SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •
Special Focus on Photonics in AI

June 2020, Vol. 63 160406:1–160406:9
https://doi.org/10.1007/s11432-019-2848-3

# Deep belief network-hidden Markov model based nonlinear equalizer for VCSEL based optical interconnect

Fukui TIAN & Chuanchuan YANG*

*State Key Laboratory of Advanced Optical Communication Systems and Networks, School of Electronics Engineering and Computer Science, Peking University, Beijing* 100871*, China*

Received 15 December 2019/Revised 1 March 2020/Accepted 18 March 2020/Published online 11 May 2020

**Abstract** The data center has developed rapidly over the past few years, leading to the demand for high speed data transmission. Vertical cavity surface emitting lasers (VCSELs) based optical interconnect is evolving to 100 Gb/s and beyond, which makes nonlinear distortions difficult to be compensated or equalized by conventional equalizers. Moreover, the challenge becomes very complicated for conventional equalizers because of the presence of inter-symbol interference (ISI) together with the nonlinear distortions. So many neural network based DSP algorithms such as artificial neural network (ANN) have been proposed to mitigate the distortions. However, ANN has the limitations that sample's relevant information is not considered, leading to degradation in ANN's performance and high computational complexity. In this paper, in order to maintain an excellent capability of mitigating nonlinear distortions like other neural network based equalizers while considering the sample's relevant information to reduce the computational complexity, we propose a deep belief network-hidden Markov model (DBN-HMM) based nonlinear equalizer which is tested in a PAM-4 modulated VCSEL and multimode fiber (MMF) optical interconnect link experimentally. The BER performance can be greatly improved compared with conventional DSP algorithms. In addition, the computational complexity of DBN-HMM based equalizer can be about 41% lower than that of ANN based method with a similar BER performance.

**Keywords** short-range optical interconnect, vertical cavity surface emitting laser (VCSEL), nonlinear equalizer, deep belief network (DBN), hidden Markov model (HMM)

**Citation** Tian F K, Yang C C. Deep belief network-hidden Markov model based nonlinear equalizer for VCSEL based optical interconnect. Sci China Inf Sci, 2020, 63(6): 160406, https://doi.org/10.1007/s11432-019-2848-3

## 1 Introduction

With the exponential increase in the employment of high definition video and cloud computing, the explosion of data transmission and service demands poses a huge challenge to the data centers, which requires high-speed data transmission [1]. Nowadays, short-distance optical interconnection links based on vertical cavity surface emitting laster (VCSEL) and multimode fiber (MMF) are commonly deployed in data centers, because VCSEL has the advantages of low energy consumption, low manufacturing cost and high stability, which makes VCSEL become the most cost-effective solution in short-range optical interconnection [2,3].

Digital signal processing (DSP) technologies are quite essential to enhance the capacity of the optical communication system, most of which mainly focus on pre-filtering as well as post-equalization and have been proved to be very effective in eliminating inter-symbol interference (ISI) [4,5]. However, with

* Corresponding author (email: yangchuanchuan@pku.edu.cn)

 

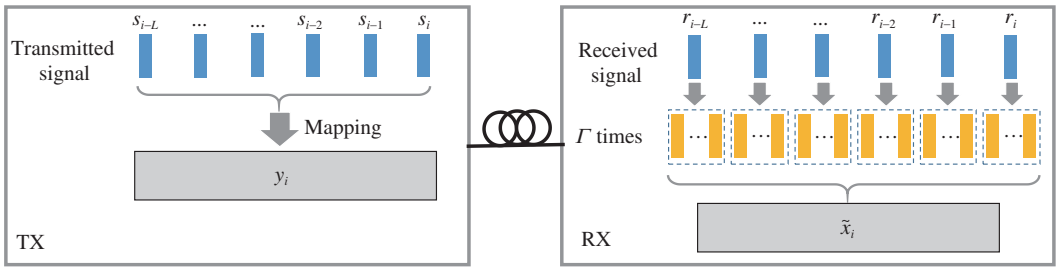


**Figure 1** (Color online) The pre-processing of signals, the current transmitted signal and its $L$ adjacent signals are mapped into a new symbol. For the received signal, the current vector after sinc interpolation and its $L$ adjacent vectors are wrapped as the final feature vector.

