Dynamic mitigation of EDFA power excursions with machine learning

Yishen Huang, Craig L. Gutterman, Payman Samadi, Patricia B. Cho, Wiem Samoud, Cédric Ware, Mounia Lourdiane, Gil Zussman, Keren Bergman

Research Article | Vol. 25, No. 3 | 6 Feb 2017 | OPTICS EXPRESS 2245

Optics EXPRESS

Dynamic mitigation of EDFA power excursions with machine learning

YISHEN HUANG,[1,*] CRAIG L. GUTTERMAN,[1] PAYMAN SAMADI,[1] PATRICIA B. CHO,[1] WIEM SAMOUD,[2,3] CÉDRIC WARE,[2] MOUNIA LOURDIANE,[3] GIL ZUSSMAN,[1] AND 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*
[*]*y.huang@columbia.edu*

Abstract: Dynamic optical networking has promising potential to support the rapidly changing traffic demands in metro and long-haul networks. However, the improvement in dynamicity is hindered by wavelength-dependent power excursions in gain-controlled erbium doped fiber amplifiers (EDFA) when channels change rapidly. We introduce a general approach that leverages machine learning (ML) to characterize and mitigate the power excursions of EDFA systems with different equipment and scales. An ML engine is developed and experimentally validated to show accurate predictions of the power dynamics in cascaded EDFAs. Recommended channel provisioning based on the ML predictions achieves within 1% error of the lowest possible power excursion over 94% of the time. We also showcase significant mitigation of EDFA power excursions in super-channel provisioning when compared to the first-fit wavelength assignment algorithm.

OCIS codes: (060.4510) Optical communications; (060.2320) Fiber optics amplifiers and oscillators; (060.4264) Networks, wavelength assignment.

References and links

1. A. S. Ahsan, C. Browning, M. S. Wang, K. Bergman, D. C. Kilper, and L. P. Barry, "Excursion-free dynamic wavelength switching in amplified optical networks," J. Opt. Commun. Netw. **7**(9), 898–905 (2015).
2. M. Fiorani, P. Samadi, Y. Shen, L. Wosinska, and K. Bergman, "Flexible architecture and control strategy for metro-scale networking of geographically distributed data centers," in *European Conference and Exhibition on Optical Communication* (2016).
3. M. Jinno, H. Takara, B. Kozicki, Y. Tsukishima, Y. Sone, and S. Matsuoka, "Spectrum-efficient and scalable elastic optical path network: architecture, benefits, and enabling technologies," IEEE Commun. Mag. **47**(11), 66–73 (2009).
4. P. Samadi, K. Wen, J. Xu, and K. Bergman, "Software-defined optical network for metro-scale geographically distributed data centers," Opt. Express **24**(11), 12310–12320 (2016).
5. P. Samadi, J. Xu, K. Wen, H. Guan, Z. Li, and K. Bergman, "Experimental demonstration of converged inter / intra data center network architecture," in *17th International Conference on Transparent Optical Networks* (2015).
6. D. C. Kilper, M. Bhopalwala, H. Rastegarfar, and W. Mo, "Optical power dynamics in wavelength layer software defined networking," in *Advanced Photonics* (2015).
7. A. K. Srivastava, Y. Sun, J. L. Zyskind, and J. W. Sulhoff, "EDFA transient response to channel loss in WDM transmission system," IEEE Photonics Technol. Lett. **9**(3), 386–388 (1997).
8. C. Tian and S. Kinoshita, "Analysis and control of transient dynamics of EDFA pumped by 1480- and 980-nm lasers," J. Lightwave Technol. **21**(8), 1728–1734 (2003).
9. D. A. Mongardien, S. Borne, C. Martinelli, C. Simonneau, and D. Bayart, "Managing channels add/drop in flexible networks based on hybrid raman / Erbium amplified spans," in *European Conference and Exhibition on Optical Communication* (2006).
10. E. A. Barboza, C. J. A. Bastos-filho, J. F. Martins-Filho, U. C. de Moura, and J. R. F. de Oliveira, "Self-adaptive Erbium-doped fiber amplifiers using machine learning," in *International Microwave & Optoelectronics Conference* (2013).
11. P. J. Lin, "Reducing optical power variation in amplified optical network," in *International Conference on Communication Technology* (2003).
12. N. Sambo, F. Cugini, G. Bottari, P. Iovanna, and P. Castoldi, "Routing and spectrum assignment for super-channels in flex-grid optical networks," in *European Conference and Exhibition on Optical Communication*

#281576 https://doi.org/10.1364/OE.25.002245
Journal © 2017 Received 28 Nov 2016; revised 19 Jan 2017; accepted 20 Jan 2017; published 27 Jan 2017

Optics EXPRESS

(2012).
13. J. Junio, D. C. Kilper, and V. W. S. Chan, "Channel power excursions from single-step channel provisioning," J. Opt. Commun. Netw. **4**(9), A1–A7 (2012).
14. K. Ishii, J. Kurumida, and S. Namiki, "Wavelength assignment dependency of AGC EDFA gain offset under dynamic optical circuit switching," in *Optical Fiber Communication Conference* (2014).
15. I. de Miguel, R. J. Durán, T. Jiménez, N. Fernández, J. C. Aguado, R. M. Lorenzo, A. Caballero, I. T. Monroy, Y. Ye, A. Tymecki, I. Tomkos, M. Angelou, D. Klonidis, A. Francescon, D. Siracusa, and E. Salvadori, "Cognitive dynamic optical networks," J. Opt. Commun. Netw. **5**(10), A107–A118 (2013).
16. Y. Huang, W. Samoud, C. L. Gutterman, C. Ware, M. Lourdiane, G. Zussman, P. Samadi, and K. Bergmen, "A machine learning approach for dynamic optical channel add / drop strategies that minimize EDFA power excursions," in *European Conference and Exhibition on Optical Communication* (2016).
17. U. Moura, M. Garrich, H. Carvalho, M. Svolenski, A. Andrade, F. Margarido, A. C. Cesar, E. Conforti, and J. Oliveira, "SDN-enabled EDFA gain adjustment cognitive methodology for dynamic optical networks," in *European Conference and Exhibition on Optical Communication* (2015).
18. C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, 2007).
19. J. D. Reis, M. Garrich, D. M. Pataca, J. C. M. Diniz, V. N. Rozental, L. H. H. Carvalho, E. C. Magalhães, U. Moura, N. G. Gonzalez, J. R. F. Oliveira, and J. C. R. F. Oliveira, "Flexible optical transmission systems for future networking," in *International Telecommunications Network Strategy and Planning Symposium (Networks)* (2014).
20. H. Zang, J. P. Jue, and B. Mukherjee, "A review of routing and wavelength assignment approaches for wavelength-routed optical WDM networks," Opt. Networks Mag. **1**(1), 47–60 (2000).

