

# An Effective Artificial Neural Network Equalizer with S-shape Activation Function for High-speed 16-QAM Transmissions using Low-cost Directly Modulated Laser

Siming Liu<sup>1,2</sup>, Peng-Chun Peng<sup>2</sup>, Chin-Wei Hsu<sup>2</sup>, Shuo Chen<sup>1</sup>, Huiping Tian<sup>1</sup> and Gee-Kung Chang<sup>2</sup>

<sup>1</sup>School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

<sup>2</sup>School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA  
hptian@bupt.edu.cn

**Abstract:** We firstly conceive an artificial neural network (ANN) equalizer with a special S-shape activation function to decimate linear/nonlinear impairments. By means of this ANN equalizer, we achieve a large improvement of ~4.67 dB in  $Q$ -factor.

**Keywords:** Fiber optics communications; Neural networks; Directly modulated laser.

## 1 Introduction

The increasing demand of broad bandwidth requires short-reach optical fiber systems in data centers and inter-connected applications to operate at a very high data rate. Such high-speed transmission systems can be implemented by quadrature-amplitude-modulation (QAM), dual polarization (DP), coherent detection and digital signal processing (DSP) [1]. However, the high cost and complexity of coherent detection technique is prohibitive in short reach applications. Intensity modulation and directly detection (IM-DD) are the widely used practical solutions for their easily implementations with a directly modulated laser (DML) or an external modulator with a single photodiode. Subcarrier multiplexed (SCM) QPSK or quadrature amplitude modulation (QAM) signaling is a good candidate for spectrally efficient direct detection in wavelength division multiplexing (WDM) systems.

DML stands out by its low cost, compact size, lower power dissipation and high output optical power in comparison to the externally modulated laser transmitter [2]. However, DMLs suffer from the chirp- and driver amplifier-induced nonlinearities, which limit their transmission distance and channel capacity [3]. When we transmit at the wavelength of 1550 nm, the DML-based systems are more seriously limited by the interplay between the large frequency chirp and fiber chromatic dispersion [4]. An adaptive feedforward nonlinear equalizer based on polynomial structure is employed to decimate the nonlinearity in the conversion of drive current into optical intensity [5]. Artificial neural network nonlinear equalizers (ANN-NLE) demonstrate their superiority in computation capability and universality. In a previous report, an ANN-NLE

with multi-saturated output regions was proposed in order to deal with QAM signals of any constellation size, but their results are limited within simulations [6]. To our knowledge, the ANN-NLE applications in DML-based transmission systems have not been reported before.

In this paper, we propose a complex-valued ANN-NLE with an S-shape activation function for  $M$ -QAM signal equalizations in a high-speed, low-cost DML-based transmission system for the first time. This S-shape activation function is capable of accommodating any level of QAM signals without modifying the characteristics of the function. With the help of the proposed ANN-NLE, the DML-based system has been demonstrated successfully to transmit 80-Gbps 16QAM Nyquist-SCM signals at a wavelength of 1550 nm through a 15-km dispersion shifted fiber (DSF) under pre-forward error correction (FEC) limit of bit error rate (BER)  $3.8 \times 10^{-3}$ .

## 2 Principle of the ANN-NLE with S-shape activation function

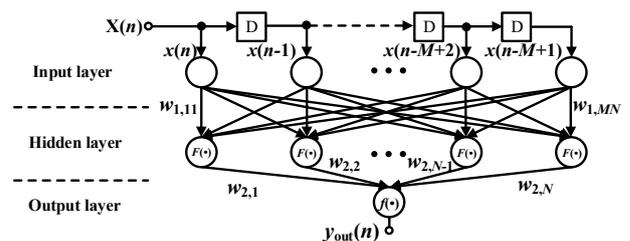


Figure 1 Structure of ANN-NLE with one hidden layer.

