OSNR Monitoring by Deep Neural Networks Trained with Asynchronously Sampled Data

Takahito Tanimura[1,2], Takeshi Hoshida[1], Jens C. Rasmussen[1], Makoto Suzuki[2], and Hiroyuki Morikawa[2]
[1] Fujitsu Laboratories Ltd., 4-1-1 Kamikodanaka, Nakahara-ku, Kawasaki 211-8588, Japan
[2] Research Center for Advanced Science and Technology, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8904, Japan
tanimura.taka@jp.fujitsu.com

Abstract: *We demonstrate a use of deep neural networks (DNN) for OSNR monitoring with minimum prior knowledge. By using 5-layers DNN trained with 400,000 samples, the DNN successfully estimates OSNR in a 16-GBd DP-QPSK system.*
Keywords: *machine learning, optical performance monitoring, coherent detection*

I. INTRODUCTION

For enabling programmable and autonomous optical networks, optical physical layer monitoring is considered as an essential part of the network systems. The networks should handle various types of optical signals modulated in different signal format and baud rate to accommodate diverse transport requirements [1]. Thus the optical physical layer monitor should enable to extract physical information from incoming optical signals with minimum prior knowledge.

Aiming at a generic monitoring without a prior knowledge, the use of an artificial neural network (ANN) has been investigated [2-4]. Although all of existing ANN-based monitors could estimate physical condition of input signals, e.g. OSNR, chromatic dispersion and differential group delay, they require careful engineering to design a feature extractor that transforms the raw data into a feature vector that is a suitable internal representation. In [2], for example, a set of parameters including Q-factor, eye-closure, RMS-jitter, and crossing amplitude of eye-diagram was used for ANN input. The parameter set was chosen by a skilled engineer who holds expertise of modulated optical signal. In other words, the conventional ANN-based monitors used in [2-4] were limited in their scalability to more general set of signals with different symbol rates and modulation formats.

Recently, a deep neural network (DNN), which is an ANN with multiple hidden layers of units between the input and output layers, has attracted attention for various applications: such as visual object recognition and speech recognition [5]. The extra layers of the DNNs would enable automatic extraction of features of input, giving the potential of modeling complex data with a fewer units than similarly performing shallow networks [5].

In this paper, we propose and experimentally demonstrate the use of the DNNs for an OSNR monitor without specific feature engineering for raw data, as shown in Fig. 1(a). The feature vectors are learned in the DNNs from asynchronously sampled raw data by coherent receiver. As the number of training dataset is increased, the DNN starts to obtain a function of feature extractor of incoming signals. Varying the number of layer of the DNN, we experimentally evaluate a required number of layers in this specific case. By using 5-layer DNN trained with 400,000 dataset, OSNR of 16 GBd DP-QPSK signal is successfully estimated with the range of 7.5 to 31 dB.

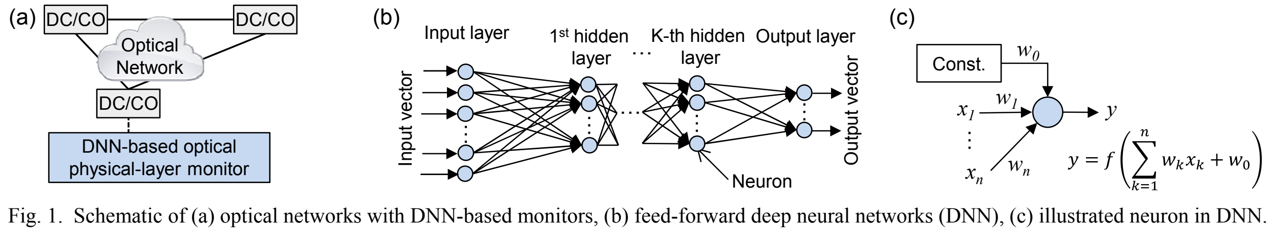


Fig. 1. Schematic of (a) optical networks with DNN-based monitors, (b) feed-forward deep neural networks (DNN), (c) illustrated neuron in DNN. DC: data center, CO: central office.

