Maximizing Fiber Cable Capacity Under A Supply Power Constraint Using Deep Neural Networks

Junho Cho[1*], Sethumadhavan Chandrasekhar[1], Erixhen Sula[2], Samuel Olsson[3], Ellsworth Burrows[1], Greg Raybon[1], Roland Ryf[1], Nicolas Fontaine[1], Jean-Christophe Antona[4], Steve Grubb[5], Peter Winzer[1], and Andrew Chraplyvy[1]

[1]Nokia Bell Labs, Holmdel, NJ, USA, [2]EPFL, Lausanne, Switzerland, [3]Nokia Corporation, Murray Hill, NJ, USA,
[4]Alcatel Submarine Networks, Nozay, France, [5]Facebook, Menlo Park, CA, USA
**email: junho.cho@nokia-bell-labs.com*

Abstract: We experimentally achieve a 19% capacity gain per Watt of electrical supply power in a 12-span link by eliminating gain flattening filters and optimizing launch powers using deep neural networks in a parallel fiber context.

1. Introduction

Massive spatial parallelism maximizes capacity and minimizes cost/bit of submarine optical cables under an electrical supply power (ESP) constraint [1]–[4]. The resulting optical power dilution among parallel fibers pushes transmission from nonlinearly-optimum launch powers into the linear regime. The logarithmic reduction in spectral efficiency resulting from a lower delivered optical signal-to-noise ratio (OSNR) per fiber is linearly over-compensated by the increased spatial multiplicity of the cable, yielding a higher total cable capacity [1], [2]. The capacity C per Watt of ESP $\mathcal{P}_E$ (both per spatial path) becomes a key figure of merit in such systems [4]: $m = C/\mathcal{P}_E$. This new metric asks for revisiting such fundamental topics as *(i)* the need for gain-flattening filters (GFFs) in optical amplifiers (as GFFs, used universally in submarine systems today, are lossy and hence waste ESP, thus potentially reducing m), and *(ii)* the optimum optical channel power allocation strategy. We address both topics and, on a 12-span 744-km straight-line system, experimentally achieve a gain of 19% in m by removing GFFs. Higher gains are expected for longer links and for pump-sharing architectures across amplifier arrays.

Predicting the received (RX) signal and noise powers from arbitrary transmit (TX) power profiles through a chain of gain-unflattened optical amplifiers is difficult, as a small change in the TX power spectral density (PSD) or in the spectral link characteristics causes a complicated evolution of signal and noise powers, making it intractable to computationally solve the problem using physics-based optical amplifier models. We therefore resort to machine learning [5] and build a deep neural network (DNN) as a digital twin of our optical fiber link. When trained with experimental link data, the DNN allows for an off-line gradient-descent (GD) optimization whose optimized results are verified experimentally.

2. Experimental Methodology and Setup

Massively parallel submarine cables will operate at low-enough optical signal powers to neglect fiber nonlinearities [2], and probabilistic constellation shaping allows to finely adapt each transponder to the SNR of a given wavelength channel [6]. This lets the delivered OSNR be a good basis for estimating polarization- and wavelength-division multiplexed (WDM) system capacities as $C = 2R_s \sum_{k=1}^{K} \log_2(1 + \eta\, SNR_k)$, where SNR_k is the OSNR of the k-th of K WDM channels (normalized to one polarization and a reference bandwidth equal to the symbol rate R_s), and $\eta \leq 1$ accounts for transponder implementation penalties. We use $\eta = 1$ in this paper without loss of generality.