The ANN with multiple layers owns a powerful ability to approximate a nonlinear input-output mapping of a general nature. Fig. 1 shows the structure of the proposed ANN-NLE with one hidden layer. The inputs are the tap delay of the received signals  $X(n) = [x(n) \ x(n-1) \ \dots \ x(n-M+1)]^T$ . The weight values between different layers are represented by  $w_{1,ij}(n)$  and  $w_{2,i}(n)$ . The neurons are inter-connected with each other by different weight values.

Every neuron occupies an activation function to process the data sent by the previous layer. For  $M$ -QAM

equalizations, we define the complex-valued function as [6]:

$$f(x) = x + \alpha \sin(\pi x) \quad (1)$$

$$F(z) = f(\text{real}(z)) + jf(\text{imag}(z)) \quad (2)$$

where  $x$  is a real value and  $z$  is a complex value at the input of the defined complex-valued activation function. This function has a generality of  $M$ -QAM signals and  $M$  can be any size. Fig. 2(a) gives the diagram of function  $f(x)$  in (1) with different  $\alpha$  values which define the nonlinearity in the activation function. Fig. 2(b) illustrates the derivatives of  $f(x)$ . The function  $f(x)$  is differentiable at all values. For all  $x$ ,  $f'(x)$  must be bigger than zero, so the weight updating is guaranteed. That is,  $0 < \alpha < 1/\pi$ . The nonlinear factor  $\alpha$  gives the ANN-NLE a nonlinear computation capability. We can define the regions approximate values -3, -1, +1 and +3 as saturation regions in  $f(x)$ , in which the values of  $f'(x)$  are small. In the standard back-propagation (BP) algorithm, when the input value approximates the saturation regions, the weight updating at each iteration is minimized due to its small gradient. This saturation characteristic makes the input value at saturation regions have less effect in updating the weights in a neural network and maximizes the influence of the input far from these regions. Also, due to the multi-saturation characteristic of the  $S$ -shape function, the points in the constellation can be compressed to their corresponding locations and robust to the input random noise. Fig. 3 shows the absolute value of the complex-valued function  $F(z)$  which is defined by (2). The  $z$ -axis corresponds to the power of the points in constellations. Saturation characteristics of  $F(z)$  is demonstrated in Fig. 3 as well.

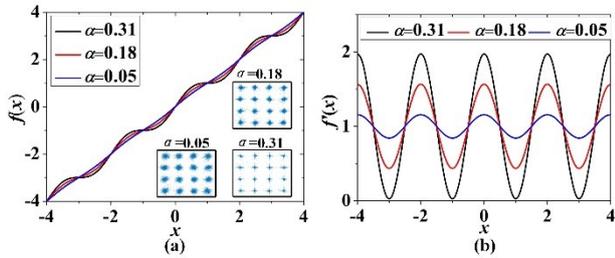


Figure 2 (a) Activation function and (b) derivative of the activation function with different  $\alpha$  values.

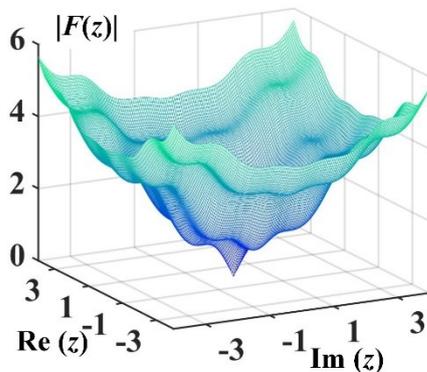


Figure 3 Absolute value of the complex function  $F(z)$ .

