Optical performance monitoring of QPSK data channels by use of neural networks trained with parameters derived from asynchronous constellation diagrams

Jeffrey A. Jargon[1,*], Xiaoxia Wu[2], Hyeon Yeong Choi[3], Yun C. Chung[3], and Alan E. Willner[2]

[1]National Institute of Standards and Technology, Boulder, CO 80305 USA
[2]University of Southern California, Los Angeles, CA 90089 USA
[3]Korea Advanced Institute of Science and Technology, Daejeon 305-701, KOREA
*jargon@boulder.nist.gov

Abstract: We demonstrate a technique for performance monitoring of quadrature phase-shift keying data channels by simultaneously identifying optical signal-to-noise ratio (OSNR), chromatic dispersion (CD), and polarization-mode dispersion (PMD) using neural networks trained with parameters derived from asynchronous constellation diagrams. A correlation coefficient of 0.987 is reported for a set of testing data from a 40 Gbps return-to-zero, quadrature phase-shift keying (RZ-QPSK) system. The root-mean-square (RMS) errors are 0.77 dB for OSNR, 18.71 ps/nm for CD, and 1.17 ps for DGD.

References and links

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. **22**(1), 294–304 (2004).
2. I. Shake, H. Takara, S. Kawanishi, and Y. Yamabayashi, "Optical Signal Quality Monitoring Method Based on Optical Sampling," Electron. Lett. **34**(22), 2152–2154 (1998).
3. N. Hanik, A. Gladisch, C. Caspar, and B. Strebel, "Application of Amplitude Histograms to Monitor Performance of Optical Channels," Electron. Lett. **35**(5), 403–404 (1999).
4. S. Ohteru, and N. Takachio, "Optical Signal Quality Monitor Using Direct Q-Factor Measurement," IEEE Photon. Technol. Lett. **11**(10), 1307–1309 (1999).
5. S. D. Dods and T. B. Anderson, "Optical Performance Monitoring Technique Using Delay Tap Asynchronous Waveform Sampling," OFC/NFOEC Technical Digest, 2006, paper OThP5.
6. B. Kozicki, A. Maruta, and K. Kitayama, "Asynchronous Optical Performance Monitoring of RZ-DQPSK Signals Using Delay-Tap Sampling," ECOC Conference Proceedings, 2007, paper P060.
7. J. A. Jargon, X. Wu, and A. E. Willner, "Optical Performance Monitoring by Use of Artificial Neural Networks Trained with Parameters Derived from Delay-Tap Asynchronous Sampling," OFC/NFOEC Technical Digest, 2009, paper OThH1.
8. H. Y. Choi, Y. Takushima, and Y. C. Chung, "Multiple-Impairment Monitoring Technique Using Optical Field Detection and Asynchronous Delay-Tap Sampling Method," OFC/NFOEC Technical Digest, 2009, paper OThJ5.
9. T. B. Anderson, A. Kowalczyk, K. Clarke, S. D. Dods, D. Hewitt, and J. C. Li, "Multi Impairment Monitoring for Optical Networks," J. Lightwave Technol. **27**(16), 3729–3736 (2009).
10. R. A. Skoog, T. Banwell, J. Gannett, S. Habiby, M. Pang, M. Rauch, and P. Toliver, "Automatic Identification of Impairments Using Support Vector Machine Pattern Classification on Eye Diagrams," IEEE Photon. Technol. Lett. **18**(22), 2398–2400 (2006).
11. J. A. Jargon, X. Wu, and A. E. Willner, "Optical Performance Monitoring Using Artificial Neural Networks Trained with Eye-Diagram Parameters," IEEE Photon. Technol. Lett. **21**(1), 54–56 (2009).
12. X. Wu, J. A. Jargon, R. A. Skoog, L. Paraschis, and A. E. Willner, "Applications of Artificial Neural Networks in Optical Performance Monitoring," J. Lightwave Technol. **27**(16), 3580–3589 (2009).
13. P. J. Winzer, and R. Essiambre, "Advanced Optical Modulation Formats," Proc. IEEE **94**(5), 952–985 (2006).
14. V. Arbab, X. Wu, A. E. Willner, and C. Weber, "Optical Performance Monitoring of Data Degradation by Evaluating the Deformation of an Asynchronously Generated I/Q Data Constellation," European Conference on Optical Communications (ECOC) 2009, paper P3.23.

