Neural Network Based Wavelength Assignment in Optical Switching

Craig L. Gutterman*, Weiyang Mo†, Shengxiang Zhu†, Yao Li†, Daniel C. Kilper†, Gil Zussman*
∗ Electrical Engineering, Columbia University, New York, NY, USA.
† College of Optical Sciences, University of Arizona, Tuscon, AZ, USA.

ABSTRACT

Greater network flexibility through software defined networking and the growth of high bandwidth services are motivating faster service provisioning and capacity management in the optical layer. These functionalities require increased capacity along with rapid reconfiguration of network resources. Recent advances in optical hardware can enable a dramatic reduction in wavelength provisioning times in optical circuit switched networks. To support such operations, it is imperative to reconfigure the network without causing a drop in service quality to existing users. Therefore, we present a system that uses neural networks to predict the dynamic response of an optically circuit-switched 90-channel multi-hop Reconfigurable Optical Add-Drop Multiplexer (ROADM) network. The neural network is able to recommend wavelength assignments that contain the power excursion to less than 0.5 dB with a precision of over 99%.

CCS CONCEPTS

• **Networks** → **Cross-layer protocols**; **Network management**; • **Computing methodologies** → **Neural networks**; • **Hardware** → **Emerging optical and photonic technologies**;

ACM Reference format:
Craig L. Gutterman*, Weiyang Mo†, Shengxiang Zhu†, Yao Li†, Daniel C. Kilper†, Gil Zussman* ∗ Electrical Engineering, Columbia University, New York, NY, USA. † College of Optical Sciences, University of Arizona, Tuscon, AZ, USA. . 2017. Neural Network Based Wavelength Assignment in Optical Switching . In *Proceedings of Big-DAMA '17, Los Angeles, CA, USA, August 21, 2017,* 6 pages.
https://doi.org/10.1145/3098593.3098600

1 INTRODUCTION

In recent years, the advent of network applications such as video streaming, Internet-of-Things (IoT), and cloud computing have contributed to exponential Internet traffic growth [2]. As traffic increases, the ability to manage capacity and provision new services in the optical layer in real-time becomes important. Provisioning in commercial systems takes minutes per wavelength, and in practice days of careful offline wavelength planning are used to mitigate unpredictable dynamic effects [1, 23]. To allow for such an increase in network traffic, optical metro and long haul networks need to be dynamic and need to utilize network resources efficiently [12, 13].

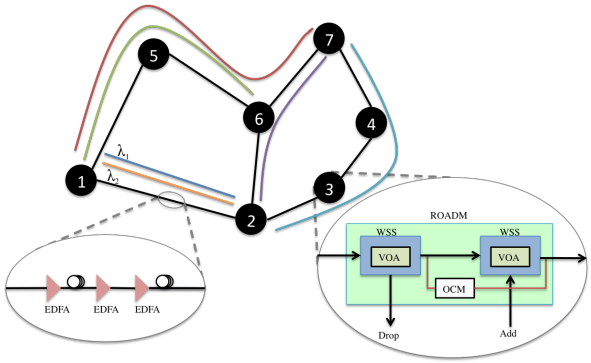


Figure 1: A 7 ROADM Mesh Network. Lightpaths across different ROADM hops are shown with different colors. Connections between ROADM nodes include transmission fiber, VOA, WSS, OCM, and EDFAs. EDFAs are used to maintain channel powers and VOAs are used to correct the power divergence due to EDFA's wavelength dependent gain and Stimulated Raman Scattering (SRS) in the transmission fiber.

The service provider provisions optical paths, for its customers, such that in Dense Wavelength Division Multiplexing (DWDM) up to 96 wavelengths (or unique optical signals) are transmitted in a single fiber. These signals are added and dropped from the network at Reconfigurable Optical Add-Drop Multiplexer (ROADM) nodes which connect to Layer 2/3 switches. Previous work investigated optical switching techniques for real time adaptation of optical layer capacity based on changing traffic conditions. A key challenge is the optical power dynamics which arise and cascade in a ROADM system in the presence of optical circuit switching [7, 8, 16, 21, 24]. Solving the issue of offering wavelengths with minimal power dynamics is key in the development of optical dynamic networks, wavelength defragmentation, and implementing real time Routing and Wavelength Allocation (RWA) algorithms.

