

# Feedback-Based Channel Frequency Optimization in Superchannels

F. Locatelli , K. Christodouloupoulos , C. Delezoide , J. M. Fàbrega , M. Svaluto Moreolo , L. Nadal ,  
A. Mahajan , and S. Spadaro 

**Abstract**—Making optical networks more efficient and reliable requires further automation of the optical layer. In this context, we propose a closed control loop that automatically performs fine frequency adjustments of the subchannels of superchannels to maintain optimal performance despite time-dependent impairments, thus achieving three main goals. At design, our scheme reduces the need for margins and guard-bands dedicated to spectrum-related impairments such as filter or channel detunings. This allows for a more efficient network at deployment. Then, during operation, the quality of transmission (QoT) of each subchannel is maximized to make the superchannel more resilient to any kind of soft failure, thus improving network resilience. Finally, considering the elastic network paradigm, performance improvements can be converted into significant capacity upgrades. We demonstrate the ability of our solution to achieve these three goals with split-step Fourier simulations. For a four-subchannel superchannel, we demonstrate robustness to  $\pm 2$  GHz frequency detunings and gains of up to 3.7 dB in quality of transmission.

**Index Terms**—Control loop, frequency margins, gradient descent method, iterative method, monitoring, superchannel optimization.

## I. INTRODUCTION

**B**ECAUSE the spectral efficiencies achieved by the latest commercial coherent transponders are very close to the Shannon limit [1], the current challenge for our industry is not to increase the capacity but to find sustainable ways to do it. This may only be achieved through an optimized usage of

Manuscript received 19 October 2023; revised 16 December 2023 and 23 January 2024; accepted 26 February 2024. Date of publication 4 March 2024; date of current version 8 July 2024. This work was supported by the ONFIRE Project of the Horizon 2020 Framework Programme under Grant 765275. (Corresponding author: K. Christodouloupoulos)

F. Locatelli was with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)/CERCA, 08860 Barcelona, Spain. He is now with the Cisco, 00142 Roma, Italy.

K. Christodouloupoulos was with Nokia Bell Labs, 70469 Stuttgart, Germany. He is now with the Department of Informatics and Telecommunication, University of Athens, 15772 Athens, Greece (e-mail: kchristodou@di.uoa.gr).

C. Delezoide is with the Nokia Bell Labs, 91620 Paris, France.

J. M. Fàbrega, M. Svaluto Moreolo, and L. Nadal are with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)/CERCA, 08860 Barcelona, Spain.

A. Mahajan was with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)/CERCA, 08860 Barcelona, Spain. He is now with the Infinera, San Jose, CA 95128 USA.

S. Spadaro is with Universitat Politècnica de Catalunya, 08034 Barcelona, Spain.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2024.3372376>.

Digital Object Identifier 10.1109/JLT.2024.3372376

the optical bandwidth, with two orthogonal approaches. The *hardware approach* seeks higher symbol rates [2] to drive costs down and minimize the proportion of unused bandwidth, i.e., guard-bands. Yet, when the capacity to deliver exceeds the maximal reach of a single channel, forming a superchannel with compacted subchannels remains the optimal solution [3]. The *software approach* seeks to optimize the spectral efficiency by minimizing the quality of transmission (QoT) margins [4]. In that context, it has been demonstrated that any effective margin reduction impacts the lightpaths' availabilities [5]. Depending on the availability cost, a margin reduction can eventually be counterproductive in terms of global network efficiency. Thus, one cannot reduce margins, nor margins-equivalents [6] such as guard-bands, without first assessing its impact on availability. For deployed networks, this can be achieved through monitoring and analyses, as done in Nokia's Wavesuite Health & Analytics [7] through the analysis of margin-availability curves. At design, however, reducing margins requires mechanisms to mitigate the uncertainties for which they are introduced [8].

In optical network design, margins are set to ensure nominal network behavior despite many uncertainty sources. Those mainly stem from a limited knowledge of the parameters of the network subsystems - fiber spans, amplifiers, transponders, nodes - and the limited accuracy of QoT estimation models. The current uncertainty on a lightpath's QoT is mostly driven by complex time-dependent effects [9], often with stochastic components. For instance, polarization-dependent loss (PDL) combined with random fluctuations of the state of polarization of the lightpath can induce significant variations of the QoT [10], [11]. With similar consequences, temperature-driven drifts of the frequency alignment between a channel and the filter cascade, give rise to filtering impairments [12]. From a design perspective, mitigating such effects allows to reduce, if not cancel, the need for associated margins and guard-bands. Due to the time-dependency, this can only be achieved through closed-loop automation, as proposed in [11], [13]. In [13], the authors automatically fine-tune the central frequency of an optical channel to maintain an optimal alignment with its filter cascade. However, this approach is only applicable to filter-bounded channels, i.e., channels with filter edges on both sides, which is not the case for subchannels in a superchannel.

To combine both hardware and software approaches to optimize bandwidth usage, we extend the work of [13] to propose a closed control loop that fine-tunes the central frequencies of the subchannels of a superchannel. We assume superchannels

formed by superpositioning single-channel transponders (several boxes, or one box containing multiples). When dealing with superchannels utilizing optical comb sources, a similar method could be applied to optimize the spacing and average frequency of the comb. Moreover, the packing of individual channels into wave-bands is considered as a scalable solution for next-generation optical networks [14], in which case the proposed solution could minimize the interference among channels allocated to the same band. Finally, filtering effects are a major limitation in optical metro and core networks [15]. To reduce that, adjacent channels that enter a Reconfigurable Optical Add-Drop Multiplexer (ROADM) at the same ingress and are forwarded to the same egress port are treated collectively as a superchannel employing a ‘superfilter’ [16], [17]. This concept finds practical applications in real deployed networks today. Thus, the necessity for automated frequency-tuning solutions becomes crucial. The proposed process probes the network, configures new frequencies to the subchannels, monitors the outcome, decides on new frequencies, and iterates until converging to optimal subchannels spacing. The novelty of this work lies in the use of monitoring information to develop a closed-loop process that is combined with an appropriate optimization method that is robust and converges to the optimum under monitoring or other random errors.

The proposed process achieves three goals. First, the ability to automatically maintain an optimal distance between subchannels, allowing to minimize guard-bands and QoT margins at design. Second, subchannel QoT’s maximization during operation effectively maximizes network resiliency to soft failures [18], [19], [21]. Third, in the elastic network paradigm [23], the extra margins gained by the automation can be converted into increased capacity without consuming more bandwidth. As a bonus, mitigating time-varying effects makes the subchannels’ instantaneous QoT’s more stable and predictable by existing physical models. In turn, this makes the network more compatible with digital twinning [24], with all related operational advantages.

The rest of the paper is organized as follows. In Section II, we formulate the optimization problem and describe the proposed solution. In Section III, we present the obtained results and Section IV concludes the paper.

