A Machine Learning Approach for Dynamic Optical Channel Add/Drop Strategies that Minimize EDFA Power Excursions

Yishen Huang[(1)], Wiem Samoud[(1)(2)], Craig L. Gutterman[(1)], Cédric Ware[(2)]
Mounia Lourdiane[(3)], Gil Zussman[(1)], Payman Samadi[(1)], Keren Bergman[(1)]

[(1)]Department of Electrical Engineering, Columbia University, New York, NY,10027, USA
[(2)]LTCI, CNRS, Télécom ParisTech, Université Paris Saclay, Paris 75013, France
[(3)]Télécom SudParis, Université Paris Saclay, Evry 91011, France
Email: y.huang@columbia.edu

Abstract *We demonstrate a machine learning approach to characterize channel dependence of power excursions in multi-span EDFA networks. This technique can determine accurate recommendations for channel add/drop with minimal excursions and is applicable to different network designs.*

Introduction

The advent of dynamic optical networks demands constant agility and configurability in response to traffic and fault handling[1]. As networks grow in both capabilities and complexity, the concept of cognitive networks – systems that can autonomously monitor, optimize, and adapt – is particularly promising to improve the networks' management and resilience[2]. As a crucial component of modern optical transport networks, Erbium Doped Fiber Amplifier (EDFA) has the ability to achieve economic regeneration of dense wavelength-division multiplexing (DWDM) channels and extend the reach of optical communication beyond the confines of cities and continents.

However, EDFA systems face an unsolved challenge of channel-dependent power excursions[1]. Modern EDFA systems employ automatic gain control (AGC) to maintain the output power levels of the amplifier within a tolerant regime[3]. In cascaded EDFA networks, upstream AGC ensures appropriate optical power levels for downstream amplifiers and receivers. However, AGC maintains the global mean gain, while each channel sees a wavelength dependent gain. If a channel with high gain is added, AGC responds to an increase in mean gain by reducing the gain on all channels. This leads to the high-gain channel effectively stealing power from lower-gain channels[4]. Conversely, adding a low-gain channel feeds power to higher-gain channels[1]. In both cases, the power excursion increases the disparity among channel powers; this discrepancy may be further exacerbated by downstream EDFA spans. We thus define undesired power excursions as the ones that increase the standard deviations (STD) of the output power levels.

Proposed solutions and limitations

The characteristics of the excursion depend on the types of EDFAs, the gain-control mechanisms, and the number of EDFA spans and light paths (LP); therefore, it is difficult to derive an analytical description that applies to all systems. Consequently, past proposed solutions focused on fully characterizing a specific EDFA system and reducing excursions by optimizing input power levels[4], balancing input channels[1], or adjusting the pumping level of the amplifier[3]. These techniques, while effective on the specific systems analyzed, are not necessarily transferrable to different networks. They also rely on the deterministic model of the gain profile, which is difficult to acquire for live-network equipment that cannot be disrupted.

Preventative approaches such as optimized wavelength assignment algorithms have been shown to reduce the excursions[5], but at the tradeoff of spectral efficiency. Case-based reasoning (CBR) is also applied to make heuristic guesses on EDFA tuning[6], but it requires a large number of past records to be effective. We present an efficient, low-overhead machine learning (ML) engine to characterize the channel dependence of power excursions in multi-span EDFA links. Historical snapshots of the network are collected and mathematically generalized. Once the ML model is trained, it is able to predict the best practices of channel add/drop to alleviate undesired excursions. The approach is non-disruptive and applies to EDFA networks of different designs. Fig. 1 illustrates the functionalities of the ML engine.

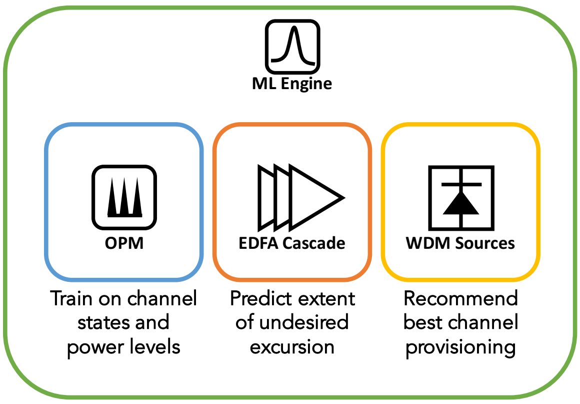

Fig. 1: Functionalities of the ML engine for minimizing undesired power excursions.

Experiment design

We construct the multi-span AGC-enabled EDFA network shown in Fig. 2. The WDM sources transmit 24 DWDM channels from ITU-T grid Ch. 21 to Ch. 44 with 100 GHz spacing, which are combined via a wavelength-selective switch (WSS). We adjusted the variable optical attenuators (VOA) and the EDFA pumping to achieve a realistic gain ripple with a maximum disparity of 6dB between the highest and lowest gains. To ensure an adequate number of channels to measure the excursions via power STD, as well as adequate number of available channel for add/drop, we maintain 10-20 ON channels at any given time, which corresponds to a spectral utilization between 42% and 83%.