For the complex-valued neural network training, the BP

algorithm is extended to the complex-valued BP algorithm [7]. The BP algorithm consists of two phases. The first phase is feed forward. This phase calculates the output vectors from the input vectors. It can be expressed as follows:

$$\begin{cases} S_1(n) = W_1(n)^T X(n) \\ Y_1(n) = F(S_1(n)) \\ S_2(n) = W_2(n)^T Y_1(n) \\ y_{out}(n) = F(S_2(n)) \end{cases} \quad (3)$$

where

$$X(n) = [x(n) \ x(n-1) \ \dots \ x(n-M+1)]^T \quad (4)$$

$$W_1(n) = \begin{bmatrix} w_{1,11}(n) & w_{1,12}(n) & \dots & w_{1,1M}(n) \\ w_{1,21}(n) & w_{1,22}(n) & \dots & w_{1,2M}(n) \\ \dots & \dots & \dots & \dots \\ w_{1,N1}(n) & w_{1,N2}(n) & \dots & w_{1,NM}(n) \end{bmatrix}^T \quad (5)$$

$$S_1(n) = [s_{1,1}(n) \ s_{1,2}(n) \ \dots \ s_{1,N}(n)]^T \quad (6)$$

$$Y_1(n) = [y_1(n) \ y_2(n) \ \dots \ y_N(n)]^T \quad (7)$$

$$W_2(n) = [w_{2,1}(n) \ w_{2,2}(n) \ \dots \ w_{2,N}(n)]^T \quad (8)$$

$F(\cdot)$  is the activation function given by (1) and (2).  $S_2(n)$  is the input of neuron in the last layer and  $y_{out}(n)$  is the output of the neural network. The symbols in (3)-(8) are defined Fig. 1. The error value is defined as,

$$e(n) = y_{out}(n) - d(n) \quad (9)$$

where  $d(n)$  is the desired signal. The second phase is back propagation. In this phase, the calculated error is back propagated to the earlier layers from the output layer. The weight values  $W_1(n)$  and  $W_2(n)$  are updated as,

$$w_{2,i}(n+1) = w_{2,i}(n) - \mu \nabla_{w_{2,i}}(E(n)) \quad (10)$$

$$w_{1,ij}(n+1) = w_{1,ij}(n) - \mu \nabla_{w_{1,ij}}(E(n)) \quad (11)$$

The objective function definition mean square error (MSE) is a little different:

$$E(n) = |e(n)e(n)^*| = e_R(n)^2 + e_I(n)^2 \quad (12)$$

where  $e_R(n)$  and  $e_I(n)$  are the real and imaginary parts of  $e(n)$ , respectively. To minimize the  $E(n)$ , according to the steepest descent technique, derivative of  $E(n)$  in the complex field is required. For the weights between the output and hidden layers, the gradient with complex values is defined as,

$$\nabla_{w_{2,i}}(E(n)) = \frac{\partial E(n)}{\partial w_{R2,i}(n)} + j \frac{\partial E(n)}{\partial w_{I2,i}(n)} \quad (13)$$

$w_{2,i}(n)$  represents the  $i^{\text{th}}$  weight value between the hidden layer and output layer, which is also defined in Fig. 1. Then according the definitions in (13), we can obtain the gradient values:

$$\frac{\partial E(n)}{\partial w_{R2,i}(n)} = 2e_R(n)f'(s_{R2}(n))y_{Ri}(n) + 2e_I(n)f'(s_{I2}(n))y_{Ii}(n) \quad (14)$$

$$\frac{\partial E(n)}{\partial w_{I2,i}(n)} = -2e_R(n)f'(s_{R2}(n))y_{Ii}(n) + 2e_I(n)f'(s_{I2}(n))y_{Ri}(n) \quad (15)$$

For simplification, define

$$\Delta_2(n) = e_R(n)f'(s_{R2}(n)) + je_I(n)f'(s_{I2}(n)) \quad (16)$$

Then we can obtain

$$\nabla_{w_{2,i}}(E(n)) = 2\Delta_2(n)y_i^*(n) \quad (17)$$

For the weights between the input and hidden layers,

$$\nabla_{w_{1,ij}}(E(n)) = \frac{\partial E(n)}{\partial w_{R1,ij}(n)} + j \frac{\partial E(n)}{\partial w_{I1,ij}(n)} \quad (18)$$