15. M. H. Hassoun, *Fundamentals of Artificial Neural Networks* (The MIT Press, 1995).
16. K. Hornik, M. Stinchcombe, and H. White, "Multilayer Feedforward Networks Are Universal Approximators," Neural Netw. 2(5), 359–366 (1989).
17. "NeuroModeler, ver. 1.5," Q. J. Zhang and His Neural Network Research Team, Department of Electronics, Carleton University, Ottawa, Canada, 2004.
18. D. Dahan, D. Levy, and U. Mahlab, "Low Cost Multi-Impairment Monitoring Technique for 43 Gbps DPSK and 86 Gbps DP-DPSK Using Delay Tap Asynchronous Sampling Method," European Conference on Optical Communications (ECOC) 2009, paper P3.01.

1. Introduction

As optical fiber transmission systems become more transparent and reconfigurable, optical performance monitoring (OPM) is essential for ensuring high quality of service [1]. Crucial impairments in optical networks include optical signal-to-noise ratio (OSNR), chromatic dispersion (CD), and polarization-mode dispersion (PMD).

The all-optical approaches for OPM have been shown to be powerful. 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 monitoring optical performance by use of off-line digital signal processing of received electrical data signals [2–12]. Three of these methods [2–4] utilize amplitude histograms or power distributions to estimate bit-error rate (BER); five [5–9] employ delay-tap plots to distinguish among impairments; and three [10–12] use parameters derived from eye diagrams 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.

With the ever increasing demands for higher capacities and longer distances in optical communications systems, advanced modulation schemes have been developed to attain higher spectral efficiencies and lower bit-error rates [13]. Many of these formats utilize phase modulation for encoding information, such as binary and quadrature phase-shift keying (BPSK and QPSK). With phase-modulated signals, constellation diagrams are often used as a visualization tool, as opposed to eye diagrams, which are typically used for amplitude-modulated signals. Initial analysis of asynchronous constellation diagrams show that the diagrams deform in a fairly predictable way with certain impairments [14]. However, the scalability of including additional impairments may be limited.

Since artificial neural networks (ANNs) have been shown to be a powerful modeling tool for identifying simultaneous impairments derived from eye-diagrams and asynchronous delay-tap plots [7, 11,12], we explore their use for the same purpose using parameters derived from asynchronous constellation diagrams. In the following sections, we briefly overview ANNs, and provide an example of our proposed method with a simulated optical channel operating at 40 Gbps using return-to-zero, quadrature phase-shift keying (RZ-QPSK).

2. Artificial neural networks

ANNs are neuroscience-inspired computational tools that are trained by use of input-output data to generate a desired mapping from an input stimulus to the targeted output [15]. ANNs consist of multiple layers of processing elements called neurons. Each neuron is linked to other neurons in neighboring layers by varying coefficients that represent the strengths of these connections. ANNs learn relationships among sets of input-output data that are characteristic 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 drop below prescribed values.

The ANN architecture used in this work is a feed-forward, three-layer perceptron structure (MLP3) consisting of an input layer, a hidden layer, and an output layer, as shown in Fig. 1.

The hidden layer allows complex models of input-output relationships. The mapping of these relationships is given by:

$$Y = g\left[W_2 \bullet g\left(W_1 \bullet X\right)\right],$$ (1)

where X is the input vector, Y is the output vector, and W_1 and W_2 are respectively the weight matrices between the input and hidden layers and between the hidden and output layers. The function $g(u)$ is a sigmoidal activation function given by:

$$g(u) = 1/\left[1 + \exp(-u)\right],$$ (2)

where u is the input to a hidden neuron. According to [16], an MLP3 with one hidden sigmoidal layer is able to model almost any physical function accurately, provided that a sufficient number of hidden neurons are available.

The most important step in neural network model development is the training process. Here, 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.

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 asynchronous constellation diagrams, and the outputs are the impairments, including OSNR, CD, and PMD. The neural network training error can be defined as:

$$E_{T_r}(w) = \frac{1}{2}\sum_{n \in T_r}\sum_{k=1}^{K}\left|y_k\left(x_n, w\right) - d_{kn}\right|^2,$$ (3)