II. CONCEPT OF DEEP NEURAL NETWORK

Figure 1(b) shows a schematic of a feed-forward, full-connected multi-layer perceptron DNN used in this study. The DNNs are neuroscience-inspired information processing models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. The first and last layers of the DNNs are an input and output layers, respectively. The other layers are called hidden layers. Each layer of the DNN contains multiple neurons that are connected to other neurons in neighboring layers by adaptive weights w_k. Each neuron has activation function $f(u)$ and calculates output of neuron from multiple inputs as shown in Fig. 1(c).

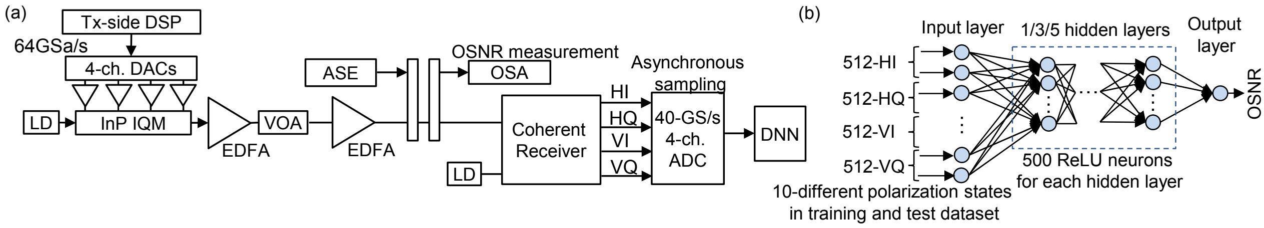
The connection weights of DNN are trained by input-output dataset, i.e. supervised learning. For example, input is measured raw values related with optical signals, and output is OSNR. The DNN learns a relationship between input and output that is characteristic of the system under consideration. After calculations of output vectors, the output vectors are compared to the desired output vectors and errors are calculated. Error derivatives are then also calculated and summed for each weight. The error derivatives are used to update the connection weights for the neurons, and training continues until the errors reach enough low values. After training, the connection weights of DNN are fixed and the DNN can be tested by other set of data.

III. Experiments

Figure 2(a) shows the experimental setup. At the transmitter, an external cavity laser (~25 kHz linewidth) was used as a light source for the channel at 193.3 THz. An integrated InP DP-IQ-modulator was driven by the drive signals generated by a four channel digital-to-analogue converters (DAC) with a sampling rate of 64 GSa/s, a physical resolution of 8 bits. The DAC generated Nyquist-filtered (roll-off factor = 0.01) 16 GBd DP-QPSK signal with pilot CW tones for carrier recovery (a detail of DSP is shown in [1]). The modulated signal was sent to an erbium-doped fiber amplifier (EDFA) and additional ASE noise was loaded to vary the received OSNR from 7.5 to 31 dB. The received OSNR was measured by optical spectrum analyzer (OSA). Chromatic dispersion was not imposed in this experiment, left for future work.

At the receiver, the local oscillator (linewidth ~25 kHz) was superimposed with the signal in a polarization-diversity optical 90° hybrid. The outputs of the hybrid were connected to four balanced photo-detectors. The resulting signals were digitized by four analog-to-digital converters (ADC) with a sample rate of 40 GS/s and a bandwidth of 16 GHz. The digital samples were processed with offline manner in desktop computer equipped with GPGPU.

The four-tributary dataset sampled by the ADCs was fed into the DNN, as shown in Fig. 2(b). Each tributary corresponding to HI/HQ/VI/VQ was 512 time-concatenated data sampled with 40 GS/s. Corresponding polarization states of both training and test dataset were randomly distributed on ten different polarization states. The DNN used in this study consists of 2,048 neurons in input layers and one neuron in output layers to predict OSNR. The activation function of output neuron was linear function. The number of hidden layer was set to 1, 3, and 5. Each hidden layer holds 500 neurons that have a activation function as rectified linear unit (ReLU), which is simply the half-wave rectifier $f(z) = max(z, 0)$, allowing training of deep supervised network without unsupervised pre-training [6]. The DNNs were trained by use of a Theano/Pylearn2 [7] software package. Batch gradient descent (BGD) and a conjugate-gradient technique were used for training. 4,000, 40,000, and 400,000 training dataset were used for training of the DNN, and another 10,000 test dataset was used for test.