the increase of transmission speed, the challenge becomes very complicated for the VCSEL-MMF link owing to the presence of ISI together with nonlinear noise. The conventional DSP algorithms, such as feed forward equalization (FFE), decision feedback equalizer (DFE) and maximum likelihood sequence estimation (MLSE) [6–9], cannot effectively compensate or equalize nonlinear noise like relative intensity noise (RIN) and modal partition noise (MPN) that exist in practical system [10].

To mitigate the distortions of nonlinearity, neural network based DSP algorithms have been proposed recently. In VCSEL-MMF based optical interconnect links, neural network based DSP algorithms, such as artificial neural networks (ANN), have been proved to achieve better performance than conventional DSP algorithms, which shows an excellent capability of mitigating nonlinear distortions [11,12]. However, ANN takes the assumption that different input samples are independent of each other as the premise, leading that the relevant information of previous sample is no longer considered when processing current sample [13]. In optical communication system, owing to the existence of inter-symbol interference, the samples are related in time domain, leading to degradation in the ANN's performance. Deep belief network-hidden Markov model (DBN-HMM) can process sequences consisting of dependent elements, which makes DBN-HMM solve many complicated signal processing problems, such as speech recognition [14]. In this paper, DBN-HMM is proposed for equalization in short-range optical interconnect. DBN-HMM based DSP algorithms not only retains an excellent capability of mitigating nonlinearity noise, but also considers the relevant information of previous sample. In addition, the number of input symbols in DBN can be greatly reduced with the aid of HMM, resulting in lower computational complexity.

Our experimental results show that the BER performance can be greatly improved by introducing DBN-HMM based equalizer compared with conventional maximum likelihood sequence estimation (MLSE). What's more, complexity analysis results show that the computational complexity of DBN-HMM can be controlled by selecting the parameters of DBN-HMM during the equalizing process. And the computational complexity of DBN based equalizer can be about 41% lower than that of ANN based method with the similar BER performance.

## 2 DBN-HMM based equalization

### 2.1 DBN-HMM

A VCSEL based optical interconnect communication system mainly uses intensity modulation and direct detection (IM/DD). PAM4 has advantages of less sensitivity to time jitter and high frequency noise, which makes it prefer among all the intensity modulation formats. The equalizer equalizes the signal by acquiring the signal's amplitude information. But the sampling rate is limited by the ability of ADC, so we have chosen to apply the feature enhancement method based on sinc interpolation that we proposed in [15] to enhance the features. Concretely, the number of features for each symbol can be increased by means of sinc interpolation. As shown in Figure 1, the received signal sequence after sinc interpolation is denoted as $\boldsymbol{r} = [\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_T]$, where the vector $\boldsymbol{r}_i$ $(i = 1, 2, \ldots, T)$ is corresponding to the $i$th

symbol, and the length of vector $\boldsymbol{r}_i$ is the interpolation multiple, which is denoted as $\Gamma$. Owing to the existence of ISI, the current vector $\boldsymbol{r}_i$ and its $L$ adjacent vectors are wrapped as the final feature vector $\widetilde{\boldsymbol{x}}_i = [\boldsymbol{r}_{i-L}, \ldots, \boldsymbol{r}_{i-1}, \boldsymbol{r}_i]$. Therefore, the observation sequence that follows chronological order can be denoted as $\widetilde{\boldsymbol{x}} = [\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{x}}_2, \ldots, \widetilde{\boldsymbol{x}}_T]$, which will be used in the training process and the equalizing process. And we denote the transmitted signal sequence as $\boldsymbol{s} = [s_1, s_2, \ldots, s_T]$, where $s_i$ $(i = 1, 2, \ldots, T)$ is corresponding to the $j$th state (for PAM-4, $j = 1, 2, 3, 4$, corresponding to four types of symbols). On the other hand, we also map current symbol $s_i$ and its $L$ adjacent symbols to a new symbol $y_i$ $(i = 1, 2, \ldots, T)$ to eliminate the impact of ISI, where $y_i \in \{q_1, q_2, \ldots, q_N\}$, $N = 4^{L+1}$, and the mapping function can be defined as