## II. PROBLEM FORMULATION AND PROPOSED SOLUTION

In this paper, we will denote scalars with italics and vectors with bold font. We assume a superchannel with an allocated spectrum  $BW_{SC}$  centered around the frequency  $f_{SC}$  and composed of a set of  $N$  subchannels, as depicted in Fig. 1. We initially assume the superchannel to be equidistant and uniform. The subchannels have symbol rate  $R_S$  and are shaped with a root raised cosine (RRC) of roll-off  $a$  (actually, the proposed solution is agnostic to the shaping, RRC is mentioned here since it was assumed in the simulations). As such, each subchannel physically occupies a bandwidth  $BW_{min} = R_S (1 + a)$ , e.g., a 32 GBd channel with 0.1 roll-off occupies 35.2 GHz, which is the minimum distance between subchannels that yields negligible crosstalk. As is the case for all connections in transparent

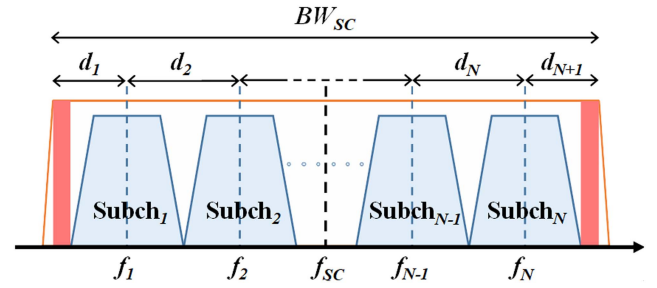


Fig. 1. Example of a superchannel with  $N$  subchannels. The two red bands at the superchannel sides represent the guard-bands.

networks, the superchannel crosses at least one filter with 3 dB bandwidth  $BW_{SC}$  and central frequency  $f_{SC}$ .

We denote by vector  $\mathbf{D} = [d_1, d_2, \dots, d_{N+1}]$  the spectral distances between subchannels including the filter edges, as shown in Fig. 1. The subchannels’ QoTs depend on several parameters but also on  $\mathbf{D}$ . In the following we use the logarithmic SNR (in dB) as the reference QoT metric, defined as the ratio of the signal power over the noise power before the decision point. We denote by  $SNR_n(\mathbf{D})$  the logarithmic SNR value (in dB) of the  $n$ -th subchannel. This, from the vector  $\mathbf{D}$  mainly depends on  $d_n$  and  $d_{n+1}$  due to crosstalk and for the outer subchannels also on filtering. We also denote by  $\mathbf{SNR}_{SC}(\mathbf{D}) = [SNR_1(\mathbf{D}), \dots, SNR_N(\mathbf{D})]$  the vector that includes the SNR values of all subchannels.

Superchannel frequency optimization consists in identifying  $\mathbf{D} = \mathbf{D}^*$  that optimizes a function of the subchannels’ SNRs. In the following, we investigate two distinct optimization strategies, each based on a unique optimization function:

- 1) the superchannel’s total (or average if divided by  $N$ ) SNR

$$h_1(\mathbf{D}) = \sum_{n=1}^N SNR_n(\mathbf{D}) \quad (1)$$

- 2) the minimum subchannel SNR

$$h_2(\mathbf{D}) = \min_n (SNR_n(\mathbf{D})) \quad (2)$$

In addition, the following set of constraints applies:

$$\sum_{n=1}^{N+1} d_n = BW_{SC} \quad n = 1, \dots, N \quad (3)$$

$$d_n \geq \begin{cases} \frac{R_S}{2}, & n = 1, N+1 \\ \frac{R_S}{2}, & 1 < n < N+1 \end{cases} \quad (4)$$

Constraint (3) keeps the subchannel distances within the filter bandwidth. Constraint (4) puts a lower limit on the distances between adjacent subchannels and outer subchannels and filter, but it is optional and is typically skipped.

The former objective, Obj#1, is directly related to the superchannel’s total capacity. According to Shannon, the total capacity is proportional to  $\sum_{n=1}^N \log_2(1 + SNR_n^{scalar}(\mathbf{D}))$ . For high SNR values, we have:  $\log_2(1 + SNR_n^{scalar}(\mathbf{D})) = \log_2(SNR_n^{scalar}(\mathbf{D})) \sim SNR_n(\mathbf{D})$ , where the last term is the logarithmic SNR metric used in our formulations. Thus, maximizing obj #1 maximizes the total superchannel’s capacity. This

objective would be the natural choice in networks where subchannels can be configured with different modulation formats. The benefits would be higher when the spectral efficiency is fine-tuned according to each subchannel physical performance, e.g., with probabilistic constellation shaping [25].

The latter objective, Obj#2, targets to increase the SNR value of the worst subchannel. This objective is particularly suited to more conventional network settings, where all subchannels have the same modulation format. In such case, using Obj#2 would make all channel SNR values to be higher than e.g., an FEC threshold, thus, making them feasible to use the highest possible common modulation format.

#### A. Subchannel Distance and SNR Function Properties

The dominant impairment that alters the SNR values of the subchannels when varying the TX laser frequencies - i.e., distances between subchannels  $\mathbf{D}$  - is the linear crosstalk between two adjacent subchannels and the amount of filtering on the outer subchannels. Varying  $\mathbf{D}$  creates small variations of the contribution of the nonlinear interference effects among subchannels within the superchannel and other network channels. Such effects can be considered as a constant degradation included in the measured SNR and thus do not affect our optimization. Regarding linear crosstalk, as we move one channel towards another, the area in which these two signals overlap increases. The crosstalk is proportional to the integral of the two signals' overlapping area. The integral of any increasing function is convex, and thus crosstalk degradation is convex as distance reduces. This is assuming that the crosstalk from the other side is negligible, i.e., the crosstalk from the side that we reduce the distance dominates. The (scalar) SNR of the channel depends inversely on the crosstalk and, thus, it is concave to the distance. Similar conclusions are drawn when moving one channel towards a filter. The logarithm of a convex function (logconvex) is also convex. Thus, the logarithm of the SNR, the metric used in our objectives, is a concave function of the distance regardless of both channel shape and filter shapes. In the generic case where we have several subchannels, the objectives described by (1) and (2) are linear combinations (summation or minimum) of concave functions, and thus the objectives are concave, making convex/concave optimization techniques based on gradients suitable to solve the related problem [27].

As an illustrative example of the convexity of the SNR with respect to distance, we created in VPI three 32 GBd single polarization quadrature phase shift keying (QPSK) modulated signals, pulse shaped with an RRC filter and 0.1 roll-off factor and launch power of 0 dBm. We denote by  $d_1$  the distance of the second subchannel from the first and by  $d_2$  the distance of the third to the second. In Fig. 2, we plot the (logarithmic) SNR values of the second channel, measured in VPI by translating the measured Error Vector Magnitude (EVM) value to SNR, as a function of its distance  $d_1$  from the first channel, while distance  $d_2$  was set to keep  $d_1 + d_2 = 69$  GHz. In simple words, we kept the first and third channels constant and shifted the second channel. We plot the SNR values for back-to-back (B2B) and transmission over two and ten standard single-mode

