channel nonlinearity for the current wavelength and the link congestion [13]. The 22 graph edges of the full network in Fig. 11 become 44 network sections, as separate communication in each direction is supported by parallel fibers. Five different sets of randomized traffic demands were routed and optimized for each network size.

Fig. 12 shows that a significant average margin benefit of 1.5 dB is available from an optimized power allocation for the full-scale network, as compared to a power allocation that uses the best possible wavelength-independent power for each section. There is a slight upward trend in the averaged margin benefit as the network size increases, which is clearly illustrated by the linear fit to the optimized power allocation benefit versus network sections.

The equal worst-case noise allocation is a $|\mathcal{S}|$-dimensional optimization, where $|\mathcal{S}|$ is the number of network sections. This sub-optimal optimization objective assumes the highest nonlinear noise occurs for all channels. This $|\mathcal{S}|$-dimensional optimization provides moderate benefits across most of the network subsets, but has a performance penalty relative to the full optimization that widens as the network size grows. If computation time is a concern for large network sizes, the $|\mathcal{S}|$-dimensional optimization is faster than the full margin optimization computation. With only one power variable per network section, the margin expression is convex and differentiable, so gradient ascent with a backtracking line search converges rapidly.

IV. Discussion

The Gaussian noise model for the Kerr nonlinearity is typically used to solve for a single optimal power level across all channels in a fiber [14]. For point-to-point links, with a uniform modulation format and coding, such a flat power allocation is nearly optimal, as demonstrated in this paper and in [8]. Current optical communication systems provide a limited degree of rate variation through modulation format adjustment among BPSK, QPSK, and 16-QAM. These modulation format changes correspond to large jumps in required SNR or, equivalently, a coarse granularity of SNR requirements. The SNR delivered by a network generally cannot be well matched by such a coarse set of SNR requirements. This mismatch leads to an inefficiency in the communication system. For systems that have essentially equal amplifier noise accumulation on all channels, such as many point-to-point links, power optimization has a negligible effect on mitigating this inefficiency. In a mesh network, adjacent channels often propagate over different distances to different destinations, with corresponding variations in amplifier noise accumulation. For these networks, an optimized power allocation can improve the minimum margin by adjusting the SNRs delivered to different channels. This effect is mediated by the inter-channel components of the Gaussian nonlinear noise, and is limited by the portion of the overall noise that is due to nonlinearity. In this way, power optimization can compensate for some of the inefficiency resulting from using a coarse set of communication rates having correspondingly coarse SNR requirements.

Fine-grained quantization of communication rates and SNRs removes most of the inefficiency resulting from a coarse rate granularity. A continuous set of data rates could remove all of the inefficiency, but a practical communication system will require a finite set of distinct rates. With a set of SNR requirements quantized from the ideal continuous set, the margin-optimizing power allocation is needed to supply a power allocation.

The continuous capacity-minus-coding-gap optimization serves as an upper bound on the system capacity with a given coding strength. With a fine-grained rate quantization, this upper bound becomes tight at more discrete SNR levels, making the bound more relevant. The power allocation that maximizes the total continuous data rate is also useful in optimizing the margin of a network changing over time. The power allocation that maximizes the minimum margin in a network for one set of requirements may reduce the capacity available for future channels that are added to the network. Using the continuous rate-maximizing allocation as a power cap will balance the needs of optimizing the network in its current state against the possible performance of the network given future demands.

The optimization methods described in this paper provide performance benefits while operating at a software layer. All that is required from a transceiver hardware standpoint is fine transmission power control, so even small performance benefits are economically attractive. Mesh network optimization requires detailed knowledge of the architecture and layout of the network: transceiver wavelengths and SNR requirements, fiber types, span lengths and power profiles, and network topology. Detailed network records are generally available, so this should not be an obstacle to putting these methods into service.

The main barrier to implementation of margin optimization is a near-quadratic scaling of computational complexity with network size. The gradient and Hessian calculations scale with the number of variables, which is proportional to the number of network sections, as well as with the number of communication requirements that fill the network. Larger networks have more communication requirements, and may have channels that traverse more sections of the network. The optimization methods of this paper were implemented in MATLAB in a serial manner. The optimization can be sped up using a native code implementation and by exploiting the inherent parallelism of the problem. Care should be taken to exploit structure and sparsity in the optimization calculations. These improvement factors should make mesh optimization practical for even large networks.

