Iterative Resource Allocation Algorithm for EONs Based on a Linearized GN Model

Citation for the original published paper (version of record):
Xu, Y., Yan, L., Agrell, E. et al (2019). Iterative Resource Allocation Algorithm for EONs Based on a Linearized GN Model. Journal of Optical Communications and Networking, 11(3): 39-51.
http://dx.doi.org/10.1364/JOCN.11.000039

(article starts on next page)

Iterative Resource Allocation Algorithm for EONs Based on a Linearized GN Model

Yuxin Xu, *Student Member, IEEE,* Li Yan, *Student Member, IEEE,*
Erik Agrell, *Fellow, IEEE,* Maïté Brandt-Pearce, *Senior Member, IEEE*

***Abstract*—Elastic optical networks (EONs) rely on efficient resource planning to meet future communication needs to avoid resource overprovisioning. Estimation of physical-layer impairments (PLIs) in EONs plays an important role in the network planning stage. Traditionally, the transmission reach (TR) and Gaussian noise (GN) models have been broadly employed in the estimation of the PLIs. However, the TR model cannot accurately estimate PLIs, whereas the GN model is incompatible with state of the art linear optimization solvers. In this paper, we propose a physical-layer estimation model based on the GN model, referred to as the conservative linearized Gaussian noise (CLGN) model. To address the routing, spectrum, and regeneration assignment problem accounting for PLIs, we introduce a link-based mixed integer linear programming formulation employing the CLGN, whose heavy computational burden is relieved by a heuristic approach referred to as the sequential iterative optimization algorithm. We show through simulation that network resources such as spectrum and regeneration nodes can be saved by utilizing the CLGN model compared with the TR model. Our proposed heuristic algorithm speeds up the optimization process and provides better resource usage, compared to state of the art algorithms on benchmark networks.**

***Index Terms*—Elastic optical networks, Routing and spectrum allocation, Gaussian noise model, Regeneration placement.**

I. Introduction

With the enormous growth of the communication industry and traffic heterogeneity, elastic optical networks (EONs) have been proposed to meet future long-haul communication demands [1]. To ensure effective provisioning in EON, routing and spectrum allocation (RSA) algorithms are applied to configure light-paths while minimizing the resource usage to allow traffic to flow despite physical-layer impairments (PLIs). Appropriate modeling of PLIs in EONs avoids network over-dimensioning and unnecessary costs.

The resource provisioning problem has been extensively studied [2]. The ultimate goal is to allocate network resources for large networks efficiently. There are two essential problems in resource allocation research for EONs: (a) approximating PLIs and (b) finding a scalable near optimal solution (number of demands, network dimension, etc.).

PLIs such as fiber loss, dispersion and nonlinearities can impair the quality of transmission (QoT) in long-haul networks [3]. The QoT identifies the network's capability to recover the transmitted information. The most common approach to account for PLIs uses the transmission reach (TR) [4], which approximates the maximum distance a signal can travel without regeneration. However, the TR model lacks sufficient flexibility and accuracy [5], [6]. A more accurate model for estimating PLIs in long-haul transport networks is the Gaussian noise (GN) model, which is a state-dependent (traffic-dependent) model [7], [8]. The GN model is incompatible with scalable resource allocation algorithms because it is a nonlinear function of the state and optimization variables, which leads to massive computational needs.

The complexity of the RSA problem, itself NP-hard (non-deterministic polynomial-time hardness) [9] [10, Chapter 2], increases exponentially as the network size expands. The optimal method for solving the RSA problem is to use a mixed integer linear programming (MILP) solver [11]. For large dimension problems, MILP engines cannot find the optimal solution within a reasonable time [12]. In order to overcome this shortcoming, heuristic algorithms have been proposed to provide a sub-optimal solution for large networks.

Scalability, non-optimality, and computation complexity remain a problem for heuristic algorithms in the published literature [3], [5], [13], [14]. In [3], [13], researchers have proposed heuristic algorithms using the GN model (or a close approximation) for networks that do not require signal regeneration. In [14], researchers combine the GN model with channel banding and a path-based algorithm. Due to the low computational complexity of path-based algorithms, the algorithm in [14] results in a high-scalability, but sacrifices optimality. In [13], Zhao et al. compare path-based meta-heuristic algorithms (e.g., the simulated annealing algorithm) with a link-based MILP algorithm. Even though simulated annealing increases the optimality of the path-based algorithm and preserves the moderate scalability, the link-based MILP algorithm achieves better optimality, albeit by sacrificing scalability. In [5], Wang et al. solve the RSA problem for a large scale EON (NSF-24) with a fast heuristic algorithm, the *re-MILP* algorithm, based on a TR constraint and taking into account regeneration node placement. The consideration of regeneration placement turns the RSA problem into a more complicated routing, spectrum, and regeneration assignment (RSRA) problem. The re-MILP algorithm is more scalable than the algorithms presented in [3] and [13], because it is able to successfully solve the RSRA problem with as many as 50 demands in the NSF-24 network, and thus it has been applied [15]. However, the aforementioned advantages of re-MILP are gained by sacrificing performance optimality for continental-scale networks.

Y. Xu (yx4vf@virginia.edu) and M. Brandt-Pearce (mb-p@virgnia.edu) are with the Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, Virginia 22904, USA

L. Yan (lyaa@chalmers.se) and E. Agrell (agrell@chalmers.se), are with Chalmers University of Technology, SE-41296 Gothenburg, Sweden

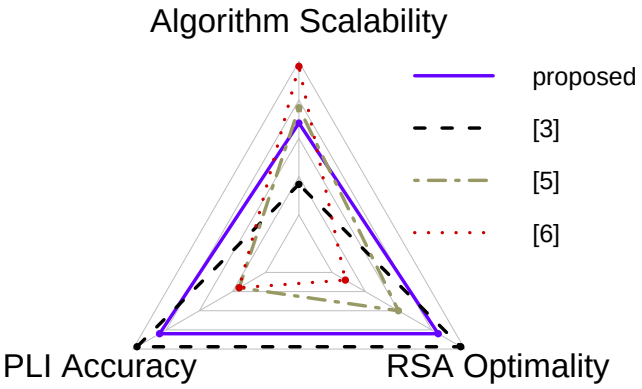

Fig. 1: State of the art of RSA showing the trade-off between competing objectives. Proposed: linearized GN model and link based SIO algorithm; [3]: finely linearized GN model and link based MILP algorithm; [5]: TR model and link based Re-MILP algorithm; [6]: TR model and path based algorithm.

In this paper we propose an improvement to the re-MILP and a more accurate PLI model that together provide a more comprehensive solution to the aforementioned problems (a) approximating PLIs accurately and (b) finding a scalable near-optimal solution [5]. The proposed RSRA algorithm is the sequential iterative optimization (SIO) algorithm, a generalization of the re-MILP that iteratively searches for a better solution, incorporating some randomness in the search process. The proposed PLI model is a linearized GN model we call, the conservative linearized Gaussian noise (CLGN) model. The SIO algorithm with the CLGN model provides a balance between complexity and estimation accuracy. Figure 1 illustrates schematically the trade-off between the three objectives (scalability, optimality, and accuracy) of our proposed algorithm and other approaches in the recent literature. Unlike path-based algorithms [6], which cannot integrate traffic-dependent PLI models, the proposed algorithm is a link-based MILP algorithm that sacrifices scalability to accommodate any PLI model that is linear in the state variables. The SIO algorithm has similar complexity and scalability as the re-MILP algorithm in [5] but results in much better performance by employing randomness to aid in convergence.

This paper is organized as follows. In Section II, we introduce the physical layer impairment models. Section III then describes the proposed models in the MILP formulation, while Section IV elaborates on the SIO algorithm. Section V provides numerical results and comparisons to [3] and [5], based on simulation. Finally, we draw conclusions in Section VI.

II. Estimation of Physical Layer Impairments

In this paper, we consider dual-polarization transmission with the same modulation format and the same PSD in both polarizations. We adopt similar assumptions as in [16]. The types of PLIs considered are nonlinear distortion, chromatic dispersion, and amplified spontaneous emission (ASE) noise [7]. Since the chromatic dispersion can be compensated by digital signal processing, we only need to consider the impairments caused by the nonlinear interference (NLI) (caused by the interaction of nonlinearity and dispersion in the fiber) and the ASE noise (caused by the erbium-doped fiber amplifiers (EDFAs)) [16], [17]. In this section, we present several methods one can use to model PLIs in RSA algorithms, including our novel proposed technique, the CLGN model.

A. Gaussian Noise Model

The ASE noise is modeled as additive Gaussian noise with power spectral density (PSD) given as

$$G_{ASE}^{span} = (e^{\alpha L} - 1) h \nu n_{sp}, \tag{1}$$

where n_{sp} represents the spontaneous emission factor, ν represents the light frequency, h represents Planck's constant, α represents the fiber power attenuation, and L represents the fiber length per span.

The GN model is applied to estimate the signal NLI [18]. The NLI effects can be divided into self-channel interference (SCI) and cross-channel interference (XCI):

$$G_{NLI,p}^{span} = G_{SCI,p}^{span} + G_{XCI,p}^{span}, \tag{2}$$

where $G_{NLI,p}^{span}$, $G_{SCI,p}^{span}$, and $G_{XCI,p}^{span}$ represent the pth channel's NLI PSD, SCI PSD, and XCI PSD per span per polarization, respectively. SCI is caused by the channel itself, only varying with the bandwidth of that channel [19]:

$$G_{SCI,p}^{span} = \mu G_p^3 \ln(\rho \Delta_p^2), \tag{3}$$

where $\rho = (\pi^2 |\beta_2|)/\alpha$, $\mu = (3\gamma^2)/(2\pi\alpha|\beta_2|)$, γ represents the fiber nonlinearity parameter, and β_2 represents the group velocity dispersion parameter. Δ_p and G_p represent the pth channel's bandwidth (being used in both polarizations) and signal PSD per polarization, respectively. We assume that the bandwidth equals the symbol rate, introducing frequency guard bands between transmissions to protect against inter-channel interference.