In order to determine SNR_k, we use the WDM channel emulation method shown in Fig. 1(a): Amplified spontaneous emission (ASE) from an Erbium-doped fiber amplifier (EDFA) is filtered by a wavelength selective switch (WSS) to generate 40 slots of 50-GHz ASE (emulating 40 signal channels, as is customary in WDM experiments [7], [8]), interleaved with 39 empty 50-GHz slots. For each channel, SNR_k can then be estimated by an optical spectrum analyzer (OSA) taking the ratio of emulated signal power S_k to ASE power N_k, interpolated between two empty slots, cf. inset to Fig. 1(b). The WSS output is boosted by a TX EDFA and attenuated by a variable optical attenuator (VOA) to produce a set of optical launch powers $\boldsymbol{P}_{1:40}$; we use the notation $\boldsymbol{X}_{1:K} \coloneqq [X_1, \ldots, X_K]$ throughout the paper. As an example, Fig. 1(b) shows a flat TX PSD across a 4-THz system bandwidth and the resulting RX PSD after 12 spans of 62-km Corning® Vascade® EX3000

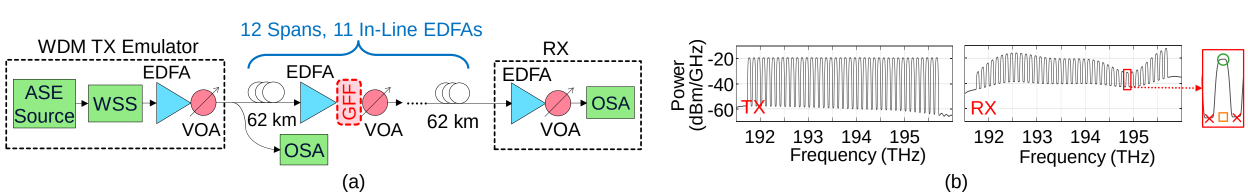

Fig. 1. (a) Experimental setup, (b) measured optical spectra at TX and RX OSAs.

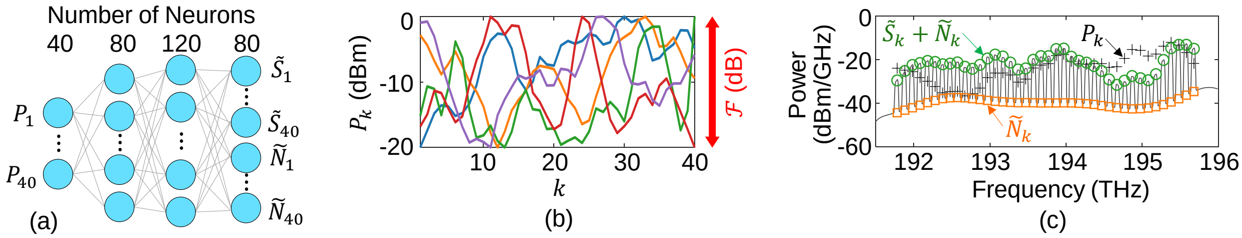

Fig. 2. (a) Structure of the DNN, (b) five random TX power profiles $\boldsymbol{P}_{1:40}$ with $\mathcal{F} = 20$ dB, and (c) an example of TX powers P_k (black pluses), measured RX PSDs (solid lines), and DNN-predicted signal+noise $\tilde{\boldsymbol{S}}_k + \tilde{\boldsymbol{N}}_k$ (green circles) and noise $\tilde{\boldsymbol{N}}_k$ (orange squares).

fiber with 0.16-dB/km loss, padded by VOAs to a span loss of 16.5 dB in order to operate our 744-km straight-line system in a lower-OSNR regime pertinent to the targeted massively parallel submarine application [1], [2]. Since launch powers are low and fiber nonlinearities are negligible (as quantified below), padding at the beginning of a span is equivalent to padding at the end. Each span is followed by a custom-designed single-stage EDFA with a removable GFF.