The question of computational complexity is also an economic one. In order to perform regular power optimization on a large reconfigurable network, a dedicated commodity server may be required. The value of 1 dB of additional worst-case margin on a single long-haul modem is more than the cost of such a commodity server. A large mesh network will contain hundreds or thousands of modems that are all controlled by the same server. In large mesh applications where fast power control is necessary, sub-optimal optimization objectives could also be used to trade off margin benefits against computational complexity.

V. Conclusion

We presented convex formulations for power optimization to maximize the minimum channel margin and to maximize a continuously variable data rate using a Gaussian noise nonlinearity model. Margin and capacity gains over a flat power allocation are minimal in homogeneous systems with flat noise spectra. Valuable gains can be obtained in systems combining multiple SNR requirements, partial wavelength fill, or variable noise spectra. The margin-maximization optimization problem has been extended to a mesh network scenario, where diverse network paths provide inhomogeneity that allows performance to be extracted through power optimization. A margin gain of 1.5 dB on average is achieved for the 14-node NSFNET network using a mesh power optimization as compared to an optimized flat power allocation for each network section.

Appendix A
Expressions for Gradients and Hessian Matrices

The nonlinear Gaussian noise function of channel n is:

$$\begin{aligned} NL_n(e^{\boldsymbol{y}}) &= \gamma^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=-1}^{1} e^{y_i+y_j+y_{i+j-n+l}+\log(D_l(i,j,n))} \\ &= \gamma^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=-1}^{1} e^{\boldsymbol{a}_{i,j,l,n}^T \boldsymbol{y}+\log(D_l(i,j,n))} \end{aligned} \tag{30}$$

where $\boldsymbol{e}_k$ is the kth unit vector, and

$$\boldsymbol{a}_{i,j,l,n} = \boldsymbol{e}_i + \boldsymbol{e}_j + \boldsymbol{e}_{i+j-n+l}. \tag{31}$$

The gradient of the nonlinear Gaussian noise function is

$$\nabla_{\boldsymbol{y}} NL_n(e^{\boldsymbol{y}}) = \gamma^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=-1}^{1} \boldsymbol{a}_{i,j,l,n} e^{\boldsymbol{a}_{i,j,l,n}^T \boldsymbol{y}+\log(D_l(i,j,n))}. \tag{32}$$

It is convenient to consider the set of gradients for each channel as a matrix $\nabla_{\boldsymbol{y}} NL$.

The Hessian matrix of the nonlinear Gaussian noise function, for a given value of n is

$$\nabla_{\boldsymbol{y}}^2 NL_n(e^{\boldsymbol{y}}) = \gamma^2 \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{l=-1}^{1} \boldsymbol{a}_{i,j,l,n} \boldsymbol{a}_{i,j,l,n}^T e^{\boldsymbol{a}_{i,j,l,n}^T \boldsymbol{y}+\log(D_l(i,j,n))}. \tag{33}$$

Equation (7), which expresses the SNR in terms of the logarithmic power variables $\boldsymbol{y}$, is log-concave but neither convex nor concave. The gradient of the SNR is useful for capacity optimization where a convex capacity expression is obtained for high SNR, where $\mathrm{SNR} + 1 \approx \mathrm{SNR}$.

$$\nabla_{\boldsymbol{y}} SNR_n(e^{\boldsymbol{y}}) = \frac{\boldsymbol{e}_n e^{y_n}}{\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})} - \frac{\nabla_{\boldsymbol{y}} NL_n e^{y_n}}{(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}}))^2}. \tag{34}$$

The gradient of the capacity (12) is

$$\nabla_{\boldsymbol{y}} C(e^{\boldsymbol{y}}) = 2\sum_{n=1}^{N} \frac{10^{\frac{-1}{10}} \nabla_{\boldsymbol{y}} \mathrm{SNR}_n(e^{\boldsymbol{y}})}{(10^{\frac{-1}{10}} \mathrm{SNR}_n(e^{\boldsymbol{y}}) + 1)) ln(2)}. \tag{35}$$

The objective of the convex barrier method for solving the minimum margin problem (24), denoted by $O(\boldsymbol{z})$, is a function of the logarithmic power vector augmented by the additional variable: $\boldsymbol{z} = [\boldsymbol{y}^T s]^T$. The expressions for the gradient and Hessian will use $\boldsymbol{y}$ and s to refer to portions of the vector $\boldsymbol{z}$ rather than complicate the notation with partial identity matrices.