. Fig. 2. (a) Experimental setup, (b) details of the DNN used in this study. InP IQM: Indium phosphide IQ modulator, LD: laser diode, VOA: variable optical attenuator, ASE: amplified spontaneous emission noise source, ReLU: rectified linear unit.

IV. Results and Discussion

Figure 3 shows the estimated OSNR by 3, 5, and 7 layers DNNs as a function of measured OSNR. Black circles in Fig. 3 show averaged value of estimated OSNR over test dataset. Each DNN was trained with 4,000, 40,000, and 400,000 dataset, respectively. Obviously, the 3 layers ANNs (or shallow neural networks) were failed to learn a relationship between input and output as shown in Fig. 3(a)-(c). Even with increased number of layer, the situation was not changed when the amount of training data was very limited such as 4,000 dataset as shown in Fig. 3 (a), (d), and (g). However, with increasing the number of both training dataset and layer, the situation was started to change as shown in Fig. 3 (e), (f), (h) and (i). The DNN might form themselves to a feature extractor by use of enormous number of training dataset such as 400,000 data. To the end, the DNN automatically learned a relationship between input raw data and OSNR as shown in Fig. 3(f) and (i).

For detailed discussion, we evaluated averaged errors between estimated and measured OSNR. Figure 4(a) shows averaged errors over test dataset or training dataset as a function of number of DNN layer with a fixed training dataset of 400,000. As the number of layer was increased, the both errors are reduced and saturated around 5 layers. According to Fig. 4(a), the 5 layer DNN seems likely to have enough deep layers in this specific case. Figure 4 (b) shows the averaged errors on both training and test dataset as a function of number of training dataset with 5 layers DNN. Since increasing of number of training dataset was beneficial for reducing the errors on test dataset, the over fitting of model may have not occurred yet in this case. Although more training dataset would provide an opportunity to additionally reduce the errors, the measured averaged error of 1.6 dB in this work (at the ranging from 7.5 to 31 dB OSNR) would be comparable to an averaged error of 1.23 dB (OSNR 12 - 32 dB) by the existing ANN-based monitor for RZ-DPSK signals with feature engineering [2].