3. Training the DNN with Experimental Data

Since each of the signal and noise powers $(\boldsymbol{S}_{1:40}, \boldsymbol{N}_{1:40})$ at the receiver depends on *all* TX powers $\boldsymbol{P}_{1:40}$ in a way that is difficult to accurately model based on amplifier physics, we resort to machine learning and construct a DNN as a digital twin of our experimental link. As shown in Fig. 2(a), our DNN has 40 input neurons $(\boldsymbol{P}_{1:40})$, two hidden layers with 80 and 120 neurons each, and 80 output neurons for the predicted signal and noise powers $(\tilde{\boldsymbol{S}}_{1:40}, \tilde{\boldsymbol{N}}_{1:40})$ at the output of the link. Linear, sigmoid, and softplus activation functions [9] are used. Numbers of neurons and activation functions are chosen to minimize the mean absolute error (MAE) [10] of measurement $(\boldsymbol{S}_{1:40}, \boldsymbol{N}_{1:40})$ and prediction $(\tilde{\boldsymbol{S}}_{1:40}, \tilde{\boldsymbol{N}}_{40})$.

The DNN training process starts by configuring one of two link setups (i.e., with and without GFFs) and choosing one of three ESPs $\mathcal{P}_E$, considering *electrical pump powers* and ignoring less fundamental overheads from amplifier control. The overall ESP is spread evenly across all 11 in-line EDFAs such that the optical output power $\mathcal{P}_O = \sum_{k=1}^{40} P_k$ is equal for all EDFAs. The TX VOA is adjusted to also provide $\mathcal{P}_O$ at the TX. The EDFAs, when operated with GFFs, have a gain ripple $<$ 2.5 dB across the 4-THz amplification band for all chosen operating conditions and operate at electrical-to-optical power conversion efficiencies of 3.1, 8.2, and 9.9% (measured after the GFF) for $\mathcal{P}_E$ =1.09, 2.27, and 7.53 W. Next, we measure $\boldsymbol{S}_{1:40}$ and $\boldsymbol{N}_{1:40}$ for 1440 randomly generated $\boldsymbol{P}_{1:40}$ at fixed $\mathcal{P}_E$, and with a peak-to-peak channel power excursion $\mathcal{F} = \max_{i,j}(|P_i - P_j|)$ that we gradually increase from 6 dB to 45 dB; 5 representative instances of TX signal powers with $\mathcal{F} =$ 20 dB are depicted in Fig. 2(b). We avoid implausibly fast changes of P_k over a narrow frequency range by applying a moving average to each TX power profile and ensure that the 1440 random power profiles uniformly fill the frequency-power rectangle. Of the 1440 recorded data sets, 90% are used for training and 10% for validation of the DNN. For all 6 test cases, the DNN rapidly converges with a minuscule MAE difference between training and validation sets, indicating the absence of overfitting [11]. Figure 2(c) shows an example of TX powers $\boldsymbol{P}_{1:40}$(black pluses), and the measured RX PSD (black line; with $\mathcal{P}_E = 2.27$ W, no GFFs). The DNN-predicted $\tilde{\boldsymbol{S}}_{1:40}$+$\tilde{\boldsymbol{N}}_{1:40}$, $\tilde{\boldsymbol{N}}_{1:40}$ (green circles, orange squares) show excellent agreement between measurement and prediction. The MAE of our DNN-predicted channel SNRs is $\leq$ 0.4 dB across a wide range of $\mathcal{F}$; for flat TX signal powers (i.e., $\mathcal{F} = 0$), our DNN yields mean and maximum absolute errors of 0.2 dB and 0.5 dB in SNR prediction, while the OASIX, a state-of-the-art EDFA design software [12], produces those of 1.9 dB and 4.7 dB, respectively.

4. Capacity Maximization and Verification

We next perform gradient descent (GD) capacity maximization off-line based on the trained DNN, cf. Fig. 3(a). The result is a capacity-maximizing TX power profile $\boldsymbol{P}_{1:40}$. Figure 3(b) shows three example optimizations ($\mathcal{P}_E = 2.27$ W, no GFFs), one starting from a flat $\boldsymbol{P}_{1:40}$ (blue) and the other two from initial conditions with poorer capacity. Irrespective of the starting condition, the DNN converges to the same optimized system capacity, a fact that is even more impressively shown in Fig. 3(c), giving initial (blue crosses) and converged (orange dots) capacities for *all* 1440 randomly chosen initial TX power profiles with varying $\mathcal{F}$ (red dots). Converged SNRs are shown in Fig. 3(d) across the system bandwidth. The converged SNR distribution is flat to within 4 dB in most cases, with minimal capacity variations between these converged solutions. The capacity for a completely flat RX SNR is 25.6 Tb/s, which is close to the experimental optimum of 25.9 Tb/s. The capacity of a flat TX signal power profile is 24.8 Tb/s and that of a flat RX signal power profile is 24.5 Tb/s. This is in contrast to the findings of Ref. [13] for systems using GFFs, where all three power profiles yield about the same capacity. Importantly, our approach does *not* subjectively favor any capacity optimization strategy based on possibly misguiding intuition, but blindly optimizes TX signal powers for maximum capacity. Note that an *experimental* GD solution is uniquely enabled by our DNN approach, as estimating only a single gradient requires 41 measurements of 4-THz RX PSDs. As in

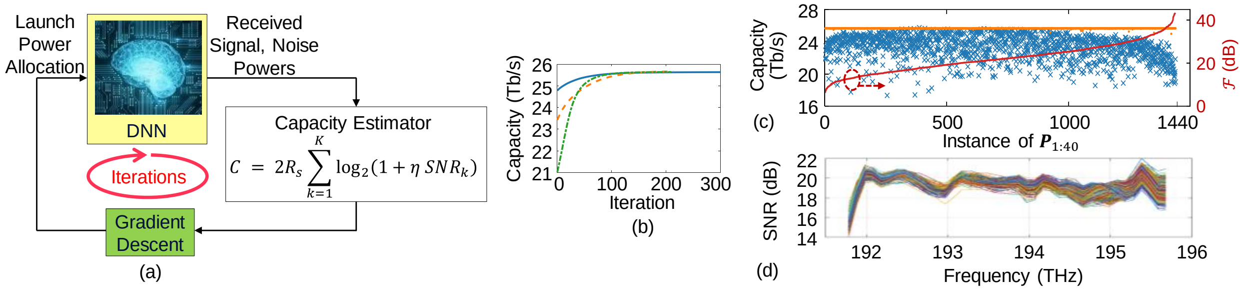

Fig. 3. (a) Optimization using DNN and GD and (b) convergence of the GD for 3 representative initial conditions; (c) capacity of the initial (blue crosses) and converged (orange dots) TX power profiles and (d) optimized RX SNRs, both for all 1440 random TX power profiles.

our fully automated system we measure 180 RX PSDs per hour, it would require >11 years to perform the full GD optimization for 1440 TX power profiles with 300 GD iterations! On the other hand, using the DNN approach, the optimization process takes only 9 hours. This >10,000× speed-up impressively reveals the power of machine learning in this application.

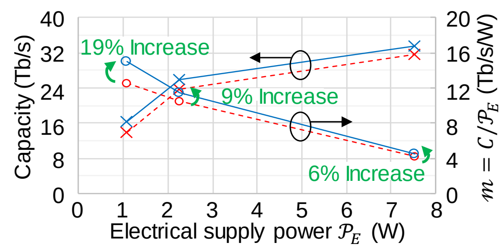

Fig. 4. Optimized C (crosses) and m (circles) as a function of the total electric pump power in systems with (dashed red) and without (solid blue) GFFs.

As a last step, we validate the results of the DNN-based GD optimization by loading the optimized TX power profiles onto the experimental system and measuring the RX power profiles. The capacity predicted by the DNN is within a 1.1% error of the experimentally measured capacity in all test cases. Figure 4 shows the actually measured capacity C (left axis) and the figure of merit m (right axis). Dashed red lines represent systems with GFFs and solid blue lines without GFFs, all with optimized TX power allocations. The experimental results show that: *(i)* systems without GFFs achieve a higher power efficiency than systems with GFFs, *(ii)* m increases with decreasing $\mathcal{P}_E$ until the EDFA pump current approaches the pump's lasing threshold, even at a significantly reduced EDFA power conversion efficiency of only 3.1% at that operating point; operating the pumps at higher power (and hence at higher efficiencies) and sharing them across multiple EDFAs will further increase m; and *(iii)* when the system operates at maximum efficiency (at largest m), both C and m can be increased by 19% *by eliminating GFFs from the system*. For the most energy-efficient case of $\mathcal{P}_E = 1.09$ W, we also verify that the maximum channel power anywhere along the link for the optimized power profile is below -4 dBm per 50 GHz, both with and without GFFs. This quantitatively justifies neglecting fiber Kerr nonlinearities, as this system operates 7 dB below a deployed 5,500-km cable that shows nonlinear peak performance at 3 dBm per 50 GHz [8].

5. Conclusion

We used experimental signal and noise data from a 12-span 744-km straight-line EDFA link to train a DNN as a digital twin of the experimental system. The DNN accurately predicts RX signal and noise powers for arbitrary TX signal powers. A gradient descent-based TX power profile optimization performed on the DNN is about 10,000× faster than what would be possible using measurements alone, and objectively optimizes TX power profiles. In the context of a massively parallel ESP-constrained system (submarine optical cables), we demonstrate a 19% capacity improvement by removing GFFs.

We acknowledge Amonics Corp for supplying the EDFAs with removable GFFs, and Corning for the loan of the EX3000 fiber used in this experiment.

References

[1] O. V. Sinkin et al., "SDM for power-efficient undersea transmission," *J. Lightw. Technol.* **36**(2), 361–371 (2017).
[2] R. Dar et al., "Cost-optimized submarine cables using massive spatial parallelism," *J. Lightw. Technol.* **36**(18), 3855–3865 (2018).
[3] P. Pecci and O. Courtois, "SDM: A revolution for the submarine industry," *Submarine Telecoms Forum* 106, 38–41 (2019).
[4] A. Turukhin et al., "Power-efficient transmission using optimized C+L EDFAs with 6.46 THz bandwidth and optimal spectral efficiency," in *Proc. ECOC*, Rome, Italy, Sep. 2018, Paper M.4.G.4.
[5] D. Zibar et al., "Machine learning techniques in optical communication," *J. Lightw. Technol.* **34**(6), 1442–1452 (2016).
[6] J. Cho and P. J. Winzer, "Probabilistic constellation shaping for optical fiber communications," *J. Lightw. Technol.* **37**(6), 1590–1607 (2019).
[7] D. J. Elson et al., "High spectral density transmission emulation using amplified spontaneous emission noise," *Opt. Lett.* **41**(1), 68–71 (2016).
[8] J. Cho et al., "Trans-Atlantic field trial using high spectral efficiency probabilistically shaped 64-QAM and single-carrier real-time 250-Gb/s 16-QAM," J. Lightw. Technol. **36**(1), 103–113 (2018).
[9] P. Ramachandran et al., "Searching for activation functions," arXiv preprint, 2017, arXiv:1710.05941.
[10] K. Taji et al., "On error backpropagation algorithm using absolute error function," in *Proc. ISMC*, Tokyo, Japan, Oct. 1999, pp. 401-406.
[11] N. Srivastava et al., "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.* **15**(1) 1929–1958 (2017).
[12] OFS, "RightWave erbium-doped optical fiber: Critical component for the amplification of optical transmission signals," Sep. 2018.
[13] A. Kam et al., "Pre-emphasis-based equalization strategy to maximize cable performance," in *Proc. SubOptic*, New Orleans, LA, USA, Apr. 2019, Paper OP14-1.