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 X and the weight vector W, neural network feed-forward computation is a process used to compute the output vector 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's neurons are fed to the next layer's neurons. During feed-forward computation, neural network weights 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.
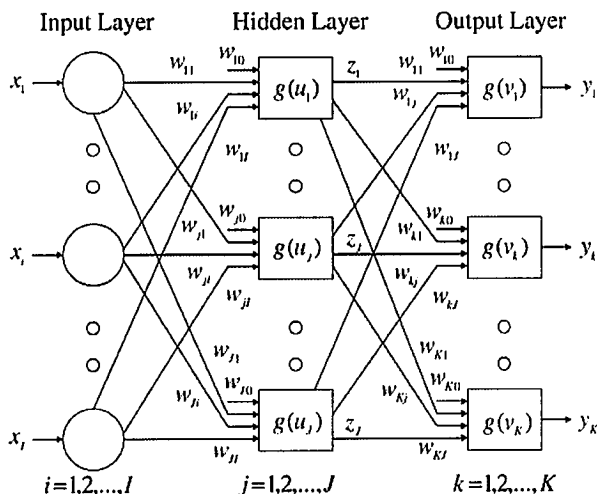

Fig. 1. Artificial neural network architecture.

3. Methodology

Asynchronous constellation diagrams (also known as transition, I-Q, and vector diagrams) are similar to regular constellation diagrams in that they both display signals in complex space, namely the in-phase signal (I) on the x-axis versus the quadrature signal (Q) on the y-axis. However, asynchronous constellation diagrams also show the signal transitions between symbols. Figure 2 illustrates simulated asynchronous constellation diagrams for a 40 Gbps RZ-QPSK signal at a few select combinations of CD and PMD for an OSNR of 32 dB. Visually, it is obvious that these impairments produce distinct features.

(a) None. (b) DGD only.
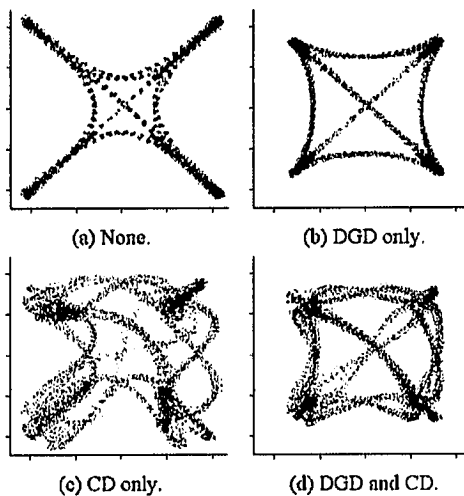
(c) CD only. (d) DGD and CD.

Fig. 2. Asynchronous constellation diagrams with various impairments.

To quantify the distinct features, we need to derive parameters that can be calculated from the asynchronous constellation diagrams. Whereas eye diagrams give rise to widely used parameters such as Q-factor, closure, jitter, and crossing amplitude, there are no such parameters available for constellation diagrams. Thus, as in [7], where we utilized delay-tap plots, we once again define new parameters that will help us capture the behavior of asynchronous constellation diagrams. One possibility is to divide them into quadrants Q1-Q4. The data pairs are divided into the quadrants as follows: $(x_i, y_i) \in Q1$ if $\{x_i < 0 \text{ and } y_i < 0\}$; $(x_i, y_i) \in Q3$ if $\{x_i > 0 \text{ and } y_i > 0\}$, as shown in Fig. 3. Potential parameters from the asynchronous constellation diagrams include the statistics of x, y, magnitude, and phase of each quadrant. More complex parameters can also be defined based on statistical calculations.
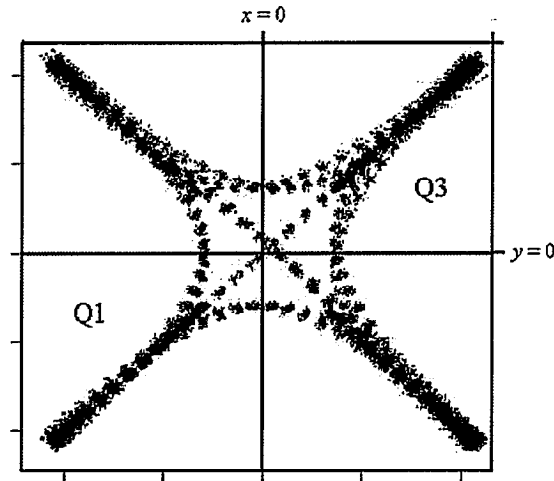

Fig. 3. Dividing the asynchronous constellation diagram into quadrants.