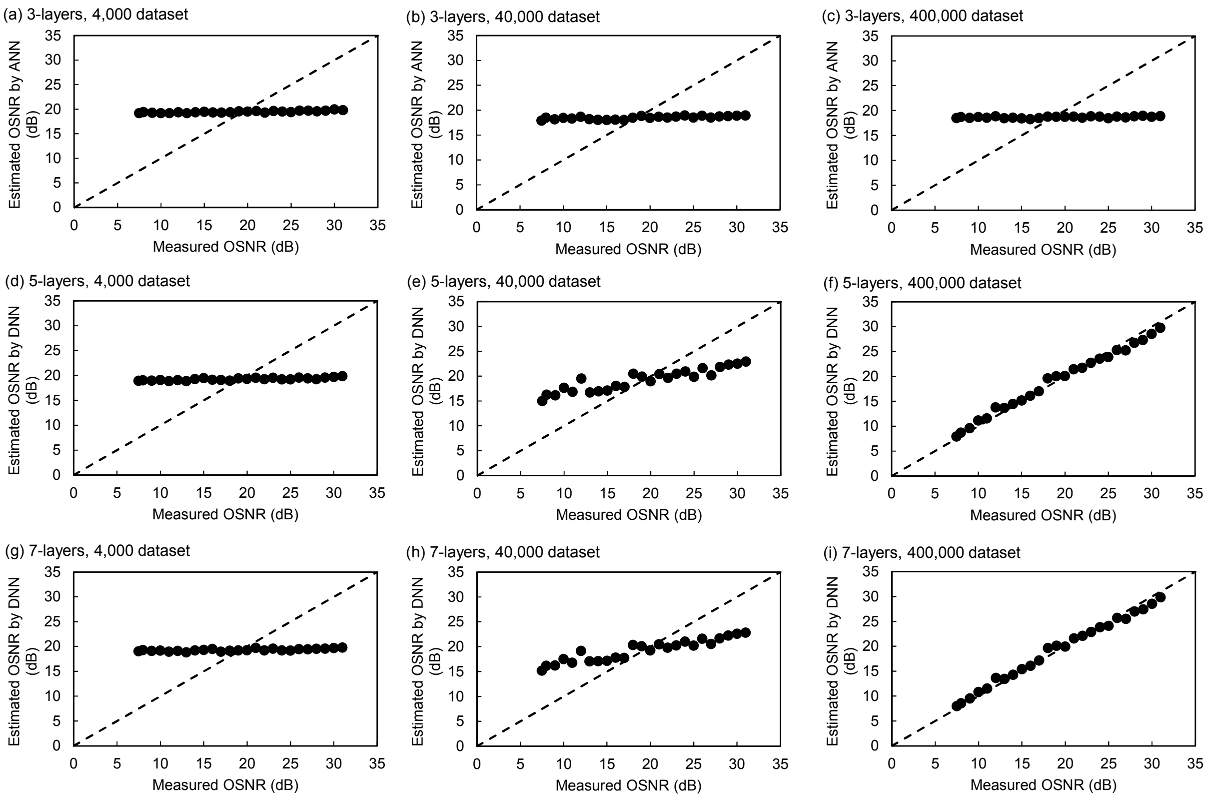


Fig. 3. Test results of OSNR estimation by DNNs.

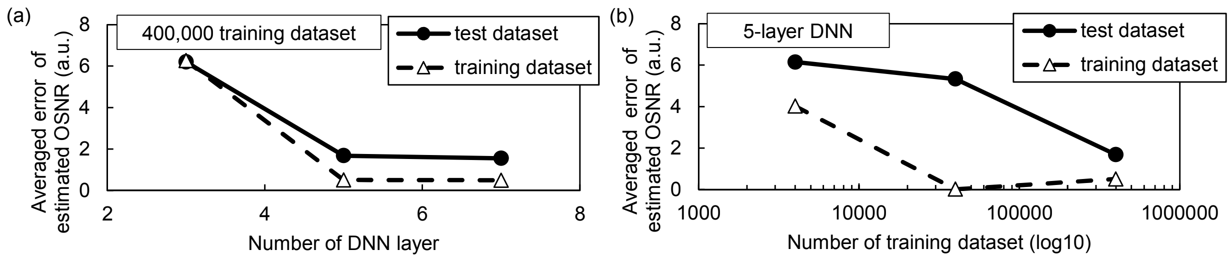


Fig. 4. Averaged error of estimated OSNR as a function of (a) number of DNN layer, and (b) number of training dataset.

V. Conclusions

We experimentally demonstrated a use of deep neural networks (DNN) trained with raw data asynchronously sampled by coherent receiver in a 16 GBd DP-QPSK system. By using 5 layers DNN trained with 400,000 dataset, the DNN obtained a function of OSNR estimation without feature engineering depending on domain expertise. The DNN-based OSNR monitors were worked well against ten different polarization states of polarization division multiplexed input signals.

References

[1] T. Tanimura, L. Dou, X. Su, T. Hoshida, Y. Aoki, Z. Tao, J. C. Rasmussen, M. Suzuki, and H. Morikawa, “Latency and bandwidth programmable transceivers with power arbitration among multi-tenanted signals,” OFC 2016, W4A.6, 2016.

[2] X. Wu, J. A. Jargon, R. A. Skoog, L. Paraschis, and A. E. Willner, “Applications of artificial neural networks in optical performance monitoring, “ IEEE J. Lightw. Technol. **27**, no. 16, 3580-3589, 2009.

[3] F. N. Khan, T. S. R. Shen, Y. Zhou, A. P. T. Lau, and C. Lu, “Optical performance monitoring using artificial neural networks trained with empirical moments of asynchronously sampled signal amplitudes,” IEEE Photon. Technol. Lett. **24**, no. 12, 982-984, 2012.

[4] J. A. Jargon, X. Wu, H. Y. Choi, Y. C. Chung, and A. E. Willner, “Optical performance monitoring of QPSK data channels by use of neural networks trained with parameters derived from asynchronous constellation diagrams,” OSA Optics Express **18**, no. 5, 4931-4938, 2010.

[5] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” nature **521**, 436-444, 2015.

[6] X. Glorot, A. Bordes, and Y. Bengui, “Deep sparse rectifier neural networks,” in Proc. of 14th International Conference on Artificial Intelligence and Statistics, 315-323, 2011.

[7] http://deeplearning.net/software/pylearn2/