1. Introduction

Dynamic workloads such as video streaming, Internet of Things (IoT), and cloud computing require optical networks to handle growing traffic demands with agility and resilience to faults [1,2]. As a result, traditional quasi-static networks must be reconfigured in real-time to deliver network resources and the required quality of service (QoS) [3]. Advances in network control planes have demonstrated dynamic adaptability of network configurations to changing traffic demands [2,4,5], but strategic operations of optical equipment that improve physical layer stability in dynamic networking have been largely overlooked. In particular, broadband optical amplifiers such as erbium doped fiber amplifiers (EDFA) are designed to operate with slow-varying channel usage [6]. Despite EDFAs' ability to achieve economic regeneration of dense wavelength-division multiplexing (DWDM) channels, they face an unsolved challenge of wavelength-dependent power excursions during rapidly changing channel configurations [1]. The fast power transients in EDFAs occur on the time scale of 10-100μs [7], and have been largely addressed by localized feedback and feedforward controls of individual EDFA instruments [8]. However, steady-state power excursions that occur across multiple EDFAs still demand a reliable and efficient solution. We focus the scope of this paper on the mitigation of steady-state power excursions due to dynamic channel add/drop operations in cascaded EDFAs. Resolving the critical issue of power excursion in EDFA systems would eradicate a major obstacle to achieving power stability in dynamic optical networking.

The cause of steady-state EDFA power excursions is the interaction between the non-flat gain tilt and the gain control mechanism of the amplifier [1]. Modern EDFA systems employ automatic gain control (AGC) to maintain the post-EDFA power levels within a tolerance window [9], with additional controls also in place to reduce power transients in response to changing input power [10]. If a channel with high gain is added, AGC responds to an increase in the mean gain by reducing the gain on all channels. This response leads to the high-gain channel effectively stealing power from lower-gain channels [11]. Conversely, adding a low-gain channel feeds power to higher-gain channels [1]. Excursions up to 2dB have been demonstrated experimentally in as few as three cascaded EDFAs through haphazard channel additions [1]. Power excursion is undesirable because it increases the discrepancy among post-EDFA channel power levels, which may be further exacerbated by downstream amplifiers.

Power excursions are also a critical challenge to the power stability in emergent flexgrid networks, in which the bandwidth of each channel is unfixed. The added variability of channel bandwidth in flexgrid networks means that EDFA systems need to respond to a

Optics EXPRESS

variety of spectral power changes [6]. In particular, adding, dropping, and shuffling flexgrid super-channels often involve provisioning multiple WDM channels at the same time [12], and therefore require more robust strategies to mitigate power impacts.

To further complicate the matter, the exact power excursion response of an EDFA depends on its specific gain tilt and gain control mechanisms. While there are analytical studies mitigating EDFA power excursions [11,13,14], these approaches cannot be generalized to other systems (see Section 2). It is critical for EDFA power excursion solutions to be transferrable among heterogeneous systems. One promising approach is to employ the concept of cognitive networks – systems that can autonomously monitor, optimize, and adapt from their historical operation performance [15]. It is especially beneficial to design a generalized cognitive workflow that applies to unique systems. The learning process of the cognitive workflow allows each system to characterize its power responses from historical data and offer tailored best practices that mitigate power excursions.

In [16] we introduced a machine learning (ML) engine to statistically characterize EDFA systems and predict best practices in fixgrid channel addition and removal. In contrast to a fixed, closed-form analytical model of the system, the ML approach flexibly adapts to the historical responses of the amplifiers by performing regression on the selection of channels and the discrepancy of their post-EDFA power levels. We showed that the ML techniques can accurately predict the system's power response to channel changes and avoid channel configurations that would trigger significant power excursions. We also demonstrated that the ML approach is directly transferrable among EDFA systems with different equipment and scales.

In this paper, we extend the work in [16] and present (i) two ML models supported by the ML engine that characterize EDFA power responses, (ii) an extension of the ML-based approach to provision WDM channels with variable spectral bandwidth to support flexgrid networks, and (iii) a thorough analysis of the performance of our proposed models in single- and super-channel provisioning scenarios. By associating the channel selection process with the knowledge of the system's overall power response, our approach shows more optimal heuristics for fast channel provisioning decisions. We experimentally show that the ML engine can recommend channels to add/drop and stay within 1% error of the lowest possible power discrepancy over 94% of the time. When performing super-channel provisioning, the ML approach shows significant mitigation of power excursions when compared to the conventional first-fit algorithm.

In Section 2, we discuss previously proposed solutions to EDFA power excursions and their limitations. Section 3 describes the experimental EDFA system constructed to validate the ML approach. Section 4 introduces the ML models and the implementation of the ML-based solution on the experimental system. Section 5 describes the recommendation capability and accuracy of the ML engine for optimized channel provisioning in single- and super-channel scenarios. Section 6 discusses the scalability of the ML engine. We discuss our conclusions in Section 7.

2. Related works

The characteristics of a power excursion depend on the specific gain-tilt of the EDFAs, EDFA gain-control mechanisms, and the number of EDFAs in a light path. Because the variability of these aspects among amplifiers makes it difficult to derive an analytical description that applies to all systems, previous work focuses on fully characterizing a specific EDFA system and reducing its excursions. In [11], different pre-emphasized power levels are set for the channels to compensate for the excursion in gains. In [1], pairs of time-division multiplexed channels are added through fast tunable lasers to achieve a balanced average power, preventing the EDFA controls from adjusting gain levels on existing channels. In [9], the pumping level of the Raman/EDFA hybrid amplifier is adjusted to reduce the power transient variations and steady-state excursions. These techniques, while effective

Optics EXPRESS

for the specific systems analyzed, are not necessarily transferrable to different networks. These techniques rely on deterministic models and full knowledge of the gain profile, details difficult to obtain for live-network equipment that cannot be disrupted.