$$f(s_{i-L}, \ldots, s_i) = y_i. \tag{1}$$

Therefore, the labeled training dataset can be denoted as $\{\widetilde{\boldsymbol{x}}_i, y_i\}, i = 1, 2, \ldots, N$, which will be used for training the model.

An HMM is mainly determined by the state transition probability distribution, the initial state distribution and the observation probability distribution. Therefore, the form of the hidden Markov model can be defined as [16]

$$\lambda = (\boldsymbol{A}, \boldsymbol{B}, \pi), \tag{2}$$

where $\boldsymbol{A}$ is the state transition probability matrix and is defined as

$$\boldsymbol{A} = [a_{mn}]_{N \times N}, \tag{3}$$

where $a_{mn}$ is the state transition probability and

$$a_{mn} = P(y_i = q_n | y_{i-1} = q_m), \quad 1 \leqslant m,\ n \leqslant N, \tag{4}$$

where $y_i$ and $y_{i-1}$ are the states at time $i$ and $i-1$, respectively. $a_{mn} \in [0, 1]$ describes the transition probability from state $q_m$ to state $q_n$, satisfying $\sum_n a_{mn} = 1$. The transition probability from one state to another is considered to be the same, which indicates that the transition probability between any two states is equivalent. Therefore, after mapping current symbol $s_i$ and its $L$ adjacent symbols to a new symbol $y_i$, the transition probability matrix for PAM-4 has the form:

$$a_{mn} = 0.25, \tag{5}$$

when

$$\left[f^{-1}(y_m)\right]_{i-j} = \left[f^{-1}(y_n)\right]_{i-j-1}, \quad j = 0, 1, \ldots, L-1. \tag{6}$$

And $\pi$ is the initial probability distribution:

$$\pi = (\pi_m), \tag{7}$$

where $\pi_m = P(y_1 = q_m), m = 1, 2, \ldots, N$ is the probability of being in state $q_i$ at the beginning time 1. $\boldsymbol{B}$ is the observation probability matrix:

$$\boldsymbol{B} = [b_m(x)]\,. \tag{8}$$

In our proposed algorithm, the observation probability distribution of features is modeled with DBN, and the elements of $\boldsymbol{B}$ can be calculated by

$$b_m(x) = P(x | y_i = q_m) = \Phi_m(x), \tag{9}$$

where $\Phi_m(x)$ represents the probability density of feature vector that is calculated by the $m$th DBN $(m = 1, 2, \ldots, N)$.

DBN can be approximated as the stack of restricted Boltzmann machines (RBMs). As shown in Figure 2(a), an RBM is an undirected graph model with a layer of visible nodes ($\boldsymbol{v}$) and a layer of