For the real part of (18)

$$\frac{\partial E(n)}{\partial w_{R1,ij}(n)} = \frac{\partial E(n)}{\partial s_{R2}(n)} \frac{\partial s_{R2}(n)}{\partial w_{R1,ij}(n)} + \frac{\partial E(n)}{\partial s_{I2}(n)} \frac{\partial s_{I2}(n)}{\partial w_{R1,ij}(n)} \quad (19)$$

$$\frac{\partial s_{R2}(n)}{\partial w_{R1,ij}(n)} = w_{R2,if}'(s_{R1,i}(n))x_R(n-j+1) - w_{I2,if}'(s_{I1,i}(n))x_I(n-j+1) \quad (20)$$

$$\frac{\partial s_{I2}(n)}{\partial w_{R1,ij}(n)} = w_{I2,if}'(s_{R1,i}(n))x_R(n-j+1) + w_{R2,if}'(s_{I1,i}(n))x_I(n-j+1) \quad (21)$$

For the imaginary part of the (18)

$$\frac{\partial E(n)}{\partial w_{I1,ij}(n)} = \frac{\partial E(n)}{\partial s_{R2}(n)} \frac{\partial s_{R2}(n)}{\partial w_{I1,ij}(n)} + \frac{\partial E(n)}{\partial s_{I2}(n)} \frac{\partial s_{I2}(n)}{\partial w_{I1,ij}(n)} \quad (22)$$

$$\frac{\partial s_{R2}(n)}{\partial w_{I1,ij}(n)} = -w_{R2,if}'(s_{R1,i}(n))x_I(n-j+1) - w_{I2,if}'(s_{I1,i}(n))x_R(n-j+1) \quad (23)$$

$$\frac{\partial s_{I2}(n)}{\partial w_{I1,ij}(n)} = -w_{I2,if}'(s_{R1,i}(n))x_I(n-j+1) + w_{R2,if}'(s_{I1,i}(n))x_R(n-j+1) \quad (24)$$

From (18)-(24), we can obtain the value of  $\nabla_{w_{1,ij}}(E(n))$ . However, the calculation process is very complicated and confusing. The expressions require some simplifications.

According to the definition of back propagation, the value of  $e(n)$  obtained from the output of the network is transferred to the previous layers of the network through the weights  $W_2(n)$ . Therefore, we define the error value transported to the hidden layer as

$$e_{h,i}(n) = w_{2,i}^*(n)\Delta_2(n) \quad (25)$$

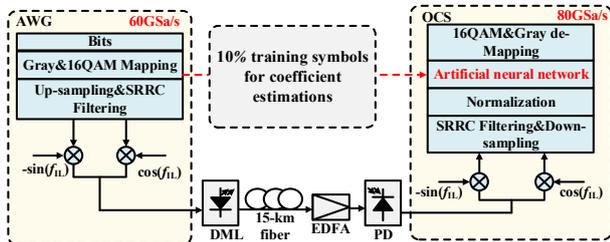
Similar to (16), we can define

$$\Delta_{1,i}(n) = e_{hR}(n)f'(s_{R1,i}(n)) + je_{hI}(n)f'(s_{I1,i}(n)) \quad (26)$$

Finally, compared with (18)-(24), we can obtain

$$\nabla_{w_{1,ij}}(E(n)) = 2\Delta_{1,i}(n)x^*(n-j+1) \quad (27)$$

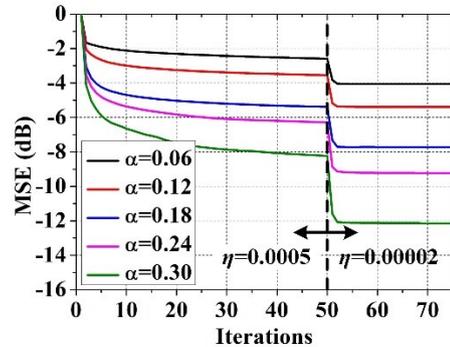
### 3 Experimental demonstration and results



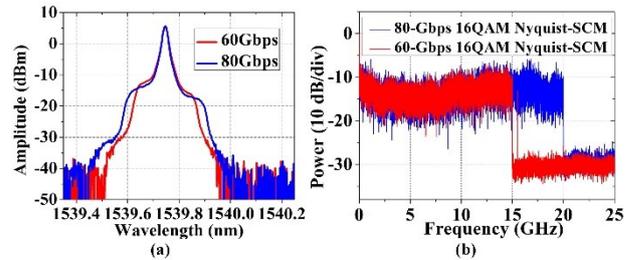
**Figure 4** Experimental setup for the DML-based 80-Gbps high-speed transmission system using the proposed ANN-NLE. AWG: arbitrary waveform generator, SRRC: square root raised cosine, DML: directly modulated laser, EDFA: Erbium doped fiber amplifier, PD: photo detector, OCS: oscilloscope.

Fig. 4 shows the experimental setup of the DML-based 80-Gbps 16QAM Nyquist-SCM transmission system based on the proposed ANN-NLE. Under the help of a square-root-raised-cosine (SRRC) filter with a roll-off

factor of zero, we realize Nyquist pulse shaping and halve the bandwidth, allowing us to implement half-cycle SCM. The output of an arbitrary waveform generator (AWG) running at 60 Gbps is fed into a 1550-nm DML. After 15-km DSF transmissions, the signal is detected by a photo detector (PD) and sampled by an analog-to-digital converter in an oscilloscope (OCS) running at 80 GSa/s. In the experiment, we try to transmit 60- and 80-Gbps 16QAM Nyquist-SCM signals. The ANN-NLE is trained for 75 epochs, of which the first 50 epochs with step size  $\mu=0.0005$  and the rest of the epochs with  $\mu=0.00002$ . Fig. 5 shows the convergence process of the ANN-NLE as a function of MSE with different  $\alpha$  values. It can be obtained that larger  $\alpha$  leads to a lower MSE floor, representing a stronger nonlinear computation capability. Therefore, in the rest of the experiment,  $\alpha$  is set to be 0.3 to provide enough nonlinearity. Fig. 6(a) and (b) give the optical and electrical spectrum of the 60-Gbps and 80-Gbps 16QAM Nyquist-SCM.



**Figure 5** Convergence of the ANN with different  $\alpha$  values measured as MSE.



**Figure 6** (a) Optical spectrum and (b) electrical spectrum of in the experiment.

Fig. 6(a) gives the optical spectrum at the output of the DML with 60/80-Gbps 16QAM SCM signals. Fig. 6(b) shows the electrical spectrum of the received signals and the fluctuated response indicates the serious inter-symbol interferences (ISI) in the system.

Fig. 7 illustrates that with increase of the driving voltage of DML, the asymmetry of the optical spectrum becomes obvious, indicating the chirp effect happening in the DML. The DML is a commercial device with a threshold current around 12 mA, as shown in Fig. 8(a). The saturated bias current of the DML is about 150 mA. Fig. 8(b) illustrates the BER performance of the 60-Gbps 16QAM signals as a function of bias current. The DML has a 3-dB bandwidth of 20 GHz when the

bias current is above 120 mA. Given in Fig. 8(b), when the bias current is larger than 120 mA, the system reaches its best performance.

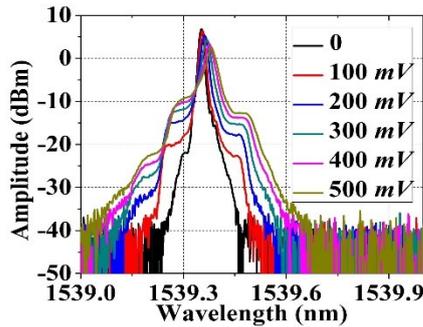


Figure 7 Optical output power of the 20-GHz DML at different bias currents.