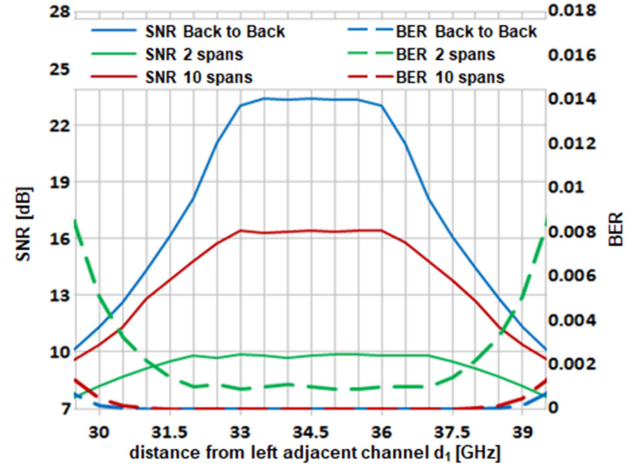


Fig. 2. Relation between the SNR and BER value of the second channel and its distance with the first one ( $d_1$ ), for back to back (B2B), two SSF spans (160 km) and ten SSF spans (800 km) scenarios.

fiber (SSMF) spans of 80 km length, with 0.2 dB/km attenuation coefficient and 16.7 ps/nm/km dispersion coefficient. Each SSF span was followed by an erbium-doped fiber amplifier (EDFA) with a 5.5 dB noise figure (NF) and gain that matched the span loss. We observe in Fig. 2 that the logarithm of the SNR of the intermediate (second) channel is a concave function of  $d_1$  around the equidistant point, that is starting at  $d_1 = d_2 = 34.5$  GHz and then having  $d_1 \pm 3$  GHz. We observed a similar shape when moving a channel to meet the side of a filter. Regarding the number of SSF spans, we observe that transmitting the signals over the spans reduces the reference SNR value and flattens the shape of the SNR curve (due to EDFA noise and widening of channels due to dispersion). Still, the concave shape is maintained. We also show in Fig. 2 the related BER values, which as expected is a convex function of the distance.

Although from Fig. 2 it might seem that there is a continuous set of distances  $d_1$  for which the maximum SNR value flattens out, if we zoomed in on that area, we would see minor SNR variations caused by random noise, wavelength-dependent penalties, and other factors. The sides are also not smooth. Furthermore, in real systems, we would also have TX/RX imperfections, EDFA gain ripples, and monitoring errors. Therefore, there are several local maxima in the flat area, and maybe at the sides. Even more, the global and local optima would vary as a function of time due to short-term impairments. Thus, all points in the flat area are considered equally good, and the goal is to reach the flat area and stop even if we do not find the global optimum. For these reasons, we decided to use the stochastic subgradient method, which is robust, can optimize in the presence of noise and uncertainties [28], and works with nonlinear and even with certain nonconcave problems [29].

#### B. Closed Control Loop for Optimizing Subchannel Frequencies

We propose a closed control loop that automatically performs fine frequency adjustments of the subchannels of a superchannel. The closed control loop leverages an iterative optimization

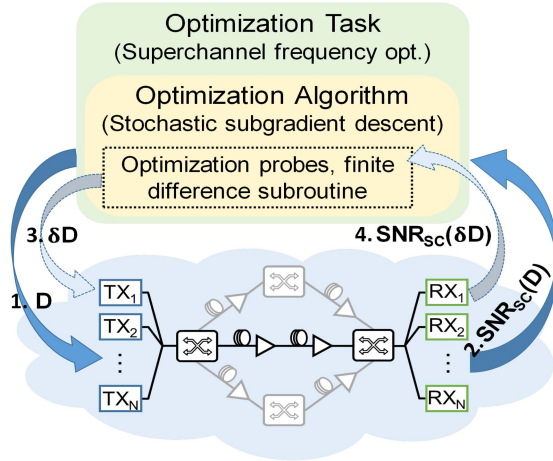


Fig. 3. Schematic representation of the implemented optimization loop with monitoring probes.

algorithm based on the stochastic subgradient method with a fixed step length. The optimization algorithm is envisioned to reside in the central SDN controller which has the appropriate interfaces to configure the TX laser frequencies and monitor the SNR (or the pre-FEC BER) values of the subchannels, e.g., at the RX [21], [22]. We employ the stochastic subgradient method to account for short-term fluctuations related to measurements and random noise. The primary issue addressed by the proposed control loop is the drifts in the laser and filter frequencies. These effects are medium- and long-term by nature, e.g., temperature-dependent, and significant changes occur over days, weeks, or more. Therefore, the proposed closed-loop process, expected to run in the order of tens of minutes, is orders of magnitude faster than the variations it corrects.

The used stochastic subgradient method finds the optimum of a noisy convex problem in a polynomial number of iterations. It requires to find the subgradients which is done using a subroutine that implements the finite difference method [30]. The subroutine provides probes, defines TX frequencies that are configured to the network through the control plane, and collects the outcomes (the monitored SNR values).

The workflow of the proposed control loop is as follows:

Input: objective function  $h$ , minibatch size  $M$ , frequency probe step  $f_{step}$ , convergence accuracy  $\varepsilon$

At the previous iteration the optimization algorithm (stochastic subgradient method) decided to configure the distance vector  $\mathbf{D}$  to the network, indicated as step 1 in Fig. 3. The algorithm monitors  $\text{SNR}_{\text{SC}}(\mathbf{D})$ , the vector of SNR values of all subchannels, indicated as step 2 in Fig. 3.

The algorithm calls the finite difference subroutine  $M$  times. The finite difference subroutine each time selects one subchannel,  $n$ , out of the  $N$ , to use in the probing. Then, the subroutine probes, by configuring the distance vector  $\delta \mathbf{D}_n$  in the network. In  $\delta \mathbf{D}_n$ ,  $f_{step}$  is subtracted from the central frequency of the  $n-1$ -th subchannel  $d_{n-1}$  and added to the frequency of the  $n$ -th subchannel  $d_n$ . This is indicated as step 3 in Fig. 3. It monitors  $\text{SNR}_{\text{SC}}(\delta \mathbf{D}_n)$ , indicated as step 4 in Fig. 3, and computes the

subgradient of  $n$ , denoted by  $g_n$ , based on objective function  $h$ ,  $\text{SNR}_{\text{SC}}(\delta \mathbf{D}_n)$  and  $\text{SNR}_{\text{SC}}(\mathbf{D})$ .

$$g_n = \frac{h(\mathbf{D}) - h(\delta \mathbf{D}_n)}{f_{step}} \quad (5)$$

The algorithm based on the  $M$  subgradients, computed using the finite difference subroutine, calculates a new distance vector  $\mathbf{D}'$  and estimates  $|h(\mathbf{D}) - h(\mathbf{D}')|$ .

The stochastic subgradient method with a fixed step length  $f_{step}$  iterates and converges to the optimum [31]. If  $f_{step} = 2\varepsilon/G$ , where  $G$  represents a bound on the gradient such as  $|h(\mathbf{D}_u) - h(\mathbf{D}_v)| < G\|\mathbf{D}_u - \mathbf{D}_v\|_2$ , valid for any two distance vectors  $\mathbf{D}_u$  and  $\mathbf{D}_v$ , then the subgradient method converges to the optimum within  $\varepsilon$ . Specifically,

$$|h_{best}^I - h^*| \leq \varepsilon, \quad (6)$$

where,  $h_{best}^I$  and  $h^*$  represent the best objective value up to the  $I$ -th iteration and the optimal solution, respectively. The required number of iterations,  $I$ , to converge to the optimum within  $\varepsilon$  is  $I = (RG/\varepsilon)^2$ , where  $R$  is the distance of the starting distance vector  $\mathbf{D}^{(0)}$  from the optimum distance vector  $\mathbf{D}^*$ ,  $R \geq \|\mathbf{D}^{(0)} - \mathbf{D}^*\|_2$ . In case of heavy noise,  $G$  will be large and, thus,  $I$  will be also large. This, high number of iterations helps to ‘average out’ the higher noise/errors [27].