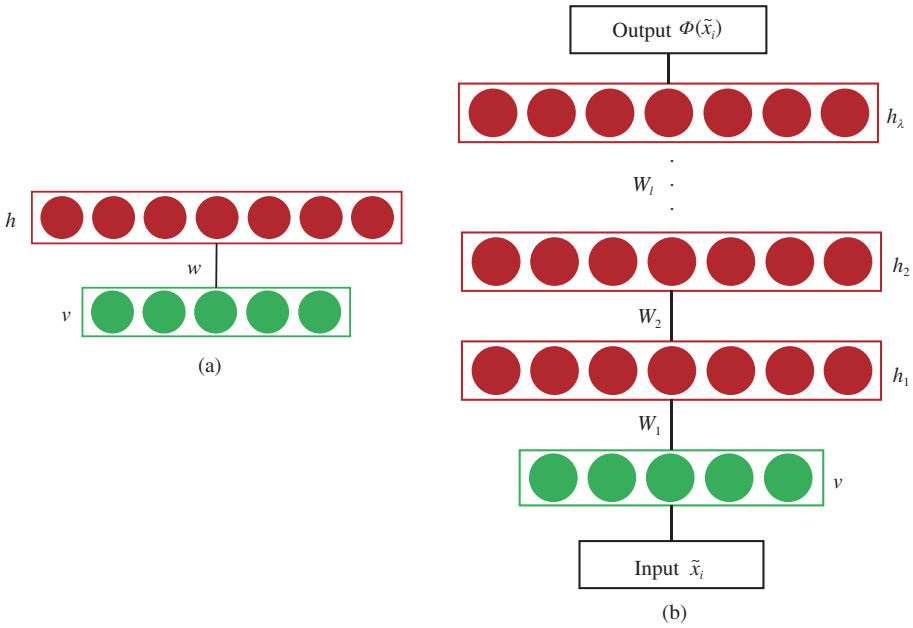


**Figure 2** (Color online) (a) The structure of RBM; (b) the structure of DBN, which can be approximated as the stack of RBMs.

hidden nodes ($\boldsymbol{h}$), which has hidden-visible connections but does not have hidden-hidden or visible-visible connections. $\boldsymbol{w}$ is the connection weights matrix between the visible layer and the hidden layer. And an RBM's joint probability can be stated as [17, 18]

$$P(\boldsymbol{v},\boldsymbol{h}) = \frac{1}{Z}\exp(-E(\boldsymbol{v},\boldsymbol{h})), \tag{10}$$

where $Z$ represents the partition function and $E(\boldsymbol{v},\boldsymbol{h})$ is the energy function with both binary visible and hidden variables, which can be defined as

$$E(\boldsymbol{v},\boldsymbol{h}) = -\sum_{i,j} v_i W_{ij} h_j - \sum_i d_i v_i - \sum_j c_j h_j, \tag{11}$$

where $W_{ij}$ are the connection weights, $d_i$ is the bias on the visible nodes and $c_j$ is also the bias on the hidden nodes, which is different from $d_i$. Because there is no within-layer connections in an RBM, the probabilities $P(\boldsymbol{v}|\boldsymbol{h})$ and $P(\boldsymbol{h}|\boldsymbol{v})$ can be calculated by [19]

$$P(v_i = 1|\boldsymbol{h}) = \sigma\left(\sum_j W_{ij} h_j + d_i\right), \tag{12}$$

$$P(h_i = 1|\boldsymbol{v}) = \sigma\left(\sum_i W_{ij} v_i + c_j\right), \tag{13}$$

where $\sigma(\cdot)$ is the sigmoid function and is defined as

$$\sigma(x) = (1 + \mathrm{e}^{-x})^{-1}. \tag{14}$$

As illustrated in Figure 2(b), the structure of DBN can be seen as the stack of RBMs. Therefore, an DBN with $\lambda$ hidden layers contains $\lambda$ connection weight matrices $\boldsymbol{W}_1, \boldsymbol{W}_2, \ldots, \boldsymbol{W}_\lambda$ and $\lambda + 1$ biases $\boldsymbol{d}_0, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_\lambda$, where $\boldsymbol{d}_0$ is the bias of the visible layer $\boldsymbol{v}$. The corresponding hidden vector $\boldsymbol{h}$ can be calculated as

$$\boldsymbol{h}_1 = \sigma(\boldsymbol{W}_1 \boldsymbol{v} + \boldsymbol{d}_1), \tag{15}$$

$$\boldsymbol{h}_l = \sigma(\boldsymbol{W}_l \boldsymbol{h}_{l-1} + \boldsymbol{d}_l), \quad l = 2, \ldots, \lambda - 1. \tag{16}$$