The XCI is caused by the interaction between channels and depends on the difference in center frequencies and bandwidths of the affecting channels:

$$G_{XCI,p}^{span} = \mu G_p \sum_{q=1;q\neq p}^{M_c} G_q^2 \ln\left(\frac{|f_p - f_q| + \Delta_q/2}{|f_p - f_q| - \Delta_q/2}\right), \tag{4}$$

where M_c represents the total number of channels on the same fiber link as the channel of interest, channel p, and f_q represents the qth channel's center frequency.

In order to guarantee the desired QoT, the signal-to-interference-plus-noise ratio (SINR) for channel p (SINR_p), over each transparent segment (light-path without regeneration) must satisfy the threshold SINR (SINR_p^{th}) [19]:

$$\text{SINR}_p = \frac{G_p}{N_p^{GN}} \geq \text{SINR}_p^{th}, \tag{5}$$

where $N_p^{GN} = \sum_{span}(G_{NLI,p}^{span} + G_{ASE,p}^{span})$ accumulates the independent noise and the interference (also assumed independent) incoherently over all spans on the transparent segment.

B. Conservative Linearized Gaussian Noise (CLGN) Model

In order to effectively use an MILP engine, we seek a linearized version of the standard GN model that follows some principles in order to preserve the properties of the GN model appropriately. First, the linearized GN model cannot underestimate the noise estimate of the standard GN model. Second, the linearized GN model should have similar QoT estimation as the GN model for the most realistic cases.

For the RSA problem, assuming fixed modulation, the bandwidth of demands Δ_q, $q = 1, ..., M_c$ are given as optimization inputs. The total number of demands on the fiber link (M_c) and the center frequencies of these demands (f_q, $q = 1, ..., M_c$) are decision variables in the optimization problem. Therefore, the SCI term in the standard GN model (3) is independent of the RSA variables M_c and f_q, $q = 1, \ldots, M_c$. The term that needs to be linearized is the XCI term. Since the expression $|f_p - f_q|$ is inside a logarithm function, we consider an upper bound on the XCI term as

$$G_{XCI,p}^{span} \leq \mu G_p \sum_{q=1;q\neq p}^{M_c} G_q^2 \ln\left(\frac{\Delta_q}{\Delta_{gb} + \Delta_p/2} + 1\right), \quad (6)$$

where Δ_{gb} is the guard band. We refer to this linearized version of the standard GN model as the conservative linearized Gaussian noise (CLGN) model since the function is linear with respect to RSA variables. In the SINR expression (5), the CLGN model uses this upper bound to the XCI term to define the noise for channel p, referred to as N_p^{CLGN}.

In the CLGN model, we consider that all channels q that contribute to the XCI for channel p are located as close as possible to the demand (without considering the actual center frequency difference $|f_p - f_q|$), yielding a conservative XCI estimate. When there are a large number of demands deployed on the same fiber link, the CLGN model provides an overestimate of XCI compared with the standard GN model. On the other hand, when there are few demands on the fiber link, the CLGN model is able to provide a similar XCI estimate compared with the standard GN model.

C. Gaussian Noise Based Transmission Reach (GNTR) Model

The TR model is applied, instead of the GN model, in most research addressing the RSA problem because of its simplicity and linearity. The TR model estimates the maximum length that a transparent segment can have and still satisfy a conservative estimate of the SINR. The disadvantage of the TR model is that it does not take the instantaneous channel state into account. Moreover, the parameters of the TR model are often obtained from experimental results [4], based on discrete experimental setups, thus leading to concerns over their universality [5].

Instead of using experimental data to derive the TR, we use a GN-model-based analytic TR algorithm, named the GN-based transmission reach (GNTR), to generate the TR in order to make a fair comparison with the CLGN model. A similar idea was proposed and validated in [3]. The GNTR is the shortest transmission reach that satisfies the worst-case QoT requirements based on the standard GN model given the bandwidth of a demand, the threshold SINR, and its input PSD [20]. In order to obtain this shortest distance, we first consider the worst-case noise level [21]

$$N_p^{GNTR} = \max_{M_c, \Delta_q, f_q} (G_{NLI,p}^{span} + G_{ASE,p}^{span}). \quad (7)$$

Given the modulation format, assuming the channel of interest p is centered in the middle of the available spectrum, and $G_p = G_q$ for all p and q, the worst noise level calculated based on the standard GN model occurs when channel p is sandwiched between two large bandwidth demands ($M_c = 3$ and maximum Δ_q), as shown in [17]. Assuming equal length spans, the GNTR of the pth channel, T_p^{GNTR}, can be calculated as:

$$T_p^{GNTR} = \frac{G_p}{\text{SINR}_p^{th} N_p^{GNTR}} \times L, \quad (8)$$

where L is the fiber length per span [17].

TABLE I: Simulation Parameters [3], [7], [17]

Δ_{gb}	12.5 GHz
G_p	0.015 W/THz
G_q	0.015 W/THz
α	0.22 dB/km
β_2	$-21.7\,\text{ps}^2/\text{km}$
γ	$1.32 \times 10^{-3}\,(\text{W} \cdot \text{m})^{-1}$
n_{sp}	1.58
L	100 km
ν	193.55 THz

D. PLI Model Validation

We simulate the standard GN, CLGN, the GNTR models on a single isolated link (not part of a network solution) in order to analyze the link-level performance of each PLI estimation model. We use the GN model as a benchmark assuming that it yields an accurate approximation to the PLIs. Fiber parameters are listed in Table I. Further simulation results can be found in [17].

In the first scenario, equal-bandwidth demands are deployed on a fiber link. The channel of interest p is centered in the middle of the spectrum, and $M_c = 3, 5, 7$, or 9. Channels are placed adjacently separated by Δ_{gb}. We define the normalized link noise estimation error for each M_c, using the standard GN model as a reference, as

$$Err^* = \frac{|N_p^* - N_p^{GN}|}{N_p^{GN}}, \quad * = CLGN, GNTR. \quad (9)$$

Figure 2 shows a comparison of the link-level noise estimation error between the GNTR and CLGN models. Err^{CLGN} is always smaller than Err^{GNTR} when $M_c \leq 7$. When $M_c = 9$ and the bandwidth of each demand exceeds 78 GHz, Err^{CLGN} is larger than Err^{GNTR}. The estimation accuracy of the CLGN model decreases as M_c increases, as expected.

In the second scenario, a probabilistic analysis is implemented by simulating random bandwidth demands deployed on the same fiber link. Each demand has a random bandwidth uniformly distributed from 30 to 100 GHz. The channel of interest is again in the middle of the spectrum and channels are

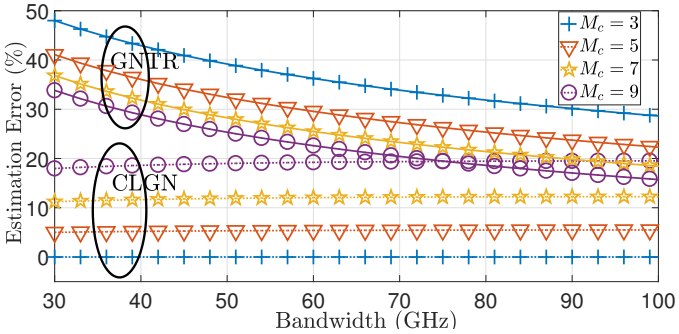

Fig. 2: Comparison of the estimation error (Err^*) generated by the GNTR, CLGN, and GN models, $\Delta_q = \Delta_p$ for $q = 1, \ldots, M_c$, for various M_c.

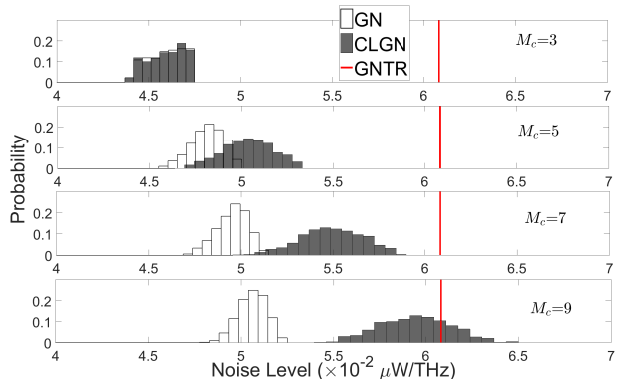

Fig. 3: Histogram of noise level for random demand bandwidths, for various M_c.

placed adjacently, separated by Δ_{gb}. The natural fragmentation that results from RSA in a network is not modeled here. After completing 10000 simulation trials for different values of M_c ($M_c = 3, 5, 7$, and 9), using similar simulation settings as [3], [5], and [13], we estimate the probability distribution of the noise level (N_p^{GN}, N_p^{GNTR}, and N_p^{CLGN}), shown in Figure 3. Again, the CLGN model is better at estimating the performance of the PLIs than the GNTR model for $M_c \leq 7$, assuming the GN model yields an accurate approximation to the PLIs. When $M_c = 9$, the CLGN model has a greater than 60% probability of being more accurate than the GNTR model.

When the number of channels on a link is large, the GNTR becomes more accurate than the CLGN. Since both of these models upper-bound the actual PLIs, the minimum of the two noise estimates can be used instead of either model.