Optimized wavelength assignment has been shown experimentally to reduce the excursions, but is limited to 5-15% reduction of gain deviations [14]. Case-based reasoning (CBR) has also been applied to make heuristic guesses on EDFA tuning [17], but is limited in its prediction capability for unknown network scenarios.

We present an efficient, low-overhead ML engine to characterize the channel dependence of power excursions in a link containing multiple EDFAs. Once the ML engine is trained on historical snapshots of the system, we show that it can predict best practices of channel add/drop to mitigate undesired excursions. Our approach is non-disruptive and applies to EDFA systems of different designs.

3. Experiment design

To emulate the complexity of a multi-span EDFA system, we developed an experimental testbed consisting of multiple cascaded EDFAs of different brands and models. The multi-span AGC-enabled EDFA system constructed is shown in Fig. 1. We vary the number of EDFAs between experiments to examine the transferability of the ML approach among different system scales. Each EDFA demonstrates a different gain tilt and responds differently to input power changes when channels are added or dropped. The channel dependence of the system's power excursion is determined by training a regression model that determines the contribution of ON/OFF channels to the overall discrepancy among the channel power levels after the final EDFA.

The WDM sources transmit 24 WDM channels from ITU-T grid 194.40THz to 192.10THz with 100 GHz spacing at a uniform laser launch power level of 13dBm per channel, which are combined via a wavelength-selective switch (WSS). Variable optical attenuators (VOA) are placed before each span's EDFA to emulate a 20dB per-span transmission loss of single-mode fibers (SMF). We adjust the EDFA pumping levels to achieve an overall gain tilt with a maximum disparity of 10dB between the highest and lowest gains; this is to ensure the receiver dynamic range is sufficient to identify channels after the EDFA cascade. Figure 2 shows the post-EDFA channel power spectra of the 2-span and 3-span systems. While the gain tilt is exaggerated compared to telecom conventions, it presents an extreme case that produces sufficient EDFA power excursions to examine the ML engine's capabilities. To ensure an adequate number of available channels for add/drop, we maintain 10-20 ON channels at any given time, which corresponds to a spectral utilization between 42% and 83%. The post-EDFA power levels are recorded with a C-band optical performance monitor (OPM), and stored in a database for analysis. The ML engine is trained on the post-EDFA power levels and channel ON/OFF states, and constructs a regression model for future predictions.

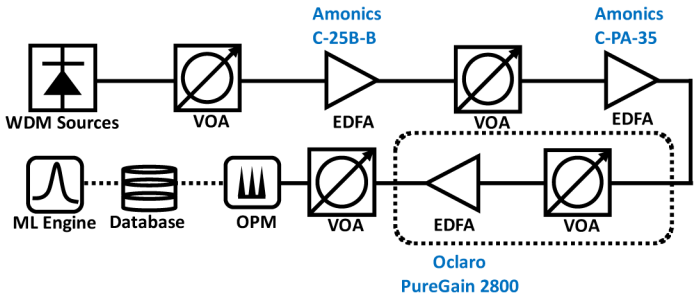

Fig. 1. Setup of the multi-span EDFA system; the additional EDFA and VOA in the dashed box are included for the 3-span system.

Optics EXPRESS

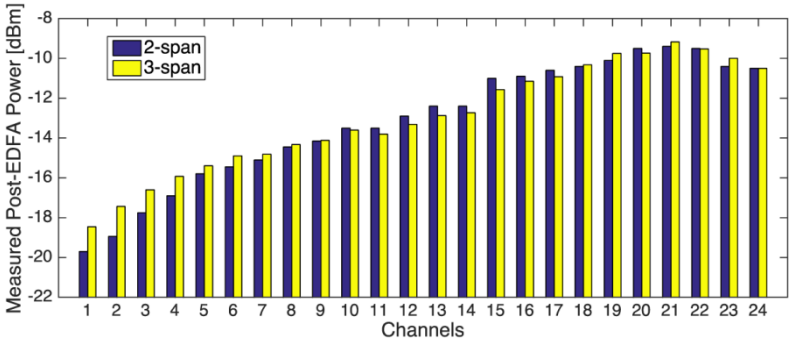

Fig. 2. Measured post-EDFA power spectra with 24 ON channels for both systems. Channels 1 to 24 correspond to ITU-T C-band 194.40THz to 192.10THz with 100GHz spacing. Channels are launched with uniform power.

4. Machine learning models and analysis

We define a regression problem with supervised ML to statistically model the channel dependence of EDFA power excursions. The predictor variable of the regression model is a 24-bit array, with each bit indicating an ON channel as 1, or an OFF channel as 0. This can be scaled up to longer arrays to accommodate more DWDM channels. The response variable of the regression model is a measurement of the post-EDFA power discrepancy. In the following analysis, we use the standard deviation (STDEV) of the channel power levels as the measurement of discrepancy. We select STDEV of post-EDFA power levels as the optimized metric instead of other figures of merit (FoM), such as the optical signal-to-noise ratio (OSNR), because STDEV directly infers the extent of undesired power excursions induced by cascaded EDFAs. STDEV is also industrially practical because it necessitates the transceivers' dynamic range, which impacts system complexity and cost. Alternatively, the power excursion measured for each ON channel can be aggregated to reflect the overall impact on the system; this has the ability to reflect the channels' power stability instead of power discrepancy. To accentuate the mitigation of the power discrepancy among channels due to EDFA power excursions, we focus on the power STDEV metric in the following analysis.

We collect 870 historical channel ON/OFF states and power STDEV values from each experimental system, which are split into a training set of size 600 and a testing set of size 270. During training with a system's training set, the ML model's parameter values are optimized for that specific system in order to achieve accurate prediction capability. We design the ML engine to streamline the training process by using historical channel ON/OFF states and post-EDFA power levels, which are operation data that can be easily collected. Because existing operation data of the system can also be used to train, the ML engine can continue to improve its accuracy with increasing size of available training data. For a larger system with complex effects, the ML engine can implement an online, continuously evolving training process to gradually capture the system's power dynamics. The ML model is evaluated with the testing set by two metrics: A) the mean square error (MSE) between the predicted and the measured STDEV of the testing set, and B) the correctness of the best channel provisioning recommended based on the predictions.