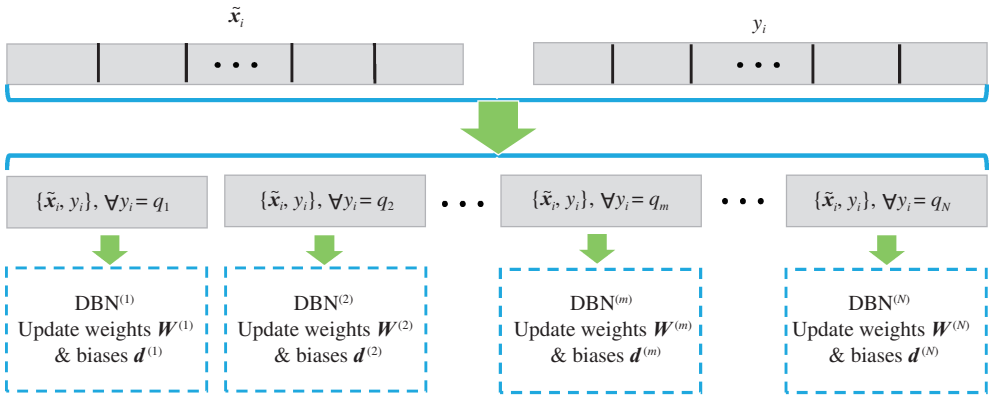


**Figure 3** (Color online) The training process of DBN-HMM based equalizer. The dataset is divided into $N$ sub-datasets, $\mathrm{DBN}^{(m)}$ is trained with the $m$th sub-dataset ($m = 1, 2, \ldots, N$).

Therefore, the output probability can be calculated by the hidden vector:

$$\Phi = \sigma(\boldsymbol{W}_\lambda \boldsymbol{h}_{\lambda-1} + \boldsymbol{d}_\lambda), \tag{17}$$

so according to (9), the elements of the observation matrix $\boldsymbol{B}$ can be derived.

Many non-ideal and nonlinear distortions that exist in practical system cannot be accurately captured by conventional algorithm, leading to the degradation of performance. But the situation becomes different for DBN, like other neural networks, DBN has a strong ability to fit the true distribution of data, which can accurately capture the nonlinear distortions that exist in practical system. So DBN is a better choice to model the received data compared with conventional DSP algorithms.

### 2.2 Training process

The training process is illustrated in Figure 3. In order to train the model, the labeled training dataset $\{\widetilde{\boldsymbol{x}}_i, y_i\}, i = 1, 2, \ldots, T$ is divided into $N$ sub-datasets $\{\widetilde{\boldsymbol{x}}_i, y_i | y_i = q_m\}, m = 1, 2, \ldots, N$ to train $N$ DBNs to calculate the observation probability.

Each DBN can be trained by approximating it to a stack of RBMs, which means that we should train RBMs firstly. At the training process, we need to find parameters in each RBM by maximizing $E_{\boldsymbol{v}\sim P_{\mathrm{data}}} \log P(\boldsymbol{v})$, which is defined as follows [18]:

$$\log(P(\boldsymbol{v})) = \log\left(\sum_{\boldsymbol{h}} P(\boldsymbol{v}, \boldsymbol{h})\right). \tag{18}$$

And the gradient update rule for a parameter $\theta_k$ is

$$\Delta\theta_k \propto \left\langle \frac{\partial E(\boldsymbol{v}, \boldsymbol{h})}{\partial \theta_k} \right\rangle_{\mathrm{data}} - \left\langle \frac{\partial E(\boldsymbol{v}, \boldsymbol{h})}{\partial \theta_k} \right\rangle_{\mathrm{model}}, \tag{19}$$

where $\theta_k$ represents all weights and biases, $\langle\cdot\rangle_{\mathrm{data}}$ represents the expected value under the empirical distribution and $\langle\cdot\rangle_{\mathrm{model}}$ represents the expected value under the model distribution [18]. However, it takes exponential time to compute $\langle\cdot\rangle_{\mathrm{data}}$, so contrastive divergence is used instead to reduce the computational complexity. The new gradient update rule can be stated as

$$\Delta\theta_k \propto \left\langle \frac{\partial E(\boldsymbol{v}, \boldsymbol{h})}{\partial \theta_k} \right\rangle_{\mathrm{data}} - \left\langle \frac{\partial E(\boldsymbol{v}, \boldsymbol{h})}{\partial \theta_k} \right\rangle_{1}, \tag{20}$$

where $\langle\cdot\rangle_1$ stands for the expectation under the distribution of samples that is from running the Gibbs sampler initialized at the data for one full step [19]. And the learning rate is fixed as $\alpha = 0.0003$ in this paper.