Quadrants 2 and 4 are not used in this particular application. With the two quadrants defined, we can perform some basic statistical calculations on the data within each quadrant, such as means and standard deviations. For quadrants 1 and 3, we calculate the means and standard deviations of the magnitudes ($\bar{r}_1, \sigma_{r1}, \bar{r}_3, \sigma_{r3}$), rather than the x's and y's separately, because constellation diagrams contain data that are roughly symmetric about the 45° axis. Additionally, we calculate the maximum and minimum values of the y's at the $x = 0$ axis (y_{max} and y_{min}), since these values vary with OSNR and DGD, and tend not to be symmetrical with CD. One final parameter we make use of is similar to the Q-factor, which we define as $Q_{31} = (\bar{r}_3 - \bar{r}_1) / (\sigma_{r1} + \sigma_{r3})$. The reasons that only quadrants 1 and 3 are used include: 1) the shape changes in quadrants 1 and 3 are more obvious as compared to quadrants 2 and 4; 2) the parameters from quadrants 1 and 3 provide good matching between testing data and ANN-modeled data, as we can see in the next few paragraphs. If additional impairments were to be included in the monitor, then the parameters from quadrants 2 and 4 would probably be required [12].

To illustrate our method, we performed 216 simulations using the following impairment combinations: OSNR – 12, 16, 20, 24, 28, and 32 dB; CD – 0, 40, 80, 120, 160, and 200 ps/nm; and PMD with values of differential group delay (DGD) equal to 0, 4, 8, 12, 16, and 20 ps.

Figure 4 shows the configuration used in the simulation. The RZ-DQPSK transmitter consisted of a continuous wave (CW) laser operating at 1550 nm with a linewidth of 100 KHz, a parallel-type DQPSK modulator, which was driven by two 20 Gbps non-return to zero (NRZ) 2^{15}-1 pseudo-random binary sequences (PRBS) and a Mach-Zehnder modulator (MZM) for pulse carving. The I/Q channels were decorrelated by use of different PRBS

orders. The pulse carver was driven by a 20 GHz sinusoidal clock signal to generate a 50% duty-cycle RZ pulse train. The generated RZ-DQPSK signal was then sent to a CD emulator, followed by a DGD (i.e. first order PMD) emulator. The output was sent to an Erbium-doped fiber amplifier (EDFA) with a variable optical attenuator in front to adjust the received OSNR. The signal was then filtered by a bandpass filter (BPF) with 0.8 nm bandwidth, and sent to the receiver, where the constellation diagrams and parameters were extracted. The receiver consisted of two out-of-phase delay-line-interferometers (DLI) followed by two balanced photo-receivers (BPDs) and the received I and Q signals were finally sampled with analog to digital converters (ADCs) asynchronously, the sampling rate of which can be much lower than the data rate. The outputs of ADCs formed the coordinates of the data in the complex plane for constructing the constellation diagrams, such that all the received samples could be plotted in one I/Q plane asynchronously during the offline processing.
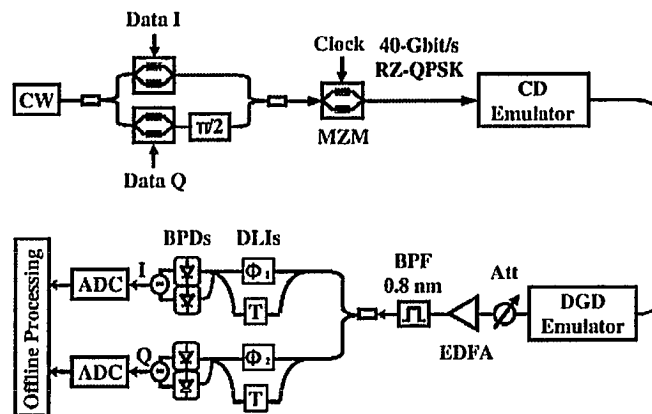

Fig. 4. Simulation model. CW: continuous-wave; I: in-phase; Q: quadrature-phase; MZM: Mach-Zehnder modulator; CD: chromatic dispersion; DGD: differential group delay; EDFA: Erbium-doped fiber amplifier; BPF: bandpass filter; Att: optical attenuator; DLI: delay-line interferometer; Φ_1: + 45°;Φ_2: -45°; T: 1-symbol time, 50 ps; BPD: balanced photo-detector; ADC: analog-to-digital convertor.

The ANN consisted of seven inputs ($\bar{r_1}, \sigma_{r1}, \bar{r_3}, \sigma_{r3}, y_{max}, y_{min}, \mathcal{Q}_{31}$), three outputs (OSNR, CD, and DGD), and 28 hidden neurons. The ANN was trained by use of a software package developed by Zhang et al. [17]. Although alternatives were explored, a conjugate-gradient technique was chosen, because it offers a nice compromise in terms of memory requirements and implementation effort.