The predictor and response variable values are preprocessed before training and testing according to Eqs. (1) and (2). The per-dimension mean is removed from the predictor variable x and the response variable y. Each predictor dimension is also standardized with a variance of 1. When used for prediction, the model takes in standardized inputs and returns offset outputs. These commonly practiced preprocessing techniques prevent dimensions with large variances or means from masking the contribution of other dimensions.

Optics EXPRESS

$$x_{ij}^{prep} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \tag{1}$$

$$y_{ij}^{prep} = y_{ij} - \bar{y}_j. \tag{2}$$

In Eqs. (1) and (2), $i = 1..n$ is for n total data points; $j = 1..d$ labels a dimension of the predictor ($d = 24$) or the response ($d = 1$); σ_j is the per-dimension STDEV of the predictor, and $\bar{x}_j$, $\bar{y}_j$ are the respective per-dimension means.

We examine the efficacy of two regression models – ridge regression (RR) and kernelized Bayesian regression (KBR) – in the context of predicting power STDEV solely from channel ON/OFF states. The rationale for RR is to examine a low-complexity model that can be trained and applied quickly. The rationale for KBR is to perform an efficient implementation of CBR with improved prediction capabilities. We explain in detail the formalism and implementation of each model below.

4.1 Ridge regression model

RR is a linear regression model with an additional ℓ_2 penalty. The weights associated with the channels are defined by Eq. (3) and determined by Eq. (4):

$$w_{RR} = \arg\min_{w} \left(\|y - Xw\|^2 + \lambda \|w\|^2 \right), \tag{3}$$

$$w_{RR} = \left(\lambda I + X^T X \right)^{-1} X^T y, \tag{4}$$

where w_{RR} is the array of weights learned by RR; y is the columnized response values of the training set of shape $n \times 1$; X is the vertically stacked predictor values of the training set of shape $n \times 25$. The dimension in addition to the 24 channels is used for the learned bias. I is the identity matrix of shape 25×25. λ is the complexity parameter. Using the cross-validation technique, in which the ML model is repeatedly validated with randomized subsets of the training set and different values of λ, the parameter is evaluated as 2.6 and 2.7 for the 2-span and 3-span systems, respectively. RR implements the λ penalty term to encourage a small variance across the determined weights. This means the model prefers to characterize contributions from every channel, instead of attributing the cause of power excursions to a few isolated channels. As a result, RR in general can better prevent over-fitting the training data than linear regression.

Figure 3 shows the weights associated with each channel in its contribution to the post-EDFA power discrepancy after training with 600 historical samples for the 2-span and 3-span EDFA systems. The magnitude of each weight indicates its influence on the post-EDFA power STDEV. The sign of each weight indicates whether the addition of the corresponding channel will increase (positive weight) or decrease (negative weight) the power STDEV. In the case of our experiment, channels close to the ends of the spectrum exacerbate the power discrepancy, while channels at the center of the spectrum mitigate the power discrepancy. We can also note that, comparing the 2-span system with the 3-span system, the same channel has different contribution weights, indicating that the two systems, while sharing some common equipment, have different power excursion responses. Channel 24, whose weights have different signs, contributes opposite effects to post-EDFA power discrepancy in the two systems, even if the magnitudes of its contributions are relatively small.

Optics EXPRESS

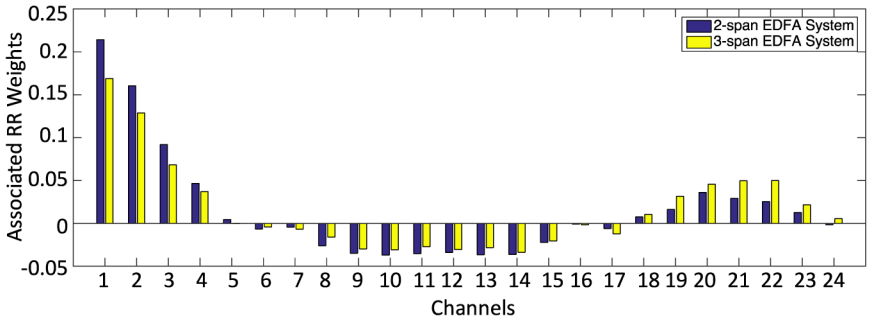

Fig. 3. Associated weights assigned to each channel by RR for the 2-span and 3-span EDFA systems, indicating each channel's contribution to the post-EDFA power discrepancy in respective systems.

4.2 Kernelized Bayesian regression model

KBR relies on mapping the training inputs to a kernel, which is equivalent to an expansion in the input feature dimensions. The regression model is trained using the kernel to determine the posterior probability distribution of the response variable. Specifically, we construct a kernel called the Radial Basis Function (RBF) shown in Eq. (5):

$$K(x,x') = \alpha \exp\left(-\frac{1}{b}\|x-x'\|^2\right), \tag{5}$$

where a and b are parameters that adjust the strength of the kernel, which we set as 0.0001 and 3.5 respectively for our analysis using cross-validation, consistent for both the 2-span and 3-span systems. The variables x and x' are two 24-dimensional predictor values; the value of the kernel function decreases exponentially with the magnitude square of their difference. The RBF kernel indicates that two network scenarios with similar ON/OFF channels are assumed to have similar extents of power excursions [17], which is leveraged to efficiently emulate the use of a database in CBR. Given a new network scenario, we can infer its predicted power STDEV based on the individual contribution of each channel, as well as how similar it is to the known network scenarios. The predictions are obtained from the linear combinations of the training outputs weighted by the kernel function values [18]. In contrary to learning the linear weights of each channel's contribution, KBR with RBF models the system as a Gaussian process by learning the mean and variance of the power STDEV given the ON/OFF channels. One advantage of KBR is its ability to pinch the variance of the Gaussian process near known data points. Hence, for a future scenario that has appeared in the training set, our KBR model can accurately recall the associated post-EDFA power STDEV, thus achieving the functionality of CBR.

4.3 Performance of the ML models

We characterize the adaptability of the models on two different systems – with two and three EDFA amplified spans respectively. Figure 4 compares the prediction MSE between KBR and RR as a function of the training set size for both systems. For each training size, the models are trained 5 times with random subsamples of the training set, and the prediction MSEs of the testing set are averaged. In both EDFA systems, RR outperforms KBR between the training sizes of 30 and 600. In addition, the ML models perform better on the 2-span system than the 3-span system. It is also noticeable that KBR's performance improves more significantly than RR with increasing training set size, despite its poorer performance when trained with a small data set. For RR on both systems, there is no significant performance improvement beyond 120 training data points.