Let

$$F_n(\boldsymbol{z}) = \log(\mathrm{SNR}_{\mathrm{req},n}) + log(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})) - y_n - s, \tag{36}$$

and let

$$\nabla_{\boldsymbol{z}} NL_n = [\nabla_{\boldsymbol{y}} NL_n(e^{\boldsymbol{y}})^T \ 0]^T. \tag{37}$$

The function $F_n(z)$ has gradient and Hessian

$$\nabla F_n(\boldsymbol{z}) = \frac{\nabla_{\boldsymbol{z}} NL_n}{\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})} - \boldsymbol{e}_n - \boldsymbol{e}_{N+1}, \tag{38}$$

$$\nabla^2 F_n(\boldsymbol{z}) = \frac{\begin{bmatrix} \nabla^2_{\boldsymbol{y}} NL_n(e^{\boldsymbol{y}}) & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}}{\sigma^2{}_n + NL_n(e^{\boldsymbol{y}})} - \frac{\nabla_{\boldsymbol{z}} NL_n \, \nabla_{\boldsymbol{z}} NL_n^T}{(\sigma^2{}_n + NL_n(e^{\boldsymbol{y}}))^2}. \tag{39}$$

Using the functions $F_n(\boldsymbol{z})$ and their derivatives, the gradient and Hessian of the objective $O(\boldsymbol{z})$ of the barrier method for the minimum margin problem are

$$\nabla O(\boldsymbol{z}) = \boldsymbol{e}_{N+1} - \left(\frac{1}{t}\right) \sum_{n=1}^{N} \frac{\nabla F_n(\boldsymbol{z})}{F_n(\boldsymbol{z})}, \tag{40}$$

and

$$\nabla^2 O(\boldsymbol{z}) = \left(\frac{1}{t}\right) \sum_{n=1}^{N} \left[\frac{\nabla F_n(\boldsymbol{z}) \nabla F_n(\boldsymbol{z})^T}{F_n(\boldsymbol{z})^2} - \frac{\nabla^2 F_n(\boldsymbol{z})}{F_n(\boldsymbol{z})} \right]. \tag{41}$$

The Newton step direction is the the usual one [10]:

$$\boldsymbol{z}_{\mathrm{Newton}} = \nabla^2 O(\boldsymbol{z})^{-1} \nabla O(\boldsymbol{z}). \tag{42}$$

References

[1] G.-H. Gho, L. Klak, and J. M. Kahn, "Rate-adaptive coding for optical fiber transmission systems," *J. Lightw. Technol.*, vol. 29, no. 2, pp. 222–233, Jan. 2011.

[2] M. Arabaci, I. B. Djordjevic, R. Saunders, and R. M. Marcoccia, "Polarization-multiplexed rate-adaptive non-binary-quasi-cyclic-LDPC-coded multilevel modulation with coherent detection for optical transport networks," *Opt. Express*, vol. 18, no. 3, pp. 1820–1832, Feb. 2010.

[3] D. A. A. Mello, A. N. Barreto, T. C. de Lima, T. F. Portela, L. Beygi, and J. M. Kahn, "Optical networking with variable-code-rate transceivers," *J. Lightw. Technol.*, vol. 32, no. 2, pp. 257–266, Jan. 2014.

[4] G.-H. Gho and J. M. Kahn, "Rate-adaptive modulation and low-density parity-check coding for optical fiber transmission systems," *J. Opt. Commun. Netw.*, vol. 4, no. 10, pp. 760–768, 2012.

[5] I. B. Djordjevic and G. T. Djordjevic, "On the communication over strong atmospheric turbulence channels by adaptive modulation and coding," *Opt. Express*, vol. 17, no, 20, pp. 18250–18262, 2009.

[6] J. M. Cioffi, G. P. Dudevoir, M. V. Eyuboglu, and J. G. David. Forney, "Mmse decision-feedback equalizers and coding-part 11: Coding results," *IEEE Trans. Commun.*, vol. 43, no. 10, pp. 2595–2604, Oct. 1995.