Although the subgradient method calculates a new distance vector,  $\mathbf{D}$ , to improve the objective at each iteration, this is an estimation based on the subgradients that it has at that point. Such a choice might not increase the measured objective when applied to the actual system because of noise. However, the algorithm is robust to such issues and converges to the optimum within  $I$  iterations, as indicated by (6). Practically, in the problem at hand, if we start from equidistant subchannels, or from a soft failure situation, we would be within  $\pm 2$  GHz from optimum, so  $R$  is known. However,  $G$  is hard to estimate. So in the above workflow and the performance evaluation we stop at iteration  $I$  when  $|h(\mathbf{D}) - h(\mathbf{D}')| \leq \varepsilon$ . Our results indicated that using this stopping criterion the algorithm was able to identify or reach close to the optimum in all examined scenarios. Different stopping criteria could be applied for different system conditions.

Apart from the nice convergence property, the choice of a fixed step size  $f_{step}$  is done mainly for practical purposes. The distance vector,  $\mathbf{D}$ , is defined by the central frequencies of the subchannel TX lasers. However, the lasers cannot be configured with infinite accuracy. Moreover, because of physical layer variations (short and medium terms) and monitoring errors, we cannot observe fine differences in the monitored SNR values. For these reasons, we use a frequency step  $f_{step}$ . Smaller steps tend to make the stochastic subgradient method slower but give better granularity, and thus approach the optima more precisely. Compared to the optima found with a brute-force approach, our results indicate good convergence performance for 0.25 GHz and 0.5 GHz frequency steps, whereas the performance is slightly worse for a step of 1 GHz.

Finally, note that the type of monitoring probes used to identify the subgradients are specific to this problem. A different optimization problem would require a different definition of

monitoring probes. Moreover, the monitoring probes approach involves several interactions with the network, which might be time-consuming. Therefore, alternative methods that use a physical layer model (PLM) / QoT estimator, such as the GN model [33], might be preferable, according to the problem. The key question is whether the PLM incorporates the key affecting impairments and how accurately it models those since the PLM inaccuracies could mislead the optimization. Such concepts are discussed in [34] for the power optimization problem and are left for future investigations for the problem at hand.

### III. SIMULATION OF OPTIMIZATION PROCESS AND RESULTS

#### A. Simulation of the Optimization Process

We implemented and verified the proposed optimization process through co-simulation in VPI [35] and MATLAB. In particular, we simulated the superchannel and its transmission by split-step Fourier method using VPI. This was chosen since the split-step Fourier method accounts for the key physical layer impairments that affect the proposed solution. It includes TX laser linewidth, nonidentical PRBS, and pulse shaping with different roll-off (r.o.) factors at the TX side. It also models linear crosstalk, wavelength-dependent deteriorations such as detailed nonlinear interference, EDFA gain ripple, and ASE noise randomization. Thus, in the simulations, there were small fluctuations in SNR values for the same set of parameters (due to random ASE noise, measurement uncertainty due to limited test sequence length, and variations in the receiver DSP). Thus, the measured SNR changed from one iteration/loop to the next, in the order of  $\pm 0.1$  dB, and the proposed process should optimize over such uncertainties. Moreover, the measured SNR values were non-flat with respect to the wavelength used, due to wavelength-dependent effects. We implemented in MATLAB the frequency optimization algorithm with the 2 objectives as discussed in Section II.

The co-simulation exchanges data between the optimization algorithm (in MATLAB) and VPI, which would be the optical network in the real world. The algorithm (in MATLAB) identifies which frequencies  $\delta \mathbf{D}_n$  to probe the superchannel, configures those to the network (changes the respective parameters in VPI), and receives the monitored SNR values  $\mathbf{SNR}_{SC}(\delta \mathbf{D}_n)$  (monitored in VPI, by translating the measured EVM value to SNR). After those probes, the algorithm calculates subgradients, uses them to calculate the next frequencies  $\mathbf{D}$ , and passes those to VPI. This process is repeated until the algorithm converges to the optimal solution. We kept track of the best-achieved solution and stopped the algorithm once the estimated improvement was below a specific threshold (0.2 dB). Unless stated differently, in the subgradient method we used a minibatch of size  $M = 2$ .

Regarding transmissions, we implemented the setup depicted in Fig. 4. We generated four 32 Gb/s, PM-QPSK or 16QAM, RRC-shaped subchannels with roll-off (r.o.) of 0.1 and 0.15, symmetrically centered around 193.1 THz. We set the launch power of each TX laser at 0 dBm. As a starting point, we consider the four subchannels equally spaced by 34.5 GHz. In addition to the equidistant configuration, we also considered 10 random starting frequencies, which were  $\pm 2$  GHz shifted

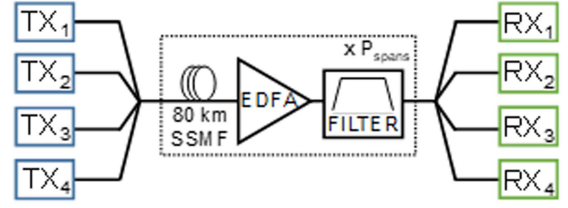


Fig. 4. VPIphotonics simulation setup. TX: transmitter, SSMF: standard single-mode fiber, EDFA: erbium-doped fiber amplifier, RX: receiver.

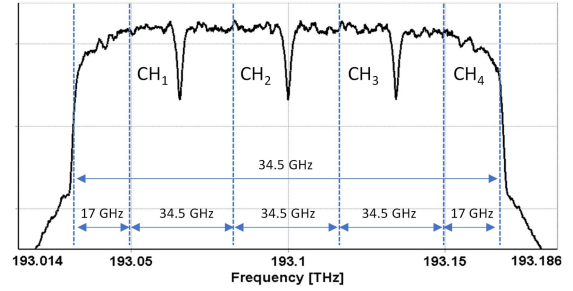


Fig. 5. VPI-generated spectrum for the default superchannel scenario, i.e.,  $a = 0.1$ , number of loops = 2,  $BW_{SC} = 137.5$  GHz.

from the equidistant setting. Then, we multiplexed the four subchannels and transmitted them in a link composed of 2 or 10 loops. Each loop was a span of 80 km length standard SMF, with 0.2 dB/km attenuation coefficient and 16.7 ps/nm/km dispersion coefficient, an EDFA with NF of 5.5 dB, followed by a tunable optical filter with 3.5th-order super-Gaussian transfer function. We set the 3 dB bandwidth of the filter with values ranging between 137.5 GHz and 200 GHz.

By varying the number of loops, we simulated propagation over different links and nodes. Before reception, we placed a dispersion-compensating optical fiber to fully compensate for the chromatic dispersion effect and avoid using DSP for faster simulation. Then, we detected each of the four subchannels at the RX side by tuning the frequency of the four coherent RX local oscillators. We set 3 dB bandwidth values for the RX filters equal to 16 GHz. We also extended our simulations up to ten subchannels, as reported at the end of this section.