Optics EXPRESS

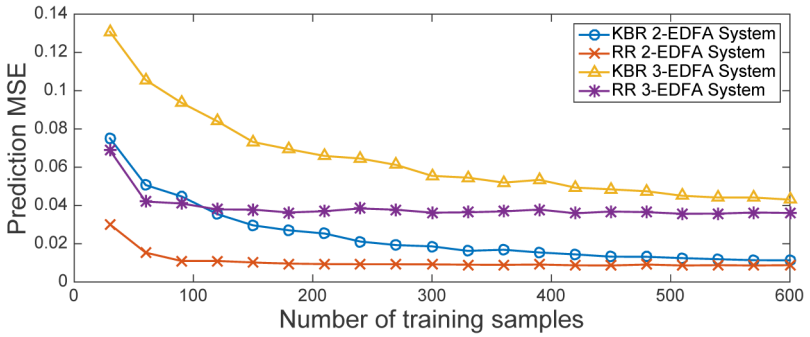

Fig. 4. Reduction of the ML models' prediction MSEs of power STDEV with increasing training data size.

RR and KBR demonstrate different training and prediction mechanisms. For RR, the training data is used to determine a set of weights associated with the input dimensions. After training, the training data can be discarded; only the set of weights is necessary to make predictions. For KBR, the training process involves building the RBF kernel; the kernel and the training data are both used when making predictions. RR has a training computation complexity of $O(N^2M)$ and single-scenario prediction computation complexity of $O(N)$, while KBR with RBF kernel is $O(N^2M)$ for training and $O(M^2)$ for single-scenario prediction, where N is the number of predictor values (number of channels + 1) and M is the number of training samples. Consequently, RR scales more optimally with increasing number of training samples, and KBR scales more optimally with increasing number of channels. If $M \gg N$, RR would be more efficient in both training and prediction. Both models are implemented in Python 3.5 and executed on a personal computer. An example of their respective training and prediction times are shown in Table 1.

Table 1. Time consumption of training and prediction for RR and KBR.

Model	RR	KBR
Time to train with 600 data points [ms]	82.2	2300
Time to predict for a single scenario [ms]	0.068	467

5. Channel provisioning recommendations

5.1 Single-channel add/drop

We demonstrate the ML engine's capability to reduce undesired post-EDFA power excursions in Fig. 5 for adding and dropping fixgrid single channels. The examples are retrieved from 3-span EDFA system with the ML engine employing KBR and trained with 600 training data points. In both cases shown in Fig. 5, we start with a randomly initialized scenario of ON/OFF channels. For adding a channel, as shown in Figs. 5(a), 5(c), and 5(e), the model predicts the power STDEV if a hypothetical channel is added to each available slot. Then the model recommends the best slots to add a channel that will result in the lowest power STDEV and the least undesired power excursions. Recommending up to four slot options provides 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 Figs. 5(b), 5(d), and 5(f). In the two tests shown, the slots recommended by the ML engine correctly align with the best slots from the actual measurements.

In Fig. 6, we examine the recommendation accuracy of the ML engine with 100 tests of single-channel addition and 100 tests of single-channel removal on both EDFA systems with randomly initialized starting conditions. We report the ML engine's ability to identify the top

Optics EXPRESS

one and top four channels to add/drop after training with various training sizes between 30 to 600, as well as the corresponding prediction MSE at these training sizes. Overall, the ML engine achieves a recommendation accuracy of 81% and 84% respectively, for identifying the correct top one and top four candidates on the 2-span system, and 75% and 89% respectively on the 3-span system. Both RR and KBR demonstrate strong correlation between the prediction MSE and the recommendation accuracy. Despite showing higher prediction MSE on both systems, KBR demonstrates better recommendation accuracy in identifying the best channels to add/drop in general.