Automatic Gain Control (AGC) is commonly used in Erbium-Doped Fiber Amplifiers (EDFAs) to maintain constant average gain for varying input conditions. EDFAs are used to boost the optical signal being transported between two end nodes. This results in power excursions, one of the main types of power dynamics in optical networks. Due to the EDFA's wavelength dependent gain, with changing input conditions, power excursions occur on active channels, and grow over multiple EDFAs along the propagation path. The AGC attempts to amplify the channel to obtain a target gain value, but adding or removing a channel can cause deviations that perturb other channels resulting in excursions. These gains can not

Big-DAMA '17, August 21, 2017, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5054-9/17/08…$15.00
https://doi.org/10.1145/3098593.3098600

be corrected until slow per-channel power controls in the optical nodes are able to re-adjust the power level, which requires repeated measurements and adjustments.

The cumulative effect of these power dynamics can increase non-linear impairments at high power, and lead to Optical Signal to Noise Ratio (OSNR) degradation at low power, both resulting in increased bit error rate. In fact, the optical power launched into each fiber span is a carefully controlled parameter that is continuously adjusted to within 1 *dB* [3]. Additional margin for wider power variations would compromise the system reach, increasing cost and energy. The combination of these effects can cascade, as commerical systems contain more than 19 amplified transmission spans and 32 EDFAs resulting in large accumulated effects [26, 28]. Even after 5 cascaded EDFAs, the excursion can be greater than 4 dB [16]. Furthermore, the impact of these dynamics remains until Variable Optical Attenuators (VOAs) or Wavelength Selective Switches (WSSs) make power adjustments, which often require seconds or even minutes of tuning, potentially leading to extended signal outages [21].

An example of these effects can be seen in the 7 node mesh network shown in Fig. 1. For simplicity, only one direction of the EDFAs and ROADM are shown. In each optical fiber there are multiple EDFAs. When looking at a ROADM node, the input signal goes in one direction and is transmitted out of the other side after going through multiple WSSs. The Optical Channel Monitor (OCM) is connected to multiple monitor ports in the node, and is used to measure the individual channel power collected from each VOA (one for each wavelength on the fiber). This information is fed back to the individual WSS so that it can remove any power excursions that occur on any individual channel.

For example, if lightpath λ_2 is dynamically added to the path from node 1 to node 2, where only λ_1 is active, then this can cause a power dynamic in the system. If the excursion is too large it will result in a decrease in signal quality and an increase in the Bit Error Rate (BER) for λ_1. Service providers need to ensure that network reconfiguration will not affect the quality of service of active clients. Therefore, it is important to accurately predict how network control systems will influence the optical physical layer.

In this work, we develop a scalable neural network and conduct experiments to evaluate its accuracy at predicting channel power excursions resulting from optical circuit switching in a multi-hop ROADM system. Scalability is tested over a full C-band 90 wavelength assignment set for 4 transmission spans and 5 ROADM nodes. 84,000 data samples are collected during optical circuit switching experiments. The data is used to develop, train, and evaluate our proposed neural network. The performance is compared with ridge regression as used in [15]. Each method is evaluated for both prediction error and effectiveness in assigning wavelengths to minimize channel power excursions during optical circuit switching.

2 RELATED WORK

The necessity of dynamic functionality in optical networks stems from increased network traffic growth as well as on-demand capacity variations. Internet measurements show that these traffic demands vary throughout the day and are wasteful by over provisioning to meet maximum demands [29]. In addition, there is a need for quickly provisioning high bandwidth links for inter-data-center bulk transfers [14, 17, 18].

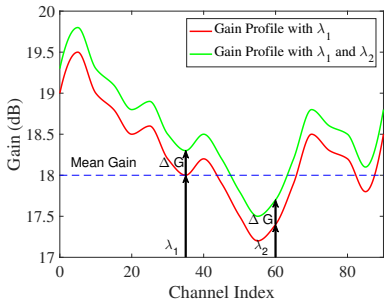


Figure 2: Theoretical gain characteristic for AGC EDFAs.

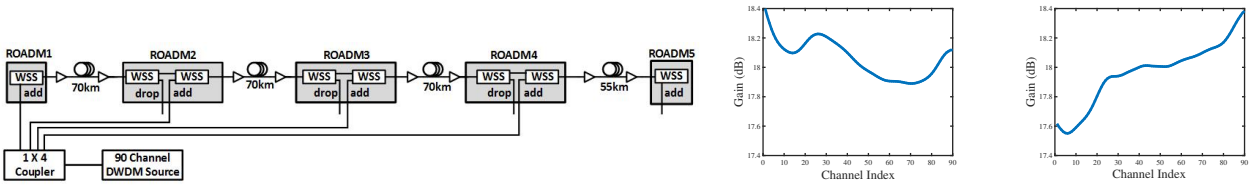
There has been extensive research on optical network management and efficient RWA algorithms, but they have mostly focused on implementation at the planning phase of optical networks [6, 10, 12, 13]. Some of this research has looked at RWA algorithms that are power aware and enable energy efficient operation [20, 27].

However, power excursions need to be considered before dynamic optical networks can be deployed. Analytical models were developed for predicting and minimizing power excursions [16, 19]. A linear EDFA gain model for real time wavelength assignment in optical circuit switching showed only $5\% - 15\%$ improvement in power excursion reductions relative to random assignment and did not include EDFA tilt and loading dependent gain variations such as Spectral Hole Burning (SHB) [16]. Other analytical models based on detailed amplifier characterization showed a 0.2 dB prediction accuracy, but are not compatible with real-time optical circuit switching [19]. In a practical multi-hop ROADM system, a lightpath may have numerous EDFAs, many of which may have different internal designs or settings and even originate from different manufacturers, thereby posing significant challenges to developing a general analytical model.

Over the past few years there has been extensive work to apply machine learning to different domains [22, 25]. These ideas have been used for control in data centers and internet traffic forecasting [5, 9, 11]. Through extensive experimental testing and data collection of the channel power response during system installation and turn-up, it is possible for a machine learning algorithm to learn to accurately distinguish between potential new wavelength assignments that will cause excursions and those that will not. Previous applications of machine learning for wavelength assignment demonstrated its effectiveness in a 24 channel, single-hop system [15]. Previously, an accurate and scalable machine learning algorithm applicable to multiple hop ROADM systems including full C-band DWDM channels has not been developed.

3 EDFA GAIN MODEL

Wavelength reconfigurations in optical networks result in power excursions of active (or provisioned) channels, even those not involved in the reconfiguration operation. The AGC in EDFAs attempts to maintain constant average gain for all active channels, however when adding a new channel it can cause power variations to active channels. By just looking at AGC and neglecting gain tilt, ASE noise,



(a) (b) (c)

Figure 3: (a) Experiment setup of a 5-ROADM amplified optical network, including 4 fiber spans and 8 EDFAs with different gain characteristics. (b) Gain characteristic of first EDFA, (c) Gain characteristic of second EDFA.

and other wavelength dependent processes, all channels should experience the same power excursion. The power excursions can be approximated by the following simplified analytical equations:

$$P^{\text{in}}_{\text{total}} = \sum_{j=1}^{N} P^{\text{in}}_{j} \tag{1}$$

$$P^{\text{out}}_{\text{total}} = G_{\text{T}} P^{\text{in}}_{\text{total}} \tag{2}$$

$$P^{\text{out}}_{\text{total}} = \sum_{j=1}^{N} P^{\text{in}}_{j} G_j \tag{3}$$

Where N is the number of active channels and G_{T} is the target gain setting of an EDFA, which is maintained as a constant by the AGC. $P^{\text{out}}_{\text{total}}$ and $P^{\text{in}}_{\text{total}}$ denote the total output power and the total input power of the EDFA, respectively. G_j denotes the wavelength dependent gain of channel j and P^{in}_{j} denotes the input power of channel j. The total output power equals the product of the total input power times the target gain G_{T}, as shown in (2), or it can be calculated as the summation of each individual input power of the N channels times the wavelength dependent gain, as shown in (3). By combining (1)-(3):

$$G_{\text{T}} = \frac{\sum_{j=1}^{N} P^{\text{in}}_{j} G_j}{\sum_{j=1}^{N} P^{\text{in}}_{j}} \tag{4}$$

When a $(N+1)^{\text{th}}$ channel wavelength with gain G_{N+1} is added into the EDFA (can be extended for dropping a channel), (4) no longer holds, so the AGC responds by adjusting the pump power to control the internal amplifier gain such that G_{T} is held constant. The amount of the response, ΔG, can be quantified by the AGC condition:

$$G_{\text{T}} = \frac{\Delta G \sum_{j=1}^{N+1} P^{\text{in}}_{j} G_j}{\sum_{j=1}^{N+1} P^{\text{in}}_{j}} \tag{5}$$

The power excursion, ΔP, is equal to the gain variation ΔG. Combining (4) and (5) leads to:

$$\Delta P = \Delta G = \frac{(\sum_{j=1}^{N} P^{\text{in}}_{j} G_j)(\sum_{j=1}^{N+1} P^{\text{in}}_{j})}{(\sum_{j=1}^{N+1} P^{\text{in}}_{j} G_j)(\sum_{j=1}^{N} P^{\text{in}}_{j})} \tag{6}$$

Assuming each channel has the same input power, (6) can be further simplified into:

$$\Delta P = \Delta G = \frac{\sum_{j=1}^{N} G_j (N+1)}{\sum_{j=1}^{N} G_j N} = \frac{N G_{\text{T}} (N+1)}{(N G_{\text{T}} + G_{N+1}) N} = \frac{N G_{\text{T}} + G_{\text{T}}}{N G_{\text{T}} + G_{N+1}} \tag{7}$$

The sign of power excursions depends on the gain of the newly added channel; if $G_{N+1} > G_{\text{T}}$ it gives $\Delta P = \Delta G < 1$, which means that the added channel has a gain higher than the target gain and to obtain the same target gain all other channels will have a negative excursion. An example of a positive excursion, where $G_{N+1} < G_{\text{T}}$ resulting in $\Delta G > 1$ can be seen in Fig. 2. When only one channel is active, λ_1, the gain curve is set in one location at G_{T}, but when a second channel, λ_2, is added that has a lower output power the EDFA AGC tries to maintain the same mean gain and it results in the gain profile moving up by ΔG.

Using the analytical solutions above shows that all previously active wavelengths should result in the same ΔG, but experimentally we see that this is not the case. Effects such as EDFA gain tilt and spectral hole burning result in other power dynamics in the system.

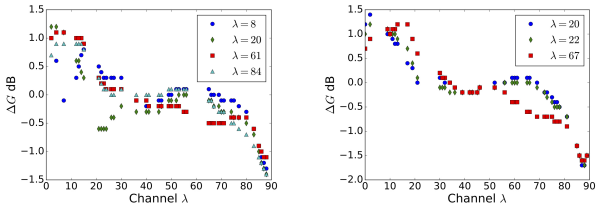
In addition, while the input power at the initial ROADM might be the same, when multiple EDFAs are used together and not all channels are active along the entire path, predicting the resulting power excursions can be difficult using analytical solutions. For example, for the experimental testbed where all the EDFAs contain the same active channels it can result in excursions greater than 3 dB.

4 EXPERIMENTAL SETUP

Fig. 3 shows the 5-ROADM experiment setup for studying the optical circuit switching wavelength assignment. The network nodes are separated by 4 standard Single Mode Fiber (SMF) spans, and each span has two EDFAs to compensate for the loss of the ROADM and the transmission fiber. Channel power excursions due to channel add/drop arise from the EDFAs' different wavelength dependent gain characteristics (Fig. 3 subset shows the two EDFAs' wavelength dependent gain of the first span).

Each ROADM node is built using 1x2 or 1x4 WSSs with per-channel VOA, which is typical for high performance gain flattened DWDM line amplifiers, and supports wavelength add and drop. Each drop port is connected to an Optical Channel Monitor (OCM) for channel power excursion measurement. A 90-channel 50 GHz-spaced DWDM source is used to create different channel configurations over the full C-band. The 90-channel DWDM source is connected to a 1x4 coupler, which allows for adding channels into the ROADMs.

With 90-channel DWDM input, the VOA of each ROADM is initially tuned to ensure launch power into the fiber span at 0 dBm per-channel and the VOA settings are stored as a reference. This configuration mitigates the channel power divergence due to wavelength dependent gain in the EDFAs and Stimulated Raman Scattering (SRS) in the transmission fiber. However, power dynamics due to



(a) (b)

Figure 4: Experimental results of adding a new channel to different network settings.

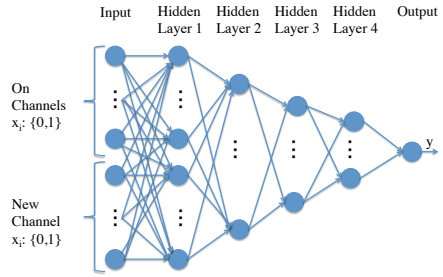


Figure 5: Neural Network structure with hidden layers 1, 2, 3, and 4 containing 180, 120, 30, and 15 neurons, respectively.

wavelength dependent gain effects described by Section 3, as well as other dynamics due to complex effects such as spectral hole burning, excited state absorption, and nonlinear interactions in the fiber spans will occur during optical circuit switching.

To see how the experimental effects differ from the analytical model, the EDFA AGC response is plotted for two different dynamic network scenarios. Fig. 4 reveals the ΔG_i that results when adding a channel to wavelength λ_j for different sets of active channels. The first example in Fig. 4(a) shows the excursions that occur when there are initially 4 random active channels at channels 8, 20, 61, and 84. It can be observed that $\lambda = 20$ behaves uniquely compared to the other three active channels when a channel is added with λ from 21 to 30. The second example in Fig. 4(b) shows the excursions that occur when there are initially 3 random active channels at channels 20, 22, and 67. It shows $\lambda = 67$ behaves distinctly when adding a channel with λ from 15 to 20 and 50 to 80 than for $\lambda = 20$ or $\lambda = 22$. These are just two examples of the numerous combinations of active channels that will result in different gain variations that would not be easily predicted by analytical solutions. Therefore, we develop a machine learning model to predict the EDFA power excursions where the analytical model could not.

To evaluate power excursion prediction and wavelength assignment accuracy when using machine learning algorithms, and particularly the neural network, we took 2,100 measurement sets with random initial channel locations and numbers. Note, that in this work all channels go through the same longest route over 4 spans, but the neural network is scalable to a larger set of input features for a multi-route case with more measurement sets. Each set contains 40 power excursion measurements as follows: 40 available wavelength channel positions for adding a new channel are randomly selected, and the dynamic power response is measured on all active channels by switching on and off these 40 positions one by one. For each measurement, only the launch power of active channels are leveled to 0 dBm per-channel, while the new channels use the pre-stored VOA reference.

5 MACHINE LEARNING TECHNIQUES AND ACCURACY MEASUREMENTS

A supervised machine learning algorithm is designed to model the dynamic optical power response caused by the EDFAs and AGC in optical circuit switching. Due to the different effects of active channels and new channels, they are separate input features for the machine learning algorithm. The output that we aim to predict, y is the maximum power excursion among all active channels, $\max\Delta P_i$. Neural networks are useful for learning when there are a large number of input features. Therefore, we develop a Neural Network (NN) architecture utilizing Tensorflow [4] to predict power excursions of the ROADM system. The problem of predicting the excursion which will occur for different network scenarios is treated as a regression problem.

The data collected in the experiments is divided into 3 classes: training data, validation data, and test data. The training data is used to train the neural network to minimize the MSE loss functions. The validation data is used to determine which parameters provide optimal performance after using the training data. We evaluate the performance of the neural network by inputting test data that has not been seen before and compare the predicted output to the actual output that occurred. For comparison we used Ridge Regression (RR) as proposed in [6] (the Kernel Bayesian Regression in [6] is not used for comparison, as it does not scale well to a large number of features or training samples). The regularization parameter was tuned to 0.01 through cross validation. The data contains 2,100 cases with 84,000 samples, where 1,680 cases (67,200 samples) are used for training, 210 cases (8,400 samples) are used for validation, and 210 cases (8,400 samples) are used for testing.

The NN depicted in Fig. 5 includes 4 hidden layers using a combination of tanh and linear activation functions. The validation data was used to compare the performance of different neural network architectures. Several neural network architectures were tested with a varying number of layers, number of neurons, and combinations of activation functions (e.g., Sigmoid and ReLU activation functions). The input to the neural network is represented with 180 features by a 180-bit binary array, where a 1 represents that the wavelength channel was active and a 0 represents that it was off. The first 90 values are used to represent the wavelength channels that were initially active. One value of the next 90 inputs is set to 1 for the new wavelength that would be added.

The model is trained by minimizing the Mean Square Errors (MSE) loss function using stochastic gradient descent with back-propagation with a mini batch size of $m = 60$ and a learning rate, α of 0.005 over a million iterations. The error function of the neural network with weights w is defined as:

$$E_{\text{MSE}}(y, w) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - f(x_i, w))^2 \tag{8}$$

Predictions of optical power excursions are useful in two main scenarios: (i) minimizing lightpath setup time for services with high priority, and (ii) recommending a wavelength channel under optical

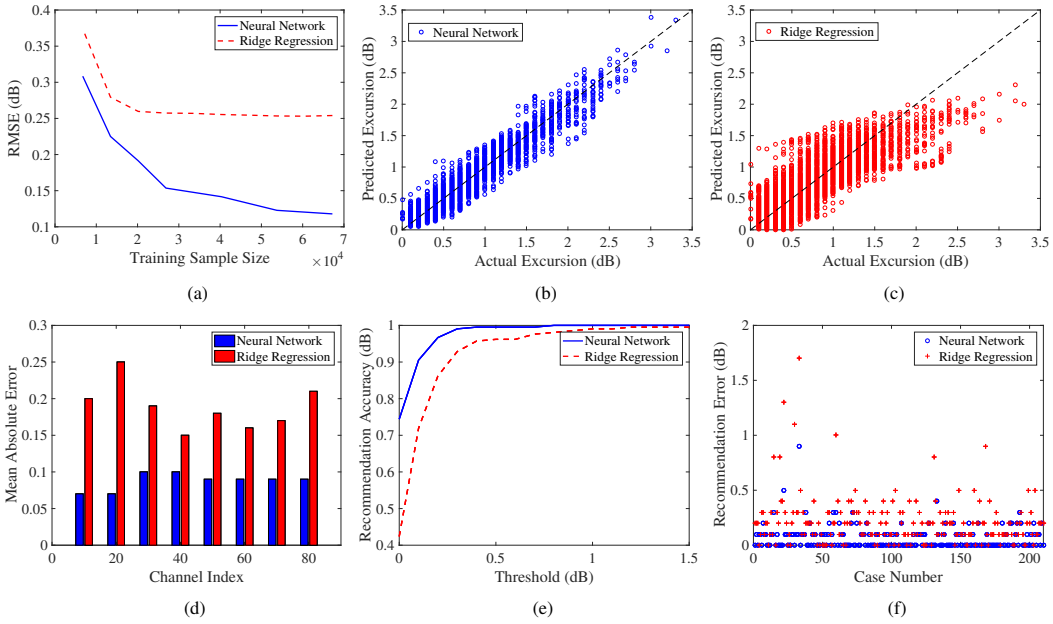


Figure 6: Experimental results: (a) RMSE of test data. (b) Neural network: Prediction vs. Actual. (c) Ridge regression: Prediction vs. Actual. (d) MAE of individual channels. (e) Recommendation accuracy vs. threshold from optimal. (f) Recommendation error.

circuit switching operations that does not generate a large excursion on active channels above a threshold as to keep power fluctuations within allocated margins. We define δ-Recommendation Accuracy as the fraction of cases in which the neural network recommends a new channel that is within δ of the optimal channel's power excursion.

The second metric is Classification Accuracy, which is the True Positive Rate (TPR) (also known as Recall) for a given False Positive Rate (FPR) and a threshold of the power excursion. TPR is the probability of making a correct prediction with an excursion below the threshold, and FPR is the probability of a prediction incorrectly giving a positive test result to an actual fault beyond the threshold.

A third metric used to measure the accuracy of the recommendation is PSRR, defined as the precision at specific recall rate and threshold for power excursion. Precision is the ratio of true positives to the number of total positive values predicted. For our system it is much more important to minimize the possibility of adding a wavelength with a large excursion (False Positive) than it is to miss a possible valid wavelength (False Negative). Thus maximizing the PSRR is important for reliable optical system operation.

6 EXPERIMENTAL RESULTS

In Fig. 6(a), the resulting test data Root Mean Square Error (RMSE) is compared versus the number of samples used for training. It can be seen that NN outperforms RR by more than a factor of two when using the full $67,200$ training samples. The error between the actual excursion and the predicted excursion for the test data can be viewed in Figs. 6(b) and 6(c), with the black dashed line indicating perfect prediction of power excursions, which varies from 0 to 3.3 dB. It can be seen that NN obtains significantly better predictions than RR for all potential excursion outputs. The latter has a higher deviation from the experimental output and the excursion predictions are underestimated when the actual excursion is above 2 dB. Another metric for

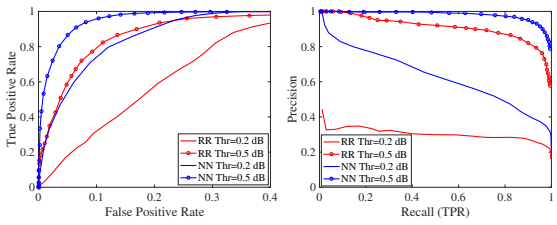


Figure 7: Performance metrics: (a) FPR vs. TPR (b) Recall vs. Precision.

prediction performance evaluation is Mean Absolute Error (MAE) on each potential wavelength channel position, which reveals the average magnitude of the errors for each individual channel. Fig. 6(d) shows the MAE of 8 channels. NN significantly outperforms RR as it keeps the average prediction error on each channel below 0.1 dB.

For evaluating the δ-Recommendation Accuracy, we consider the performance in predicting the channel with the minimal power excursion for the 210 test cases. In Fig. 6(e), the δ threshold is varied and the resulting δ-Recommendation accuracy is viewed for both NN and RR. When the threshold δ is set to 0, the NN was able to pick the best channel of 40 candidates over 70% of the time, while RR did it only 42%. When δ is set to 0.3 dB the accuracy of RR is 92%, while NN is over 99% with only 2 exceptional cases. The recommendation error for each of the 210 cases is in Fig. 6(f) for which NN outperforms RR in both the recommendations and the maximum error.

To evaluate the Classification Accuracy, in Fig. 7(a) the FPR is plotted versus the corresponding TPR for different thresholds. Fig. 7(a) shows that the NN is able to better minimize the false positives, while still determining a large portion of the available channels. The NN can contain the excursions within 0.5 dB threshold

with the TPR of 76% while obtaining a FPR less than 1%. On the other hand, the RR only has a 25% TPR for the same FPR and threshold.

The evaluation of the precision as a function of the recall (for different thresholds), is shown in Fig. 7(b). Fig. 7(b) reveals that as we cast a larger net to increase the recall of finding valid wavelenghts it also causes the system to obtain more false positives. *The NN can predict that all excursions will be less than* 0.5 *dB with a precision of over* 99%*, while obtaining a recall of greater than* 55%. To obtain this precision, only channels with predicted excursions below 0.33 dB are used. This results in a few false positives, but also in missing roughly 45% of potential available channels. In order to improve the recall rate a reduction in the RMSE is needed. In comparison for RR to obtain the same precision it would only have a recall of less than 14%. If the system is looking to predict channels that will cause an excursion of less than 0.2 dB, NN can obtain a precision of 80%, with a recall of approximately 15%, while RR can not obtain that precision.

7 CONCLUSION AND FUTURE WORK

A neural network is developed to predict the dynamic optical power response of a 90-channel multi-hop ROADM system, containing 4 spans and 8 EDFAs with different gain characteristic. The neural network has an RMSE and MAE below 0.15 dB. Both are a factor two improvement over ridge regression. The neural network performs efficiently for the large data set of a full C-band multi-hop ROADM system with wavelength assignment precision of 99%. The neural network can be implemented in dynamic optical networks to allow for quick wavelength assignment. This work is a first step in applying neural networks to rapidly route dynamic traffic demands. In future work, we will investigate machine learning in optical mesh and ring topologies, as well as implementation in software defined optical networks.

8 ACKNOWLEDGMENTS

This work was supported by the NSF Center for Integrated Access Networks (CIAN) EEC-0812072 and NSF grants CNS-1423105, CNS-1650685, PFI-1601784, and DGE-16-44869.

REFERENCES

[1] 2013. *Infinera and DANTE Set Guinness World Record on GÉANT Network Record set for fastest provisioning of long haul optical transmission capacity*. Technical Report. Infinera.

[2] 2016. *The Zettabyte Era - Trends and Analysis - Cisco*. Technical Report. Cisco.

[3] 2017. openroadm. (Mar 2017). Retrieved March 10, 2017 from http://www.openroadm.org/

[4] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv:1603.04467* (2016).

[5] Pedro Amaral, Joao Dinis, Paulo Pinto, Luis Bernardo, Joao Tavares, and Henrique S Mamede. 2016. Machine learning in software defined networks: data collection and traffic classification. In *Proc. IEEE ICNP'16*.

[6] Marianna Angelou, Siamak Azodolmolky, Ioannis Tomkos, Jordi Perelló, Salvatore Spadaro, Davide Careglio, Kostas Manousakis, Panagiotis Kokkinos, Emmanouel Varvarigos, Dimitri Staessens, et al. 2012. Benefits of implementing a dynamic impairment-aware optical network: results of EU project DICONET. *IEEE Commun. Mag.* 50, 8 (2012).

[7] Berk Birand, Howard Wang, Keren Bergman, Dan Kilper, Thyaga Nandagopal, and Gil Zussman. 2014. Real-time power control for dynamic optical networks-Algorithms and experimentation. *IEEE J. Sel. Areas Commun.* 32, 8 (2014), 1615–1628.

[8] Berk Birand, Howard Wang, Keren Bergman, and Gil Zussman. 2013. Measurements-based power control-a cross-layered framework. In *Proc. OFC'13*.

[9] Peter Bodík, Rean Griffith, Charles A Sutton, Armando Fox, Michael I Jordan, and David A Patterson. 2009. Statistical machine learning makes automatic control practical for internet datacenters. In *Proc. USENIX HotCloud'09*.

[10] Angela L Chiu, Gagan Choudhury, George Clapp, Robert Doverspike, Mark Feuer, Joel W Gannett, Janet Jackel, Gi Tae Kim, John G Klincewicz, Taek Jin Kwon, et al. 2012. Architectures and protocols for capacity efficient, highly dynamic and highly resilient core networks. *IEEE J. Opt. Commun. Netw.* 4, 1 (2012), 1–14.

[11] Paulo Cortez, Miguel Rio, Miguel Rocha, and Pedro Sousa. 2006. Internet traffic forecasting using neural networks. In *Proc. IEEE IJCNN'06*.

[12] Robert D Doverspike and Jennifer Yates. 2012. Optical network management and control. *Proc. IEEE* 100, 5 (2012), 1092–1104.

[13] Steven Gringeri, Nabil Bitar, and Tiejun J Xia. 2013. Extending software defined network principles to include optical transport. *IEEE Commun. Mag.* 51, 3 (2013), 32–40.

[14] Chi-Yao Hong, Srikanth Kandula, Ratul Mahajan, Ming Zhang, Vijay Gill, Mohan Nanduri, and Roger Wattenhofer. 2013. Achieving high utilization with software-driven WAN. In *Proc. ACM SIGCOMM'13*.

[15] Yishen Huang, Craig L Gutterman, Payman Samadi, Patricia B Cho, Wiem Samoud, Cédric Ware, Mounia Lourdiane, Gil Zussman, and Keren Bergman. 2017. Dynamic mitigation of EDFA power excursions with machine learning. *Optics Express* 25, 3 (2017), 2245–2258.

[16] Kiyo Ishii, Junya Kurumida, and Shu Namiki. 2016. Experimental investigation of gain offset nehavior of feedforward-controlled WDM AGC EDFA under various dynamic wavelength allocations. *IEEE Photon. J.* 8, 1 (2016), 1–13.

[17] Sushant Jain, Alok Kumar, Subhasree Mandal, Joon Ong, Leon Poutievski, Arjun Singh, Subbaiah Venkata, Jim Wanderer, Junlan Zhou, Min Zhu, Jonathan Zolla, Urs Holzle, Stephen Stuart, and Amin Vahdat. 2013. B4: Experience with a globally-deployed software defined WAN. In *Proc. ACM SIGCOMM'13*.

[18] Xin Jin, Yiran Li, Da Wei, Siming Li, Jie Gao, Lei Xu, Guangzhi Li, Wei Xu, and Jennifer Rexford. 2016. Optimizing bulk transfers with software-defined optical WAN. In *Proc. ACM SIGCOMM'16*.

[19] Joseph Junio, Daniel C Kilper, and Vincent WS Chan. 2012. Channel power excursions from single-step channel provisioning. *IEEE J. Opt. Commun. Netw.* 4, 9 (2012), A1–A7.

[20] Daniel Kilper, Kyle Guan, Kerry Hinton, and Robert Ayre. 2012. Energy challenges in current and future optical transmission networks. *Proc. IEEE* 100, 5 (2012), 1168–1187.

[21] Daniel C Kilper, M Bhopalwala, H Rastegarfar, and W Mo. 2015. Optical power dynamics in wavelength layer software defined networking. In *Proc. Photonic Networks and Devices'15*.

[22] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. 2013. Playing atari with deep reinforcement learning. *arXiv:1312.5602* (2013).

[23] Lynn E. Nelson, Guodong Zhang, Narayan Padi, Craig Skolnick, Kathleen Benson, Troy Kaylor, Satoshi Iwamatsu, Robert Inderst, Fabio Marques, Daniel Fonseca, Mei Du, Terry Downs, Thor Scherer, Christopher Cole, Y Zhou, Paul Brooks, and Andreas Schubert. 2017. SDN-Controlled 400GbE end-to-end service using a CFP8 client over a deployed, commercial flexible ROADM system. In *Proc. OFC'17*.

[24] Masaki Shiraiwa, Hideaki Furukawa, Takaya Miyazawa, Yoshinari Awaji, and Naoya Wada. 2015. Concurrently establishing and removing multi-wavelength channels reconfiguration system: Implementation for a dynamic and agile next-generation optical switching network. In *Proc. IEEE PS'15*.

[25] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529, 7587 (2016), 484–489.

[26] Thomas A. Strasser and Jefferson L. Wagener. 2010. Wavelength-selective switches for ROADM applications. *IEEE J. Sel. Topics Quantum Electron.* 16, 5 (2010), 1150–1157.

[27] Yong Wu, Luca Chiaraviglio, Marco Mellia, and Fabio Neri. 2009. Power-aware routing and wavelength assignment in optical networks. In *Proc. IEEE ECOC'09*.

[28] Tiejun J. Xia, Glenn Wellbrock, Bert Basch, Scott Kotrla, Wang Lee, Tsutomu Tajima, Kiyoshi Fukuchi, Milorad Cvijetic, Jim Sugg, Yin K. Ma, et al. 2010. End-to-end native IP data 100G single carrier real time DSP coherent detection transport over 1520-km field deployed fiber. In *Proc. OFC'10*.

[29] Ying Zhang and Ake Årvidsson. 2012. Understanding the characteristics of cellular data traffic. In *Proc. ACM SIGCOMM'12 workshop on Cellular networks: operations, challenges, and future design*.