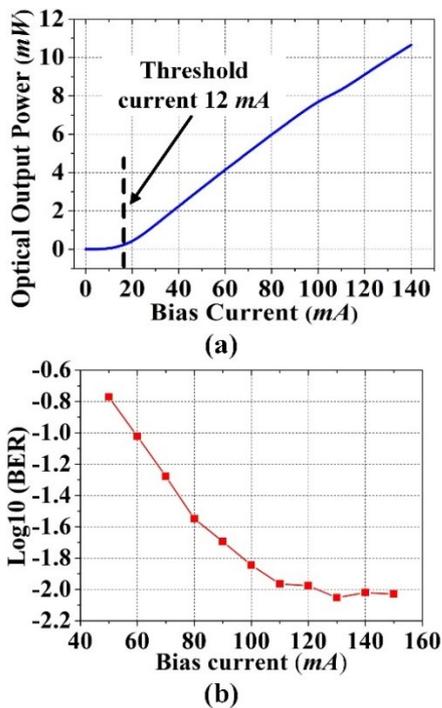


Figure 8 (a) Optical output power of the 20-GHz DML at different bias currents. (b) BER performance of the system with changing bias current.

Fig. 9(a) investigates the relationship of the training data set sizes and BER performances of the ANN-NLE. With increasing training data set size, better BER performance can be obtained due to the sufficient training. Insufficient training data set causes overfitting problem of the neural network. A good characteristic of a machine learning model is its ability to generalize accurately from the training data to any future data previously unseen by the model. In the case of overfitting, the ANN-NLE may work well with training data but may exhibit poor performances with previously unseen data. With increasing training data set size, the overfitting problem can be avoided.

The input tap number highly depends on the bandwidth of the signals and frequency response of the channels. Increasing the tap number can help to mitigate the ISI,

as shown in Fig. 9(b). With the tap number larger than 10, we have optimal BER performances with vibration smaller than 10%.

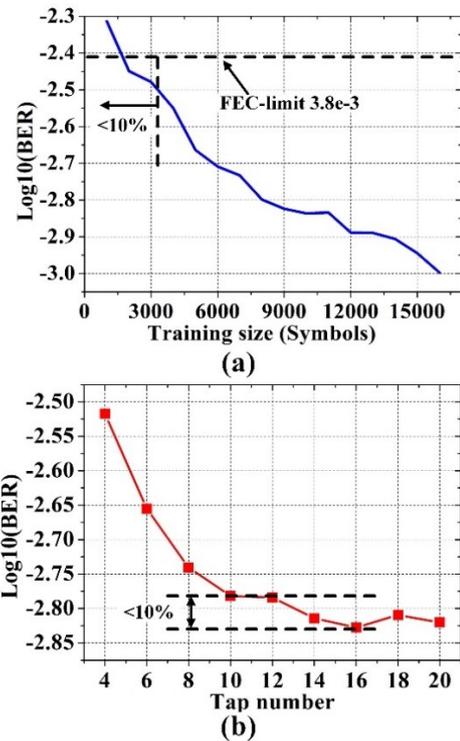


Figure 9 BER performances with varying (a) training data size and (b) input tap numbers.

Fig. 10 shows the  $Q$ -factor performances of the 60-Gbps signals with/without the help of ANN-NLE as functions of driving voltage of the DML. In this experiment, there are 20, 20 and 1 neurons in the input, hidden and output layers of the network. The training set is a sequence of symbols with length 3200, 10% of the whole sequences. The  $Q$ -factor is calculated from the BER value by  $Q = 20 \log_{10}[\sqrt{2} \operatorname{erfc}^{-1}(2BER)]$ . With small driving voltage, the signals suffer from low signal-to-noise ratio (SNR). The continual increasing driving voltage leads to the nonlinearities caused by drivers and chirp in the DML. We can get a  $\sim 4.67$  dB  $Q$ -factor enhancement at 180 mV with ANN equalizations.

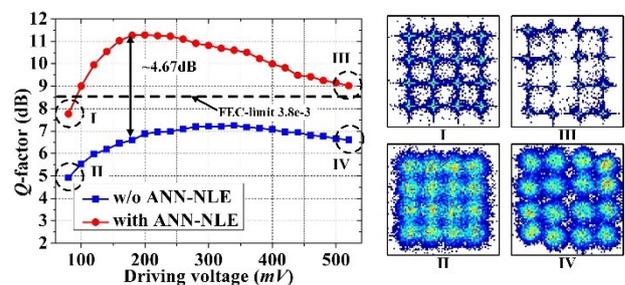
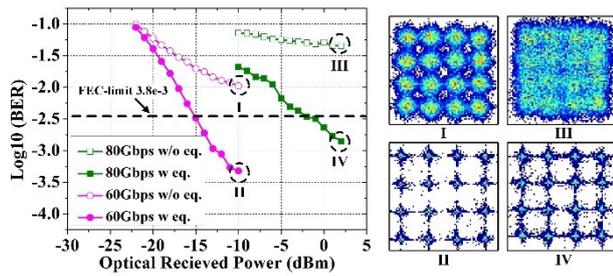


Figure 10  $Q$ -factor of the system as functions of the driving voltages with and without the ANN-NLE.

Fig. 11 shows the system sensitivity using 60/80-Gbps 16QAM Nyquist-SCM. After equalizations, power sensitivity  $-15/-2$  dBm at the FEC limit of  $BER 3.8 \times 10^{-3}$  can be observed in the cases of 60/80-Gbps,

respectively.



**Figure 11** BER performances of 16QAM signals as functions of the received power.

## 4 Conclusions

We have proposed a novel ANN-NLE with a complex-valued *S*-shape activation function for delivering high-speed Nyquist-SCM 16QAM signals in a DML-based transmission system. The multi-saturated characteristics of the activation function enables the ANN-NLE to possess a nonlinear computation capability and makes it robust to the input noise. With the help of the proposed ANN-NLE, we have achieved a  $\sim 4.67$  dB *Q*-factor improvement at the driving voltage 180 mV in the 16QAM transmission experiments. We also observe power sensitivity of the 60/80-Gbps signals at BER  $3.8 \times 10^{-3}$  to be -15 and -2 dBm, respectively.

## Acknowledgements

This research was supported by the fund of NSFC under Grant No. 61372038 and 61431003 and the project of Joint

Laboratory for Undersea Optical networks, China.

## References

- [1] Abdullah S. Karar and John C. Cartledge. Generation and detection of a 56 Gb/s signal using a DML and half-cycle 16-QAM Nyquist-SCM. *IEEE PHOTONICS TECHNOLOGY LETTERS*, 2013, 25(8):757-760.
- [2] Abdullah S. Karar and John C. Cartledge. Electronic post-compensation of dispersion for DML systems using SCM and direct detection. *IEEE PHOTONICS TECHNOLOGY LETTERS*, 2013, 25(9):825-828.
- [3] Nobuhiko Kikuchi, Riu Hirai and Takayoshi Fukui. Nonlinear compensation of high-speed PAM4 signals from directly-modulated laser at high extinction ratio. *European Conference on Optical Communication*. 2016:1-3.
- [4] S. H. Bae, Hoon Kim and Y. C. Chung. Transmission of 51.56-Gb/s OOK signal using 1.55- $\mu$ m directly modulated laser and duobinary electrical equalizer. *OPTICS EXPRESS*, 2016, 24(20):22555-22562.
- [5] Yasuhiro Matsui, Thang Pham, William A. Ling, Richard Schatz, Glen Carey, Henry Daghighian, Tsurugi Sudo and Charles Roxlo. 55-GHz bandwidth short-cavity distributed reflector laser and its application to 112-Gb/s PAM-4. *The Optical Networking and Communication Conference & Exhibition*. 2016:Th5B.4.
- [6] Cheolwoo You and Daesik Hong. Nonlinear blind equalization schemes using complex-valued multilayer feedforward neural networks. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, 1998, 9(6): 1442-1455.
- [7] N. Benvenuto and F. Piazza. On the complex backpropagation algorithm. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, 1992, 40(4): 967-96