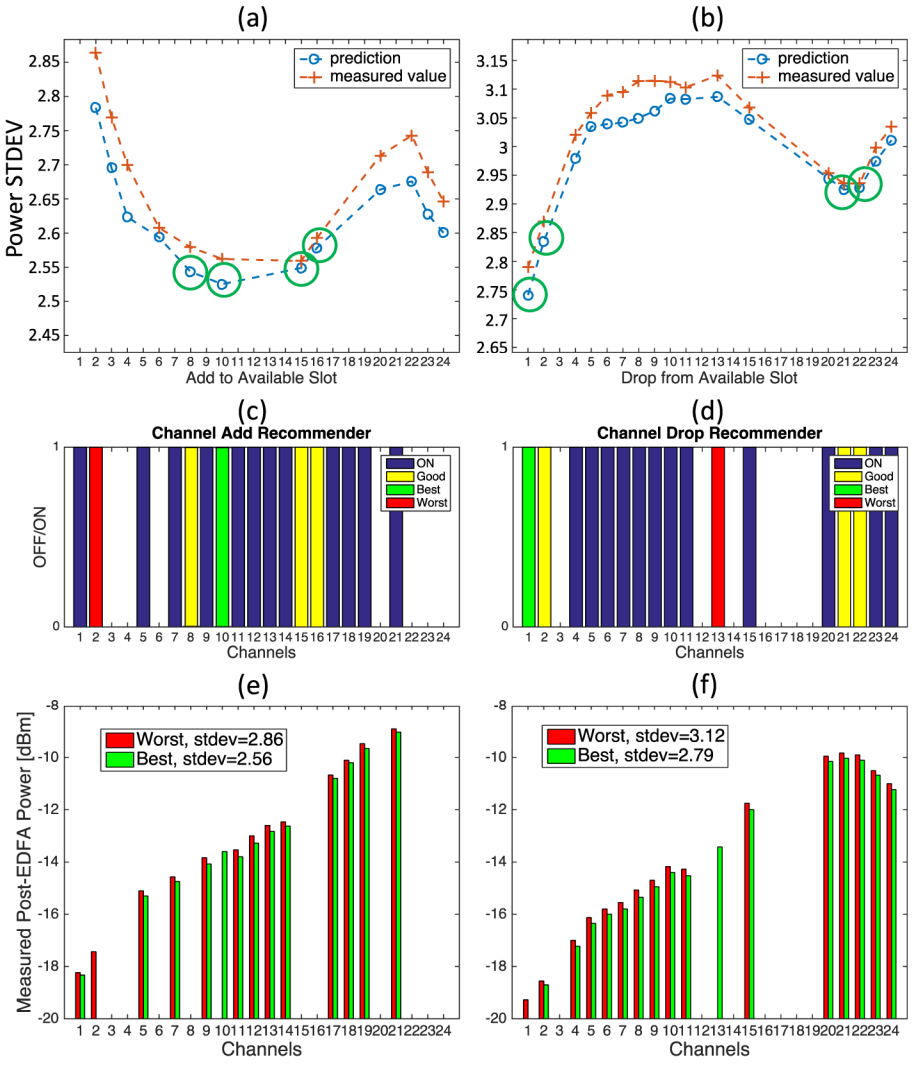

Fig. 5. Comparisons between predictions and measurements of post-EDFA power discrepancy for single-channel add (a) and drop (b). The top four slot options with the lowest predicted power STDEV are circled. (c) and (d) illustrate the good channels to add/drop, and the worst channels to avoid. (e) and (f) compare the post-EDFA power spectra of ON channels between the best and worst cases in channel add and drop, respectively.

Optics EXPRESS

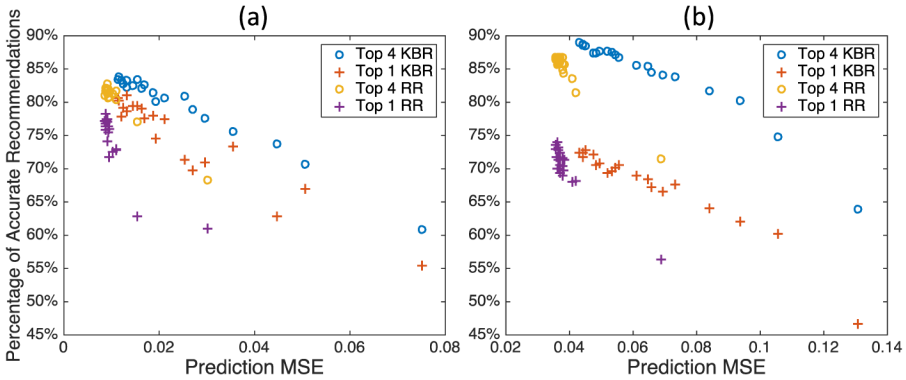

Fig. 6. Correlation between the prediction MSE (horizontal axis) and the percentage of accurate recommendations (vertical axis) for KBR and RR for the (a) 2-span and (b) 3-span EDFA systems. The percentage of accurate recommendations reflects the number of tests in which the ML engine correctly recommends the top channels, out of the 200 tests performed.

Even when the ML engine does not recommend the actual best channel, the recommendations would still result in post-EDFA power discrepancy comparable to the lowest achievable power discrepancy, known as the "ground truth". This makes it feasible to provision channels solely based on the ML engine's recommendations. Figure 7 shows the cumulative distribution of the difference between the ML recommendations' power discrepancy and the ground truths. Each cumulative distribution function (CDF) plot is generated from 200 single-channel add/drop cases with randomly initialized ON/OFF channels. The vertical axis represents the fraction of the channel recommendations in the 200 tests that fall within a certain percent error from the lowest possible post-EDFA power discrepancy. In the 2-span system, over 95% of the ML recommendations are within 1% error of the lowest achievable power discrepancy for both RR and KBR. In the 3-span system, over 94% of the ML recommendations are within 1% error for both RR and KBR. This means that even if the ML engine does not give the exact best channel for add/drop, its recommendation would still be almost as good as the best outcome. If we were to randomly pick a channel to add/drop, the choice would be within 1% error of the best provisioning only 16% of the time. Figure 8 shows a systematic analysis of the training set size and recommendation accuracy for RR and KBR on the 2-span and 3-span systems. KBR shows monotonic improvement in recommendation accuracy with larger training set, while RR shows less correlation between recommendation accuracy and different training set sizes from 100 to 600, with the best result given by a training set size of 200.

Optics EXPRESS

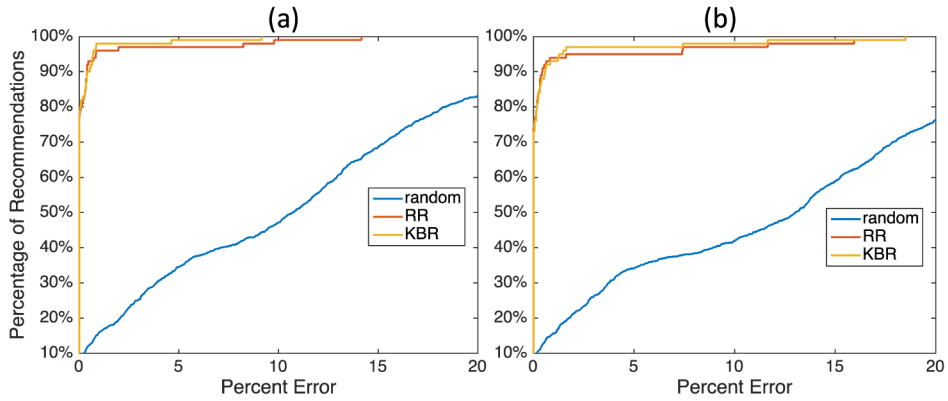

Fig. 7. CDF of the percent error between the power discrepancy of recommended options and the actual lowest power discrepancy, plotted for single channel candidates based on KBR, RR, and random selections of channels for the (a) 2-span and (b) 3-span EDFA systems. Compared to random selections, ML recommendations with KBR and RR show significant improvement in approaching the lowest possible post-EDFA power discrepancy.