[7] P. Poggiolini, "The GN model of non-linear propagation in uncompensated coherent optical systems," *J. Lightw. Technol.*, vol. 30, no. 24, pp. 3857–3879 , Dec. 2012.

[8] D. J. Ives, P. Bayvel, and S. J. Savory, "Adapting transmitter power and modulation format to improve optical network performance utilizing the gaussian noise model of nonlinear impairments," *J. Lightw. Technol.*, vol. 32, no. 21, pp. 3485–3494, Nov. 2014.

[9] D. J. Ives, P. Bayvel, and S. J. Savory, "Routing, modulation, spectrum and launch power assignment to maximize the traffic throughput of a nonlinear optical mesh network," *Photon. Netw. Commun.*, vol. 29, pp. 244–256, Mar. 2015.

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

[11] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimization Eng.*, vol. 8, no. 1, pp. 67–127, 2007.

[12] S. Boyd and J. Park, *Subgradient Methods*, *Notes EE364b, Stanford, CA, USA: Stanford Univ.*, 2014.

[13] S. J. Savory, "Congestion aware routing in nonlinear elastic optical networks," *IEEE Photon. Technol. Lett.*, vol. 26, no. 10, pp. 1057–1060, Aug. 2014.

[14] P. Poggiolini, G. Bosco, A. Carena, V. Curri, Y. Jiang, and F. Forghieri, "The GN model of non-linear propagation and its applications," *J. Lightw. Technol.*, vol. 32, no. 4, pp. 694–721, Feb. 2014.

Ian Roberts received the Bachelor's degree from the University of Ottawa, Ottawa, ON, Canada, and the Master's degree from the Stanford University, Stanford, CA, USA, both in electrical engineering. He is working toward the Ph.D. degree at the Stanford University. He worked at Ciena on DSP hardware development for WaveLogic 3 Extreme prior to beginning graduate studies. His research interests include optical communication, optimization, and compensation of nonlinear propagation.

Joseph M. Kahn (F'00) received the A.B., M.A., and Ph.D. degrees in physics from the University of California, Berkeley, CA, USA, in 1981, 1983, and 1986, respectively. From 1987–1990, he was at the ATT Bell Laboratories, Crawford Hill Laboratory, in Holmdel, NJ, USA. He demonstrated multi-Gbit/s coherent optical fiber transmission systems, setting world records for receiver sensitivity. From 1990 to 2003, he was on the faculty of the Department of Electrical Engineering and Computer Sciences, University of California performing research on optical and wireless communications. Since 2003, he has been a Professor of electrical engineering at the Stanford University, where he heads the Optical Communications Group. From 1993–2000, he served as a Technical Editor of *IEEE Personal Communications Magazine*. From 2009–2014, he was an Associate Editor of IEEE/OSA JOURNAL OF OPTICAL COMMUNICATIONS AND NETWORKING. In 2000, he helped found StrataLight Communications, where he served as Chief Scientist from 2000–2003. StrataLight was acquired by Opnext, Inc. in 2009. His current research interests include fiber-based imaging, spatial multiplexing, rate-adaptive and spectrally efficient modulation and coding methods, coherent detection and associated digital signal processing algorithms, digital compensation of fiber nonlinearity, and free-space systems. He received the National Science Foundation Presidential Young Investigator Award in 1991.

David Boertjes received the Bachelor's degree in physics from the University of New Brunswick, Fredericton, NB, Canada, the Master's degree in physics from the Dalhousie University, Halifax, NS, Canada, and the Ph.D. degree in electrical engineering from the University of Alberta, Edmonton, AB, Canada. His Ph.D. research was in the area of optical telecommunications at TRLabs in Edmonton focusing on applications of active and passive polymer optics for low cost photonic components. He is currently the Principal Architect for Ciena's photonic line transport solutions and also one of the driving forces behind the Liquid Spectrum initiative. He joined Nortel's Optical Networks division in 1998 where he worked on fibre optic transmission products with a particular focus on photonics. He has been involved in the OFC/NFOEC Conference series acting in several capacities including as an Invited Speaker and as the Chair of the Subcommittee on Optical Processing and Analogue Subsystems. His work within Ciena's RD has resulted in more than 20 patents in the area of photonic transport. He received the Ciena's Distinguished Engineer Award in 2014.