#### B. Default Network Scenario

For the four-subchannel scenario, we considered a default scenario with  $BW_{SC}$  equal to 137.5 GHz, 0.1 roll, and two loops. Fig. 5, shows the spectrum at the receiver observed in VPI. Considering such default parameters, the received four subchannel SNR values were equal to  $\mathbf{SNR}_{SC}(\mathbf{D}_{eq-dis}) = [13.55, 16.86, 16.91, 13.15]$  dB. This resulted in an average superchannel SNR value of 15.12 dB, with 13.15 dB being the minimum among the four subchannel SNR values. Once optimized using the first objective (maximizing the superchannel average SNR - Section II) with a step size of 0.25 GHz, we found the frequencies  $\mathbf{D}^*_{obj1} = [18.25, 33.25, 33.25, 33, 19.75]$  GHz. The superchannel average SNR value associated with this scenario was 15.48 dB, yielding a 0.36 dB improvement with respect to the equally spaced initial distances  $\mathbf{D}_{eq-dis}$ . Optimizing with the second objective

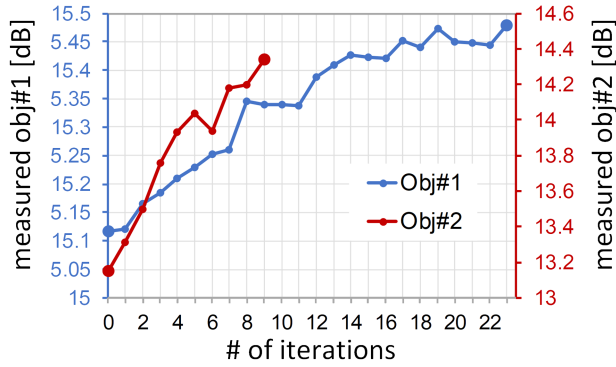


Fig. 6. Evolution of the SNR related to the two optimization problems as a function of the iteration number for the default scenario.

(maximization of the minimum subchannel SNR), we obtained the frequencies  $\mathbf{D}^*_{\text{obj2}} = [19, 33, 31.75, 32.75, 21]$  GHz and a corresponding set of SNR values  $\mathbf{SNR}^*_{\text{obj2}}(\mathbf{D}^*_{\text{obj2}}) = [14.34, 14.58, 14.37, 14.38]$  dB. We see that the algorithm flattened the SNR values of the subchannels, expected in such an optimization objective, where the algorithm tries to improve the minimum, eventually bringing all the subchannels to a similar level. The optimal minimum SNR value was found to be 14.34 dB, yielding a 1.19 dB improvement with respect to equidistant initial setting  $\mathbf{D}_{\text{eq-dis}}$ . The iterations (denoted by  $I$ ) needed to obtain these solutions were 23 for the superchannel average SNR maximization, and 9 for the minimum SNR maximization. Note that the number of iterations affects (linearly) the number of probes and monitoring calls and thus the execution time.

The evolution of the two optimization objectives measured in the system as a function of the iteration number for the default scenario and equidistant subchannels as a starting point is shown in Fig. 6. We observe that the stochastic subgradient method is not an ascent method, like gradient ascent (note that gradient descent is used in minimizing convex problems while gradient ascent is the same method used for maximizing concave problems, as the problem at hand). The algorithm calculates a new configuration that it expects to improve the objective, however, when that configuration is applied to the system, the measured outcome in the system, which is plotted in Fig. 6, could differ due to noise. The used method is robust and not affected by such issues. The convergence of Obj#1 which involves all channels is typically slower than that of Obj#2 which involves the worst channel. Regarding the stopping criterion, as discussed in Section II.B, the implemented process stops at iteration  $I$  when it estimates that it cannot improve the objective, that is, when  $|h(\mathbf{D}) - h(\mathbf{D}')| \leq \varepsilon$ , with  $\varepsilon$  set to 0.2 dB. Although it seems that the measured objective is still rising for Obj#2 (red line) at iteration 9 in Fig. 6, the last measurement can include a high measurement error/noise. The method at iteration 9 estimated that it could not further (higher than  $\varepsilon$ ) improve the objective and thus it stopped. We conducted an exhaustive search using the brute force method, exploring all possible solutions (with a specific granularity). Our observations revealed that the proposed stochastic subgradient method, coupled with the employed stopping criterion, was able to find the optimum or

approximate it to a very good degree in all examined scenarios, as discussed next. The stopping criterion could be chosen differently under different system conditions.

The proven convergence property of the stochastic subgradient method under any starting conditions and random noise [27] was verified by repeating the same simulations for 10 initial random conditions ( $\pm 2$  GHz from the equidistant setting). When we initiated the process from a distinct initial random condition, the algorithm followed different steps, reached different intermediate solutions, and required a different number of iterations to converge to the same or a very close solution. The convergence was also verified in the results presented in the next section, where in each case we consider soft-failures we simulated 10 initial random conditions and averaged the results.

### C. Improvement Over Equidistant Superchannel

In addition to the default scenario discussed above, we also considered eight other scenarios. The results for the nine in total scenarios (sc.) are summarized in Table I for obj #1 (maximizing superchannel's average SNR), and in Table II for obj #2 (maximizing the minimum subchannel SNR). With respect to the default sc. 1, in sc. 2 to 5, we considered variations of the: modulation format (sc. 2), superchannel filter bandwidth (sc. 3), roll-off (r.o.) (sc. 4), and span number (sc. 5). In general, the results reported in Tables I and II showcase that, independent from the scenario variations, the proposed optimization solution was always able to improve the considered objectives. We observed for obj #1 average SNR value improvements ranging from 0.28 to 0.75 dB (the total SNR improvement of the superchannel that employs 4 subchannels is four times that). The improvements with respect to the minimum subchannel SNR value ranged from 0.97 to 2.04 dB for obj #2. In addition, we ran simulations considering ten different random starting frequencies for sc. 2, 3, 4, and 5. In all such cases, the algorithm always converged on the same optimal solution.

The sc. 6-9 correspond to B2B configurations. We considered those to understand the contribution of the filtering and crosstalk effects. In sc. 7 and 9, where we set a very broad filter (3 dB bandwidth equal to 200 GHz), almost no improvement was observed after the optimization with respect to the equidistant superchannel configuration. This happens because our starting conditions (i.e., 34.5 GHz distanced subchannels – see Fig. 2) already give an impairment-free transmission. On the other hand, when considering B2B and tighter filter bandwidth (i.e., sc. 6 and 8), crosstalk- and filter-related penalties appear, and therefore our solution can improve the considered objectives. As expected, in sc. 8, where we put a higher r.o. factor, has a lower initial (average and total) SNR value than sc. 6. However, its improvement with respect to equidistant configuration is relatively lower since both crosstalk and filtering are strong, and not enough optimization space is present.

### D. Soft-Failure Resilience

To provide a benchmark for the optimality performance of the proposed method but also to measure the improvement over soft failures, we also report in Tables I and II the best and the worst

TABLE I  
RESULTS WHEN OPTIMIZING THE SUPERCHANNEL AVERAGE SNR VALUE (OBJ #1) WITH A STEP SIZE OF 0.25 GHz

| Scenario                                       | Average SNR<br>equidistant<br>subch. [dB] | #<br>Iterations | Average SNR<br>(Obj#1) [dB] | Brute-force $\pm 2$ GHz around opt.<br>solution |                           | Improvement<br>wrt worst<br>case [dB] | Improvement<br>wrt equi-<br>distanced<br>subch. [dB] |
|------------------------------------------------|-------------------------------------------|-----------------|-----------------------------|-------------------------------------------------|---------------------------|---------------------------------------|------------------------------------------------------|
|                                                |                                           |                 |                             | Best average<br>SNR [dB]                        | Worst average<br>SNR [dB] |                                       |                                                      |
| 1 (DEF) QPSK, 0.1 r.o., 2 spans, 1 fil. 135GHz | 15.12                                     | 23              | 15.48                       | 15.48                                           | 13.01                     | 2.47                                  | 0.36                                                 |
| 2 16QAM, 0.1 r.o., 2 spans, 1 fil. 135 GHz     | 15.65                                     | 17              | 16.13                       | 16.13                                           | 12.10                     | 4.03                                  | 0.48                                                 |
| 3 QPSK, 0.1 r.o., 2 spans, 1 fil. 150 GHz      | 15.40                                     | 43              | 16.15                       | 16.28                                           | 13.56                     | 2.59                                  | 0.75                                                 |
| 4 QPSK, 0.15 r.o., 2 spans, 1 fil. 135 GHz     | 15.20                                     | 9               | 15.48                       | 15.48                                           | 12.71                     | 2.77                                  | 0.28                                                 |
| 5 QPSK, 0.1 r.o., 10 spans, 2 fil. 135 GHz     | 8.56                                      | 24              | 9.06                        | 9.06                                            | 7.79                      | 1.27                                  | 0.50                                                 |
| 6 QPSK, 0.1 r.o., 0 span/B2B, 1 fil. 135 GHz   | 21.25                                     | 31              | 21.62                       | 21.62                                           | 15.55                     | 6.07                                  | 0.37                                                 |
| 7 QPSK, 0.1 r.o., 0 span/B2B, 1 fil. 200 GHz   | 23.29                                     | 36              | 23.33                       | 23.33                                           | 16.08                     | 7.25                                  | 0.04                                                 |
| 8 QPSK, 0.15 r.o., 0 span/B2B, 1 fil. 135 GHz  | 20.92                                     | 47              | 21.04                       | 21.04                                           | 16.12                     | 4.92                                  | 0.12                                                 |
| 9 QPSK, 0.15 r.o., 0 span/B2B, 1 fil. 200 GHz  | 22.86                                     | 43              | 22.86                       | 22.86                                           | 18.87                     | 3.96                                  | 0.00                                                 |

TABLE II  
RESULTS WHEN OPTIMIZING THE MINIMUM SUPERCHANNEL SUBCHANNEL SNR VALUE (OBJ #2) WITH A STEP SIZE 0.25 GHz

| Scenario                                       | Minimum SNR<br>equidistant<br>subch. [dB] | #<br>Iterations | Minimum<br>SNR (Obj#2)<br>[dB] | Brute-force $\pm 2$ GHz around opt.<br>solution |                        | Improvement<br>wrt worst<br>case [dB] | Improvement<br>wrt equi-<br>distanced<br>subch. [dB] |
|------------------------------------------------|-------------------------------------------|-----------------|--------------------------------|-------------------------------------------------|------------------------|---------------------------------------|------------------------------------------------------|
|                                                |                                           |                 |                                | Best min. SNR<br>[dB]                           | Worst min. SNR<br>[dB] |                                       |                                                      |
| 1 (DEF) QPSK, 0.1 r.o., 2 spans, 1 fil. 135GHz | 13.15                                     | 9               | 14.34                          | 14.34                                           | 10.61                  | 3.73                                  | 1.19                                                 |
| 2 16QAM, 0.1 r.o., 2 spans, 1 fil. 135 GHz     | 13.63                                     | 8               | 14.86                          | 14.86                                           | 10.03                  | 4.83                                  | 1.23                                                 |
| 3 QPSK, 0.1 r.o., 2 spans, 1 fil. 150 GHz      | 13.72                                     | 23              | 15.76                          | 15.78                                           | 10.61                  | 5.15                                  | 2.04                                                 |
| 4 QPSK, 0.15 r.o., 2 spans, 1 fil. 135 GHz     | 13.21                                     | 8               | 14.17                          | 14.18                                           | 10.50                  | 3.67                                  | 0.96                                                 |
| 5 QPSK, 0.1 r.o., 10 spans, 2 fil. 135 GHz     | 6.19                                      | 17              | 8.10                           | 8.10                                            | 6.60                   | 1.50                                  | 1.91                                                 |
| 6 QPSK, 0.1 r.o., 0 span/B2B, 1 fil. 135 GHz   | 18.97                                     | 10              | 20.08                          | 20.08                                           | 12.84                  | 7.24                                  | 1.11                                                 |
| 7 QPSK, 0.1 r.o., 0 span/B2B, 1 fil. 200 GHz   | 23.04                                     | 26              | 23.19                          | 23.19                                           | 15.89                  | 7.30                                  | 0.15                                                 |
| 8 QPSK, 0.15 r.o., 0 span/B2B, 1 fil. 135 GHz  | 18.68                                     | 10              | 19.41                          | 19.41                                           | 13.58                  | 5.83                                  | 0.73                                                 |
| 9 QPSK, 0.15 r.o., 0 span/B2B, 1 fil. 200 GHz  | 22.47                                     | 26              | 22.58                          | 22.66                                           | 18.64                  | 3.94                                  | 0.11                                                 |

SNR values obtained by employing a brute-force approach. We considered a  $\pm 2$  GHz interval around the optimal solutions found by the algorithm (i.e.,  $\mathbf{D}^*_{\text{obj1}}$  and  $\mathbf{D}^*_{\text{obj2}}$ ) and collected the SNR values related to all the possible distance combinations in such interval, with a frequency step size of  $f_{\text{step}}^{\text{BF}} = 1$  GHz. Considering an explored spectrum range  $\Delta F = 4$  GHz (i.e.,  $\pm 2$  GHz) and a  $f_{\text{step}}^{\text{BF}} = 1$  GHz, we have to configure and monitor the network  $[1 + (\Delta F / f_{\text{step}}^{\text{BF}})]^N$  times. So, for each considered scenario listed in Tables I and II, the total number of brute-force iterations was  $5^4$ . A smaller step, e.g., 0.25 GHz, would result in an extremely large number of combinations and prohibitive time, as also discussed later.

In a real network setting, the worst SNR value found in this frequency range would represent the result of a soft failure occurring while the network is operating (e.g.,  $\pm 2$  GHz drift of each of the lasers). This worst case should be considered in the planning phase with a margin for the SNR so as to select an adequate modulation format. Instead, with our solution, we reduce this need. Regarding the default scenario, with respect to the worst soft failure, we achieved improvements of 2.47 dB and 3.73 dB for the two considered objectives, respectively. The proposed method would run periodically or on demand when e.g., we observe through monitoring that the performance (SNR or BER) has fallen below a specified threshold. Thus, the proposed method requires a margin, which can, however, be set to e.g., 1 dB from the optimum, and still achieve considerable savings of 1.47 dB and 2.73 dB, respectively. Such savings

would reduce the network design cost, meeting the first goal of the proposed method as discussed in the introduction. Such margin can also take the form of a higher network resilience or an increased capacity, meeting the second and third goals, introduced in Section I. Note that the results discussed in the previous Section III.C correspond to improvements achieved over an equidistant superchannel, and do not account for possible soft failures.

From Tables I and II, we can see that the proposed method reached almost always the optimal SNR values identified by the brute-force approach. However, when this did not happen (i.e., for a certain initial starting point in the 10 simulation runs for scenario 9 in Table I), the found solution was almost equal to the best one. Note that the brute-force approach is unsuitable to be used in a real network, mainly for two reasons. Assuming a monitoring time of 5 minutes, this would take more than two days to complete, even for  $N = 4$  subchannels, not to mention having a finer granularity. Secondly, because it operates blindly, without considering the subchannel SNR values, would also configure the worst case, thus requiring the related margin. Comparing that with a few tens of monitoring calls (i.e.,  $FM$ , with  $M = 2$ ) used by the proposed solution, we see a clear benefit of employing our approach instead of the brute-force one.

### E. Larger Superchannels

To further assess the effectiveness of the proposed method, we also integrated the optimization algorithm in scenarios simulated

TABLE III  
RESULTS COMPARISON BETWEEN DIFFERENT SUBCHANNEL CONFIGURATIONS FOR OBJ#2 (STEP SIZE 0.5 GHz)

| Scenario               | Min SNR<br>equidistant<br>subch. [dB] | # Iterations | Min. SNR<br>(Obj#2)<br>[dB] | Brute-force $\pm 1$ GHz around<br>opt. solution |                        | Improvement<br>wrt worst case<br>[dB] | Improvement wrt<br>equidistant subch.<br>[dB] |
|------------------------|---------------------------------------|--------------|-----------------------------|-------------------------------------------------|------------------------|---------------------------------------|-----------------------------------------------|
|                        |                                       |              |                             | Best Min.<br>SNR [dB]                           | Worst Min.<br>SNR [dB] |                                       |                                               |
| 2 spans<br>0.1 r.o.    |                                       |              |                             |                                                 |                        |                                       |                                               |
| 4 subch.,<br>137.5 GHz | 13.15                                 | 4            | 14.28                       | 14.34                                           | 10.61                  | 3.67                                  | 1.13                                          |
| 6 subch.,<br>200 GHz   | 11.90                                 | 16           | 13.57                       | 13.57                                           | 9.53                   | 4.04                                  | 1.67                                          |
| 8 subch.,<br>275 GHz   | 13.07                                 | 11           | 14.40                       | 14.40                                           | 9.84                   | 4.56                                  | 1.33                                          |
| 10 subch.,<br>340 GHz  | 13.11                                 | 14           | 14.30                       | 14.44                                           | 9.81                   | 4.49                                  | 1.19                                          |

with six, eight, and ten subchannels. Apart from the TXs and RXs number, the rest of the setup was exactly as that depicted in Fig. 4 and the default scenario (i.e., scenario 1 in Tables I and II). The key difference was that we set the filter 3 dB bandwidth, to 200 GHz for  $N = 6$ , 275 GHz for  $N = 8$ , and 340 GHz for  $N = 10$  channels, while in each case we started with equally spaced subchannels. The minibatch size was set to  $M = N/2$  and we used a frequency step size equal to 0.5 GHz for all the considered configurations and objectives.

For both objectives, the achieved results always improved the figure of merit related to the starting conditions. Concerning Obj#1, we improved the superchannel average SNR value, with respect to the equally spaced scenario, by 0.29 dB, 0.18 dB, and 0.06 dB, for the cases of six, eight, and ten subchannels, respectively (the results for Obj#1 are not shown, due to size limitations). Note that the total SNR value of the superchannel should be multiplied by  $N$ . Moreover, for Obj#2, the improvements were equal to 1.67 dB, 1.33 dB, and 1.19 dB, respectively. We summarize the main results of the simulations related to Obj#2 in Table III, where for the sake of comparison, we also report the default scenario with four subchannels. Note that the number of iterations required to reach the optimal solution depends on the considered subchannel starting positions and thus it is not meaningful to compare the different cases. For instance, in Table III, the iterations needed to find Obj#2 optimum in the case of six subchannels were higher compared to those needed to reach the optimum with eight subchannels. In addition, we ran the brute-force approach around the optimal found solution only within an interval of  $\pm 1$  GHz with a step  $f_{step}^{BF} = 1$  GHz. This resulted in a total number of brute-force iterations equal to  $3^N$ . Such reduction in the interval was deemed necessary to complete the simulations in a reasonable time (i.e., within hours). This is also a further indication of the infeasibility of the brute-force approach in real scenarios. For Obj#1 and the three considered configurations (i.e., six, eight, and ten subchannels), when comparing our results to those related to a soft failure condition (i.e., the worst case found within the brute-force interval), we achieved improvements equal to 2.79 dB, 2.60 dB, and 1.82 dB, respectively (the results for Obj#1 are not shown due to size limitations). Furthermore, as per the previous four-subchannel scenario, also in these cases, our optimal found solutions almost always matched the brute-force found optima. The inaccuracies

were quite low, i.e., in the ten subchannels scenario with Obj#2, our solution was 0.14 dB far from the optimum found with the brute-force approach.

#### IV. CONCLUSION

Typically, superchannels are established with equidistant subchannels and a margin to account for laser drifts, filter misalignments, and wavelength-dependent effects. We proposed a closed control loop process that leverages the network feedback to correct the above issues, trace the real optimum, reduce the margin, and increase the efficiency of the superchannel. The proposed solution probes and monitors the network and then uses the stochastic subgradient method to maximize either i) the total capacity of the superchannel or ii) the minimum subchannel SNR value. We implemented and verified the proposed process through co-simulations in VPI and MATLAB. The obtained results showed that we were able to approach the optimum in all considered scenarios. Considering a superchannel with four sub-channels and a soft failure of  $\pm 2$  GHz subchannel frequency drifts, we would need a 2.47 dB and 3.73 dB margin, respectively. This can be reduced by employing the proposed method, yielding high savings in the network design, and/or higher network resilience, and/or increased capacity. The proposed method was shown to operate properly and achieve substantial savings for a higher number of subchannels.

#### REFERENCES

- [1] R. - J. Essiambre, G. Kramer, P. J. Winzer, G. J. Foschini, and B. Goebel, "Capacity limits of optical fiber networks," *J. Lightw. Technol.*, vol. 28, no. 4, pp. 662–701, Feb. 2010.
- [2] T. Richter et al., "1 Tb/s and 800 Gb/s real-time transmission at 138 Gbd over a deployed ROADM network with live traffic," in *Proc. Opt. Fiber Commun. Conf. Exhib.*, 2023, pp. 1–3.
- [3] M. Imran, P. M. Anandarajah, A. Kaszubowska-Anandarajah, N. Sambo, and L. Potí, "A survey of optical carrier generation techniques for terabit capacity elastic optical networks," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 211–263, Jan.–Mar. 2018.
- [4] Y. Pointurier, "Design of low-margin optical networks," *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. A9–A17, Jan. 2017.
- [5] C. Delezoide, P. Ramantanis, and P. Layec, "Leveraging field data for the joint optimization of capacity and availability in low-margin optical networks," *J. Lightw. Technol.*, vol. 38, no. 24, pp. 6709–6718, Dec. 2020, doi: [10.1109/JLT.2020.3022107](https://doi.org/10.1109/JLT.2020.3022107).

- [6] C. Delezoide, P. Ramantanis, and P. Layec, "Machine learning and data science for low-margin optical networks: The ins and outs of margin optimization," in *Machine Learning for Future Fiber-Optic Communication Systems*. Cambridge, MA, USA: Academic, 2022, pp. 281–315.
- [7] Nokia WaveSuite applications, available: [Online]. Available: <https://www.nokia.com/networks/optical-networks/wavesuite/>
- [8] P. Ramantanis, C. Delezoide, P. Layec, and S. Bigo, "Revisiting the calculation of performance margins in monitoring-enabled optical networks," *J. Opt. Commun. Netw.*, vol. 11, no. 10, pp. C67–C75, Oct. 2019.
- [9] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B60–B71, Apr. 2021.
- [10] T. Duthel, C. R. S. Fludger, J. Geyer, and C. Schulien, "Impact of polarisation dependent loss on coherent POLMUX-NRZ-DQPSK," in *Proc. IEEE Conf. Opt. Fiber Commun./Nat. Fiber Optic Engineers Conf.*, 2008, pp. 1–3, doi: [10.1109/OFC.2008.4528544](https://doi.org/10.1109/OFC.2008.4528544).
- [11] C. Delezoide, H. Akbari, P. Ramantanis, F. Boitier, and P. Layec, "Automated mitigation of quality of transmission fluctuations induced by PDL anomalies," in *Proc. Int. Conf. Opt. Netw. Des. Model.*, 2023, pp. 1–3.
- [12] J. M. Fabrega et al., "On the filter narrowing issues in elastic optical networks," *J. Opt. Commun. Netw.*, vol. 8, no. 1, pp. A23–A33, Jan. 2016.
- [13] C. Delezoide, P. Layec, and S. Bigo, "Automated alignment between channel and filter cascade," in *Proc. IEEE Opt. Fiber Commun. Conf. Exhib.*, 2019, pp. 1–4.
- [14] H. Hasegawa, S. Subramaniam, and K. Sato, "Node architecture and design of flexible waveband routing optical networks," *J. Opt. Commun. Netw.*, vol. 8, no. 1, pp. 734–744, Jan. 2016.
- [15] T. Zami, I. Ruiz, A. Ghazisaeidi, and B. Lavigne, "Growing impact of optical filtering in future WDM networks," in *Proc. IEEE Opt. Fiber Commun. Conf.*, 2019, pp. 1–3.
- [16] F. Paolucci et al., "Superfilter technique in SDN-controlled elastic optical networks," *J. Opt. Commun. Netw.*, vol. 7, no. 2, pp. A285–A292, Feb. 2015.
- [17] K. Hayashi, Y. Mori, and H. Hasegawa, "Cost-effective network capacity upgrade by heterogeneous wavelength division multiplexing density with bandwidth-variable virtual direct links," *J. Opt. Commun. Netw.*, vol. 15, no. 9, pp. D23–D32, Sep. 2023.
- [18] H. Lun et al., "Soft failure identification in optical networks based on convolutional neural network," in *Proc. IEEE 45th Eur. Conf. Opt. Commun.*, 2019, pp. 1–3.
- [19] C. Delezoide et al., "Marginless operation of optical networks," *J. Lightw. Technol.*, vol. 37, no. 7, pp. 1698–1705, Apr. 2019.
- [20] N. Sambo et al., "Field trial: Demonstrating automatic reconfiguration of optical networks based on finite state machine," *J. Lightw. Technol.*, vol. 37, no. 16, pp. 4090–4097, Aug. 2019.
- [21] K. Christodoulopoulos et al., "Toward efficient, reliable, and autonomous optical networks: The ORCHESTRA solution," *J. Opt. Commun. Netw.*, vol. 11, no. 9, pp. C10–C24, Sep. 2019.
- [22] P. Ramantanis, C. Delezoide, P. Layec, and S. Bigo, "Uncertainty aware real-time performance monitoring for elastic optical networks," in *Proc. IEEE Eur. Conf. Opt. Commun.*, 2021, pp. 1–4.
- [23] M. Jinno, "Elastic optical networking: Roles and benefits in beyond 100-Gb/s era," *J. Lightw. Technol.*, vol. 35, no. 5, pp. 1116–1124, Mar. 2017.
- [24] Q. Zhuge et al., "Building a digital twin for intelligent optical networks [invited tutorial]," *J. Opt. Commun. Netw.*, vol. 15, no. 8, pp. C242–C262, Aug. 2023.
- [25] W. Idler et al., "Field trial of a 1 Tb/s super-channel network using probabilistically shaped constellations," *J. Lightw. Technol.*, vol. 35, no. 8, pp. 1399–1406, Apr. 2017.
- [26] B. Collings, "New devices enabling software-defined optical networks," *IEEE Comm. Mag.*, vol. 51, no. 3, pp. 66–71, Mar. 2013.
- [27] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K., MA, USA: Cambridge Univ. Press, 2004.
- [28] N. Z. Shor, *Nondifferentiable Optimization and Polynomial Problems*, 1st ed., New York, NY, USA: Springer, 1998.
- [29] V. Patel and S. Zhang, "Stochastic gradient descent on nonconvex functions with general noise models," 2021. [Online]. Available: <https://arxiv.org/abs/2104.00423>
- [30] J. R. R. A. Martins and J. T. Hwang, "Review and unification of methods for computing derivatives of multidisciplinary computational models," *AIAA J.*, vol. 51, pp. 2582–2599, 2013.
- [31] R. Tibshirani, R. Chen, D. Schwartz, and S. Yao, "Convex optimization," 2015. [Online]. Available: <http://www.stat.cmu.edu/~ryantibs/convexopt-F15/lectures/07-sg-method.pdf>
- [32] F. Paolucci, A. Sgambelluri, F. Cugini, and P. Castoldi, "Network telemetry streaming services in SDN-based disaggregated optical networks," *J. Lightw. Technol.*, vol. 36, no. 15, pp. 3142–3149, Aug. 2018.
- [33] G. Bosco, V. Curri, A. Carena, P. Poggiolini, and F. Forghieri, "On the performance of nyquist-WDM terabit superchannels based on PM-BPSK, PM-QPSK, PM-8QAM or PM-16QAM subcarriers," *J. Lightw. Technol.*, vol. 29, no. 1, pp. 53–61, Jan. 2011.
- [34] A. Mahajan, K. Christodoulopoulos, R. Martinez, R. Munoz, and S. Spadaro, "Quality of transmission estimator retraining for dynamic optimization in optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B45–B59, Apr. 2021.
- [35] VPIphotonics, 2020. [Online]. Available: <https://vpiphotonics.com>