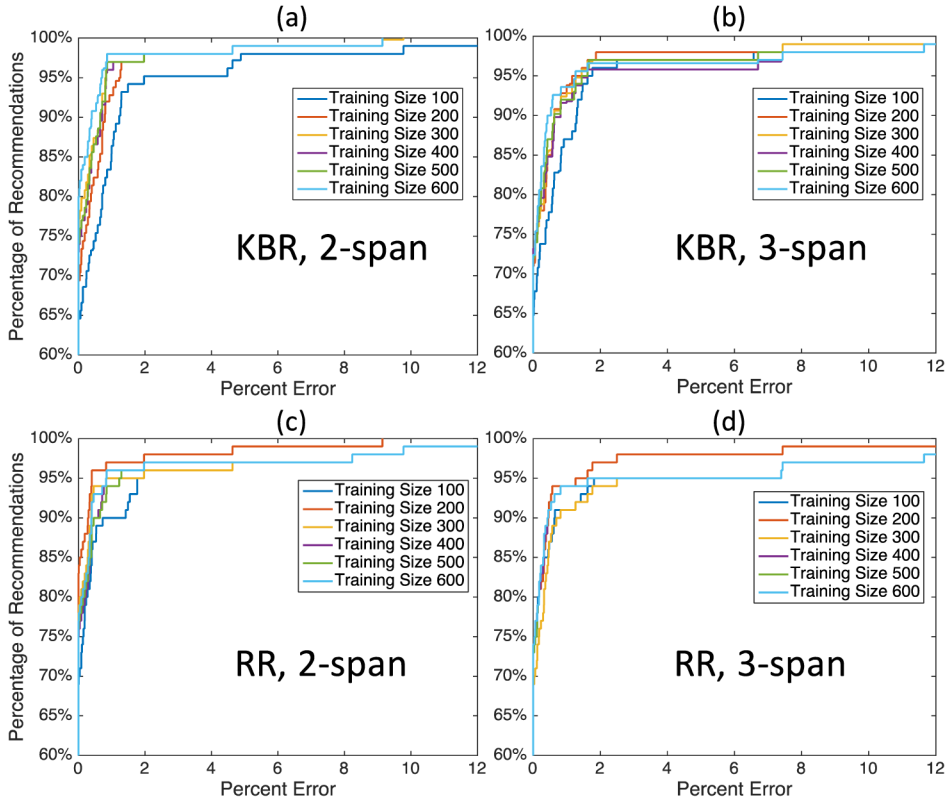

Fig. 8. CDF of the percent error between the power discrepancy of recommended options and the actual lowest power discrepancy, at different training data sizes. (a) and (b) illustrate for the KBR model on the 2-span and 3-span EDFA systems respectively. (c) and (d) illustrate for the RR model on the 2-span and 3-span EDFA systems respectively.

Optics EXPRESS

5.2 Super-channel provisioning

In order to further support the dynamicity of traffic demands, flexgrid optical networks employ variable channel bandwidth to accommodate diverse modulation formats and data rates [19]. One implementation of flexgrid networks employs super-channels consisting of multiple contiguous sub-channels to enable higher data bandwidth and greater spectral efficiency [3]. In the context of EDFA-induced power excursions, the addition of a super-channel results in a greater input power change than the addition of a regular channel. Applying the ML engine to predict and avoid undesired power excursion induced by a super-channel would improve the power stability of the EDFA system.

We experimentally demonstrate that the ML engine can assign two to three contiguous sub-channels to form an optimal super-channel while mitigating undesired power excursions. Figure 9 illustrates the workflow of the ML engine for super-channel provisioning. Figure 10 shows two examples of ML recommendations for super-channel addition on the 3-span EDFA system. Figures 10(a) and 10(b) demonstrate the close fit between the predicted and measured post-EDFA power STDEV values of the super-channel at all available locations. Figures 10(c) and 10(d) illustrate the recommendation view for the best and worst locations. The ML engine is trained with KBR and the same 600 training data sets used in the previous single-channel provisioning.

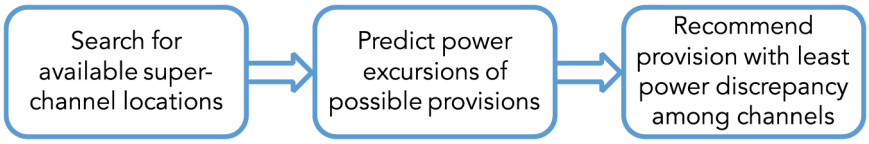

Fig. 9. Functional workflow of the ML engine that enables super-channel addition with the least induced power excursions and post-EDFA power discrepancies.

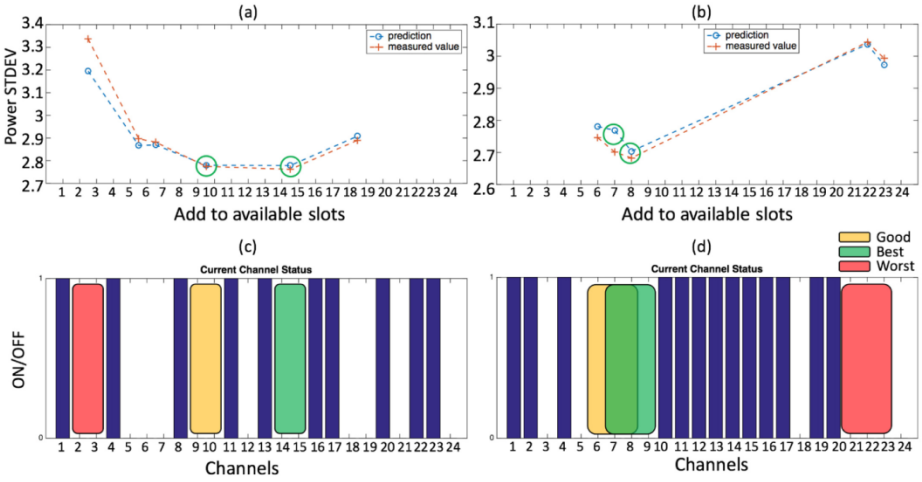

Fig. 10. Comparisons between predictions and measurements of post-EDFA power discrepancy for super-channel addition consisting of (a) two contiguous sub-channels and (b) three contiguous sub-channels. The top two super-channel candidates with the lowest predicted power STDEV are circled. (c) and (d) illustrate the best, good, and worst super-channel candidates.

We see a significant mitigation in post-EDFA power discrepancy when using the ML engine to determine the optimal locations for the super-channels. Figure 11 shows a comparison of the post-EDFA power STDEV among the best provisioning options, the ML engine recommendations, and super-channel allocation based on the first-fit algorithm [20], with which the first available location starting from the lower end of the spectrum is selected.

Optics EXPRESS

Overall, the ML engine recommendations agree closely with the best provisioning options based on actual measurements. For cases where the first-fit algorithm results deviate from the best options, the ML recommendations demonstrate clear improvements over the first-fit algorithm results for mitigating the post-EDFA power discrepancy.

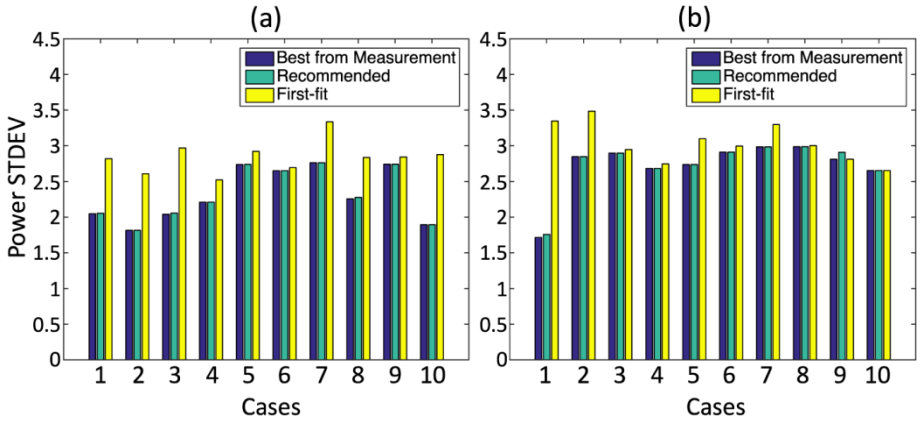

Fig. 11. Twenty randomly initialized scenarios to compare the resultant power STDEV values among super-channel additions based on actual measurements, ML predictions, and the first-fit allocation algorithm for (a) two-channel-wide and (b) three-channel-wide super-channels in the 3-span EDFA system.

6. Scalability of the ML engine

While the ML engine's efficacy in mitigating post-EDFA power discrepancy is demonstrated experimentally for single- and super-channel provisioning, it is important to discuss how the ML engine would scale with increasing dimensions and complexity in an optical network. For a greater number of amplifiers and spans of fibers in a light path, the ML engine's training and prediction processes are unaltered. This is because the ML models are trained on the overall channel ON/OFF states, as well as the power discrepancy after all cascaded amplifiers, which are independent from the number of EDFAs and fiber spans in a light path.

A more complex optical network may employ wavelength add/drop capability at intermediate nodes along a light path, as shown in Fig. 12. In this case, the set of channels entering the first EDFA, labeled A, are different from the set of channels exiting an intermediate EDFA, labeled B. To operate the ML engine in this scenario, the predictor variable to the ML models would capture the ON/OFF states of all 4 channels. The response variable of the ML models is determined as the post-EDFA power STDEV calculated from the power levels of Channels 1 and 2 after EDFA C, and the power levels of Channels 3 and 4 after EDFA B. The recommendation process of the ML engine would take into account the channel assignment constraint – only Channels 1 and 2 are available from EDFA A to EDFA C, and Channels 3 and 4 are available for add/drop at EDFA B. This represents a limited search space for the ML engine to make the channel add/drop recommendations.

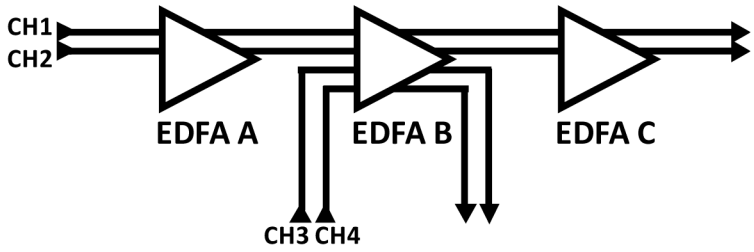

Fig. 12. Illustration of channel add/drop operations at the intermediate EDFA node.

Optics EXPRESS

For more complex networks implementing Reconfigurable Optical Add/Drop Multiplexers (ROADM), network edges that connect pairs of ROADMs may carry different wavelength channels. Each edge may also contain multiple EDFAs for optical power regeneration. In this case, multiple ML engines can be deployed to individual edges and operate in a distributed fashion. This allows individual ML engines to mitigate the power excursions in each edge of the network – consistency of channel power levels on one edge would be optimized before handing off to the next edge. However, the distributed ML engines will need to coordinate for light paths that traverse multiple network edges to take into account wavelength continuity and system-wide QoS. These additional functionalities are out of the scope of this paper and would be suitable for future explorations to augment the ML engine's capabilities.

7. Conclusion

Channel dependent power excursions in gain controlled EDFAs present critical challenges to system performance and agility in dynamic optical networking. The strong dependence of the excursion on the gain profile of an EDFA makes it infeasible to transfer analytical solutions to different systems. We introduce an ML engine based on a regression approach that characterizes the channel dependence of power excursions in a WDM network with multiple cascaded EDFAs. Two machine learning models – RR and KBR with RBF kernel – are trained with 600 historical network usage data points and the discrepancy of post-EDFA channel power levels. Channel provisioning options recommended by the trained ML engine achieve within 1% error of the lowest possible post-EDFA power discrepancy in over 94% of randomized network scenarios. We also demonstrate that the ML engine's workflow can be transferred to systems of different EDFAs and achieve similar performance, and is applicable to both single- and super-channel provisioning to support flexgrid optical networking. By using ML to accurately predict the power excursions in WDM channel provisioning, network operators can make quick and precise decisions to address network demands and optimize EDFA power dynamics. For future explorations, the capabilities of the proposed ML engine can be further augmented with optimizations of OSNR and additional QoS metrics. For larger networks, distributed and coordinated operations of multiple ML engines can be explored to optimize power consistency for individual network edges, while ensuring wavelength continuity and overall system level QoS.

Funding

Center for Integrated Access Networks (CIAN) NSF ERC (EEC-0812072); National Science Foundation (NSF) (CNS-1423105); Department of Energy (DoE) Advanced Scientific Computing Research (ASCR) under Turbo Project (DE-SC0015867); NSF Graduate Research Fellowship (DGE-16-44869).

Acknowledgments

The authors would like to thank AT&T Labs for their generous loan of the equipment.