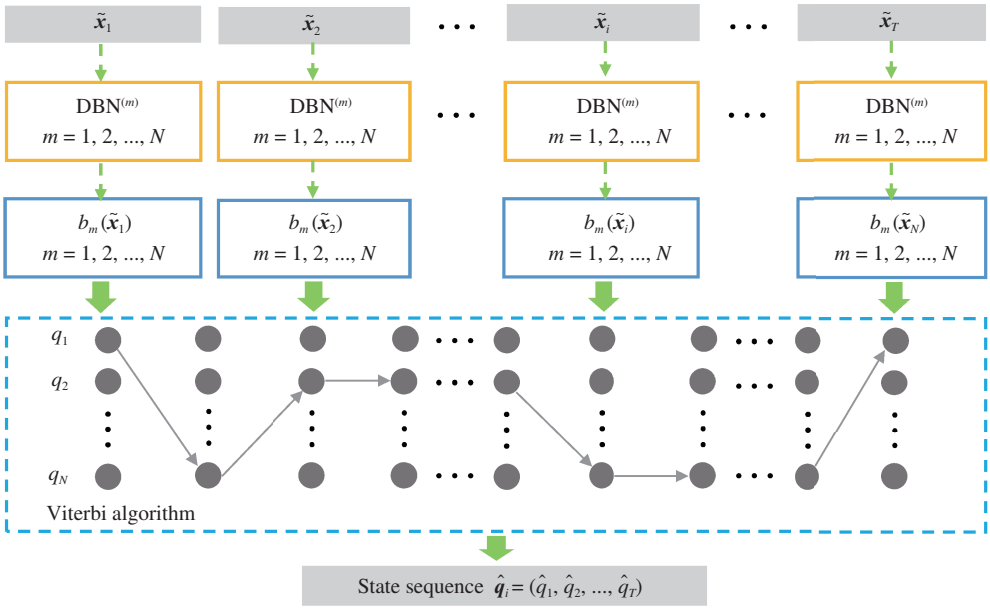


**Figure 4** (Color online) The equalizing process of DBN-HMM based equalizer.

Each DBN can be trained by using hierarchical training, which means that the trained RBM's output is used as the input of RBM in the upper layer. After an RBM that makes up the DBN components has been trained, the parameters in the trained RBM is used as those of the first layer in DBN. Then the RBM in the second layer of DBN can be trained by maximizing [20]:

$$E_{\boldsymbol{v}\sim p_{\text{data}}} E_{\boldsymbol{h}_1\sim p_1(\boldsymbol{h}_1|\boldsymbol{v})} \log P_2(\boldsymbol{h}_1), \tag{21}$$

where $p_1$ represents the probability distribution of the first RBM and $p_2$ represents the probability distribution of the second RBM. The gradient update rule is the same as (20), which can get the second trained RBM. This process is repeated, adding as many layers as you want to DBN. Each DBN can be considered as a whole neural network to calculate the probability density of received symbols $b_m(\widetilde{\boldsymbol{x}}_i), i = 1, 2, \ldots, T$.

### 2.3 Equalizing process

The equalizing process is illustrated in Figure 4. After training, the parameters in the neural networks are fixed and remain unchanged during equalizing process. During the equalizing process, the observation probability $b_m(\widetilde{\boldsymbol{x}}_i)$ $(m = 1, 2, \ldots, N, i = 1, 2, \ldots, T)$ is calculated by $N$ DBNs, and the transition matrix can be calculated according to (5). Then the probability-based Viterbi algorithm is conducted for decision in HMM.