III. Link-Based MILP For EON Networks

We first introduce the basic MILP for the RSRA problems and then apply physical layer constraints (both the CLGN and GNTR models) to it. We adopt notations and formulations from [5], [11] and more detailed information can be found in [17]. In order to optimize resources used by EONs, which in this work consist of routes, spectrum, and regeneration nodes, the optimization objective of the MILP is a linear combination of the total spectrum usage, C, and the number of regeneration nodes used, T.

A. Basic MILP without Signal Regeneration

The objective function in this scenario is the total spectrum usage C. The EON is formulated as a connected graph $(\mathbb{N}, \mathbb{L})$ with nodes denoted by $\mathbb{N}$ and unidirectional links denoted by $\mathbb{L}$. $\mathbb{D}$ is the set of demands. $\Delta_{s,d}$ is the bandwidth of demand $D_{s,d}$, corresponding to its modulation format.[1] The node parameter $S_{n;s,d} = 1$ if $n = s$; $S_{n;s,d} = -1$ if $n = d$; otherwise, $S_{n;s,d} = 0$. $F_{s,d} \in \mathbb{R}$ represents the lowest frequency allocated to demand $D_{s,d}$. $\mathrm{U}_{i,j;s,d} \in \{0, 1\}$ represents the link usage corresponding to demand $D_{s,d}$ on link $L_{i,j} \in \mathbb{L}$. $\delta_{s,d,\hat{s},\hat{d}} \in \{0, 1\}$ represents the order of the frequency index between the spectrum allocated to demands $D_{s,d}$ and $D_{\hat{s},\hat{d}}$. If $\delta_{s,d,\hat{s},\hat{d}} = 1$, then $F_{s,d} < F_{\hat{s},\hat{d}}$; if $\delta_{s,d,\hat{s},\hat{d}} = 0$, then $F_{s,d} > F_{\hat{s},\hat{d}}$.

The basic MILP constraints include the total spectrum usage constraint,

$$C \geq F_{s,d} + \Delta_{s,d}, \quad \forall D_{s,d} \in \mathbb{D}, \tag{10}$$

used to enforce the relationship between the spectrum usage and the highest frequency used in the EON. The flow conservation constraint,

$$S_{n;s,d} = \sum_{i=n;L_{i,j}\in\mathbb{L}} \mathrm{U}_{i,j;s,d} - \sum_{j=n;L_{i,j}\in\mathbb{L}} \mathrm{U}_{i,j;s,d}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, \tag{11}$$

ensures that each demand has only one path from source to destination without bifurcations, loops, or dead-ends during the transmission through intermediate nodes. The non-overlapping spectrum constraints,

$$\delta_{s,d;\hat{s},\hat{d}} + \delta_{\hat{s},\hat{d};s,d,} = 1, \quad \forall D_{s,d}, D_{\hat{s},\hat{d}} \in \mathbb{D}, \tag{12}$$

$$(F_{s,d} - F_{\hat{s},\hat{d}} + \Delta_{s,d} + \Delta_{gb}) \leq \overline{L} \times (3 - \delta_{s,d,\hat{s},\hat{d}} - \mathrm{U}_{i,j;s,d} - \mathrm{U}_{i,j;\hat{s},\hat{d}}), \quad \forall D_{s,d}, D_{\hat{s},\hat{d}} \in \mathbb{D}, L_{i,j} \in \mathbb{L}, \tag{13}$$

ensure the lowest frequencies of each demand are far enough to prevent overlapping. $\overline{L}$ is a fixed large number.

B. MILP with Signal Regeneration

Additional constraints at the signal regeneration nodes allow the GNTR-model-based MILP [constraints (14)-(18)] and the CLGN-model-based MILP [constraints (18)-(22)] to function for practical continental-scale networks. With regeneration nodes, the optimization objective becomes a multi-objective function with a weighting factor, either $C + \varepsilon T$ or $T + \varepsilon C$, where ε is a small number compared with C/T and T/C. C is the spectrum used, and T is the number of regeneration nodes. Using $C + \varepsilon T$, the prime objective is to optimize the total spectrum usage while trying to reduce the number of regeneration nodes. Using $T + \varepsilon C$, the prime objective is to minimize the number of regeneration nodes while controlling the total spectrum usage with a lower priority.

[1] Extending from the link level to the network level, this section uses source and destination nodes (s, d) to refer to a demand instead of the channel index p used in Section II.

1) GNTR Model: If demand $D_{s,d}$ routes through node n, $\mathrm{Y}^{\mathrm{TR}}_{n;s,d} \in \mathbb{R}$ is its physical propagation distance from the beginning node of the transparent segment to node $n \in \mathbb{N}$; otherwise $\mathrm{Y}^{\mathrm{TR}}_{n;s,d} = 0$. If demand $D_{s,d}$ is assigned on link $L_{i,j}$, $\mathrm{Z}^{\mathrm{TR}}_{i,j;s,d} \in \mathbb{R}$ is its accumulated propagation distance from the beginning node of the transparent segment on the light-path to node i, i.e., $\mathrm{Z}^{\mathrm{TR}}_{i,j;s,d} = \mathrm{Y}^{\mathrm{TR}}_{i;s,d}$; otherwise $\mathrm{Z}^{\mathrm{TR}}_{i,j;s,d} = 0$. $\ell_{i,j}$ is the physical length of link $L_{i,j}$. If a regeneration circuit for demands $D_{s,d}$ is allocated on node n, the binary variable $I_{n;s,d} = 1$; otherwise $I_{n;s,d} = 0$. If no regeneration circuit is allocated on node n, $I_n = 0$; otherwise $I_n = 1$, which means node n is a regeneration node. The regeneration nodes contain a limited number of regeneration circuits, denoted as $I^{\max}$. If $I_{n;s,d} = 1$ and $i = n$, then $\mathrm{K}^{\mathrm{TR}}_{i,j;s,d} = 0$, representing that the regeneration circuit has regenerated demand $D_{s,d}$ at node n; otherwise $\mathrm{K}^{\mathrm{TR}}_{i,j;s,d} = \mathrm{Z}^{\mathrm{TR}}_{i,j;s,d}$, representing that demand $D_{s,d}$ is not regenerated at node n and the accumulated length is not affected.

To accommodate the GNTR model, the following constraints must be added to the basic MILP [5]. The link usage and accumulated traveling distance constraint,[2]

$$\mathrm{Z}^{\mathrm{TR}}_{i,j;s,d} = \mathrm{Y}^{\mathrm{TR}}_{n=i;s,d} \times \mathrm{U}_{i,j;s,d}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, L_{i,j} \in \mathbb{L}, \tag{14}$$

builds the relationship between $\mathrm{Y}^{\mathrm{TR}}_{n;s,d}$ and $\mathrm{Z}^{\mathrm{TR}}_{i,j;s,d}$. The transmission reach constraint (QoT constraint of the GNTR model),

$$\mathrm{Y}^{\mathrm{TR}}_{n;s,d} \leq \mathrm{T}^{GNTR}_{s,d}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, \tag{15}$$

ensures that a demand is transmitted with the desired QoT. $\mathrm{T}^{GNTR}_{s,d}$ represents the TR for demand $D_{s,d}$, defined in (8). The accumulated traveling distance constraint,

$$\mathrm{Y}^{\mathrm{TR}}_{n;s,d} = \sum_{L_{i,j} \in \mathbb{L}; j=n} \mathrm{K}^{\mathrm{TR}}_{i,j;s,d} + \mathrm{U}_{i,j;s,d} \times \ell_{i,j}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, \tag{16}$$

is used to obtain $\mathrm{Y}^{\mathrm{TR}}_{n;s,d}$ by a recursive accumulation of the propagation distance along the route of the demand. The constraint

$$\mathrm{K}^{\mathrm{TR}}_{i,j;s,d} = (1 - I_{n=i;s,d}) \times \mathrm{Z}^{\mathrm{TR}}_{i,j;s,d}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, L_{i,j} \in \mathbb{L}, \tag{17}$$

is used to ensure the relationship between regeneration circuits and the accumulated length. Lastly, the constraint

$$I_n \times I^{\max} \geq \sum_{\forall D_{s,d} \in \mathbb{D}} I_{n;s,d}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, \tag{18}$$

ensures that the number of regeneration circuits on one regeneration node is bounded by the maximum allowed, $I^{\max}$. We define the total number of regeneration circuits as $I_t = \sum_{n \in \mathbb{N}} \sum_{D_{s,d} \in \mathbb{D}} I_{n;s,d}$ and the number of regeneration nodes as $T = \sum_{n \in \mathbb{N}} I_n$.

2) CLGN Model: If demand $D_{s,d}$ routes through node n, $\mathrm{Y}^{CL}_{n;s,d} \in \mathbb{R}$ represents the accumulated noise, both ASE and NLI, along the route from the beginning node of the transparent segment; otherwise $\mathrm{Y}^{CL}_{n;s,d} = 0$. If demand $D_{s,d}$ is assigned to link $L_{i,j}$, $\mathrm{Z}^{CL}_{i,j;s,d}$ is the accumulated noise from the beginning node of the transparent segment on the light-path to node i; otherwise, $\mathrm{Z}^{CL}_{i,j;s,d} = 0$. If $I_{n;s,d} = 1$ and $i = n$, then $\mathrm{K}^{\mathrm{CL}}_{i,j;s,d} = 0$, representing that the regeneration circuit has canceled the noise of demand $D_{s,d}$ at node n; otherwise $\mathrm{K}^{\mathrm{TR}}_{i,j;s,d} = \mathrm{Z}^{CL}_{i,j;s,d}$, representing that the accumulated noise is not affected.

The following constraints must be added to the basic MILP, analogously to the GNTR model. The link usage and accumulated noise constraint,

$$\mathrm{Z}^{CL}_{i,j;s,d} = \mathrm{Y}^{CL}_{n;s,d} \times \mathrm{U}_{i,j;s,d}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, L_{i,j} \in \mathbb{L}, \tag{19}$$

is similar to the constraint in (14). The QoT constraint,

$$\mathrm{Y}^{CL}_{n;s,d} \leq \frac{G_{s,d}}{\mathrm{SINR}^{th}_{s,d}}, \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, \tag{20}$$

ensures that each demand transmitted satisfies the desired QoT. $G_{s,d}$ is the signal PSD of demand $D_{s,d}$. $\mathrm{SINR}^{th}_{s,d}$ denotes the required SINR for a given demand $D_{s,d}$ and a specified QoT, corresponding to the modulation format. The accumulated noise constraint,

$$\mathrm{Y}^{CL}_{n;s,d} = \sum_{L_{i,j} \in \mathbb{L}; j=n} \mathrm{K}^{CL}_{i,j;s,d} + \mathrm{U}_{i,j;s,d} N^{CLGN}_{i,j,s,d} \quad \forall n \in \mathbb{N}, D_{s,d} \in \mathbb{D}, \tag{21}$$

recursively accumulates the total noise along the route. $N^{CLGN}_{i,j,s,d}$ is the PLIs estimated by the CLGN model using (5) and (6), as discussed in Section II-B. In our link-based MILP algorithm, $G_{XCI,i,j;s,d}$ defined in (6) sums all the XCI noise contributed by all the demands shared with $D_{s,d}$ on link $L_{i,j}$,

$$N^{CLGN}_{i,j,s,d} = G_{ASE,i,j;s,d} + G_{SCI,i,j;s,d} + \underbrace{\sum_{L_{i,j} \in \mathbb{L}} G^2_{s,d} \ln\left(\frac{\Delta_{\hat{s},\hat{d}}}{\Delta_{gb} + \Delta_{s,d}/2} + 1\right) \mathrm{U}_{i,j;\hat{s},\hat{d}}}_{G_{XCI,i,j;s,d}}. \quad \forall D_{s,d} \in \mathbb{D}, L_{i,j} \in \mathbb{L} \tag{22}$$

The regeneration circuits and accumulated noise constraint,

$$\mathrm{K}^{CL}_{n;s,d} = (1 - I_{n;s,d}) \times \mathrm{Z}^{CL}_{i,j;s,d}, \quad \forall n \in \mathbb{N}, \quad D_{s,d} \in \mathbb{D}, L_{i,j} \in \mathbb{L}, \tag{23}$$

and the constraint that limits the number of circuits (18) are also needed to ensure the regeneration placement.

As networks become more congested, the number of simultaneous demands on one link may increase, making the GNTR constraint more accurate on some links, and the CLGN more accurate on others. We introduce a new upper bound, $N^{UB}_{i,j,s,d}$, that gives the tighter of the two constraints, within the CLGN formulation. An extra constraint is needed in the MILP,

$$N^{UB}_{i,j,s,d} = \min(N^{CLGN}_{i,j,s,d}, N^{GNTR}_{s,d}), \quad \forall D_{s,d} \in \mathbb{D}, L_{i,j} \in \mathbb{L}, \tag{24}$$

which can easily be linearized [22]. $N^{CLGN}_{i,j,s,d}$ in constraint (21) can thus be replaced by $N^{UB}_{i,j,s,d}$. $N^{GNTR}_{i,j,s,d}$ can be precalculated by (7), because this worst case noise is network

[2] A linearization of (14) can be found in [22]

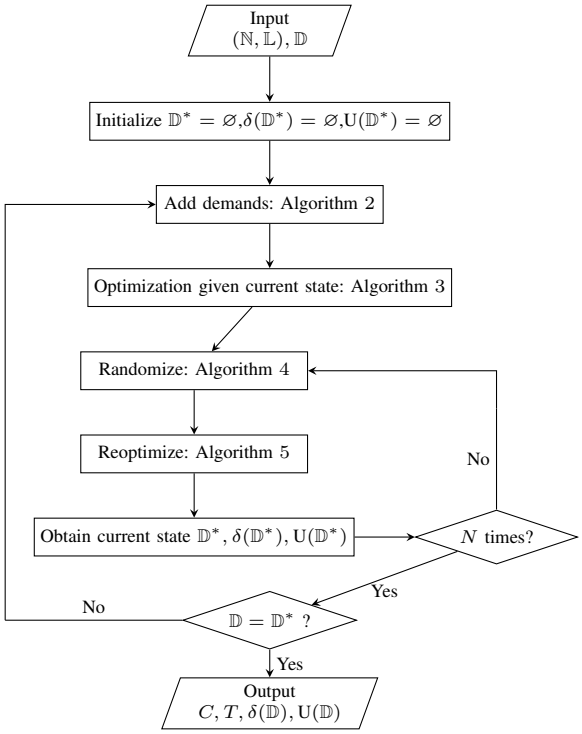

Fig. 4: Flowchart of SIO algorithm

state-independent.

IV. Heuristic Algorithm: Sequential Iterative Optimization

The complexity of the RSRA problem goes beyond the capacity of current MILP solvers. Even though the CLGN model provides less complexity than the standard GN model, the RSRA problem cannot be properly solved within a reasonable time using the standard MILP algorithm. Therefore, we propose a heuristic approach that we call the sequential iterative optimization (SIO). The SIO algorithm optimizes the routing of the demands and the assignment of spectral and regeneration resources simultaneously.

For optimal MILP methods, during the solving process, an unacceptably long time is spent solving for the integer values in dead-end intermediate results. The SIO algorithm can mostly avoid these stubborn local minima. The SIO algorithm reduces the number of integer variables by using iterations to converge to a near-optimal solution. In addition, the SIO algorithm applies randomness in order to exit stubborn local minima and enhance the performance of each iteration. The randomness allows the algorithm to explore many non-contiguous areas of the variable space, resulting in a closer approach to the optimal solution, and providing a near-optimal starting point for subsequent iterations.

The process of the SIO algorithm is shown as a flowchart in Figure 4 and described in detail in Algorithms 1-5. Demands are introduced a few at a time, recursively added to the existing state that has been partially shuffled and solved iteratively.

In Algorithm 1, when the number of allocated demands $\mathbb{D}^*$ is less than the number of demands needing to be optimized (i.e., $\mathbb{D}^* \neq \mathbb{D}$), we recursively expand the size of the optimization problem (i.e., $\mathbb{D}^* \cup \mathbb{D}_{\text{new}} = \texttt{AddNewDemands}(\mathbb{D}^*, \mathbb{D}, m)$) and optimize it, until all demands have been allocated. In addition, at each stage (i.e., while the number of demands $|\mathbb{D}^*|$ remains unchanged), we iteratively reoptimize the problem N times while applying randomness (Algorithm 5) each time.

In Algorithm 2, when $\mathbb{D}^* \neq \mathbb{D}$, we add m more demands into the optimization problem. The demands to be used in the optimization process (i.e., Algorithm 3) become $\mathbb{D}^* \cup \mathbb{D}_{\text{new}}$, where $|\mathbb{D}_{\text{new}}| = m$.

In Algorithm 3, $\mathbb{D}^* \cup \mathbb{D}_{\text{new}}$ represents the set of demands considered by the MILP algorithm. $\delta(\mathbb{D}^*)$ and $\mathrm{U}(\mathbb{D}^*)$ represent the optimized spectrum and routing information, respectively, obtained from the last iteration when the set of demands was $\mathbb{D}^*$. These resources are considered unavailable. This MILP process optimizes resources assigned to $\mathbb{D}_{\text{new}}$ given that those assigned to $\mathbb{D}^*$ are fixed.

In Algorithm 4, we randomly select a set of demands, $\mathbb{D}_{\text{out}}$ from the set of allocated demands, $\mathbb{D}^*$. The size of the selected demand set is $1/\eta$ of the allocated demands (i.e., $|\mathbb{D}_{\text{out}}| = \lfloor |\mathbb{D}^*|/\eta \rfloor$). η can be adjusted to ensure the scalability of the algorithm. In Algorithm 5, we delete the routing $\mathrm{U}(\mathbb{D}_{\text{out}})$ and spectrum assignment information $\delta(\mathbb{D}_{\text{out}})$ of this set from the last iteration. Based on the remaining spectrum and routing information, i.e., $\delta(\mathbb{D}^* \setminus \mathbb{D}_{\text{out}})$ and $\mathrm{U}(\mathbb{D}^* \setminus \mathbb{D}_{\text{out}})$, the resources for demands $\mathbb{D}^*$ are reoptimized. In this reoptimization process, which is performed N times, demands are allocated based on partially optimized information from the previous iteration. There is a trade-off between scalability and optimality of the SIO algorithm. When either η increases or m decreases, the scalability of the SIO system increases and the optimality decreases.

V. Numerical Results

Before simulating in continental-scale networks, we validate the performance of the proposed SIO algorithm on small networks by taking the optimal MILP as a benchmark for comparison. We apply the GNTR model and the CLGN model separately for both the optimal MILP and the SIO algorithm. As detailed in [17, Section 4.6], we first tested the SIO algorithm on a variety of small networks without regeneration nodes ($T = 0$), and the objective C was always identical to the results of the optimal MILP for simulation settings similar to common assumptions used in the literature [3], [5], [13]. For larger networks, we expect the MILP to perform better than the SIO algorithm. Therefore, we simulate our proposed algorithm in realistic benchmark networks. We validate the advantages brought by the CLGN model compared with typically-used TR-based model. We also show the advantages of the proposed SIO algorithm compared with published algorithms [3], [5].

A. Simulation Settings

The simulation parameters are listed in Table I. The network topologies we tested are the NSF-24 network with 24 nodes and 86 unidirectional links [5] and the DT-14 network with 14 nodes and 46 unidirectional links [3]. These two network topologies are used for testing algorithms of different scales.

Algorithm 1 Sequential Iterative Optimization

Input:

- Network topology $(\mathbb{N}, \mathbb{L})$
- Set of demands $\mathbb{D}$
- The number of iterations per stage N
- The increment (granularity) of the number of demands m at the beginning of each stage

Set definitions:

- $\mathbb{D}^*$ is the set of demands processed in the current iteration, with initial value $\mathbb{D}^* = \varnothing$
- $\delta(\mathbb{D}^*) = \{\delta_{s,d;\hat{s},\hat{d}} | D_{s,d} \in \mathbb{D}^*\}$ is the pair-wise spectral ordering of demands in $\mathbb{D}^*$, with an initial value $\delta(\mathbb{D}^*) = \varnothing$
- $\mathrm{U}(\mathbb{D}^*) = \{\mathrm{U}_{i,j;s,d} | D_{s,d}, D_{\hat{s},\hat{d}} \in \mathbb{D}^*, L_{i,j} \in \mathbb{L}\}$ is the link usage of demands in $\mathbb{D}^*$, with an initial value $\mathrm{U}(\mathbb{D}^*) = \varnothing$
- k is the index of the current iteration, with an initial value $k = 1$
- $\mathrm{Obj}(\mathbb{D}^*)$ is the optimization objective (C, T) for demands $D_{s,d} \in \mathbb{D}^*$

while $\mathbb{D}^* \neq \mathbb{D}$ **do**
  $\mathbb{D}^* \cup \mathbb{D}_{\text{new}} = \texttt{AddNewDemands}(\mathbb{D}^*, \mathbb{D}, m)$
  $[\delta(\mathbb{D}^* \cup \mathbb{D}_{\text{new}}), \mathrm{U}(\mathbb{D}^* \cup \mathbb{D}_{\text{new}}), \mathrm{Obj}(\mathbb{D}^* \cup \mathbb{D}_{\text{new}})] =$ $\texttt{MILP}(\mathbb{D}^* \cup \mathbb{D}_{\text{new}}, \delta(\mathbb{D}^*), \mathrm{U}(\mathbb{D}^*))$
  $\mathbb{D}^* \leftarrow \mathbb{D}^* \cup \mathbb{D}_{\text{new}}$
  $\delta(\mathbb{D}^*) \leftarrow \delta(\mathbb{D}^* \cup \mathbb{D}_{\text{new}})$
  $\mathrm{U}(\mathbb{D}^*) \leftarrow \mathrm{U}(\mathbb{D}^* \cup \mathbb{D}_{\text{new}})$
  $k = 1$
  while $k \leq N$ **do**
    $\mathbb{D}_{\text{out}} = \texttt{Randomize}(\mathbb{D}^*)$
    $\delta(\mathbb{D}^*), \mathrm{U}(\mathbb{D}^*), \mathrm{Obj}(\mathbb{D}^*) =$ $\texttt{Reoptimize}(\mathbb{D}_{\text{out}}, \mathbb{D}^*, \delta(\mathbb{D}^*), \mathrm{U}(\mathbb{D}^*))$
    $k \leftarrow k + 1$
  end while
end while
Output:

- The spectral ordering of all the demands $\delta(\mathbb{D})$
- The link usage of all the demands $\mathrm{U}(\mathbb{D})$
- The allocation of regeneration nodes
- The optimization objective $\mathrm{Obj}(\mathbb{D})$

Algorithm 2 `AddNewDemands`

Input: $\mathbb{D}^*$, $\mathbb{D}$, m
if $\mathbb{D}^* \neq \mathbb{D}$ **then**
  $M = \max\{m, |\mathbb{D} \setminus \mathbb{D}^*|\}$
  Randomly choose a subset $\mathbb{D}_{\text{new}} \subseteq \mathbb{D} \setminus \mathbb{D}^*$ such that $|\mathbb{D}_{\text{new}}| = M$
else
  $\mathbb{D}_{\text{new}} = \varnothing$
end if
Output: $\mathbb{D}^* \cup \mathbb{D}_{\text{new}}$

Algorithm 3 `MILP`

Input: $\mathbb{D}^* \cup \mathbb{D}_{\text{new}}, \delta(\mathbb{D}^*), \mathrm{U}(\mathbb{D}^*)$
Allocate resources to each demand $D_{s,d} \in \mathbb{D}^* \cup \mathbb{D}_{\text{new}}$ by MILP subject to the following constraints

- Flow conservation
- Non-overlapping spectrum
- QoT requirements and allocation of regeneration nodes
- Update $\delta(\mathbb{D}^* \cup \mathbb{D}_{\text{new}})$ based on $\delta(\mathbb{D}^*)$
- Update $\mathrm{U}(\mathbb{D}^* \cup \mathbb{D}_{\text{new}})$ based on $\mathrm{U}(\mathbb{D}^*)$

Output: $\delta(\mathbb{D}^* \cup \mathbb{D}_{\text{new}}), \mathrm{U}(\mathbb{D}^* \cup \mathbb{D}_{\text{new}}), \mathrm{Obj}(\mathbb{D}^* \cup \mathbb{D}_{\text{new}})$

Algorithm 4 `Randomize`

Input: $\mathbb{D}^*$
Randomly choose $\mathbb{D}_{\text{out}} \subset \mathbb{D}^*$ such that $|\mathbb{D}_{\text{out}}| = \lfloor |\mathbb{D}^*| / \eta \rfloor$
Output: $\mathbb{D}_{\text{out}}$

Algorithm 5 `Reoptimize`

Input: $\mathbb{D}_{\text{out}}, \mathbb{D}^*, \delta(\mathbb{D}^*), \mathrm{U}(\mathbb{D}^*)$
Allocate routes, spectrum, regeneration nodes, and regeneration circuits to each demand $D_{s,d} \in \mathbb{D}^*$ by MILP subject to the following constraints

- Flow conservation
- Non-overlapping spectrum
- QoT requirements and allocation of regeneration nodes
- Update $\delta(\mathbb{D}^*)$ based on $\delta(\mathbb{D}^* \setminus \mathbb{D}_{out})$
- Update $\mathrm{U}(\mathbb{D}^*)$ based on $\mathrm{U}(\mathbb{D}^* \setminus \mathbb{D}_{out})$

Output: $\delta(\mathbb{D}^*), \mathrm{U}(\mathbb{D}^*), \mathrm{Obj}(\mathbb{D}^*)$

All plots show the mean values of the quantity measured for 40 trials (spectrum usage, number of regeneration nodes and circuits), and 90% confidence intervals. All simulations use the following common assumptions: we generate static traffic demands between node pairs, wherein each pair consists of a randomly selected source and a destination node; the bandwidth of these demands is uniformly distributed from 30 to 100 GHz [5]; all demands use the same modulation format, either polarization-multiplexed binary phase-shift keying (PM-BPSK) or polarization-multiplexed quadrature phase-shift keying (PM-QPSK), and have equal input PSDs; no modulation conversion or wavelength conversion is considered [17]; each regeneration circuit serves one light-path; and an upper-bound on the number of circuits per regeneration node of $I^{\max} = 10$ is assumed.[3] The increment in the number of demands allocated is $m = 5$, and we set $\eta = 2$ to ensure scalability for the parameter values tested.

Note that for the network topologies and simulation parameters used, the PLI computed for each link using the CLGN model resulted in a lower noise estimate than the GNTR equivalent noise, and therefore constraint (24) was never invoked.

B. Performance of SIO Algorithm

To test our algorithms in a continental-sized network, we simulate the NSF-24 topology. This network has been

[3]We ran our simulations on the Rivanna research computing cluster provided by the University of Virginia. The MILP engine used for simulation is the Gurobi Optimization [23].

used extensively in literature to evaluate network planning algorithms [5], [24], [25]. We test our proposed algorithm with diverse optimization objectives. In addition, we simulate our proposed model and algorithm with a realistic resource shortage.

1) RSRA with Multi-Optimization Objectives: Spectrum and regeneration nodes are valuable resources when considering the capital expenditures of deploying long-haul optical networks. In different scenarios, there are different priorities for these resources. We simulate two representative scenarios to show the performance of our proposed CLGN model and the SIO algorithm: (i) $\min C + \varepsilon T$ and (ii) $\min T + \varepsilon C$ objective functions. In both cases, ε is chosen as a small number. In objective (i), the primary goal is optimizing the total spectrum usage. In objective (ii), the primary goal is optimizing the number of regeneration nodes. In both cases, we test both the GNTR and the CLGN models, as described in Section III, and solve the RSRA problem by the SIO algorithm in the NSF-24 network.

In Figure 5 (a), for the $C + \varepsilon T$ case and BPSK modulation, the total spectrum usage for the GNTR and CLGN models is similar because of the loose QoT requirements, i.e., the seldom need for regeneration (the SINR threshold for BPSK is much smaller than the actual SINR). In Figure 5 (b), for the $C + \varepsilon T$ case using QPSK, the total spectrum usage of the GNTR and CLGN models is also similar, but for a different reason than for the BPSK case: the similar spectrum usage for QPSK happens because the total spectrum used is the primary optimization objective. Both the GNTR and the CLGN models sacrifice other resources, such as the number of regeneration nodes, to ensure the optimality of the spectrum usage. In other words, in this scenario, the difference in the performance caused by using different estimates of the PLIs is compensated by the allocation of regeneration nodes.

In Figure 5 (a), for the $T + \varepsilon C$ scenario with BPSK, the loose QoT requirements again result in similar spectrum used by both the GNTR model and the CLGN model. Comparing the different optimization objectives, $C + \varepsilon T$ and $T + \varepsilon C$, using BPSK, the spectrum usage is also similar since there is seldom need for regeneration.

In Figure 5 (b), for the $T + \varepsilon C$ scenario with QPSK, the total spectrum usage of the GNTR model with less than 20 demands is higher than that of the CLGN model. The reason for this is that the main objective of this scenario is to minimize the number of regeneration nodes. Therefore, when there are fewer than 20 demands, the CLGN model has the potential to save more spectrum because it is state-dependent, thus has a more accurate approximation of the noise level. However, the GNTR model is a worst-case approximation, overestimating the PLIs when the situation is far from the worst case (the actual length a signal can propagate while satisfying the SINR threshold is much longer than the length obtained using the GNTR model). When the number of demands increases, the value of the total spectrum usage of both the GNTR model and the CLGN model closely align. When there are more than 20 demands, both the GNTR and the CLGN models result in the same routing solution in order to save on the number of regeneration nodes, resulting in the curve of the GNTR model closely following the curve of the CLGN model. Comparing the different optimization objectives, $C + \varepsilon T$ and $T + \varepsilon C$, using QPSK, the spectrum usage of $T + \varepsilon C$ is higher than that of $C + \varepsilon T$ because $T + \varepsilon C$ sacrifices routing flexibility, resulting in a higher spectrum utilization, in exchange for fewer regeneration nodes.

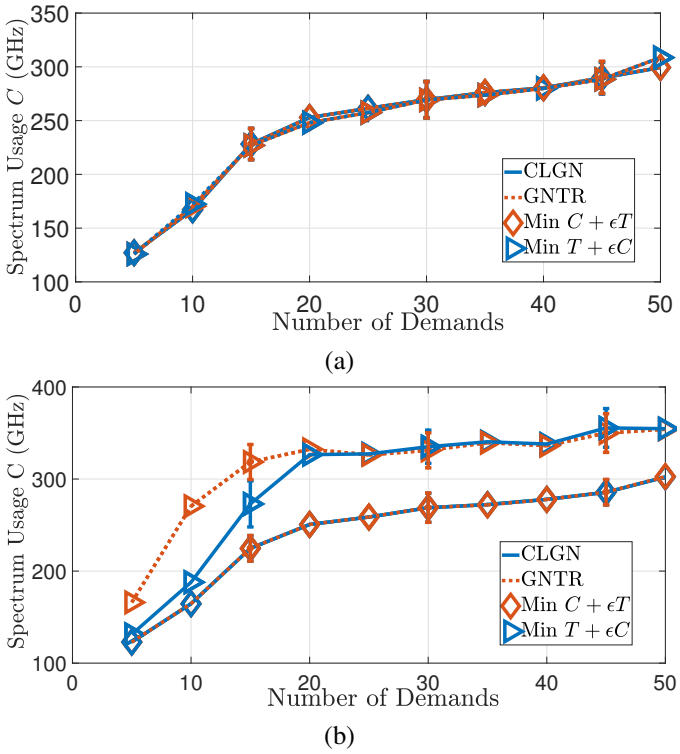

Fig. 5: Optimized spectrum usage with $C + \varepsilon T$ and $T + \varepsilon C$ as functions of $|\mathbb{D}|$ with (a) BPSK modulation, (b) QPSK modulation.

In Figure 6 (a), for the $C + \varepsilon T$ case, the number of regeneration nodes for the GNTR model is higher than that for the CLGN model for any number of demands. For the $T + \varepsilon C$ case, the primary optimization objective is to minimize the number of regeneration nodes T. Thus, the magnitude of separation is less compared with the $C + \varepsilon T$ case, although the curve for the GNTR model is still higher than for the CLGN model. Note that, for the $T + \varepsilon C$ case, when there are fewer than 25 demands, the magnitude of separation between the GNTR and CLGN model curves is greatest. This results from the fact that the GNTR model has a higher PLI approximation error when the actual noise level is far from the worst-case. Thus, we conclude that the CLGN model is able to reduce the number of regeneration nodes used compared with the GNTR model.

The number of regeneration circuits is not the optimization objective in either the $C + \varepsilon T$ case or the $T + \varepsilon C$ case. However, regeneration circuits are also an expensive and limited resource in EONs, in addition to regeneration nodes. As shown in Figure 6 (b), the GNTR model requires more regeneration circuits compared with the CLGN model for both objective functions. Additionally, the magnitude of separation between the CLGN and GNTR model curves in the $C + \varepsilon T$

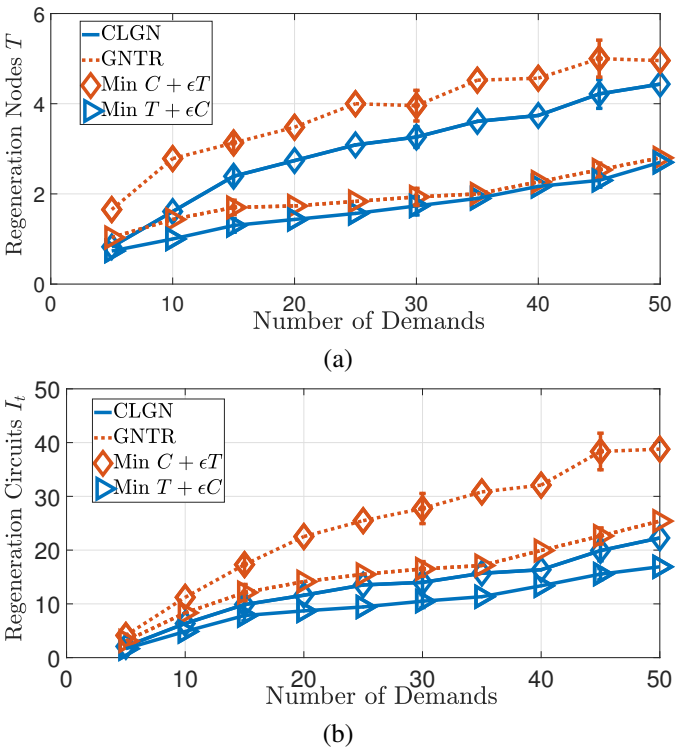

Fig. 6: (a) Optimized number of regeneration nodes as functions of $|\mathbb{D}|$ with $C+\varepsilon T$ and $T+\varepsilon C$, (b) number of regeneration circuits as functions of $|\mathbb{D}|$ with $C+\varepsilon T$ and $T+\varepsilon C$. QPSK modulation.

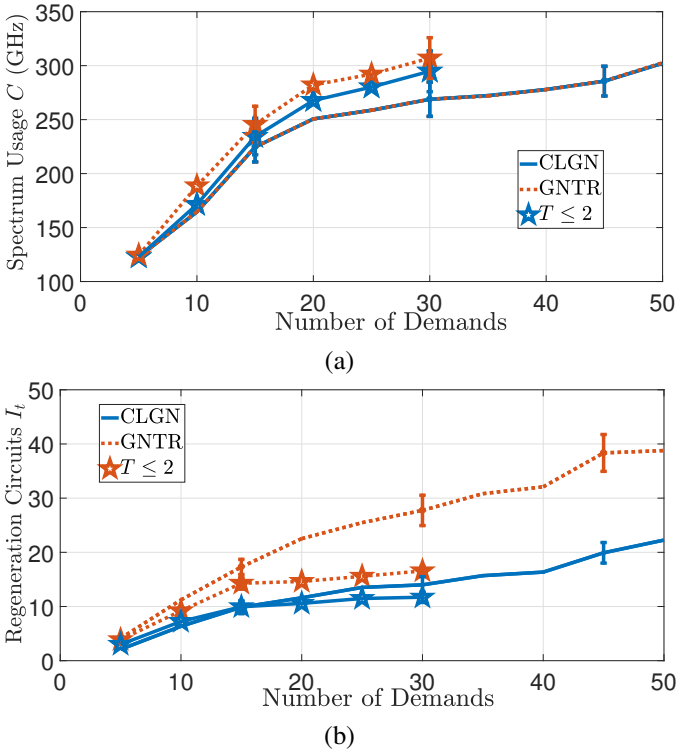

Fig. 7: (a) Optimized spectrum usage and (b) number of regeneration circuits with $C+\varepsilon T$ as functions of $|\mathbb{D}|$ with and without limited regeneration nodes. QPSK modulation.

case is larger than in the $T+\varepsilon C$ case. The reason is that in both the $C+\varepsilon T$ and $T+\varepsilon C$ cases, the GNTR model results in a higher PLI estimation error accumulated in the network when the number of demands increases to 50, leading to extra regeneration circuit expenditures. The benefit of the CLGN model in saving regeneration circuits, compared with the GNTR model, is substantial.

2) RSRA with Limited Regeneration Nodes: In this scenario, we simulate the RSRA problem with a limited number of nodes that can be assigned as regeneration nodes; this requires an extra constraint in the MILP described in Section III.B. As we discussed above, regeneration nodes are a limited resource because the allocation of regeneration nodes, as well as their maintenance, is expensive. Consequently, regeneration nodes should be carefully allocated in the RSRA problem.

We maintain the optimization objective $C+\varepsilon T$, for both the GNTR model and the CLGN model, and solve the RSRA problem using the SIO algorithm. We simulate this scenario in order to compare the performance of the total spectrum usage between the GNTR model and the CLGN model with limited regeneration nodes. In addition, we observe the effects of utilizing more or fewer regeneration nodes. After simulating different cases ($T\leq 2$, $T\leq 3$, and $T\leq 4$) the case of $T\leq 2$ is most representative for the NSF-24 network. In this scenario, because of the QoT requirements and congestion in the network, we stop the simulation at 30 demands.