Once the model was trained, we validated its accuracy with a different set of testing data. We used 125 simulations with the following impairment combinations: OSNR – 14, 18, 22, 26, and 30 dB; CD – 20, 60, 100, 140, and 180 ps/nm; and DGD – 2, 6, 10, 14, and 18 ps. The software reported a correlation coefficient of 0.987 for the testing data. The root-mean-square (RMS) errors were 0.77 dB for OSNR, 18.71 ps/nm for CD, and 1.17 ps for DGD. Figure 5 compares the testing and ANN-modeled data for OSNR, CD, and DGD.
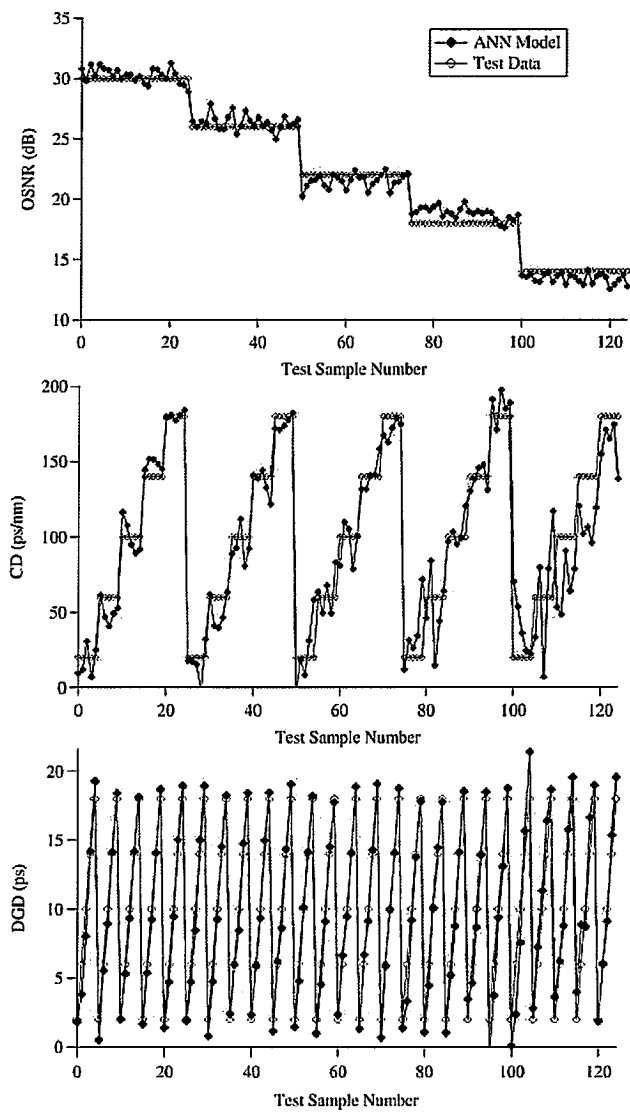
Fig. 5. Comparison of testing and ANN-modeled data for the 40 Gbps RZ-QPSK channel.

4. Conclusions

We have shown how ANN models trained with parameters derived from asynchronous constellation diagrams can be used to simultaneously identify levels of OSNR, CD, and DGD for 40 Gbps RZ-QPSK signals. This method provides a powerful technique for monitoring the performance of optical QPSK data channels.

Although not discussed here, it should be possible to identify CD, PMD, and OSNR in the presence of nonlinearities using this technique. In a previous publication [12], we were able to do so with 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 accumulated fiber nonlinearity. Although additional input parameters were required for the ANN, we were able to achieve a correlation coefficient of 0.97.

Furthermore, this neural network approach should also work for other modulation formats, such as NRZ-QPSK, although it would require another set of training data for generating a new neural network.

An initial demonstration using neural networks trained with parameters derived from delay-tap asynchronous sampling plots was recently performed to identify multiple impairments for an 86 Gbps polarization-multiplexed DPSK signal [18]. It should be mentioned that this approach of using ANNs trained with parameters derived from asynchronous constellations could also potentially be used in the case of polarization-multiplexed QPSK (PM-QPSK), since this will likely be a format used with 100 Gbps signals detected with coherent receivers. Such a monitor would probably require a different set of training parameters given the fact that the polarization dependent impairments, such as polarization dependent loss, may play a more important role in the polarization-multiplexed system.

Acknowledgments