The Viterbi algorithm is stated as follows [21].

(1) Initialization:

$$\delta_1(m) = \pi_m b_m(\widetilde{\boldsymbol{x}}_1), \ m = 1, 2, \ldots, N, \tag{22}$$

$$\psi_1(m) = 0, \quad m = 1, 2, \ldots, N. \tag{23}$$

(2) Recursion:

$$\delta_i(m) = \max_{1\leqslant n\leqslant N} [\delta_{i-1}(n) a_{mn}]\, b_m(\widetilde{\boldsymbol{x}}_i), \quad m = 1, 2, \ldots, N, \ i = 2, \ldots, T, \tag{24}$$

$$\psi_i(m) = \arg \max_{1\leqslant n\leqslant N} [\delta_{i-1}(n) a_{mn}], \quad m = 1, 2, \ldots, N. \tag{25}$$

(3) Termination:

$$\hat{P} = \max_{1\leqslant m\leqslant N} \delta_T(m), \tag{26}$$

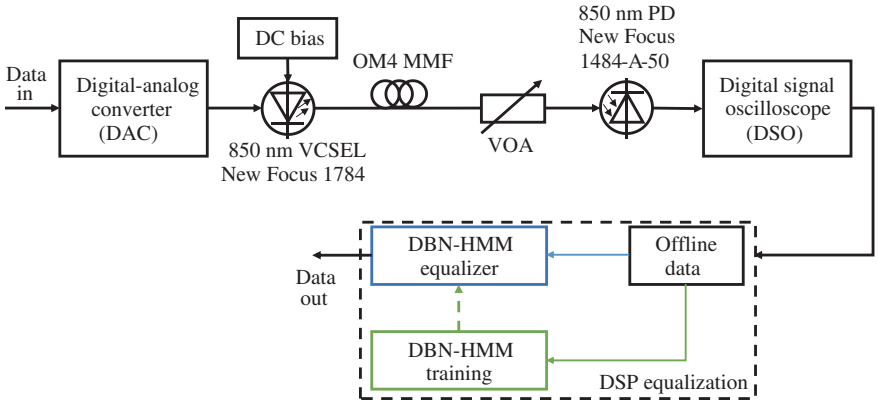


**Figure 5** (Color online) Experiment block diagram of VCSEL based optical interconnect link with DBN-HMM. The input data is generated with bit-pattern generator (BPG) using random pattern (56 Gb/s).

$$\hat{q}_T = \underset{1 \leqslant m \leqslant N}{\arg\max} \left[\delta_T(m)\right]. \tag{27}$$

(4) Path (state sequence) backtracking:

$$\hat{q}_i = \psi_{i+1}(\hat{q}_{i+1}), \quad i = T-1, T-2, \ldots, 1. \tag{28}$$

Finally the optimal path $\hat{\boldsymbol{q}} = (\hat{q}_1, \hat{q}_2, \ldots, \hat{q}_T)$ has been obtained.

For convenience, DBN-HMM based equalizer is denoted as DBN-HMM $(\Gamma, L, R)$, where $\Gamma, L, R$ are the interpolation multiple, adjacent number, and the number of neurons in each DBN's hidden layer.

## 3 Experiment results and discussion

### 3.1 Experiment setup and results

In order to evaluate the performance of DBN-HMM based equalizer, the experiments have been conducted with a 56 Gb/s PAM4-modulated VCSEL-MMF optical link, which is shown in Figure 5. The experimental system mainly consists of a 850 nm VCSEL that is directly modulated by a bit pattern generator (BPG) of SHF 12104A, 100-m OM4 MMF and a photodiode (PD). The PAM-4 modulated optical signal is coupled into MMF by FC/PC connector. At the end of MMF, PD detects the optical signal and transforms the optical signal into baseband electrical signal. Then the high speed real time digital signal oscilloscope (DSO) samples the baseband electrical signal for the offline DSP algorithm. The adjacent number $L$ mainly depends on the inter-symbol interference, in this experiment, the BER performance can get the optimal result when $L = 2$, and a larger $L$ cannot lead to a better BER. Therefore, $L$ is fixed to be 2 in our DSP algorithm. Each DBN is composed of 2 RBMs, which can get a good result, and more RBMs in DBN cannot get better BER performance.