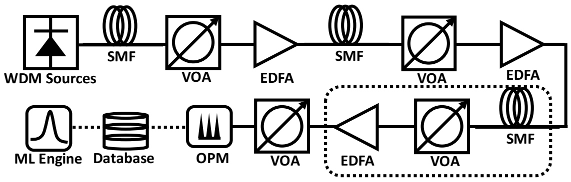

Fig. 2: Setup of multi-span EDFA network; the components in dashed box are included for the 3-span link.

Single-mode fibers (SMF) and VOAs are placed before each span's EDFA to emulate a 20dB transmission loss. The output power levels are recorded with a C-band optical performance monitor (OPM), and stored in a database for analysis. The ML engine ingests the power levels and channel ON/OFF states, and constructs a kernelized Bayesian regression (KBR) model for future predictions.

Machine learning and statistical analysis

We define a regression problem with supervised ML to statistically model the channel dependence of EDFA power excursions. The input is represented by a 24-bit array, each indicating an ON channel as 1, or an OFF channel as 0. This can be scaled up to 40-bit or 80-bit to accommodate the full DWDM C-band. The output is the power STD of the ON channels after the EDFA spans. We collected historical channel ON/OFF states and power STD values, which are split into a training set and a testing set. The regression model learns from the training set, and is evaluated against the testing set. The accuracy of the model is evaluated by two metrics: A) the mean square error (MSE) between the predicted and the measured STD, and B) correctness of the best channel provisioning identified based on the predictions.

The input and output values are preprocessed before training and testing according to Eq. (1) and (2). The DC bias is removed from each dimensions of the input and the output. The input dimensions are also standardized with an STD of unity. The prediction process takes in standardized inputs and returns offset outputs.

$$x'_{ij} = \frac{x_{ij}-\bar{x}_j}{\sigma_j}, \qquad (1)$$

$$y'_{ij} = y_{ij} - \bar{y}_j, \qquad (2)$$

where $i = 1..n$ for n total data points; $j = 1..d$ labels a dimension of input ($d = 24$) or output ($d = 1$); σ_j is the per-dimension STD of the input, and $\bar{x}_j$, $\bar{y}_j$ are respective per-dimension means.

Two network states with similar ON/OFF channels are assumed to have similar extents of power excursions[6]. We leverage this aspect with an input kernel, which can efficiently replace the need for a database, as is the case of CBR. Given a new network state, we can infer its predicted power STD from how similar it is to the known network states. Specifically, we construct a kernel called the Radial Basis Function (RBF) shown in Eq. (3).

$$K(x, x') = a \exp\left\{-\frac{1}{b}\|x - x'\|^2\right\}, \qquad (3)$$

where a and b are arbitrary factors to adjust the strength of the kernel, which we set as 0.0001 and 3.5 respectively for our analysis using cross-validation methods. x and x' are two 24D inputs; the value of the kernel function decreases exponentially with the L2 distance between the two inputs. The predictions are obtained from the linear combinations of the training outputs weighted by the kernel function values[7]. Fig. 3 compares the prediction MSE between KBR and linear regression (LR). It shows that KBR's predictions have lower errors when converged.

We study the adaptability of the model on two different networks – one with two EDFA spans and one with three EDFA spans. Fig. 3 shows, for both networks, the model's prediction accuracy improves with increasing size of the training data set; the improvement levels off after 400 training data points. This shows the training data does not need to grow much beyond this size to leverage the ML engine.

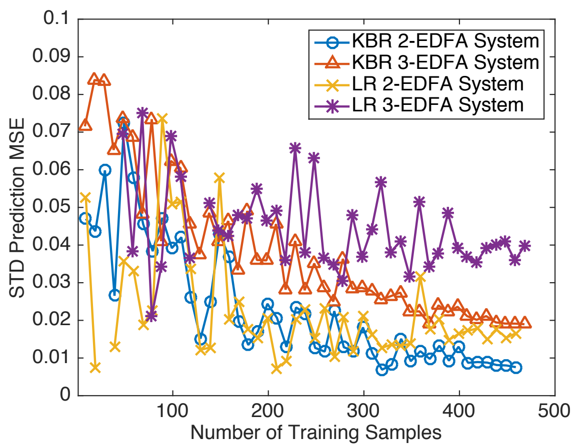

Fig. 3: Prediction MSEs converge after 400 training data points; KBR has lower converged MSEs than LR in both 2-EDFA and 3-EDFA networks.

Tab. 1: Training set size and KBR prediction MSE for two EDFA network designs.

Networks	2-EDFA Spans	3-EDFA Spans
Training set size	459	468
Prediction MSE	0.0076	0.019

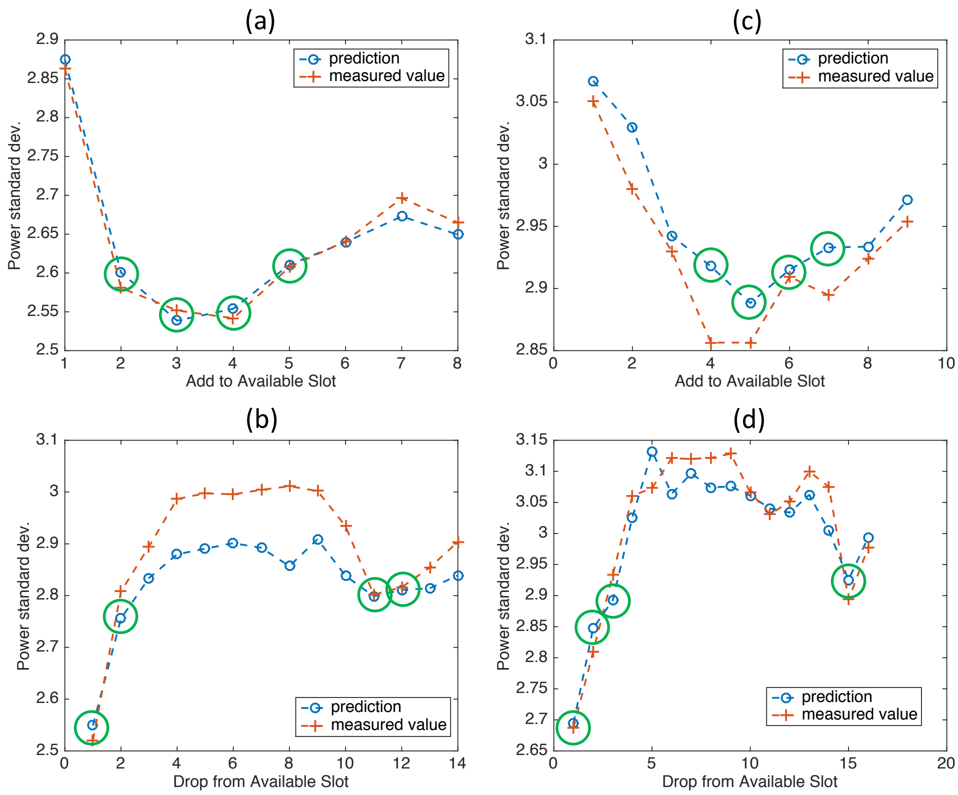

Fig. 4: Comparisons between predictions and measurements for single channel add/drop in 2-span network (a,b) and 3-span network (c,d). We recommend the top four slot candidates based on minimum predicted power STDs for adding a channel (a,c) and dropping a channel (b,d); the recommended slots circled in green agree with the best slots based on measured values.

Channel add/drop recommendations

We demonstrate the model's prediction accuracy and recommendation correctness in Fig. 4 for two network scenarios. The training set sizes and the MSE values of both are shown in Tab. 1. In each, we examine how the model can facilitate channel provisioning to reduce undesired power excursions. In all four cases shown in Fig. 4, we start with a randomly generated initial state of ON/OFF channels. For adding a channel, as shown in Fig. 4(a,c), the model predicts the power STD after a hypothetical channel is added to one of the available slots. Then the model recommends the best slots to add a channel that will result in the lowest power STD and therefore the least undesired power excursions. Recommending multiple slots provides educated guesses with flexibility for network operators. We perform the actual channel additions over the span of a week to verify the accuracy of the predictions and their tolerance to system variations over time. This test is repeated for dropping a channel, shown in Fig. 4(b,d). In all four tests shown, the slots recommended by the ML engine correctly align with the best slots from the actual measurements. Since the recommendations are based on the relative STD ranking, they are robust under slight deviations of the exact STD values predicted.

Conclusion

We introduce a machine learning engine to characterize the channel dependence of power excursions in a WDM network with multiple EDFA spans. A KBR model with RBF is trained with past channel states and power STDs. We show experimentally that it can give accurate recommendations on channel add/drop strategies to minimize undesired power excursions, and is applicable to different EDFA network designs.

Acknowledgement

This project is supported by CIAN NSF ERC under grant EEC-0812072 and NSF CNS-1423105. The authors would also like to thank AT&T Labs for their generous donation of the equipment.

References

[1] A. S. Ahsan et al., "Excursion-Free Dynamic Wavelength Switching in Amplified Optical Networks," J. Opt. Commun. Netw., Vol. **7**, no. 9, p. 898 (2015).

[2] I. de Miguel et al, "Cognitive Dynamic Optical Networks," J. Opt. Commun. Netw., Vol. **5**, no. 10, p. A107 (2013).

[3] D. A. Mongardien, "Managing Channels Add/Drop in Flexible Networks Based on Hybrid Raman/Erbium Amplified Spans," Proc. ECOC (2006).

[4] P. J. Lin, "Reducing Optical Power Variation in Amplified Optical Network," Proc. ICCT (2003).

[5] K. Ishii et al., "Wavelength Assignment Dependence of AGC EDFA Gain Offset under Dynamic Optical Circuit Switching," Proc. OFC, W3E.4 (2014).

[6] U. Moura et al., "SDN-enabled EDFA Gain Adjustment Cognitive Methodology for Dynamic Optical Networks," Proc. ECOC, (2015).

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer (2006).