In Figure 7 (a), for the $T\leq 2$ case the total spectrum usage required by the CLGN model is less than that of the GNTR model. In addition, the spectrum usage without limiting the number of regeneration nodes is lower than the spectrum usage with a limitation on regeneration nodes. However, this lower spectrum usage comes at the cost of more regeneration circuits, as seen in Figure 7 (b). The results also show that, for $T\leq 2$, the number of regeneration circuits required by the CLGN model is less than that of the GNTR model. We thus conclude that the CLGN model is better at saving network resources compared with the GNTR model when there is a limited number of regeneration nodes.

C. Comparison with the Finely Linearized GN Model [3]

In this section, the proposed algorithm that is composed of the CLGN model and the SIO algorithm is compared to the algorithm described in [3]. The published algorithm models the RSA problem by a finely linearized GN model, which is more accurate than (6), and solves the problem by MILP. The finely linearized GN model has a linearization error less than 1%. We consider the published algorithm as a benchmark and compare the proposed algorithm with it. We apply the same topology as in [3], the DT-14 network, and the exact same demands for both algorithms. The DT-14 network is chosen over the NSF-24 network in this section due to the computational limitations of the benchmark algorithm, which does not consider signal regeneration ($T=0$). The objective here for both algorithms is to minimize the total spectrum usage, C.

The total spectrum usage required by both algorithms is shown in Figure 8 (a). When the number of demands is less

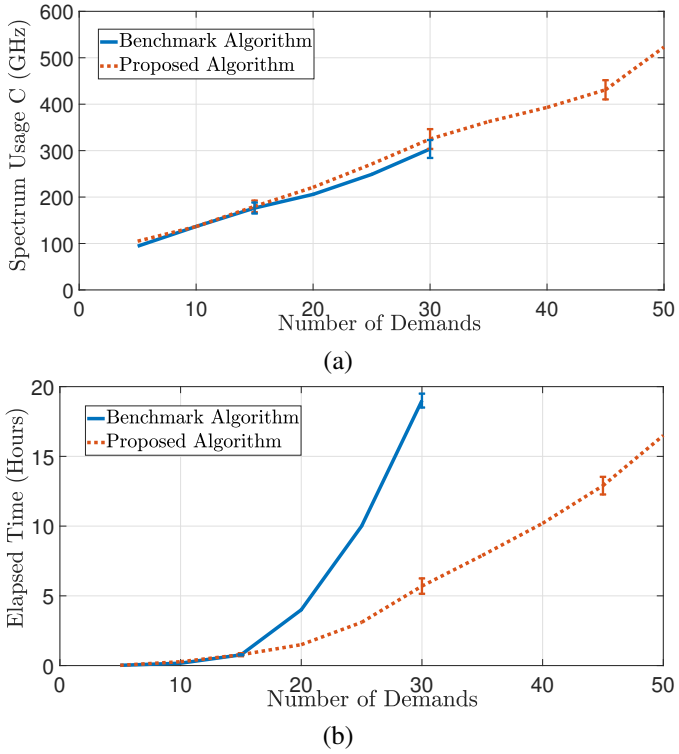

Fig. 8: Comparison between the proposed algorithm and the algorithm in [3] with BPSK modulation using the DT-14 network; (a) total spectrum usage as functions of $|\mathbb{D}|$, (b) elapsed time as functions of $|\mathbb{D}|$.

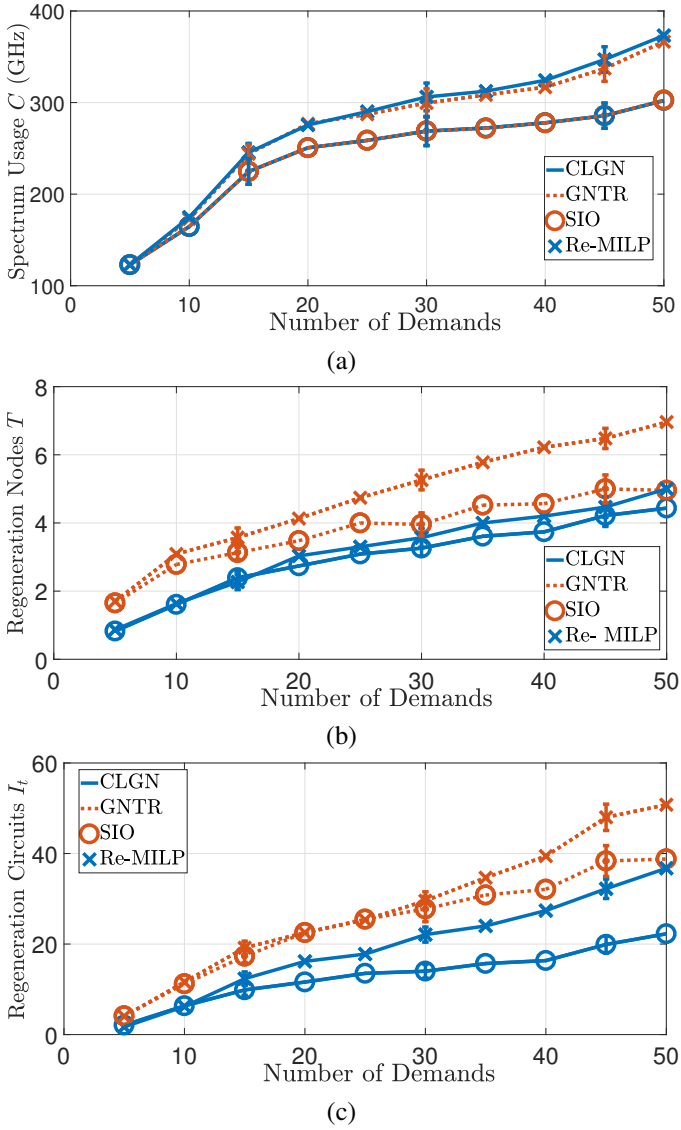

Fig. 9: Resource usage of proposed algorithm compared with the algorithm in [5], using QPSK, for the NSF-24 network (a) optimized spectrum usage as functions of $|\mathbb{D}|$ (b) regeneration nodes as functions of $|\mathbb{D}|$ (c) regeneration circuits as functions of $|\mathbb{D}|$.

than or equal to 15, the proposed and benchmark algorithms have similar spectrum usage. When the number of demands is between 15 to 30, the spectrum usage of our algorithm is slightly higher than the benchmark, less than 10%. When the number of demands exceeds 30, the published algorithm fails to give any result within a reasonable time (i.e., 36 hours). However, the proposed algorithm is able to solve the RSA problem. Although the benchmark algorithm provides slightly better results for up to 30 demands, the proposed algorithm is more scalable, uniquely able to provide results for up to 50 demands in less than 36 hours of computation time.

In Figure 8 (b), the total elapsed computation time for both algorithms is shown to increase approximately exponentially as a function of the number of demands. However, the computational efficiency of the proposed algorithm is much greater than the benchmark algorithm, due to the SIO algorithm appropriately breaking down the RSA problem into smaller problems. These results make the proposed algorithm more applicable to practical scenarios. We thus conclude that there is a trade-off where a small sacrifice in spectrum usage results in large savings on computation time.

D. Comparison with the Re-MILP Algorithm [5]

In this section, we compare the performance of our proposed heuristic algorithm, the SIO, with another published algorithm, the re-MILP algorithm [5]. The optimization objective is to minimize $C + \varepsilon T$. We simulate and compare the performance of both the CLGN and the GNTR models solved by both algorithms separately on the NSF-24 topology, with QPSK modulation.

The spectrum usage of both methods is shown in Figure 9 (a). Compared with the re-MILP algorithm, the SIO algorithm achieves a significant spectrum efficiency gain, which increases as the number of demands grows and reaches 19.0% at 50 demands. The spectrum usage of the CLGN model and the GNTR model is similar since the primary optimization objective here is the spectrum usage. The optimization thus sacrifices other resources such as regeneration nodes to ensure the optimality of the spectrum usage.

Even though the SIO algorithm has a complexity of the same order as the re-MILP algorithm, the resource savings gained by using SIO are notable. SIO takes approximately twice the computation time than the re-MILP to find a solution.

The SIO algorithm is therefore as scalable as the re-MILP algorithm because they can both allocate 50 demands in the NSF-24 network.

The usage of regeneration nodes and circuits is illustrated in Figure 9 (b) and (c), respectively. For both resources, the optimization solution from the SIO algorithm is better than that of the re-MILP algorithm in all cases tested. The number of regeneration nodes required by the SIO algorithm is up to 28.5% lower than that of re-MILP when there are 50 demands. In addition, the number of regeneration circuits needed by the SIO algorithm is up to 38.8% lower than that of the re-MILP algorithm at 50 demands. The reason is that the SIO algorithm can avoid local optima in the optimization process and iteratively pursues a better result. Moreover, the advantage of the SIO algorithm over the re-MILP algorithm increases when the complexity of the RSRA problem increases.

Figure 9 (b) and (c) also show that using the CLGN, a state-dependent PLI model, instead of a worst-case constraint such as the GNTR, can significantly reduce the number of regeneration nodes and circuits required. Notably, the numbers of regeneration nodes and regeneration circuits of the SIO algorithm with the CLGN model are 37.1% and 56.8%, respectively, less than those of the re-MILP with the GNTR model.

VI. Conclusions

In this paper we consider the RSRA problem for EONs with PLIs. Based on the standard GN model, we propose a novel estimation of PLIs, the CLGN model. In addition, we implement a GN-based transmission reach algorithm, referred to as the GNTR model, in order to make the comparison with the CLGN model fair.