In our experiments, the 850 nm VCSEL is New Focus® 1784 [1)], and PD is New Focus® 1484-A-50 [2)], while the −3 dB bandwidth of VCSEL and PD are 18 GHz and 22 GHz respectively. The OM4 MMF is chosen as YOFC® MaxBand OM4 bend insensitive multimode fiber with an over filled launch (OFL) bandwidth of 4394 MHz×km. At the receiver side, the DSO is Agilent DSAX96204Q with sampling rate of 160 GSa/s. The offline DSP algorithm is conducted via Tensorflow, which is a neural network API written in Python. At the beginning of the training process, all parameters in our model are initialized by using the default implemented in Tensorflow [3)].

In this experiment, 3 sets of PAM-4 symbols have been generated with a bit pattern generator (BPG) of SHF 12104A. Therefore, we can get 3 datasets (denoted as set 1, set 2, set 3) in this experiment, each

1) Newport, 1784 VCSEL Datasheet. 2017. https://www.newport.com.cn/p/1784.
2) Newport, 1484-A-50-Newport. 2017. https://www.newport.com/p/1484-A-50.
3) Tensorflow official documentation. https://tensorflow.google.cn/.

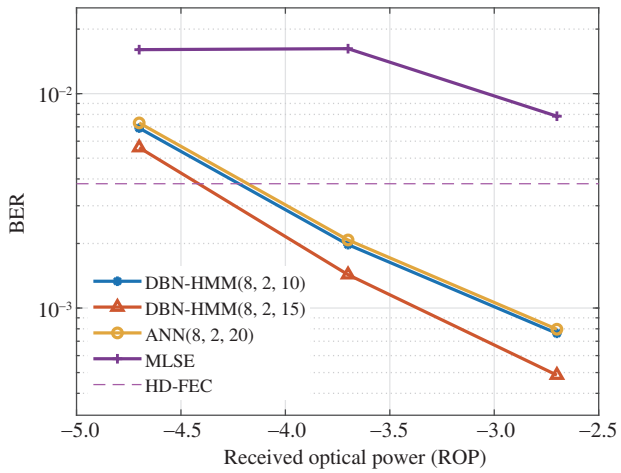


**Figure 6** (Color online) Measured BER vs. ROP with different equalizing strategies. The adjacent number $L = 2$, the BER performance of MLSE and ANN are also given. In ANN(8, 2, 20), 8, 2, 20 represent the interpolation multiple, adjacent number, and the number of neurons in hidden layer respectively. Compared with MLSE, the BER performance of DBN-HMM(8, 2, 15) can be greatly improved. DBN-HMM(8, 2, 15) and ANN(8, 2, 20) have the same computational complexity.

of which contains 1048576 PAM-4 symbols. In order to train the model, each dataset is divided into the training data and the testing data. For DBN-HMM training, the training data contains 50% data in set 1, set 2 and set 3 while the testing data contains the rest 50% data in set 1, set 2 and set 3.

Figure 6 presents the BER performance after DBN-HMM based equalization. We compare DBN-HMM based method with conventional MLSE, and the experiment results show that the BER performance can be greatly improved by introducing DBN-HMM. In order to strengthen our conclusion, the BER performance of ANN is also displayed in Figure 6. And conclusion can be reached that the computational complexity of DBN-HMM based equalizer can be about 41% lower than that of ANN based method with a similar BER performance. Lower computational complexity represents less number of additions and multiplications, which means lower energy consumption in the hardware. As the number of additions and multiplications decreases, the execution time of the algorithm can be reduced for the same hardware and the area of the chips can be reduced, leading to high energy efficiency. