Applications of Artificial Neural Networks in Optical Performance Monitoring

Xiaoxia Wu, *Student Member, IEEE*, Jeffrey A. Jargon, *Senior Member, IEEE*, Ronald A. Skoog, *Member, IEEE*, Loukas Paraschis, *Senior Member, IEEE*, and Alan E. Willner, *Fellow, IEEE*

***Abstract*—Applications using artificial neural networks (ANNs) for optical performance monitoring (OPM) are proposed and demonstrated. Simultaneous identification of optical signal-to-noise-ratio (OSNR), chromatic dispersion (CD), and polarization-mode-dispersion (PMD) from eye-diagram parameters is shown via simulation in both 40 Gb/s on-off keying (OOK) and differential phase-shift-keying (DPSK) systems. Experimental verification is performed to simultaneously identify OSNR and CD. We then extend this technique to simultaneously identify accumulated fiber nonlinearity, OSNR, CD, and PMD from eye-diagram and eye-histogram parameters in a 3-channel 40 Gb/s DPSK wavelength-division multiplexing (WDM) system. Furthermore, we propose using this ANN approach to monitor impairment causing changes from a baseline. Simultaneous identification of accumulated fiber nonlinearity, OSNR, CD, and PMD causing changes from a baseline by use of the eye-diagram and eye-histogram parameters is obtained and high correlation coefficients are achieved with various baselines. Finally, the ANNs are also shown for simultaneous identification of in-phase/quadrature (I/Q) data misalignment and data/carver misalignment in return-to-zero differential quadrature phase shift keying (RZ-DQPSK) transmitters.**

***Index Terms*—Neural networks, optical fiber communication, optical performance monitoring, phase modulation.**

I. Introduction

HIGH-PERFORMANCE optical networks are susceptible to various degrading effects that can change over time. Knowledge of the data channel degradation can be used to diagnose the network, repair the damage, drive a compensator/equalizer, and/or reroute traffic around a non-optimal link [1]–[3]. Therefore, it is valuable to monitor the channels for many types of impairments, such as optical signal-to-noise-ratio (OSNR), chromatic dispersion (CD), polarization-mode-dispersion (PMD), and fiber nonlinearity, which can change with temperature, plant maintenance, and path reconfiguration. Key features of any optical performance monitors are simplicity in implementation and the ability to accommodate different modulation formats and impairments.

Recently, optical networks have been evolving from closed systems to open systems, in which the optical layer is designed to allow transmitter/receiver add and drop without affecting the current structure. This trend has been reflected in service provider requirements for "alien wavelengths" and in the standards—most notably, ITU-T G.698.2 [4]. Associated with changes in the number of channels are the power transients in the surviving channels arising from cross saturation in optical amplifiers and the nonlinear interactions among channels. To maintain system performance, agile optical performance monitoring (OPM) and automatic system control become increasingly important.

OPM can be performed by measuring changes to the data and determining "real-time" changes resulting from various impairments, such that a change in a particular effect will change a measured parameter. This can employ: (i) optical techniques to monitor changes in a radio frequency (RF) tone power or in the spectral channel power distribution [5], or (ii) electrical post-processing techniques in the specific case of coherent detection [6], [7].

The optical approaches have been shown to be powerful for OPM. However, the electrical distortions that are crucial for the signal quality at the decision point tend to be neglected in the optical approaches. Several techniques have been proposed for OPM using off-line digital signal processing of received electrical data signals [8]–[21]. Four of these methods [8]–[11] utilize amplitude histograms, power distributions or asynchronous sampling to estimate bit error rate (BER); four [12]–[15] employ delay-tap plots to distinguish among impairments; three [16]–[18] use pattern recognition techniques to identify multiple impairments; and the rest [19]–[21] use parameters derived from eye diagrams and histograms for the same purpose. The latter approach is to probe the network upon initialization and train each receiver to record a specific data eye-diagram pattern that corresponds to a specified range of potential physical parameters. These eye diagrams can be generated either from a synchronized sampler, or by a technique that regenerates such diagrams from asynchronous samples [11]. Once the network is fully operational, variations in the received eye diagram from the ideal formation can then be attributed to specific physical parameters derived from the prior network/receiver training.

Recently, we have made use of a neural network approach to "train" receivers in an optical network to distinguish between resultant shapes of the data channel's eye diagrams and the degrading effects of OSNR, CD, PMD [19], [20]. The ANN approach has further been applied to monitor accumulated fiber nonlinearity in addition to OSNR, CD, PMD [21]. By use of

Manuscript received January 15, 2009; revised May 20, 2009. First published June 02, 2009; current version published July 24, 2009. This work was supported by the DARPA CORONET PARAGON Program, the U.S. Department of Commerce, and Cisco Systems, and is not subject to U.S. copyright.

X. Wu, and A. E. Willner are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: xiaoxia@usc.edu).

J. A. Jargon is with the Department of Commerce, National Institute of Standards and Technology (NIST), Boulder, CO 80305 USA.

R. A. Skoog is with Optical Networking Research, Telcordia Technologies, Inc., Red Bank, NJ 07701 USA.

L. Paraschis is with Cisco, San Jose, CA 95134 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JLT.2009.2024435

this method, the coefficients of the neural network algorithm are iteratively derived prior to live traffic being sent through the network. A similar technique has also been used for time misalignment monitoring in return-to-zero differential quadrature phase shift keying (RZ-DQPSK) transmitters [22], which extends the applications of our ANN approach to a broader sense of OPM.

In this paper, we show various applications of ANNs in OPM. In Section II, the concept and structure of ANNs are introduced. The popularly used multilayer perceptron (MLP) neural network and various steps involved in the development of neural network models are described. In Section III, simultaneous identification of OSNR/CD/PMD is demonstrated in 40 Gb/s on-off keying (OOK) and DPSK systems via simulation. Subsequent experimental verification is performed to simultaneously identify OSNR and CD. In Section IV, we add accumulated channel nonlinear effects to CD, PMD, and OSNR. We demonstrate this technique in a 3-channel 40 Gb/s RZ-DPSK WDM system. Furthermore, we propose using our ANN approach to monitor impairment causing changes from a baseline instead of the absolute values. Simultaneous identification of accumulated fiber nonlinearity, OSNR, CD, and PMD introducing changes from a baseline by use of the eye-diagram and eye-histogram parameters in a 3-channel 40 Gb/s DPSK WDM system is obtained with various baselines. In Section V, ANNs are used for simultaneous identification of in-phase/quadrature (I/Q) data and data/carver misalignments in RZ-DQPSK transmitters, which indicates the applications of ANNs in a broader sense of OPM.

II. Artificial Neural Networks

A. ANN Concepts

As bit rates increase, it becomes more difficult to predict the data degradation mechanisms in optical networks. In order to enable robust and cost-effective "self-managed" operation, it would be desirable for the network itself to agilely monitor the physical impairments and the quality of the data signals, and automatically diagnose and feed back information to control the network. By incorporating trained receivers, a simple structure of a self-managed network is shown in Fig. 1(a). Impairments are indentified by the trained receivers in the optical network element (ONE) and error signals are generated and sent to the routers. Further actions can be taken so that the network controller can agilely control and manage the network.

To illustrate how the trained receivers work, we introduce the concepts of ANNs. ANNs are information-processing systems that learn from observations and generalize by abstraction [23], [24], which are attractive alternatives to conventional methods such as numerical modeling methods, analytical methods, or empirical modeling solutions. ANNs have the ability to model multi-dimensional nonlinear relationships and are simple to use. Furthermore, the neural network approach is generic (i.e., the same modeling technique can be re-used for passive/active devices/systems) and the response is fast. Due to these features, the ANN approach has gained much attention as a powerful tool in a number of areas such as pattern recognition, speech processing, control, and bio-medical engineering, and recently been applied in RF modeling, microwave design, and optical

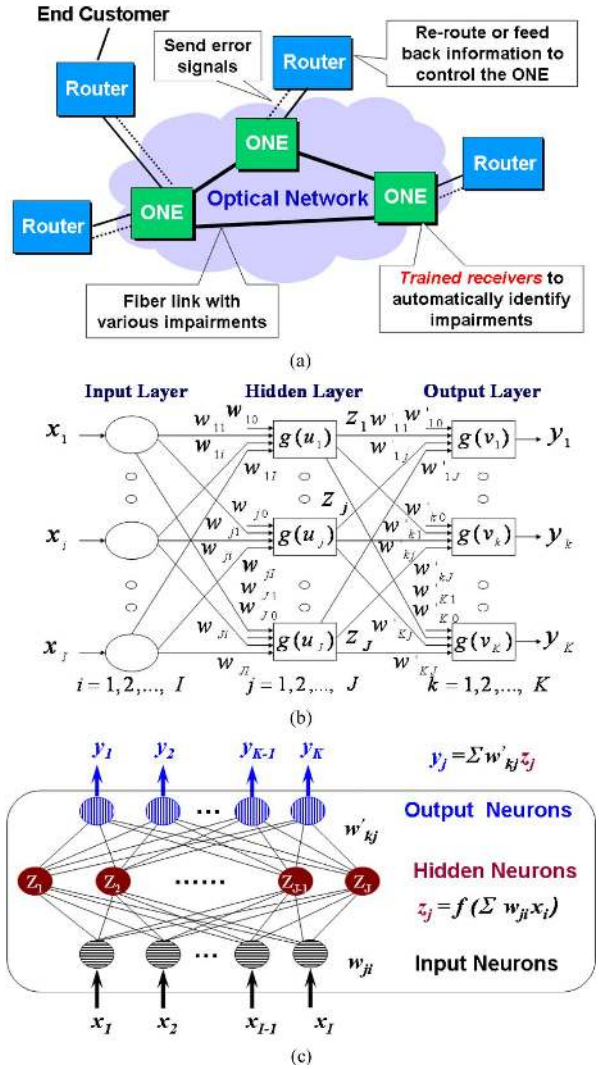


Fig. 1. Concepts of ANNs. (a) Self-managed optical networks. ONE: optical network element; (b) the structure of an artificial neutral network (ANN); (c) a 3-layer perceptron (MLP3) ANN model.

performance monitoring. Oftentimes, neural networks are first trained to model the electrical/optical behavior of passive and active components/circuits/systems. These trained neural networks can then be used in high-level simulation and design, providing fast answers to the task they have learned [25], [26].

An ANN consists of multiple layers of processing elements called neurons. Each neuron is connected to other neurons in neighboring layers by varying coefficients that represent the strengths of these connections, as shown in Fig. 1(b). ANNs learn the relationships among sets of input-output data that are characteristics of the device or system under consideration. After the input vectors are presented to the input neurons and output vectors are computed, the ANN outputs are compared to the desired outputs, and errors are calculated. Error derivatives are then calculated and summed for each weight until all of the training sets have been presented to the network. The error derivatives are used to update the weights for the neurons, and training continues until the errors reach prescribed low values.

MLP is the basic and most frequently used structure. In the MLP neural network, the neurons are grouped into layers. The

first and last layers are called input and output layers, respectively, and the remaining layers are called hidden layers. Typically, an MLP neural network consists of an input layer, one or more hidden layers, and an output layer. For example, an MLP neural network with an input layer, one hidden layer, and an output layer, is referred to as 3-layered MLP or MLP3, as shown in Fig. 1(c). The hidden layer allows complex models of input-output relationships. The mapping of these relationships is given by $\mathbf{Y} = g[\mathbf{W}' \bullet g(\mathbf{W} \bullet \mathbf{X})]$, where $\mathbf{X}$ is the input vector, $\mathbf{Y}$ is the output vector, and $\mathbf{W}$ and $\mathbf{W}'$ are the weight matrices between the input and hidden layers and between the hidden and output layers, respectively. The function $g(u)$ can be the smooth switch-type activation functions, such as sigmoid, arc-tangent, and hyperbolic-tangent, which are bounded, continuous, monotonic and continuously differentiable. In our analysis, a nonlinear sigmoidal activation function given by $g(u) = 1/[1+\exp(-u)]$ is used, where u is the input to a hidden neuron or an output neuron.

In addition to MLP, there are other ANN structures [27], such as radial basis function (RBF) networks, wavelet networks, and recurrent networks. The universal approximation theorem [28] states that there always exists a 3-layer MLP neural network that can approximate any arbitrary, nonlinear, continuous, multidimensional function to any desired accuracy. The number of hidden neurons depends upon the degree of nonlinearity of the function and the dimensionality of the model. Highly nonlinear systems require more neurons, while smoother systems require fewer neurons. In our work, the number of hidden neurons is optimized via adaptive processes, which add/delete neurons during training.

B. ANN Training and Testing

The most important step in neural network model development is the training process. In this sub-section, we will explain the ANN training and testing processes in more details.

The neural network weight parameters (w) are initialized so as to provide a good starting point for training. The widely used strategy for MLP weight initialization is to initialize the weights with small random values (e.g., in the range $[-0.5, 0.5]$). To improve the convergence of training, one can use a variety of distributions (e.g., Gaussian distribution), and/or different ranges and different variances for the random number generators used in initializing the ANN weights [29].

The training data consists of sample pairs, $\{(x_n, d_n),\ n \in T_r\}$, where x_n and d_n are I- and K-vectors representing the inputs and the desired outputs of the neural network and T_r represents the index set of the training data. In our work, the inputs are the parameters derived from eye diagrams or other sources, e.g. RF tone power, and asynchronous diagrams, and the outputs are the impairments, e.g. OSNR, CD, and PMD. We define the neural network training error as [30]

$$E_{T_r}(w) = \frac{1}{2} \sum_{n \in T_r} \sum_{k=1}^{K} |y_k(x_n, w) - d_{kn}|^2 \qquad (1)$$

where d_{kn} is the kth element of d_n and $y_k\ (x_n, w)$ is the kth neural network output for input x_n. The purpose of neural network training is to adjust w such that the error function $E_{Tr}(w)$ is minimized. The error between training data and ANN outputs is fed back to the ANN to guide the internal weight update of the network. Here, $\Delta w = \eta h$ is called the weight update, and η is a positive step size known as the learning rate. Gradient based iterative training techniques determine update direction h based on error information $E_{Tr}(w)$ and error derivative information $\partial E_{Tr}(w)/\partial w$. Step size η can be determined in one of the following ways: (1) small value, either fixed or adaptive during training; or (2) line minimization to find best value of η.

The time needed for training depends on the amount of training data involved, the structure of the neural network, and also the training algorithm. There are several gradient-based iterative training algorithms, including back propagation, conjugate gradient and quasi-Newton. Back propagation is relatively slow in converging, so second-order training algorithms, such as conjugate gradient and quasi-Newton, are oftentimes preferred for their increased efficiency. The quasi-Newton approach is relatively fast due to its quadratic converge property, although more computer memory is required since it relies on the Hessian matrix whose inverse needs to be calculated. The conjugate gradient method is a nice compromise in terms of memory and implementation effort, since the descent direction runs along the conjugate direction, which can be determined without matrix computations.

We use feed-forward computation in our work. Given the input vector $\mathbf{X}$ and the weight vector $\mathbf{W}$, neural network feed-forward computation is a process used to compute the output vector $\mathbf{Y}$. It is useful not only during neural network training but also during the usage of the trained neural model. The external inputs are first fed to the input neurons and the outputs from the input neurons are fed to the hidden neurons. Continuing this way, the outputs of one layer neurons are fed to the next layer neurons [30]. During feed-forward computation, neural network weights $\mathbf{W}$ remain fixed.

After training, the ANN can be tested by use of other sets of data. The correlation coefficient, which represents how close the ANN model outputs to the testing data, can be used as the quality measurement factor.

III. ANNs for CD/PMD/OSNR Monitoring

A. CD/PMD/OSNR

With the increase of system capacity, optical networks will be highly susceptible to deleterious and data-degraded fiber-based impairments. CD, PMD, and OSNR are among a few of the most important impairments due to the broad spectra of high-rate signals. Therefore, the ability of the network to identify the amount of the impairments is quite important to maintain system performance.

Fig. 2 shows the simulated eye diagrams for a 40 Gb/s RZ-OOK signal at a few select combinations of OSNR, CD and first-order PMD (i.e., differential group delay (DGD)). The simulated DGD emulation assumes that the signal polarization principle states have worst-case alignments with 50:50 power in the fast and slow axes. Visually, it is obvious that these impairments produce distinct features. To quantify these attributes, we can calculate various eye-diagram parameters. For this example, we choose four such parameters, including Q-factor, eye closure, root-mean-square (RMS) jitter, and crossing amplitude. Q-factor is defined as the difference of

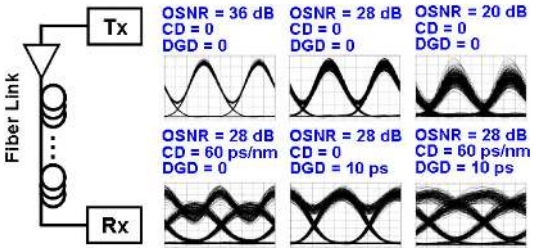


Fig. 2. The impact of degradation effects on eye diagrams of an RZ signal. Tx: transmitter; Rx: receiver; DGD: differential group delay (first-order PMD).

the mean upper and lower levels divided by the sum of the upper and lower level standard deviations; eye closure is the ratio of the outer eye height to the inner eye height; crossing amplitude is the point on the vertical scale where the rising and falling edges intersect; and RMS jitter is usually defined as the standard deviation of the time data calculated in a narrow window surrounding the crossing amplitude. These four inputs are chosen because they change significantly with varying impairment combinations.

The ANN architecture used in this work is a feed-forward, three-layer perceptron structure. The ANN consists of four inputs (Q-factor, closure, jitter, and crossing-amplitude), three outputs (OSNR, CD, and DGD), and twelve hidden neurons. The ANN is trained by use of a software package developed by Zhang *et al.* [31]. We first verify the concept via simulation in 40 Gb/s RZ-OOK and RZ-DPSK systems. The conjugate gradient method is used for training. The training data are obtained from the eye diagrams by use of one set of 125 samples ($\mathrm{OSNR} = 32, 28, 24, 20, 16$ dB; $\mathrm{CD} = 0, 15, 30, 45, 60$ ps/nm; $\mathrm{DGD} = 0, 2.5, 5, 7.5, 10$ ps). Another set of 64 samples ($\mathrm{OSNR} = 30, 26, 22, 18$ dB; $\mathrm{CD} = 7.5, 22.5, 37.5, 52.5$ ps/nm; $\mathrm{DGD} = 1.25, 3.75, 6.25, 8.75$ ps) is used for testing.

The simulated fiber channel includes a laser with a full width at half maximum (FWHM) line-width of 10 MHz; a 40 Gb/s logic source; a single-arm, Mach-Zehnder modulator (MZM) biased at the quadrature point with V_π driving voltage for generating OOK and at minimum point with $2V_\pi$ driving voltage for generating DPSK, where V_π is the half-wave voltage of the MZM, followed by another MZM for RZ pulse carving. Impairments are added through emulators in the link and then the signals are detected by using a single photodiode for RZ-OOK and a balanced receiver following a delay line interferometer (DLI) for RZ-DPSK, where the eye diagrams are recorded and the eye diagram parameters are extracted.

Fig. 3(a) shows the training error versus the epochs. An epoch is defined as a stage of ANN training that involves presentation of all the samples in the training data set to the neural network once for the purpose of learning. The testing and ANN-modeled data are compared in Fig. 3(b) and (c). The ANN reports a correlation coefficient of 0.97 and 0.96 for OOK and DPSK systems, respectively. The measured average errors for OSNR, CD and DGD are 0.57 dB, 4.68 ps/nm, and 1.53 ps, respectively for 40 Gb/s RZ-OOK, and are 0.77 dB, 4.74 ps/nm, and 0.92 ps, respectively for 40 Gb/s RZ-DPSK.

B. Experimental Verification

The experimental setup is shown in Fig. 4. 40 Gb/s RZ-DPSK or RZ-OOK signals are generated using two cascaded MZMs.

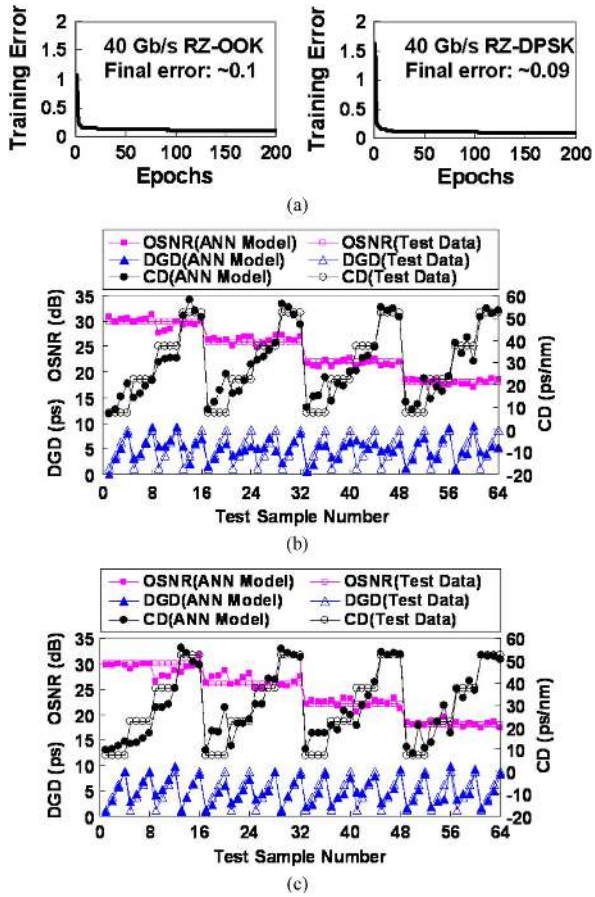


Fig. 3. Simulation results for OSNR/CD/PMD monitoring in 40 Gb/s OOK and DPSK systems. (a) Training error; (b) 40 Gb/s RZ-OOK testing results; (c) 40 Gb/s RZ-DPSK testing results.

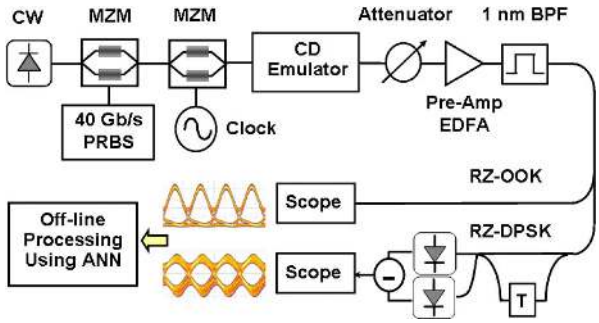


Fig. 4. Experimental setup. CW: continuous wave.

The signal then goes through a tunable dispersion compensating module (TDCM) with $+/-400$ ps/nm tuning range and 10 ps/nm tuning resolution, which serves as the CD emulator. The output of the TDCM is sent to an erbium-doped fiber amplifier (EDFA) with a variable optical attenuator (VOA) in front to adjust the received OSNR. The noise-loaded signal is then filtered by a bandpass filter (BPF) with 1 nm bandwidth, and sent to a scope, where the eye diagram parameters are extracted.

In our experiment, we vary OSNR and CD to get two sets of eye diagram parameters for 40 Gb/s RZ-DPSK and RZ-OOK signals, respectively, including extinction ratio, eye opening factor and signal-to-noise ratio. One set with 20 samples ($\mathrm{OSNR} = 32, 28, 24, 20, 16$ dB; $\mathrm{CD} = 0, 10, 30, 50$ ps/nm) is sent to the ANN model for training, and the other set with 12 samples ($\mathrm{OSNR} = 30, 26, 22, 18$ dB; $\mathrm{CD} = 10, 20, 40$ ps/nm)

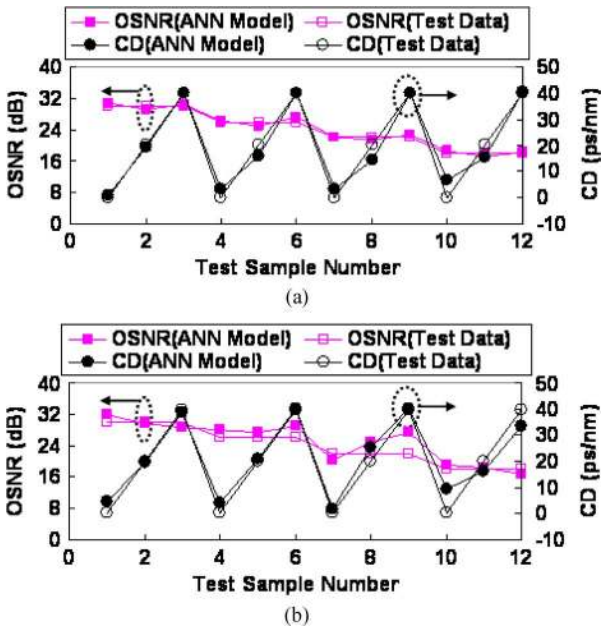


Fig. 5. Experimental results for OSNR/CD monitoring in 40 Gb/s OOK and DPSK systems. (a) 40 Gb/s RZ-OOK testing results; (b) 40 Gb/s RZ-DPSK testing results.

is used for testing. The final training errors for the OOK and DPSK data are ∼0.03 and ∼0.04, respectively. Fig. 5 shows testing results with the experimental data. For the RZ-DPSK signal, we use the eye of the destructive port of the DLI to extract parameters since we cannot estimate balanced eye diagrams with the scope. The ANN reports a correlation coefficient of 0.99 for both of the 40 Gb/s RZ-OOK and RZ-DPSK systems. Fig. 5 compares the testing and ANN-modeled data. The measured average errors for OSNR and CD are 0.58 dB, 2.53 ps/nm, respectively for 40 Gb/s RZ-OOK and are 1.85 dB, 3.18 ps/nm, respectively for 40 Gb/s RZ-DPSK.

The OSNR considered in this experimental work is 16∼32 dB for illustration purpose. In real optical systems, the OSNR can be lower, such as 10–12 dB in 40 Gb/s DPSK systems, which is validated via simulation is the next sub-section.

C. Monitoring Low OSNR

OSNR values in real optical networks may degrade to levels as low as 10–12 dB for 40 Gb/s DPSK systems. Here, we perform a simulation for 40 Gb/s RZ-DPSK using parameters similar to that in the experiment above. Only OSNR and CD are varied for illustration purposes. We use 49 samples (OSNR $= 34, 30, 26, 22, 18, 14, 10$ dB; CD $= 0, 10, 20, 30, 40, 50, 60$ ps/nm) for training and 36 samples (OSNR $= 32, 28, 24, 20, 16, 12$ dB; CD $= 5, 15, 25, 35, 45, 55$ ps/nm) for testing. The eye-diagram parameters include Q-factor, eye closure, and RMS jitter. Fig. 6 compares the testing and ANN-modeled data. The ANN reports a correlation coefficient of 0.99, which shows the effectiveness of using ANNs for identification of lower OSNRs. In this case, the measured average errors for OSNR and CD are 1.23 dB, and 4.56 ps/nm, respectively.

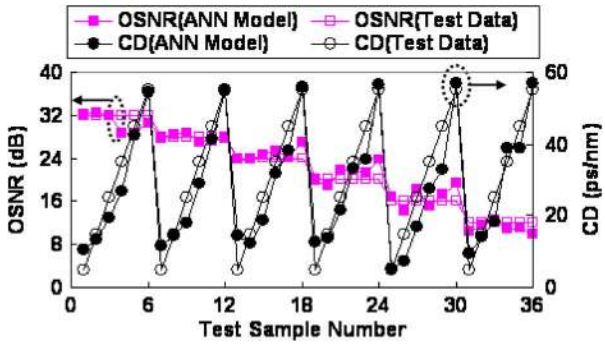


Fig. 6. Simulation results for OSNR/CD monitoring in a 40 Gb/s DPSK system.

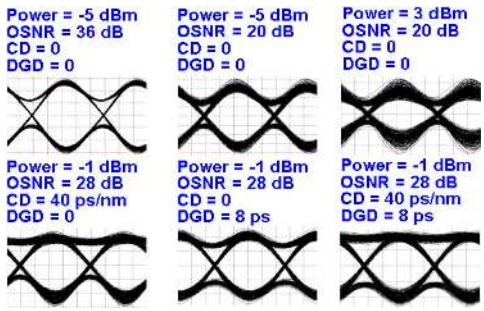


Fig. 7. The impact of degradation effects on the eye diagrams of RZ-DPSK.

IV. ANNs for CD/PMD/OSNR/Accumulated Nonlinearity Monitoring

A. CD/PMD/OSNR/Accumulated Nonlinearity

One parameter that has not been explored much in OPM has been the accumulation of nonlinear impairment on the data channels, which has typically been one of the most difficult parameters to monitor in an optical network. Adding accumulated nonlinearity is also a challenge in terms of the neural network approach, due to its specific signatures on the eye diagrams.

Fig. 7 shows simulated eye diagrams for the middle channel of a 3-channel 40 Gb/s RZ-DPSK WDM system at a few select combinations of OSNR, CD, DGD and optical power. We can clearly see that different impairment combinations imprint different signatures on the eye diagrams. In this case, the four outputs are input optical power, OSNR, CD, and PMD, and the eight inputs include Q-factor, eye-closure, RMS jitter, '0'-level crossing amplitude, mean of '1's and '0's, standard derivation (SD) of '1's and '0's.

Fig. 8 shows the 3-channel WDM configuration used in the simulation. The 40 Gb/s RZ-DPSK signals are generated by two cascaded MZMs and then coupled together with a channel spacing of 0.8 nm. The channels are decorrelated by use of logic sources with different pseudo-random bit sequence (PRBS) orders. The WDM signals then pass through 2 km of highly nonlinear fiber (HNLF) with a nonlinear coefficient of 18 $W^{-1} \cdot km^{-1}$, zero dispersion wavelength of λ_0 (1550 nm), and dispersion slope of 0.05 $ps/nm^2/km$, following by a CD emulator and a PMD emulator. The output is sent to an EDFA with a variable optical attenuator in front to adjust the received OSNR. The signal is then filtered by a BPF with 0.64 nm bandwidth, and sent to an oscilloscope, where the eye diagram and eye histogram parameters are extracted. A 3-channel case is chosen to

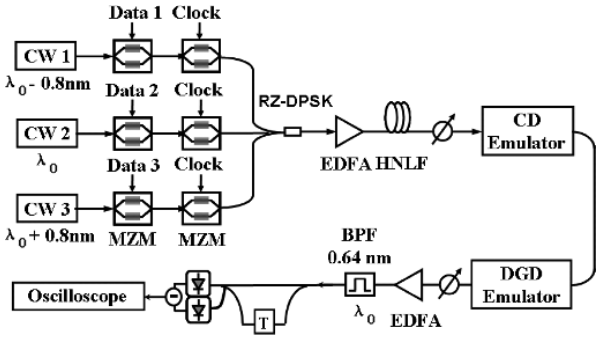


Fig. 8. Simulation setup. λ_0: the wavelength of the channel of interest. $\lambda_0 = 1550$ nm.

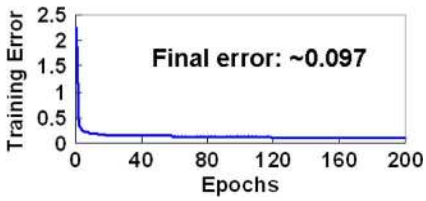


Fig. 9. Training error versus number of epochs for a 40 Gb/s RZ-DPSK channel (middle channel) in a 3-channel WDM system.

illustrate the concept, although this approach is also applicable to WDM networks with more channels.

The middle channel is chosen for the analysis because it experiences the strongest interchannel nonlinearity. The training data are obtained from the eye diagrams by use of a set of 135 samples (optical power $= -5, -3, -1, 1, 3$ dBm; OSNR $= 36, 28, 20$ dB; CD $= 0, 20, 40$ ps/nm; DGD $= 0, 4, 8$ ps). Note that a few training samples are used in this work. In practical networks, a much larger amount of data will be required for training. Fig. 9 shows the training error versus epochs. The final training error is $\sim$0.1 in our case.

Once the model is trained, we validate its accuracy by use of a different set of testing data that includes 32 samples (optical power $= -4, -2, 0, 2$ dBm; OSNR $= 32, 24$ dB; CD $= 10, 30$ ps/nm; DGD $= 2, 6$ ps). Again, the simulated DGD emulation assumes that the signal polarization principle states have worst-case alignments with 50:50 power in the fast and slow axes. The ANN reports a correlation coefficient of 0.97.

Fig. 10 compares the testing and ANN-modeled data for optical power, OSNR, CD, and DGD. The measured average errors for optical power, OSNR, CD and DGD are 0.46 dB, 1.45 dB, 3.98 ps/nm, and 0.65 ps, respectively. It is shown that the ANN models, trained with parameters derived from eye diagrams and eye histograms, can potentially be used to simultaneously identify accumulated fiber nonlinearity, OSNR, CD, and PMD in WDM channels.

B. ANNs for Identification of Impairment Causing Changes from a Baseline

Normally, when considering a system to be monitored, we assume the system is impairment-free; then different impairments, such as CD and PMD, are added for the purpose of testing the monitoring approaches. However, systems are not perfect, and inevitably contain a certain amount of impairments. Thus, starting from a baseline is more practical in terms of performance monitoring.

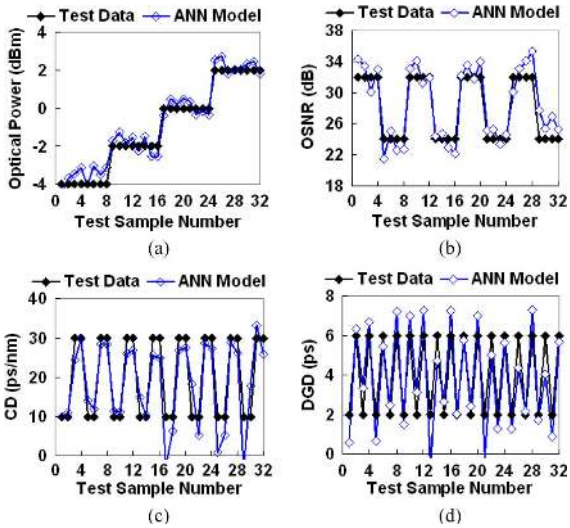


Fig. 10. Simulation results for comparison of testing and ANN-modeled data for a 40 Gb/s RZ-DPSK channel (middle channel) in a 3-channel system. (a) Optical power; (b) OSNR; (c) CD; (d) first-order PMD.

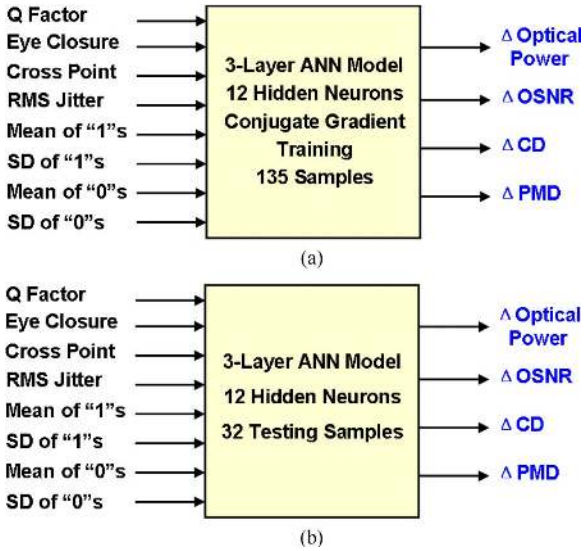


Fig. 11. Block diagrams of using ANN for monitoring changes from a baseline. RMS: root-mean-square; SD: standard derivation. (a) Training; (b) testing.

From the analyses and demonstrations in the former sections, ANNs have been shown to be a potentially powerful tool for OPM. Continuing with this ANN approach, we make use of changes in optical power, OSNR, CD and PMD as the outputs of the neural network, rather than absolute values. Similarly, different impairment combinations imprint different signatures on the obtained eye diagrams, where the input parameters are extracted. Again, the ANN used is an MLP3 with 12 hidden neurons. Fig. 11 shows a block diagram for the training and testing, where 135 samples are used for training (Δ optical power $= -4, -2, 0, 2, 4$ dB; ΔOSNR $= -8, 0, 8$ dB; ΔCD $= -20, 0, 20$ ps/nm; ΔDGD $= -4, 0, 4$ ps) and 32 samples are used for testing (Δoptical power $= -3, -1, 1, 3$ dB; ΔOSNR $= -4, 4$ dB; ΔCD $= -10, 10$ ps/nm; ΔDGD $= -2, 2$ ps).

Similar to Fig. 8, Fig. 12 shows the simulation setup for monitoring impairment causing changes from a baseline. The initial system has a certain amount of impairments, which is added

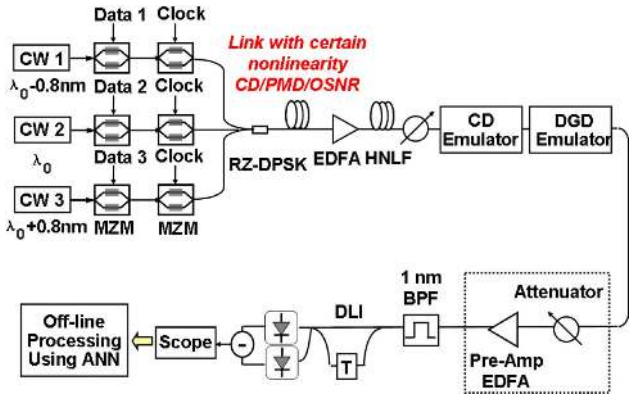


Fig. 12. Simulation setup. λ_0: the wavelength of the channel of interest. $\lambda_0 = 1550$ nm. DLI: delay-line interferometer.

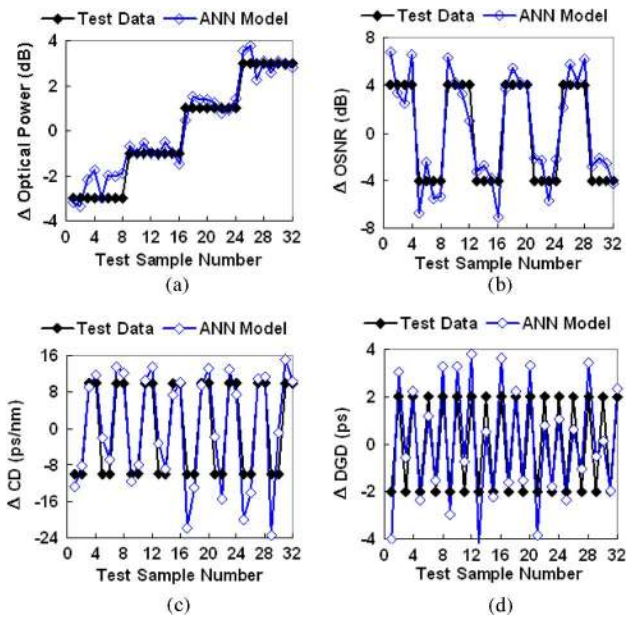


Fig. 13. Simulation results for comparison of testing and ANN-modeled data for a 40 Gb/s RZ-DPSK channel (middle channel) in a 3-channel WDM system. (a) Optical power change; (b) OSNR change; (c) CD change; (d) First-order PMD change.

with a spool of optical fiber. The rest of the system stays the same, as shown in Fig. 8.

The baseline is set to optical power $= -1$ dBm, OSNR $= 28$ dB, CD $= 20$ ps/nm and DGD $= 4$ ps for initial training and testing. The final training error is $\sim$0.1 in this case. The ANN reports a correlation coefficient of 0.93. Fig. 13 compares the testing and ANN-modeled data for the optical power, OSNR, CD, and DGD changes. It is shown that the ANN models, trained with parameters derived from eye diagrams and eye histograms, can potentially be used to simultaneously identify the accumulated fiber nonlinearity, OSNR, CD, and PMD causing changes in WDM channels.

To consider other cases, we vary the baseline arbitrarily and repeat the training and testing. Fig. 14 shows a plot of correlation coefficients for various baselines. We can clearly see that high coefficients are achieved regardless of the baseline, which shows that the ANN approach is largely independent of the system reference. This technique could potentially be valuable for performance monitoring in optical systems with dynamic traffic.

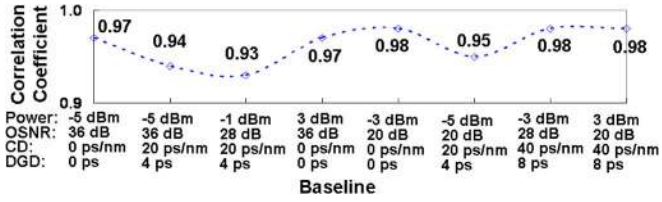


Fig. 14. Variation of the correlation coefficient with various baselines.

The inputs of ANNs so far are derived from eye-diagrams, which in general need clock recovery and are considered high cost. Recently we are able to derive parameters from the asynchronously generated delay-tap plots and train the ANNs to simultaneously identify OSNR, CD and DGD [32]. Moreover, the inputs of ANNs can be any other types of parameters that reflect the changes of impairments. In the following section, we extend the monitoring work to identify the time misalignments in RZ-DQPSK transmitters, in which case the inputs of ANNs are the RF tone/low frequency power levels.

V. ANNs for Time Misalignment Identification in RZ-DQPSK Transmitters

As the modulation format becomes more advanced, the transmitter tends to become more complex in terms of number of components and time synchronization among the components. Due to unavoidable optical/electronic device aging, imperfections and temperature variations, maintaining the correct timing within the transmitter is quite difficult and yet crucial to maintain system performance. Therefore, a laudable goal would be to monitor the time misalignment in order to provide a feedback signal and maintain proper synchronization. For an RZ-QPSK transmitter, the following are important: (i) I and Q data must be aligned with each other, and (ii) the RZ pulse carver must be synchronized to the data.

There have been reports of measurements of time misalignment/synchronization for serial and parallel types of RZ-DQPSK transmitters [33]. In these techniques, a specific parameter is measured, such as power in an RF tone or power in one part of the spectrum. These parameters will either increase or decrease with a particular temporal misalignment. One could use a simple feedback loop that would either maximize or minimize these measured values. However, it would be more valuable if the transmitter could be "trained" to recognize and directly relate RF tone power or spectral power to a specific temporal misalignment cause and value.

Since ANNs have the ability to learn the relationships among sets of input-output data that are characteristic of the device or system under consideration and then apply the relationship to any testing data within the range of interest, we apply this technique to identify the time misalignments in both parallel and serial types of RZ-DQPSK transmitters.

Fig. 15(a) shows the concept of misalignments in parallel-type RZ-DQPSK transmitters. When data streams I and Q are misaligned, the clock tone power at the symbol rate decreases with the increase of the misalignment. When data I/Q are aligned, the RF power at low frequencies increases with the data/pulse carving misalignment. Fig. 15(b) shows the misalignments in a serial-type RZ-DQPSK transmitter. By monitoring the optical clock tone at the symbol rate, we can

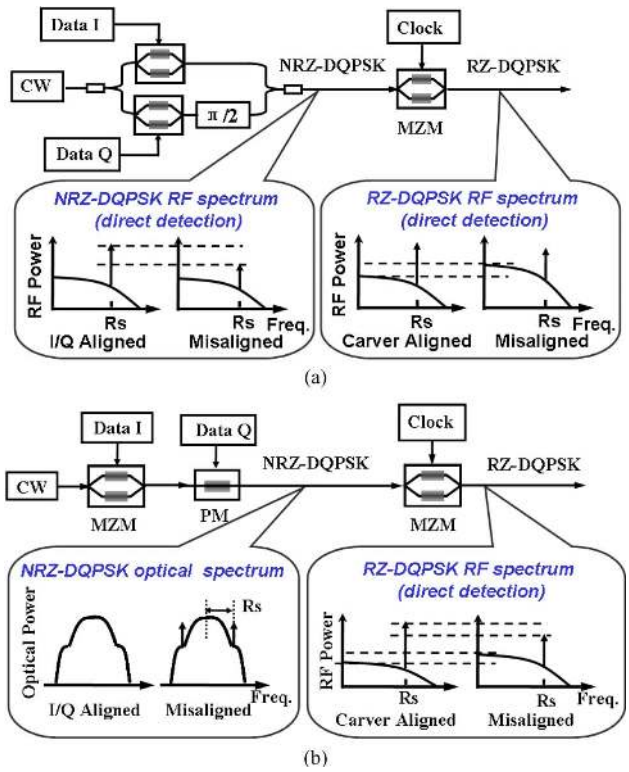


Fig. 15. Conceptual diagram of the misalignments in parallel/serial-type RZ-DQPSK transmitters. (a) Parallel-type. (b) Serial-type.

determine the I/Q misalignment, and the misalignment between data and carver can be monitored by measuring the power change in RF clock tone.

The ANN architecture used in this work is a feed-forward, three-layer perceptron structure. The hidden layer consists of 8 hidden neurons. The conjugate gradient method is used for training. We choose the RF clock tone power after direct detection of DQPSK and low-frequency RF power after direct detection of RZ-DQPSK for the parallel case, while for the serial case, we use the optical clock tone power of DQPSK (which can be filtered by an optical filter and detected by a photodiode to convert to RF power) and the RF clock tone power after direct detection of RZ-DQPSK for the inputs to the ANNs. Note that after directly detecting RZ-DQPSK in the parallel case and the low-frequency RF power in the serial case, the clock power can serve as a third parameter for training.

The training data is a set of 121 samples (I/Q misalignment $= 0 - 50$ ps in steps of 5 ps; carver misalignment $= 0 - 50$ ps in steps of 5 ps). Fig. 16(a) shows the training error versus epochs for the 20-Gb/s parallel RZ-DQPSK transmitter. The final training error is $\sim$0.087 when two inputs are used and $\sim$0.03 when three inputs are used. Once the model is trained, we validate its accuracy by use of a different set of testing data that includes 100 samples (I/Q misalignment $= 2.5 - 47.5$ ps in steps of 5 ps; carver misalignment $= 2.5 - 47.5$ ps in steps of 5 ps). The ANN reports a correlation coefficient of 0.97 and 0.99 for 2-input and 3-input, respectively. Fig. 16(b) and (c) compare the testing and ANN-modeled data for the 2-input and 3-input models. We observe that the 3-input case gives a better prediction.

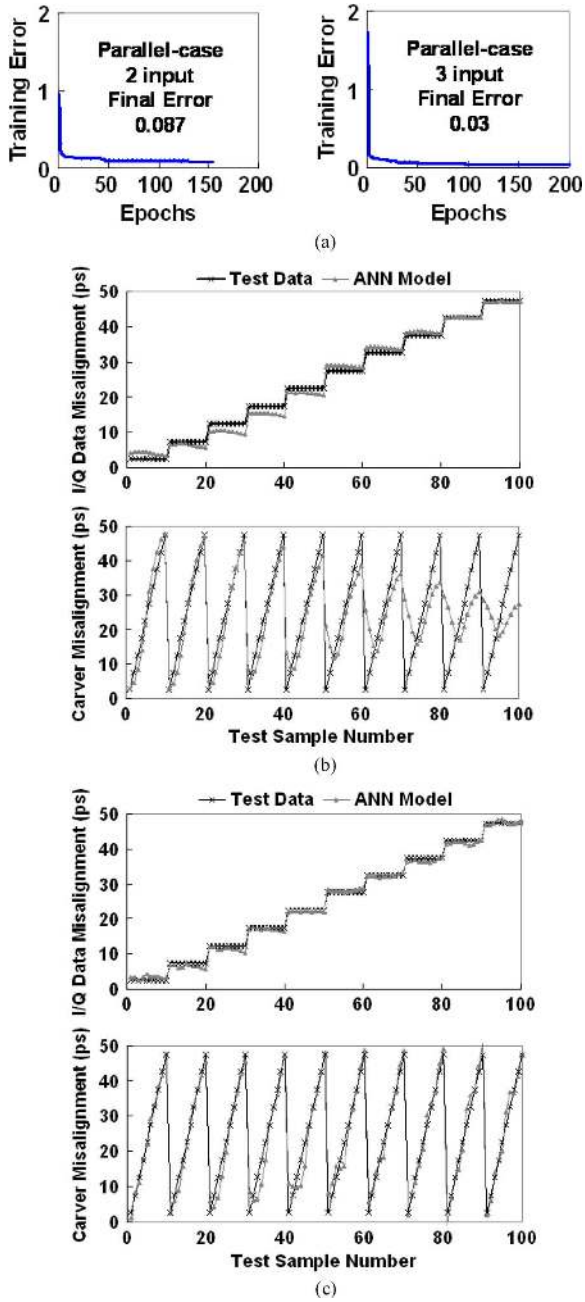


Fig. 16. Simulation results for comparison of testing and ANN-modeled data for 20-Gb/s parallel RZ-DQPSK. (a) Training error; (b) with 2-input ANN model; (c) with 3-input ANN model.

Fig. 17 shows the results for the 80-Gb/s serial-type RZ-DQPSK transmitter. A set of 121 samples (I/Q misalignment $= 0 - 12.5$ ps in steps of 1.25 ps; carver misalignment $= 0 - 12.5$ ps in steps of 1.25 ps) is used for training and another set of 100 samples (I/Q misalignment $= 0.625 - 11.875$ ps in steps of 1.25 ps; carver misalignment $= 0.625 - 11.875$ ps in steps of 1.25 ps) is used for testing. We observe that the 2-input model gives a good prediction, with a correlation coefficient of 0.99. In contrast, the 2-input model in the parallel case does not do as well. The reason is that the RF low-frequency power, which serves as the second input in the parallel-type transmitter depends not only on the carver misalignment but also on the

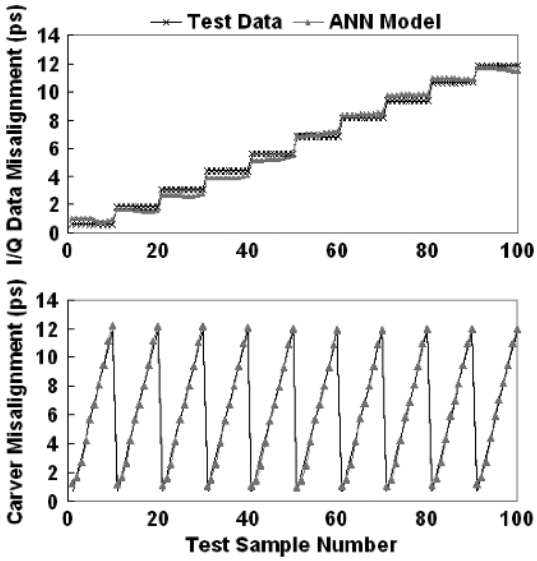


Fig. 17. Simulation results for comparison of testing and ANN-modeled data for 80-Gb/s serial RZ-DQPSK.

I/Q misalignment, while for the serial case, the second input RF clock tone power depends only on the carver misalignment due to the previous phase modulation.

This technique is shown for direct-detection systems, but should also work for coherent systems, since coherent and noncoherent setups can share the same types of transmitters.

VI. Conclusion

In this paper, we proposed and demonstrated a technique of using artificial neural networks for optical performance monitoring. The concept and structure of our neural networks were introduced. Simultaneous identification of OSNR, CD and PMD from eye-diagram parameters was demonstrated in 40 Gb/s OOK and DPSK systems with high correlation coefficients. The technique was extended to identify accumulated channel nonlinear effects in addition to CD, PMD, and OSNR from eye-diagram and eye-histogram parameters in a 3-channel 40 Gb/s DPSK WDM system. A correlation coefficient of 0.97 was obtained for a set of testing data. Furthermore, we proposed using our ANN approach to monitor impairment causing changes from a baseline. Simultaneous identification of accumulated fiber nonlinearity, OSNR, CD, and PMD causing changes from baseline was obtained and high correlation coefficients were achieved with various baselines. ANNs were also used for the simultaneous identification of I/Q data misalignment and data/carver misalignment in both parallel-type and serial-type RZ-DQPSK transmitters. A correlation coefficient of 0.99 was obtained by using a 3-input ANN for the parallel case and a 2-input ANN for the serial case.

We have shown that ANNs are a powerful tool for performance monitoring in optical fiber communication systems. Because ANNs have the ability to model arbitrary relationships between inputs and outputs, they have the potential to be useful in other aspects of optical system and device design.

Acknowledgment

The authors thank Prof. Qi-Jun Zhang from Carleton University for providing them with the NeuroModeler software. They also thank Dr. Anjali Agarwal and Dr. Janet Jackel from Telcordia Technologies, Inc. for valuable discussions.

References

[1] D. C. Kilper, R. Bach, D. J. Blumenthal, D. Einstein, T. Landolsi, L. Olstar, M. Preiss, and A. E. Willner, "Optical performance monitoring," *J. Lightwave Technol.*, vol. 22, no. 1, pp. 294–304, Jan. 2004.

[2] A. Saleh and J. M. Simmons, "Evolution toward the next-generation core optical network," *J. Lightwave Technol.*, vol. 24, no. 9, pp. 3303–3321, Sep. 2006.

[3] Y. C. Chung, "Optical performance monitoring techniques; current status and future challenges," presented at the European Conference on Optical Communication (ECOC) '08, Sep. 2008, paper We.1.D.1.

[4] O. Gerstel, R. Cassata, L. Paraschis, and W. Wakim, "Operational solutions for an Open DWDM layer," presented at the Optical Fiber Communications (OFC) and the National Fiber Optic Engineers Conference (NFOEC) '09, Mar. 2009, paper NThF1.

[5] T. Luo, Z. Pan, S. M. R. Motaghian Nezam, L. S. Yan, A. Sahin, and A. E. Willner, "Chromatic-dispersion-insensitive PMD monitoring using optical off-center bandpass filtering," presented at the Conference on Optical Fiber Communications (OFC) '03, Mar. 2003, paper ThY3.

[6] H. Sun, K. Wu, and K. Roberts, "Real-time measurements of a 40 Gb/s coherent system," *Opt. Express*, vol. 16, no. 2, pp. 873–879, Jan. 2008.

[7] F. N. Hauske, J. C. Geyer, M. Kuschnerov, K. Piyawanno, T. Duthel, C. R. S. Fludger, D. van den Borne, E.-D. Schmidt, B. Spinnler, H. de Waardt, and B. Lankl, "Optical performance monitoring from FIR filter coefficients in coherent receivers," presented at the Conference on Optical Fiber Communications (OFC) '08, Feb. 2008, paper OThW2.

[8] I. Shake, H. Takara, S. Kawanishi, and Y. Yamabayashi, "Optical signal quality monitoring method based on optical sampling," *Electron. Lett.*, vol. 34, no. 22, pp. 2152–2154, Oct. 1998.

[9] N. Hanik, A. Gladisch, C. Caspar, and B. Strebel, "Application of amplitude histograms to monitor performance of optical channels," *Electron. Lett.*, vol. 35, no. 5, pp. 403–404, Mar. 1999.

[10] S. Ohteru and N. Takachio, "Optical signal quality monitor using direct Q-factor measurement," *IEEE Photon. Technol. Lett.*, vol. 11, no. 10, pp. 1307–1309, Oct. 1999.

[11] I. Shake, H. Takara, and S. Kawanishi, "Simple measurement of eye diagram and BER using high-speed asynchronous sampling," *J. Lightwave Technol.*, vol. 22, no. 5, pp. 1296–1302, May 2004.

[12] S. D. Dods and T. B. Anderson, "Optical performance monitoring technique using delay tap asynchronous waveform sampling," presented at the Conference on Optical Fiber Communications (OFC) '06, Mar. 2006, paper OThP5.

[13] S. D. Dods, T. B. Anderson, K. Clarke, M. Bakaul, and A. Kowalczyk, "Asynchronous sampling for optical performance monitoring," presented at the Conference on Optical Fiber Communications (OFC) '07, Mar. 2007, paper OMM5.

[14] B. Kozicki, A. Maruta, and K. Kitayama, "Experimental demonstration of optical performance monitoring for RZ-DPSK signals using delay-tap sampling method," *Opt. Express*, vol. 16, no. 6, pp. 3566–3576, Mar. 2008.

[15] B. Kozicki, A. Maruta, and K. Kitayama, "Transparent performance monitoring of RZ-DQPSK systems employing delay-tap sampling," *J. Opt. Netw.*, vol. 6, no. 11, pp. 1257–1269, Oct. 2007.

[16] R. A. Skoog, T. C. Banwell, J. W. Gannett, S. F. Habiby, M. Pang, M. E. Rauch, and P. Toliver, "Automatic identification of impairments using support vector machine pattern classification on eye diagrams," *IEEE Photon. Technol. Lett.*, vol. 18, no. 22, pp. 2398–2400, Nov. 2006.

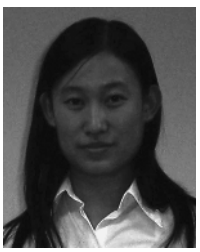
[17] T. B. Anderson, S. D. Dods, K. Clarke, J. Bedo, and A. Kowalczyk, "Multi-impairment monitoring for photonic networks," presented at the European Conference on Optical Communication (ECOC) '07, Sep. 2007, paper 3.5.1.

[18] T. B. Anderson, K. Clarke, D. Beaman, H. Ferra, M. Birk, G. Zhang, and P. Magill, "Experimental Demonstration of Multi-Impairment Monitoring on a commercial 10 Gbit/s NRZ WDM channel," presented at the Conference on Optical Fiber Communications (OFC)'09, Mar. 2009, paper OThH7.

