

Machine-learning-based EDFA gain estimation [Invited]

JIAKAI YU,^{1,*} SHENGXIANG ZHU,¹ CRAIG L. GUTTERMAN,² GIL ZUSSMAN,² AND DANIEL C. KILPER¹

¹College of Optical Sciences, The University of Arizona, Tucson, Arizona 85721, USA

²Department of Electrical Engineering, Columbia University, New York, New York 10027, USA

*Corresponding author: jiakaiyu@email.arizona.edu

Received 14 December 2020; revised 11 January 2021; accepted 12 January 2021; published 12 February 2021 (Doc. ID 417584)

Optical transmission systems with high spectral efficiency require accurate quality of transmission estimation for optical channel provisioning. However, the wavelength-dependent gain effects of erbium-doped fiber amplifiers (EDFAs) complicate precise optical channel power prediction and low-margin operation. In this work, we examine supervised machine learning methods using multiple artificial neural networks (ANNs) to build models for gain spectra prediction of optical transmission line EDFAs under different operating conditions. Channel-loading configurations and channel input power spectra are used as an a posteriori knowledge data feature for model training. In a hybrid learning approach, estimated gain spectra calculated by an analytical model are added as an a priori input data feature to further improve the EDFA ANN model performance in terms of prediction accuracy, training time, and quantity of training data. Using these methods, the root mean square error and maximum absolute error of the predicted channel output power can be as low as 0.144 dB and 1.6 dB, respectively. © 2021 Optical Society of America

<https://doi.org/10.1364/JOCN.417584>

1. INTRODUCTION

The rapidly growing traffic and ultra-low latency demands in 5G networks, stemming from various applications and services, such as mobile edge cloud computing, augmented/virtual reality, and self-driving vehicles, require increased capacity of the underlying optical networks. As a result, advanced modulation formats, such as 400 Gb/s dual-polarization quadrature phase-shift keying (DP-QPSK) and 800 Gb/s dual-polarization 32 quadrature amplitude modulation (DP-32QAM), are used to increase the optical fiber channel capacity. In addition, optically amplified transparent optical networks, in which the signals are kept in the optical domain as much as possible, reduce dependence on higher latency and higher cost optical-electrical-optical conversion and electronic processing along the signal path [1].

However, higher-order modulation formats need tight quality-of-transmission (QoT) control to maintain error-free operation in optical transmission. Transparent optical networks contend with physical impairments, such as fiber nonlinearities, dispersion, and optical amplifier (OA) noise, which degrades the QoT [2]. The QoT of the optical signal needs to be estimated before an optical channel is provisioned. Generally, QoT models are based on a physical model that tracks the signal power, linear noise, and nonlinear noise levels for an end-to-end light path, to estimate the optical signal-to-noise ratio (OSNR), generalized OSNR (gOSNR), Q factor,

or bit error rate (BER). The accuracy of the QoT model is a key factor in determining the engineering margins that must be achieved to ensure error-free operation. These margins typically account for the difference between the actual QoT and the recoverable error-free signal threshold for the forward error correction (FEC) coding, plus the uncertainties in the QoT estimation and other factors. Larger margins are taken into account for the estimation error either from the QoT model itself or the input parameters to the model [3,4]. Thus, reducing uncertainties in the QoT estimation is an important strategy for achieving low-margin operation.

The erbium-doped fiber amplifier (EDFA) is one of the most widely used OAs in optical transmission systems [5]. EDFAs can have a strong impact on QoT estimation because their gain spectrum determines the power levels of each individual channel in a transmission fiber. The launched channel power spectrum in each fiber span is the main factor in determining the magnitude of nonlinear fiber impairments. In addition, amplified spontaneous emission noise from the EDFAs is the main determinant of the OSNR. In general, system design must balance these two effects, increasing power to improve the OSNR, while reducing it to avoid nonlinear fiber impairments. This leads to an optimal launch channel power for each signal, depending on its modulation properties and other performance requirements. Furthermore, the optical channel power excursions induced by EDFA dynamic gain

effects require active control in each node to maintain channel powers near the desired design point. Therefore, the ability to accurately predict the output channel power spectrum of each EDFA is a critical factor with potential to enable low-margin optical networks.

There are many studies on the characterization of dynamic EDFA wavelength-dependent gain effects. Gain-flattening filters are used in wideband line EDFAs, but significant wavelength-dependent gain, often as high as ± 1 dB, can persist over the full range of operating conditions. Furthermore, spectral hole burning and other nonlinear gain effects, as well as intentional gain tilt can also contribute to this gain variation. Numerical models exist to account for the detailed transient time-dependent effects and nonlinear gain phenomena [6]. However, the steady-state wavelength dependent gain effects in constant gain-controlled amplifiers can be characterized using analytical models including different levels of detail [7,8]. Due to their computational efficiency, these models are attractive for use in QoT estimation during channel-provisioning operations. Each of these models, including the full numerical treatment, are dependent on the detailed data of the internal (inaccessible to users) and external (accessible to users) amplifier characteristics. Lookup tables of the amplifier properties under different operating conditions can be developed for particular design operating points, but the full range of potential operating conditions across all of the different channel configurations and gain settings [9] are problematic to test exhaustively. Recently, machine-learning-based modeling has been studied in optical networks [10]. Deep neural networks were used to reduce network margins through end-to-end OSNR prediction [11]. Rather than using inaccurate EDFA characterization data, this work considers amplifiers as black-boxes and instead learns the end-to-end behavior. Several studies using machine learning have focused on amplifier models for dynamic channel power estimation and power divergence prediction in EDFAs [12–15] and Raman amplifiers [16,17]. Machine learning (ML) efficiently makes use of extensive data collection on the amplifiers in a lab setting (e.g., prior to deployment) to determine a more accurate, but computationally simple model for use in QoT estimation. However, these ML models are built solely from a posteriori knowledge abstracted by the models themselves. Existing a priori knowledge such as mathematical and physical equations that can include lightweight or heavyweight relationships between individual data features is ignored.

We have reported our EDFA artificial neural network (ANN) models in previous conference papers [18,19]. In this work, we describe in more detail how we developed the ML algorithms to determine the channel configuration and input-power-dependent EDFA gain spectra. The modular functionality of the controller design for efficient automated EDFA data collection is reported here along with additional experimental details. As described in previous papers, the ANN is built and trained to predict the gain spectrum based on the input power spectrum and gain settings. Further, as previously reported, we developed a hybrid machine learning (HML) model for EDFAs, which utilizes an analytical model as an additional input feature to achieve higher prediction accuracy, while reducing the training complexity in both training time

and training sample size. The trained ANN model for each individual EDFA provides a computational tool for use in QoT estimation. Thus, by employing the ANN HML model to characterize each amplifier prior to deployment, accurate wavelength-dependent-gain models are made available for improving the QoT estimation.

2. MODEL FUNDAMENTALS

This section introduces the fundamentals of the mathematical model and artificial neural networks that we used to build the ML EDFA output power spectrum model.

A. Wavelength Division Multiplexing Channel Gain Models

The addition or deletion of wavelength channels can affect channels already provisioned in the network. Most EDFAs use automatic gain control (AGC) to maintain a constant target gain, which relies on the total optical power gain, not the individual channel gains. With varying channel configurations, the total channel power gain is adjusted to reach the target level through the AGC action. The individual channel power gain deviates from the target due to the gain variations in the spectrum. As the EDFA gain changes, the wavelength-dependent gain will tilt as the internal gain deviates from the design gain used to create the gain-flattening filter. Thus, the gain spectrum varies as the channel loads and the gain settings change. Furthermore, an internal variable optical attenuator (VOA) is often used to change the internal gain, while maintaining the total gain at the same level to tilt the gain spectrum. To illustrate this, we show in Fig. 1 the measured EDFA gain spectra for nine different two-stage gain-flattened, AGC-controlled dense wavelength division multiplexing (WDM) line JDS Uniphase EDFAs (P/N 22068325, 2014 model) with the 18 dB target gain, 7 dB mid-stage loss, and -20 dBm per channel input power. The internal VOAs of the amplifiers were adjusted by setting the tilt parameter for each amplifier to 0 dB. Although the amplifiers are of identical make and model, their gain profiles or gain ripple functions vary by roughly ± 0.5 dB. The output of EDFA #4 is lower by 0.5 dB, and therefore in service, the gain would be increased, modifying the operating conditions. As a result, channel power dynamics due to the wavelength-dependent gain of each EDFA can cause provisioned light paths to vary in optical power [7] and result in inaccurate quality of transmission estimation.

The EDFA gain dynamics during the AGC control can be well characterized with accurate and detailed model parameters. The gain excursion $\hat{g}(\lambda_i)$ of wavelength λ_i from the target AGC gain due to the input power P_j configuration in the wavelength channels λ_j can be written as [7]

$$\hat{g}(\lambda_i) = \frac{G_{TC}}{G_M} \left[\frac{\sum_j P_j + N_I + N_C}{\sum_j P_j g_j t_j + g_R N_R + g_I N_I} \right] g_m(\lambda_i), \quad (1)$$

where G_{TC} is the target gain, g_j is the residual gain ripple about the mean gain G_M , and t_j is the tilt for each channel j . $g_m(\lambda_i)$ is the original measured gain of channel λ_i before the new input powers P_j are applied. The average incident noise gain ripple

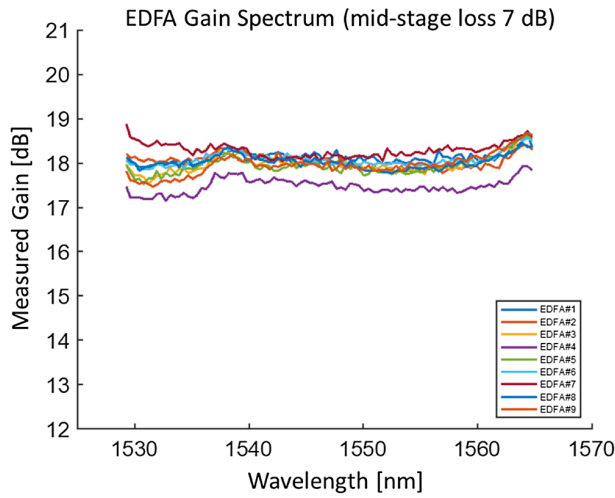


Fig. 1. Measured EDFA gain spectra at the 18 dB target gain with a 7 dB mid-stage loss when all 90 channels are lit.

is g_i , average input-referred noise gain ripple is g_R (the noise generated by the EDFA itself), the total input noise is N_I , the total amplifier input-referred noise is N_R , and the amplifier AGC noise compensation factor is N_C . Many of these factors including g_i and t_i are dependent on the input channel configuration through the internal amplifier gain and its nonlinear channel power-dependent effects such as spectral hole burning and excited state absorption [6].

Although the model Eq. (1) can characterize the dynamic EDFA gain spectrum with a worst-case accuracy ~ 0.2 dB and within 0.1 dB under most conditions, it requires many input parameters that are difficult to characterize. Equation (1) describes a center of mass (CM)-balancing effect of the different wavelength channels about the mean output power level. A simplified version of Eq. (1) was developed based on a set of basic characterization measurements and is given by [8]

$$\hat{g}(\lambda_i) = g(\lambda_i) + \frac{\sum_{j=1}^n [g_s(\lambda_j) - g(\lambda_j)]}{n}, \quad (2)$$

where $\hat{g}(\lambda_i)$ is the gain of wavelength λ_i when a set of wavelengths $\{\lambda_1, \dots, \lambda_n\}$ are input to the EDFA, $g(\lambda_i)$ represents the gain of channel λ_i when all WDM channels are lit, and $g_s(\lambda_j)$ denotes the gain spectrum when only a single channel λ_j is present. Although the CM model Eq. (1) is a more accurate physical model compared with the simplified CM model Eq. (2), model Eq. (2) conveniently provides a method to estimate the EDFA gain spectrum from easily measured configurations. An ANN model potentially offers a way to use similar additional collected data to obtain a more accurate model. However, the training time to cover a wide range of operating conditions could still be a factor, and therefore, more efficient training methods are needed. The simplified CM model Eq. (2) can also be a solution to improve the ANN model performance as it describes the CM effect in the EDFA gain spectra.

B. Artificial Neural Networks

The ANN is a mathematical model widely used in ML that mimics the human brain's learning and memory functions provided by neurons to solve problems with complex nonlinearity [20]. It is generally a flexible parametric model for regression or classification to capture data distributions when correct data features and sufficient datasets are available. The ANN can automatically fit the data distribution by tuning the neuron weights in each model layer with simple hyperparameters such as the number of layers, number of neurons in each layer, activation function, and learning rate. Commonly, the structure of the ANN model is composed of three types of layers: input layer, hidden layers, and output layer. The input layer consists of many neurons corresponding to distinct input data features. Multiple hidden layers are connected with the input layer to bring together all features from the input layer and form neural networks. An activation function is applied in each neuron from the input layer to the output layer. Thus, a non-linear transformation is added on the weighted linear combinations of all inputs from previous layers. Most commonly used activation functions include sigmoid, hyperbolic tangent, and rectified linear unit (ReLU). The output layer generates the target values following the generated marginal distribution probability from the network of neurons.

To obtain an accurate ANN model, a learning process is executed to determine the ANN parameters. As Fig. 2 shows, supervised learning is a commonly used training process in which input data features and target output data values are used to adjust the weights of neurons in each layer and find the optimal weights by supervising a loss function. A supervised ANN model is initialized with a predefined structure where random or customized weights are given for each layer. When the input training data forward propagate through the neural networks, the results are generated by the output layer. The loss function will first calculate the deviation between generated outputs and the given reference outputs. Then, the total loss is minimized by decreasing the loss contributed by each weighted neuron where the derivative of the loss function regarding all neurons' weights is evaluated from the last to the first layer. Stochastic gradient descent (SGD) is implemented to tune neurons' weights and minimize the prediction error of the ANN model. The feedforward and backpropagation of the training datasets across the neural networks are repeated so that neurons' weights are updated to obtain the least global loss at each learning process iteration. After certain learning process

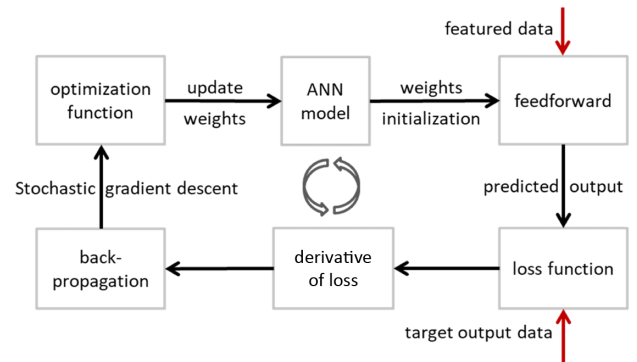


Fig. 2. Artificial neural networks supervised learning process.

iterations, the complete ANN model is applied to new input datasets to predict the target values.

3. EXPERIMENTAL SETUP AND EDFA ANN ML MODELS

In this section, we present the experimental testbed in which the EDFA input/output power spectrum data are collected, and the automated data collection scheme is implemented. The collected data samples include the EDFA power spectrum with varying channel-loading configurations (channel on/off status, input channel power). Then two supervised ANN ML regression models are constructed for model training to estimate the output channel power levels based on input channel power levels. The original EDFA ANN model is trained with only the input channel power spectrum data as the input model parameters. The hybrid EDFA ANN model is trained with the addition of the channel gain calculated by the simplified CM analytical model.

A. Experimental Setup and Data Collection

Figure 3 shows the EDFA data collection setup built using a Nistica wavelength selective switch (WSS) (FFLB-C58L200-1NI) and JDS Uniphase EDFAs (P/N 22068325). The experimental setup consists of a comb source that generates a full set of optical channels (90 channels) with a flat output spectrum. The Nistica WSS turns individual channels on/off and tunes the power of individual channels to within 0.1 dB accuracy. A two-stage DWDM line amplifier (EDFA) is the device under test (DUT). Two optical channel monitors (OCM) that are integrated with the WSS are used to capture the EDFA input and output power spectra, respectively [19]. The OCMs are calibrated using separate measurements with a powermeter and optical spectrum analyzer to take into account the tap (and output attenuator not shown) and connector losses, to obtain the powers at the input and output of the DUT EDFA. The channel gains are calculated from the linear ratio of these powers and converted to decibels. An additional EDFA shown with dashed connections is inserted to emulate input noise when measuring the input noise dependence. Using WSS controls and the line system OCMs to train the ANNs ensures that they are trained and tested with the same devices used in the commercial line systems.

The comb source paired with a WSS—to turn on/off specific channels and tune the optical power for each “on” channel—and an EDFA—for adding dynamic input amplified spontaneous emission (ASE) noise together—provides arbitrary input-channel-loading configurations, where the channel on/off status and its power level are accurately controlled.

A controller is designed to obtain automatic data collection for ML model training and evaluation. There are several functional modules in the controller, including comb source power control, EDFA gain/tilt adjustment, WSS channel on/off control, WSS attenuation setting, and OCM data collection. The automatic data collection process controlled by the controller is illustrated with the following steps: (1) turn on the comb source to ensure that each output optical channel has power above -10 dBm; (2) set the target gain/tilt for the EDFA

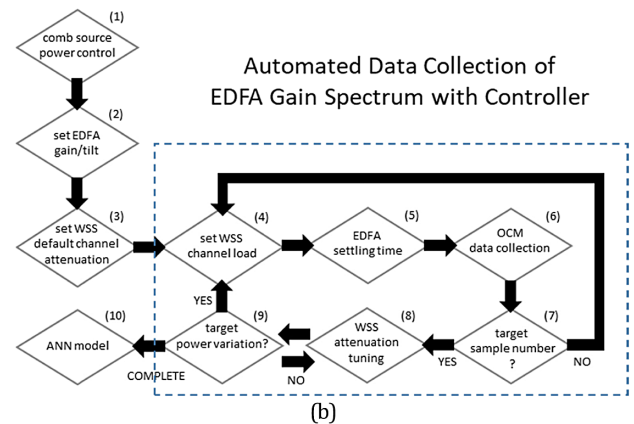
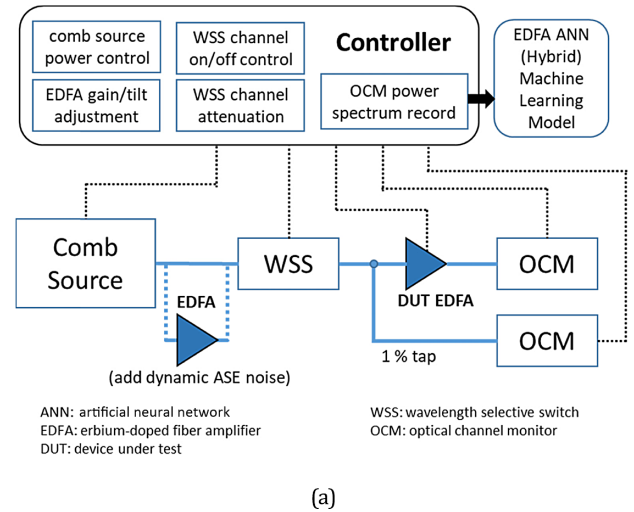


Fig. 3. (a) Experimental setup for automatic data collection. (b) Control logic of automatic EDFA data collection.

(e.g., 18 dB gain with 3 dB tilt); (3) set the WSS to a default channel attenuation (i.e., 6 dB attenuation for each channel) to leave enough space for power adjustment; (4) set the WSS with 90 channels lit once, a single channel lit once for each channel, and then random channel loading configurations; (5) wait 30 s for the EDFA gain to settle; (6) collect the EDFA input/output power spectrum from the OCMs; (7) detect if the total number of collected data reaches the target (i.e., 2500 samples); (8) adjust the attenuation in an adaptive manner to generate the channel input power variation; (9) detect if the actual power reaches the target value (i.e., the difference between the actual power and target power is below a threshold such as 0.1 dB); and (10) when the data collection is completed, all data are sent to the ANN ML model for model training and evaluation. Steps (4)–(7) are repeated to collect data with various channel loads at a given channel power-level until the total number of collected samples reaches the target. Then Steps (8)–(9) are triggered to collect data with various channel input power levels.

For the DUT EDFA configured with the 18 dB gain, there are 15,000 collected samples with random channel-loading configurations with channel input power variations, 90 samples for the single-channel gain function, and one sample for the fully loaded gain function. Often, EDFAs are operated

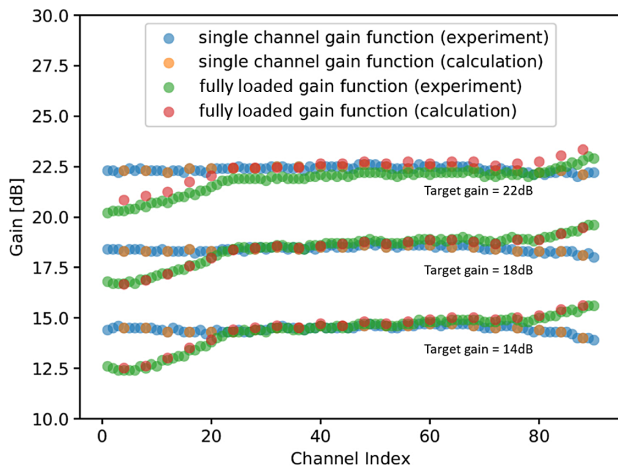


Fig. 4. Measured and calculated (using the simplified CM analytical model on 22 of the 90 channels) EDFA gain spectrum with target gain values of 14, 18, and 22 dB. The single channel gain function is the gain spectrum of individual channels (only the measured individual channel is turned on while the others are all turned off), which are measured one by one and then plotted together. The fully loaded gain function is the gain spectrum with an input of fully loaded optical channels (all channels are turned on).

with a tilted gain to compensate for stimulated Raman scattering. Here, the EDFA tilt was set to 3 dB to evaluate the performance for a relatively high tilt setting under which the amplifier-channel-loading dependence would be maximized. Lower tilt settings would exhibit less variability with channel loading based on previous studies [7]. For the simplified CM model, the g and g_s gain parameters must be obtained for different target gain settings, either by measurement or interpolation of measurements at other gain settings. Figure 4 shows examples of measured gain spectra for an EDFA with 14, 18, and 22 dB target gain values, each including a single loaded channel at the channel index varying from 1 to 90 and the fully loaded gain function for which the input is the full spectral load of 90 channels. To obtain the single-channel gain spectrum, we did 90 gain measurements where only a single channel was lit. The fully loaded gain spectrum was measured when all 90 channels were lit. The simplified CM model calculation results using these spectra are also presented for 22 of the 90 channels, singly and with all 22 turned on. Note that the calculation with 22 uniformly distributed channels is slightly different from the 90-channel case because the effective center of mass of the 22-channel case differs from the 90-channel case, mostly from the gain difference between the channels at the two ends of the spectrum. In each case (Fig. 4), the input power level is fixed at -18 dBm for all lit channels. For the 18 dB gain case, the total output power has a maximum value of approximately 20 dBm (channels fully loaded with 0 dBm output per channel). For the maximum gain of 22 dB, the total output power is approximately 22 dBm, which is 3 dB lower than the maximum output power of the DUT EDFA. Since Eq. (2) does not take the input power level into consideration, it implicitly assumes that the input power level is within a typical range, which is here around -18 dBm/channel.

B. EDFA ANN (H)ML Models

The ANN ML method for EDFA gain profile characterization has been investigated in optical transmission systems in recent years [11–15]. The supervised algorithm is preferred where the EDFA input spectrum is fed into the ANN input layer, and the output power spectrum or the gain spectrum is generated in the output layer during the training process. When the training samples represent the full data distribution and the model tuning parameters are appropriately adjusted, the EDFA ANN ML models can obtain better prediction accuracy than the analytical models [18].

In our work, TensorFlow [21] is used to train the EDFA ANN ML model. The ANN architecture is shown in Fig. 5(a). We developed an ANN ML model containing 90 sub-models where each had 90 input features corresponding to the input power levels (mW) from the 90 channels and the targeted channel's output power (mW). This sub-model approach (use 90 independent neural networks for each output channel) was chosen because this model resulted in a better prediction accuracy than a model that had 90 input and output features within a single neural network. The max–min normalization and a linear scaler with a factor 300 are used for data preprocessing. The four hidden layers and the final output layer use ReLU, Linear, ReLU, Linear, and ReLU activation functions, respectively. The input layer and four hidden layers are fully connected with 90 neurons at each layer. The mean square error (MSE) function and SGD are chosen as the loss function and optimizer, respectively.

Although we can achieve a higher prediction accuracy using neural networks instead of the simplified CM analytical model, the knowledge from the simplified CM analytical model is lost. If we can instead combine the two models (ML and the simplified CM analytical model), we may have a comprehensive model that can combine the knowledge from both models and obtain better performance. As a result, other than the power of each channel and channel loadings (which channels are on or off) as two features of the input spectrum, we improve our EDFA ANN ML model with predicted output power levels (estimated gain spectra) calculated by the simplified CM analytical model as a third set of input features in the neural network. The inaccuracy of the simplified CM model will not degrade the enhanced ANN model since the backpropagation will automatically neglect the corresponding inputs by minimizing the weights during the training process. As Fig. 5(b) shows, 90 extra neurons were added to the input layer to form the enhanced ANN hybrid HML model. Each hidden layer contained 90, 90, 90, and 45 neurons, respectively. All the activation functions and data preprocessing methods are the same as in the previous ANN model. Based on previous studies of different ML methods for describing amplifier behavior, we chose an ANN and optimized its structure. We tried different configurations for the ANN models considering model complexity and accuracy. The implemented configurations are the best architecture for the EDFA power prediction in our trials.

All collected data from the DUT EDFA with the 18 dB target gain and 3 dB tilt were divided into the training, validation, and test datasets with the ratio 8:1:1. The learning rate was selected as 0.00025. For the given 15,000 data samples, the

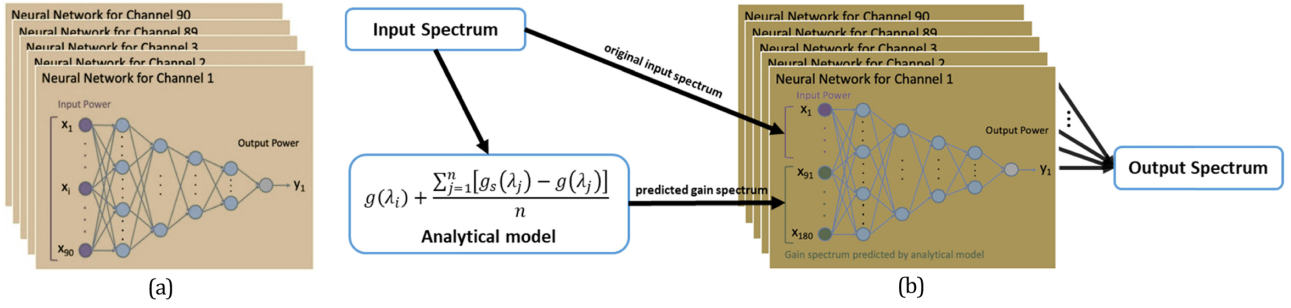


Fig. 5. Design of the ML model based on an ANN and the simplified CM analytical model.

training process was done with 12,000 training samples and over 15,000 epochs to obtain models with converged prediction accuracy (high accuracy after 5000 epochs). Although a large number of epochs and small learning rate were used in our ANN model training process, the high prediction accuracy for both evaluation and testing datasets ensures no overfitting was present.

4. RESULTS AND ANALYSIS

A. EDFA ANN ML Models with Power Spectrum Only

The EDFA ANN ML model was completed with 15,000 epochs. First, we compared the performance between the simplified CM analytical model and the original EDFA ANN ML model.

Further, we analyzed the normalized frequency density (NFD) of the prediction error for channel gain using the trained ANN ML model. Figure 6 shows the NFD of the prediction error using the simplified CM analytical model and the ANN model. The dynamic range is defined as the range over which the input channel power is varied around the target power at -18 dBm in a uniform random distribution. It shows that a wider dynamic range in the input power leads to higher error in the prediction of the gain spectrum for the analytical model. The simplified CM analytical model cannot predict the EDFA gain spectrum accurately because the actual input channel loading (lit input channels and their power levels) affects the gain spectrum.

Figures 6(a), 6(c), and 6(e) present the results compared with the simplified CM analytical model. With the ± 3 dB channel input power dynamics, 99.95% and 94.9% of test datasets are within the 0.5 dB prediction error for the ANN ML model and the simplified CM analytical model, respectively. At the low input power dynamic range, both models show similar performance. However, as the input power dynamics increase, the ANN ML model shows much better accuracy than the simplified CM analytical model. For the ± 6 and ± 9 dB power dynamics, the ANN ML model has the prediction error < 0.5 dB for 98.37% and 93.57% of test datasets, while the simplified CM analytical model achieves the same accuracy only for 85.11% and 81.07% of test datasets, respectively. As the predicted error decreases, the ANN ML model has higher NFD than the simplified CM analytical

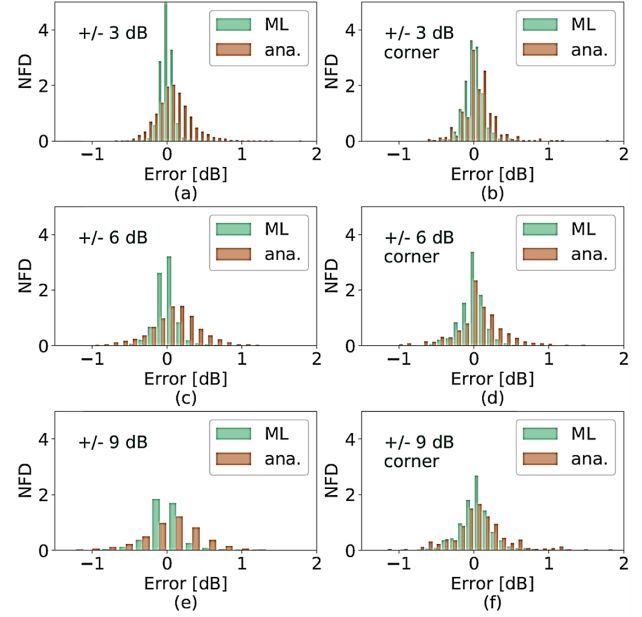


Fig. 6. NFD versus prediction error of the simplified CM analytical model and ANN ML model with a dynamic range of ± 3 , ± 6 , and ± 9 dB. The EDFA is set with a 3 dB tilt and 18 dB target gain.

model, regardless of power dynamics. This means more evaluation datasets obtained a lower predicted error using the ANN model.

More severe optical channel power excursions may occur when fewer channels are active in the system [22]. To validate the ANN ML model performance, the prediction accuracy for these corner cases should be considered. Thus, we took these corner case samples where only two channels were activated to verify the model. Figures 6(b), 6(d), and 6(f) show the results. Similarly, the ANN ML model had better performance than the simplified CM analytical model regarding different power dynamics. The ANN ML model had more accurate prediction for the EDFA output power spectrum that 99.38%, 95.87%, and 95.62% of test datasets were distributed within the 0.5 dB prediction error, when the simplified CM analytical model can only obtain the same prediction accuracy for 95.88%, 87.78%, and 83.18% with the power dynamics of ± 3 , ± 6 , and ± 9 dB, respectively. The results show that the EDFA ANN ML model can handle various channel loading configurations, including some extreme cases in which channel power has the highest power excursion.

Table 1. RMSE of the Simplified CM Analytical Model and ANN ML Model with Various Channel Noise Levels

Noise Level (OSNR Level) [dBm]	Analytical [dB]	ANN ML [dB]
−30 (19.0)	0.3136	0.2840
−35 (25.7)	0.3511	0.2346
−40 (29.3)	0.4904	0.2368

In commercial optical transmission systems, several EDFAs can be used on multiple spans between nodes in a cascading way. In this scenario, the ASE noise can become large in portions of the spectrum that are not occupied by channels. For this situation, the channel input power level includes the optical signal, the input ASE noise internally generated from the target EDFA itself, and input ASE noise from previous EDFAs. To apply the ASE from prior EDFAs, we add another EDFA between the comb source and WSS, as Fig. 3(a) shows. By using this first EDFA, we can adjust its gain and the WSS attenuation to monitor the dynamic channel noise levels into the second EDFA, which is maintained at the 18 dB gain and 3 dB tilt (the DUT EDFA). Additional 1500 data samples were collected in the noise loaded EDFA setup and merged with the previous datasets without input noise for the EDFA ANN ML model training, validation, and test. The model prediction performance is shown in Table 1. Compared with the simplified CM analytical model, the ANN ML prediction results show the root mean square error (RMSE) is reduced by 51.71%, 33.18%, and 9.44% with unoccupied channel noise levels at −40, −35, and −30 dBm per channel, respectively. Although the EDFA ANN model was trained with signal power data at a certain OSNR level, the results show that the re-trained model using a few new data points can still achieve high prediction accuracy for commercial networks where signals experience different ASE noise levels along a path or series of spans with EDFAs.

B. EDFA ANN HML Models with Extra Predicted Gain Spectra

The EDFA ANN HML model is enhanced by adding the third data feature, the predicted output power spectrum using the simplified CM analytical model. We compare the model performance using datasets with ± 3 , ± 6 , and ± 9 dB input channel power dynamic ranges in this section. Table 2 shows the prediction accuracy in mean absolute error (MAE) with input channel power variations. The ANN models outperform the simplified CM analytical model in terms of different input channel power dynamics. The ANN ML (HML) model can reduce the RMSE by 67% (61%), 63% (63%), and 45% (58%) compared with the simplified CM analytical model where the dynamic ranges are ± 3 dB, ± 6 , and ± 9 dB, respectively. The corresponding NFDs and cumulative distribution functions (CDFs) versus the prediction error of three models are shown in Figs. 7(a)–7(c), respectively. Both ANN ML and HML models were built with 5000 epochs and had better performance than the simplified CM analytical model.

For trained ANN models with 25,000 epochs, the RMSE of the predicted channel power was 0.362 dB for the simplified CM analytical model, 0.160 dB for the ANN ML model, and

Table 2. Predicted MAE Results of the Simplified CM Analytical Model and ANN Models with Various Channel Input Power Dynamics

Dynamic Range [dB]	Analytical [dB]	ANN ML [dB]	ANN HML [dB]
± 3	0.18	0.06	0.07
± 6	0.27	0.10	0.10
± 9	0.31	0.17	0.13

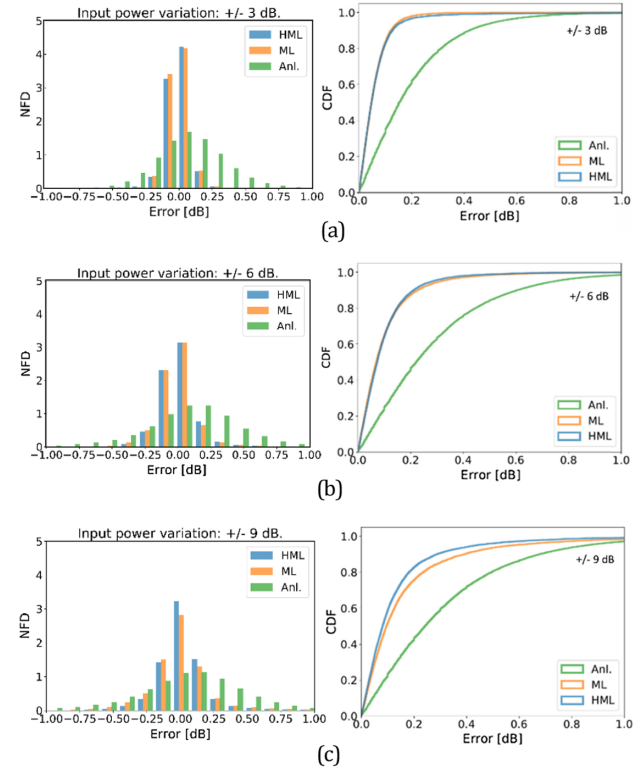


Fig. 7. NFD/CDF versus the prediction error of the simplified CM analytical and ANN models after 5000 epochs for ± 3 , ± 6 , and ± 9 dB channel input power dynamics. The EDFA is set with a 3 dB tilt and 18 dB target gain.

0.144 dB for the ANN HML model, as shown in Table 3. The ANN HML model has a 10.5% RMSE reduction compared with the ANN ML model. Although the ANN HML has almost the same prediction accuracy as the original ANN ML for the overall view, the ANN HML has improved the worst-case prediction by decreasing the maximum error from 2.6 dB (the simplified CM analytical model and ANN ML model) to 1.6 dB based on multiple testing results. We have investigated the input power spectral data that correspond to the maximum prediction error. As expected, these worst-case configurations exhibit a small number of channels (< 10), with a subset of channels at the gain curve extrema (here primarily at the low channel indices; see Fig. 4) and with a large variation in the relative input powers of the channels.

In addition, other EDFA gain configurations were collected and trained with the ANN model to evaluate its performance. Figure 8 shows the CDF versus prediction error distribution of three prediction models for the same EDFA. The target gain

Table 3. Predicted RMSE and Max Error Results of the Simplified CM Analytical Model and ANN (H)ML Models for ± 6 dB Power Dynamics

Model	RMSE [dB]	Max Error [dB]
Analytical	0.362	2.6
ANN ML	0.160	2.6
ANN HML	0.144	1.6

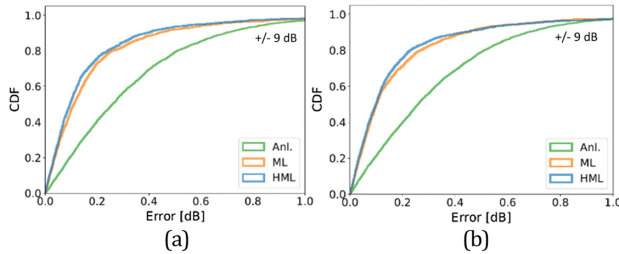


Fig. 8. CDF versus prediction error of the simplified CM analytical model and ANN models for the ± 9 dB channel input power dynamics. The EDFA is set with 3 dB tilt and target gain of (a) 14 dB and (b) 22 dB, respectively.

values were set as 14 dB and 22 dB with the power dynamics of ± 9 dB. A total of 5000 data samples were used for the model training, validation, and test. Our model can still achieve high accuracy. The overall MAE values are 0.195 (0.172) dB and 0.195 (0.183) dB for the ANN ML (HML) models for the 14 dB and 22 dB configurations, respectively.

Figure 9(a) shows the training epoch reduction versus the model accuracy of the ANN MHL model compared with the ANN ML model using the same 12,000 training samples. Although both ANN models can achieve the prediction accuracy as low as 0.14 dB after 20,000 epochs, the ANN HML has a much faster converged training speed. If a 0.4 dB RMSE is the model target, ANN HML requires 1400 fewer training epochs than the original ANN ML model. For 0.14 dB target RMSE, the ANN HML model requires 5300 fewer epochs. Overall, the enhanced ANN model reduces at least 20% of the epochs compared to the original model to achieve the same prediction accuracy. In Fig. 9(b), the training sample size reduction versus model accuracy is shown for the case of 25,000 training epochs processed to complete the ANN models. The same test datasets are used to examine the prediction RMSE with varying training data sample sizes for both models. The ANN HML can always outperform the original ANN model. Especially, the ANN HML model can achieve the best prediction accuracy of 0.14 dB RMSE with 3050 fewer data samples, which is a 46% sample size reduction. The results show that the extra ANN inputs of predicted EDFA gain spectra using the simplified CM analytical model improve the model performance. Although this feature uses estimated parameters, these inputs indeed induce more heavyweight couplings between varying input channel configurations to speed up the ANN model built with less datasets during the neurons' weight tuning.

The results indicate that very similar mean error is obtained without using the hybrid model, so the main savings come

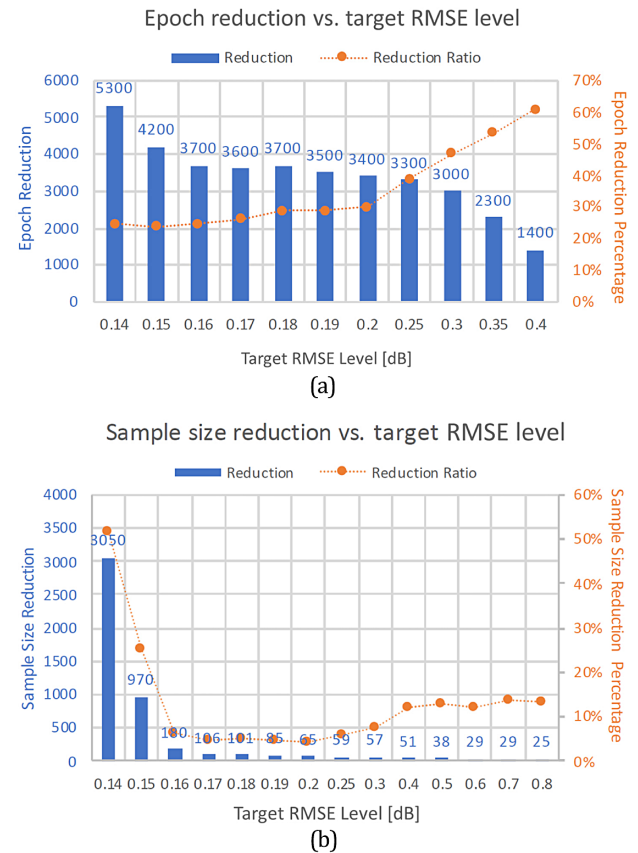


Fig. 9. ANN MHL model performance compared with the ANN ML model: (a) training epoch reduction versus model accuracy using the same 12,000 training samples, (b) training sample size reduction versus model accuracy when there are 25,000 training epochs.

from the maximum error and the reduction in training time, the value of which will depend on the manufacturer.

5. CONCLUSION

We examined the use of ANN ML models to predict a single EDFA output power spectrum as a means of accurate channel power estimation for use in QoT estimation. The model was constructed from multiple, separate ANNs to predict the output power of each channel in a 90-channel WDM transmission system. A method for rapid, automated pre-deployment testing was developed. The use of a posteriori knowledge derived from a simplified CM analytical model as input features to the ANNs to further improve the accuracy and convergence time was studied in a hybrid ML approach. Each of these were compared with the CM-based analytical model and shown to be more accurate under a wide range of operating conditions, including variable noise loading, channel powers, and amplifier gain settings. The hybrid approach was particularly effective in reducing the maximum power error and the convergence time, and training dataset size, which are important for efficient amplifier characterization.

Funding. National Science Foundation (0812072, 1650669, 1650685, 1737453, 1827923).

Acknowledgment. We gratefully acknowledge Weiyang Mo, Yao Li, Alan Diaz Montiel, and Marco Ruffini for their help with this work.

REFERENCES

1. A. A. Saleh and J. M. Simmons, "All-optical networking—evolution, benefits, challenges, and future vision," *Proc. IEEE* **100**, 1105–1117 (2012).
2. Y. Pointurier, "Design of low-margin optical networks," *J. Opt. Commun. Netw.* **9**, A9–A17 (2017).
3. E. Seve, J. Pesic, C. Delezoide, S. Bigo, and Y. Pointurier, "Learning process for reducing uncertainties on network parameters and design margins," *J. Opt. Commun. Netw.* **10**, A298–A306 (2018).
4. M. Filer, J. Gaudette, Y. Yin, D. Billor, Z. Bakhtiari, and J. L. Cox, "Low-margin optical networking at cloud scale," *J. Opt. Commun. Netw.* **11**, C94–C108 (2019).
5. M. Tachibana, R. I. Laming, P. R. Morkel, and D. N. Payne, "Erbium-doped fiber amplifier with flattened gain spectrum," *IEEE Photon. Technol. Lett.* **3**, 118–120 (1991).
6. M. Bolshtyansky, N. King, and G. Cowle, "Dynamic behavior of spectral hole burning in EDFA with 980nm pumping," in *Optical Fiber Communication Conference (OFC)* (2007).
7. J. Junio, D. C. Kilper, and V. W. Chan, "Channel power excursions from single-step channel provisioning," *J. Opt. Commun. Netw.* **4**, A1–A7 (2012).
8. K. Ishii, J. Kurumida, and S. Namiki, "Experimental investigation of gain offset behavior of feedforward-controlled WDM AGC EDFA under various dynamic wavelength allocations," *IEEE Photon. J.* **8**, 7901713 (2016).
9. Viavi Solutions, <https://www.viavisolutions.com/en-us/>.
10. Y. Huang, C. L. Gutterman, P. Samadi, P. B. Cho, W. Samoud, C. Ware, M. Lourdiane, G. Zussman, and K. Bergman, "Dynamic mitigation of EDFA power excursions with machine learning," *Opt. Express* **25**, 2245–2258 (2017).
11. A. D'Amico, S. Straullu, A. Nespola, I. Khan, E. London, E. Virgillito, S. Piciaccia, A. Tanzi, G. Galimberti, and V. Curri, "Using machine learning in an open optical line system controller," *J. Opt. Commun. Netw.* **12**, C1–C11 (2020).
12. F. Musumeci, C. Rottondi, A. Nag, I. Macaluso, D. Zibar, M. Ruffini, and M. Tornatore, "An overview on application of machine learning techniques in optical networks," *IEEE Commun. Surv. Tutorials* **21**, 1383–1408 (2018).
13. Y. You, Z. Jiang, and C. Janz, "OSNR prediction using machine learning-based EDFA models," in *European Conference on Optical Communication* (2019).
14. A. Mahajan, K. Christodouloupoloulos, R. Martinez, S. Spadaro, and R. Muñoz, "Modeling EDFA gain ripple and filter penalties with machine learning for accurate QoT estimation," *J. Lightwave Technol.* **38**, 2616–2629 (2020).
15. D. E. M. Alba, J. M. C. Herrera, and G. A. P. Leguizamón, "Design of a flattening filter using fiber Bragg gratings for EDFA gain equalization: an artificial neural network application," *Cienc. Ing. Neogranadina* **29**, 25–36 (2019).
16. D. Zibar, A. M. R. Brusin, U. C. de Moura, F. Da Ros, V. Curri, and A. Carena, "Inverse system design using machine learning: the Raman amplifier case," *J. Lightwave Technol.* **38**, 736–753 (2019).
17. A. M. R. Brusin, U. C. de Moura, V. Curri, D. Zibar, and A. Carena, "Introducing load aware neural networks for accurate predictions of Raman amplifiers," *J. Lightwave Technol.* **38**, 6481–6491 (2020).
18. S. Zhu, C. L. Gutterman, W. Mo, Y. Li, G. Zussman, and D. C. Kilper, "Machine learning based prediction of erbium-doped fiber WDM line amplifier gain spectra," in *European Conference on Optical Communication* (2018).
19. S. Zhu, C. L. Gutterman, A. D. Montiel, J. Yu, M. Ruffini, G. Zussman, and D. Kilper, "Hybrid machine learning EDFA model," in *Optical Fiber Communication Conference (OFC)* (2020).
20. B. Yegnanarayana, *Artificial Neural Networks* (PHI Learning, 2009).
21. TensorFlow, <https://www.tensorflow.org/>.
22. W. Mo, S. Zhu, Y. Li, and D. C. Kilper, "Dual-wavelength source based optical circuit switching and wavelength reconfiguration in multi-hop ROADMs systems," *Opt. Express* **25**, 27736–27749 (2017).

Jiakai Yu is currently pursuing his Ph.D. degree in electrical and computer engineering from the University of Arizona. He received a B.S. degree from the Physics College at Nankai University, Tianjin, China, in 2013, and an M.S. degree in electrical engineering from the State University of New York at Buffalo, NY, USA, in 2015. His research interests include software-defined networking, 5G wireless and optical converged networks, edge cloud, and machine learning.

Shengxiang Zhu completed his Ph.D. degree in electrical and computer engineering from the University of Arizona, advised by Professor Daniel C. Kilper. His research interests include advanced optical transmission systems, software defined networks, and ML-based modeling for optical networks.

Craig Gutterman received his B.S. degree (summa cum laude) in electrical engineering from Rutgers University in May 2012 and an M.S. degree in electrical engineering from Columbia University in February 2014. He received his Ph.D. degree in electrical engineering from Columbia University in summer 2020, where he was advised by Professor Gil Zussman. His research interests include improving the performance of future networks and systems by developing ML-based network systems and data-driven network algorithms.

Gil Zussman received the Ph.D. degree in electrical engineering from the Technion in 2004 and was a postdoctoral associate at MIT in 2004–2007. He is currently a professor of electrical engineering at Columbia University. He is a co-recipient of seven paper awards, including the ACM SIGMETRICS'06 Best Paper Award, the 2011 IEEE Communications Society Award for Advances in Communication, and the ACM CoNEXT'16 Best Paper Award. In addition, he received the Fulbright Fellowship, the DTRA Young Investigator Award, and the NSF CAREER Award, and was a member of a team that won first place in the 2009 Vodafone Foundation Wireless Innovation Project competition.

Daniel C. Kilper is a research professor in the College of Optical Sciences at the University of Arizona, Tucson. He holds a joint appointment in electrical and computer engineering at the University of Arizona and an adjunct faculty position in electrical engineering at Columbia University. He is the founder and chief technology officer of Palo Verde Networks, Inc. He received a Ph.D. in physics from the University of Michigan in 1996. From 2000 to 2013, he was a member of the technical staff at Bell Labs. In addition, he is a senior member of IEEE, an editor for *IEEE Transactions on Green Communications and Networking*, and a steering committee member for the IEEE Green ICT Initiative. His work has been recognized with the Bell Labs President's Gold Medal Award, and he served on the Bell Labs Presidents Advisory Council on Research. He holds seven patents and has authored five book chapters and more than 130 peer-reviewed publications.