The RSRA problem for networks suffering from PLIs is then modeled as an MILP formulation. We propose a heuristic algorithm, referred to as the sequential iterative optimization (SIO) algorithm, to solve the RSRA problem. We show through simulation that the CLGN model is better than the GNTR model in estimating PLIs, thus saving EONs resources, namely, the spectrum, regeneration nodes, and regeneration circuits. Moreover, the SIO algorithm outperforms the re-MILP algorithm [5]. It can be used on larger scale networks and assign more demands than the algorithm in [3]. We conclude that the proposed algorithm provides an effective balance between performance and scalability.

In future work, the proposed algorithm can easily be extended to address the enormous expected growth in the number and size of traffic demands by accommodating super-channels. The proposed RSRA approach can also be easily extended to include modulation selection, wavelength conversion, and modulation conversion, as in [5].

Acknowledgment

This work was supported in part by NSF grants CCF-1422871 and CNS-1718130, and by the Swedish Research Council (VR) under grant no. 2013-5271.

References

[1] O. Gerstel, M. Jinno, A. Lord, and S. Yoo, "Elastic optical networking: A new dawn for the optical layer?" *IEEE Commun. Mag.*, vol. 50, no. 2, pp. 12–20, 2012.

[2] B. C. Chatterjee, N. Sarma, and E. Oki, "Routing and spectrum allocation in elastic optical networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1776–1800, thirdquarter 2015.

[3] L. Yan, E. Agrell, M. N. Dharmaweera, and H. Wymeersch, "Joint assignment of power, routing, and spectrum in static flexible-grid networks," *J. Lightw. Technol.*, vol. 35, no. 10, pp. 1766–1774, May 2017.

[4] A. Klekamp, R. Dischler, and F. Buchali, "Limits of spectral efficiency and transmission reach of optical-ofdm superchannels for adaptive networks," *IEEE Photon. Technol. Lett.*, vol. 23, no. 20, pp. 1526–1528, Oct. 2011.

[5] X. Wang, M. Brandt-Pearce, and S. Subramaniam, "Impact of wavelength and modulation conversion on translucent elastic optical networks using MILP," *J. Opt. Commun. Netw.*, vol. 7, no. 7, pp. 644–655, July 2015.

[6] M. Klinkowski and K. Walkowiak, "On performance gains of flexible regeneration and modulation conversion in translucent elastic optical networks with superchannel transmission," *J. Lightw. Technol.*, vol. 34, no. 23, pp. 5485–5495, Dec. 2016.

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

[8] P. Johannisson and E. Agrell, "Modeling of nonlinear signal distortion in fiber-optic networks," *J. Lightw. Technol.*, vol. 32, no. 23, pp. 4544–4552, Dec. 2014.

[9] Y. Wang, X. Cao, and Y. Pan, "A study of the routing and spectrum allocation in spectrum-sliced elastic optical path networks," in *2011 Proc. IEEE INFOCOM*, Apr. 2011, pp. 1503–1511.

[10] J. Van Leeuwen and J. Leeuwen, *Handbook of theoretical computer science*. Elsevier, 1990, vol. 1.

[11] K. Christodoulopoulos, I. Tomkos, and E. A. Varvarigos, "Elastic bandwidth allocation in flexible OFDM-based optical networks," *J. Lightw. Technol.*, vol. 29, no. 9, pp. 1354–1366, May 2011.

[12] S. Talebi, F. Alam, I. Katib, M. Khamis, R. Salama, and G. N. Rouskas, "Spectrum management techniques for elastic optical networks: A survey," *Optical Switching and Networking*, vol. 13, pp. 34–48, 2014.

[13] J. Zhao, H. Wymeersch, and E. Agrell, "Nonlinear impairment-aware static resource allocation in elastic optical networks," *J. Lightw. Technol.*, vol. 33, no. 22, pp. 4554–4564, Nov. 2015.

[14] R. Wang, S. Bidkar, R. Nejabati, and D. Simeonidou, "Load and nonlinearity aware resource allocation in elastic optical networks," in *Opt. Fiber Commun. Conf. Exhib. (OFC)*, March 2017, pp. 1–3.

[15] F. M. Madani, "Scalable framework for translucent elastic optical network planning," *J. Lightw. Technol.*, vol. 34, no. 4, pp. 1086–1097, 2016.

[16] L. Yan, E. Agrell, H. Wymeersch, and M. Brandt-Pearce, "Resource allocation for flexible-grid optical networks with nonlinear channel model," *J. Opt. Commun. Netw*, vol. 7, no. 11, pp. B101–B108, 2015.

[17] Y. Xu, "Resource allocation in elastic optical networks with physical-layer impairments," Master's thesis, Dept. of Elect. & Comp. Eng., Univ. of Virginia, https://doi.org/10.18130/V3SW8F, Aug. 2017.

[18] L. Yan, E. Agrell, H. Wymeersch, P. Johannisson, R. Di Taranto, and M. Brandt-Pearce, "Link-level resource allocation for flexible-grid nonlinear fiber-optic communication systems," *IEEE Photon. Technol. Lett.*, vol. 27, no. 12, pp. 1250–1253, June 2015.

[19] L. Yan, "Resource allocation in flexible-grid optical networks with nonlinear interference," Licentiate thesis, Chalmers Univ. of Technol., Sweden, Dec. 2015.

[20] E. Palkopoulou, G. Bosco, A. Carena, D. Klonidis, P. Poggiolini, and I. Tomkos, "Nyquist-WDM-based flexible optical networks: Exploring physical layer design parameters," *J. Lightw. Technol.*, vol. 31, no. 14, pp. 2332–2339, July 2013.

[21] D. J. Ives, P. Bayvel, and S. J. Savory, "Physical layer transmitter and routing optimization to maximize the traffic throughput of a nonlinear optical mesh network," in *Proc. IEEE International Conference of Optical Network Design and Modeling (ONDM)*, Stockholm, Sweden, May 2014, pp. 168–173.

[22] FICO Xpress Optimization Suite, "MIP formulations and linearizations quick reference," https://www.fico.com/en/latest-thinking/brochures/mip-formulations-quick-reference, 2009.

[23] Gurobi Optimization, Inc., "Gurobi optimizer reference manual," http://www.gurobi.com, 2015.

[24] C. F. Hsu, Y. C. Chang, and S. C. Sie, "Graph-model-based dynamic routing and spectrum assignment in elastic optical networks," *J. Opt. Commun. Netw.*, vol. 8, no. 7, pp. 507–520, July 2016.
[25] A. Cai, J. Guo, R. Lin, G. Shen, and M. Zukerman, "Multicast routing and distance-adaptive spectrum allocation in elastic optical networks with shared protection," *J. Lightw. Technol.*, vol. 34, no. 17, pp. 4076–4088, Sept. 2016.

Yuxin Xu was born in Hangzhou, Zhejiang, China. He received a B.S. in Electrical Engineering from Zhejiang Gongshang University, Hangzhou, China in 2015. In 2017, he received his M.S in Electrical and Computer Engineering from the University of Virginia, Charlottesville, U.S.A, where he is currently working toward the Ph.D. His current research interests include resource allocation for elastic optical networks and physical layer impairments in optical networks.

Li Yan received the M. S. degree in photonics in 2013 from Beijing University of Post and Telecommunications, Beijing, China and the Ph.D. degree in electrical engineering in 2018 from Chalmers University of Technology, Gothenburg, Sweden. He is currently a postdoctoral researcher in the department of electrical engineering in Chalmers University of Technology. His research focuses meainly on resource allocation in optical networks, combinatorial optimizations, and convex optimizations. He is also interested in optical communication simulations and optical channel modelling.

Erik Agrell (M'99–SM'02–F'18) received the Ph.D. degree in information theory in 1997 from Chalmers University of Technology, Sweden.

From 1997 to 1999, he was a Postdoctoral Researcher with the University of California, San Diego and the University of Illinois at Urbana-Champaign. In 1999, he joined the faculty of Chalmers University of Technology, where he is a Professor in Communication Systems since 2009. In 2010, he cofounded the Fiber-Optic Communications Research Center (FORCE) at Chalmers, where he leads the Electrical Engineering research area. He was a Visiting Professor at University College London in 2014–2017. His research interests belong to the fields of information theory, coding theory, and digital communications, and his favorite applications are found in optical communications.

Prof. Agrell served as Publications Editor for the IEEE Transactions on Information Theory from 1999 to 2002 and as Associate Editor for the IEEE Transactions on Communications from 2012 to 2015. He is a recipient of the 1990 John Ericsson Medal, the 2009 ITW Best Poster Award, the 2011 GlobeCom Best Paper Award, the 2013 CTW Best Poster Award, the 2013 Chalmers Supervisor of the Year Award, and the 2015 JLT Best Paper Award.

Maïté Brandt-Pearce is a professor of Electrical Engineering and vice provost for faculty affairs at the University of Virginia. She joined UVA after receiving her Ph.D. in Electrical Engineering from Rice University in 1993. Her research interests include free-space optical communications, visible light communications, nonlinear effects in fiber-optics, and cross-layer design of optical networks subject to physical layer degradations. Dr. Brandt-Pearce is the recipient of an NSF CAREER Award and an NSF RIA. She is a co-recipient of Best Paper Awards at ICC 2006 and GLOBECOM 2012. She had served on the editorial board of IEEE Transaction of Communications, IEEE Communications Letters, IEEE/OSA Journal of Optical Communications and Networks and Springer Photonic Network Communications. She was Jubilee Professor at Chalmers University, Sweden, in 2014. After serving as General Chair of the Asilomar Conference on Signals, Systems & Computers in 2009, she served as Technical Vice-Chair of GLOBECOM 2016. She is a member of Tau Beta Pi, Eta Kappa Nu, and a Senior Member of the IEEE. In addition to co-editing a book entitled Cross-Layer Design in Optical Networks, Springer Optical Networks Series, 2013, Prof. Brandt-Pearce has over two hundred technical publications.