[19] J. Jargon, X. Wu, and A. E. Willner, "Optical performance monitoring using artificial neural networks trained with eye-diagram parameters," *IEEE Photon. Technol. Lett.*, vol. 21, no. 1, pp. 54–56, Jan. 2009.

[20] X. Wu, J. Jargon, and A. E. Willner, "Off-line monitoring of CD/PMD/OSNR degradation effects using neural-network-based training sequences," presented at the European Conference on Optical Communication (ECOC) '08, Sept. 2008, paper We.3.D.6.

[21] X. Wu, J. Jargon, L. Christen, and A. E. Willner, "Training of neural networks to perform optical performance monitoring of a combination of accumulated signal nonlinearity, CD, PMD, and OSNR," presented at the IEEE LEOS Annual Meeting 2008, Nov. 2008, paper WP 4.
[22] X. Wu, J. Jargon, L. Christen, L. Paraschis, and A. E. Willner, "Monitoring I/Q data and pulse carving misalignments in RZ-DQPSK transmitters using a neural network approach," presented at the IEEE LEOS Annual Meeting 2008, Nov. 2008, paper WP 2.
[23] M. H. Hassoun, *Fundamentals of Artificial Neural Networks*. : the MIT Press, 1995.
[24] Q. J. Zhang and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Norwood, MA: Artech House, 2000.
[25] K. C. Gupta, "Emerging trends in millimeter-wave CAD," *IEEE Trans. Microwave Theory Tech.*, vol. 46, pp. 747–755, 1998.
[26] V. K. Devabhaktuni, M. Yagoub, Y. Fang, J. Xu, and Q. J. Zhang, "Neural networks for microwave modeling: Model development issues and nonlinear modeling techniques," *Int. J. RF Microwave CAE*, vol. 11, pp. 4–21, 2001.
[27] F. Wang, V. K. Devabhaktuni, C. Xi, and Q. J. Zhang, "Neural network structures and training algorithms for microwave applications," *Int. J. RF Microwave CAE*, vol. 9, pp. 216–240, 1999.
[28] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.
[29] G. Thimm and E. Fiesler, "High-order and multilayer perceptron initialization," *IEEE Trans. Neural Networks*, vol. 8, pp. 349–359, 1997.
[30] Q. J. Zhang, K. C. Gupta, and V. K. Devabhaktuni, "Artificial neural networks for RF and microwave design: from theory to practice," *IEEE Trans. Microwave Theory Tech.*, vol. 51, pp. 1339–1350, Mar. 2003.
[31] NeuroModeler, ver. 1.5, Q.J. Zhang and His Neural Network Research Team. Ottawa, Canada, Department of Electronics, Carleton University, 2004. Ottawa, Canada.
[32] J. A. Jargon, X. Wu, and A. E. Willner, "Optical performance monitoring using artificial neural networks trained with parameters derived from delay-tap asynchronous sampling," presented at the Conference on Optical Fiber Communications (OFC) '09, Mar. 2009, paper OThH1.
[33] X. Wu, L. Christen, B. Zhang, W.-R. Peng, J.-Y. Yang, L. Zhang, S. R. Nuccio, J. A. Jargon, A. E. Willner, and L. Paraschis, "Synchronization monitoring of I/Q data and pulse carving misalignment for a parallel-type RZ-DQPSK transmitter by measuring RF Clock tone/low frequency power," *IEEE Photon. Technol. Lett.*, vol. 20, no. 24, pp. 2138–2140, Dec. 2008.



Xiaoxia Wu (S'06) received the B.E. degree from Jilin University, Changchun, China, in 2004, and the M.E degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2006. She is currently working toward the Ph.D. degree in electrical engineering at University of Southern California, Los Angeles.

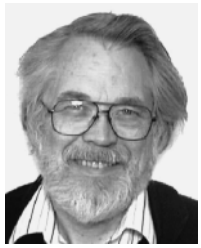
Her current research interests include high-speed optical/electronic signal processing, optical performance monitoring and advanced modulation formats. She has authored or coauthored more than 60 papers in prestigious international journals and conferences.

Ms. Wu is also a student member of the Optical Society of America (OSA) and the International Society for Optical Engineering (SPIE).

Jeffrey A. Jargon (M'98–SM'01) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the University of Colorado at Boulder in 1990, 1996, and 2003, respectively.

He has been with the National Institute of Standards and Technology (NIST), Boulder, CO, since 1990, and has conducted research in the areas of vector network analyzer calibrations and microwave metrology. He is presently a member of the High-Speed Measurements Project in the Optoelectronics Division. He has published over 60 technical articles

Dr. Jargon has received four best paper awards, an URSI Young Scientist Award, and a Department of Commerce Silver Medal Award. He is a member of Tau Beta Pi and Eta Kappa Nu, a Senior Member of IEEE, and is a registered Professional Engineer in the State of Colorado.



Ronald A. Skoog (M'89) received the B.S. degree from Oregon State University, Corvallis, and the M.S. and Ph.D. degrees in electrical engineering (control and systems theory) from the Massachusetts Institute of Technology (MIT), Cambridge.

Prior to joining Telcordia, he spent 29 years at Bell Laboratories/AT&T Bell Laboratories/AT&T Laboratories working in the areas of transport network design; signaling network design, protocols, and performance/reliability studies; and circuit-switched network systems engineering and performance/reliability studies. He has been a Senior Scientist at Telcordia Technologies, Inc., Red Bank, NJ, since October 1998, and during that time he has worked in the areas of optical networking architectures, Internet protocol/wavelength-division- multiplexing network architectures and evolution studies, optical network management and control, emerging network technology studies, and reliability studies for optical networks and optical network elements.

Dr. Skoog is a Member of the IEEE Communications Society (COMSoc), the IEEE Lasers & Electro-Optics Society (LEOS), the Optical Society of America (OSA), and Sigma Xi.

Loukas Paraschis (SM'06) was born in Athens, Greece, where he completed his undergraduate studies. He received the M.S.E.E. degree in 1998 and the Ph.D. degree in applied physics in 1999 from Stanford University, Stanford, CA.

He is Business Development Manager at Cisco, responsible for next generation core networks in the emerging markets, service provider, solution architectures and scaling group. At Cisco, he has worked also on optical transport technologies, WDM systems, multi-service metro and IP-over-WDM architectures, and the associated business models and market development efforts. Prior to his current role, he worked as an R&D engineer, product manager, and technical leader in optical networks and core routing. He has (co)authored more than 50 peer-reviewed publications, invited, and tutorial presentations, a book chapter, multiple technical reports, and three patent applications.

Dr. Paraschis has been Associate Editor for Optical Networks of the *Journal of Communication and Networks*, guest editor of the IEEE JOURNAL OF LIGHTWAVE TECHNOLOGY, and member of the IEEE, the OSA, and multiple conference organizing committees.

Alan E. Willner (S'87–M'88–SM'93–F'04) received the Ph.D. degree in electrical engineering from Columbia University, New York, NY, in 1988.

He has worked at AT&T Bell Laboratories and Bellcore. He is currently a Professor at the University of Southern California, Los Angeles. He has 775 publications, including two books and 25 patents.

Prof. Willner is a Fellow of the Optical Society of America (OSA) and was a Fellow of the Semiconductor Research Corporation. His professional activities have included the following: President of the IEEE Lasers and Electro-Optics Society (LEOS), Editor-in-Chief of the IEEE/OSA JOURNAL OF LIGHTWAVE TECHNOLOGY, Editor-in-Chief of OSA OPTICS LETTERS, Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN QUANTUM ELECTRONICS, Co-Chair of the OSA Science and Engineering Council, General Co-Chair of the Conference on Lasers and Electro-Optics (CLEO), Chair of the IEEE TAB Ethics and Conflict Resolution Committee, General Chair of the LEOS Annual Meeting Program, Program Co-Chair of the OSA Annual Meeting, and Steering and Program Committee Member of the Conference on Optical Fiber Communications (OFC). He has received the National Science Foundation (NSF) Presidential Faculty Fellows Award from the White House, the Packard Foundation Fellowship, the NSF National Young Investigator Award, the Fulbright Foundation Senior Scholars Award, the IEEE LEOS Distinguished Traveling Lecturer Award, the USC University-Wide Award for Excellence in Teaching, and the Eddy Award from Pennwell for the Best Contributed Technical Article.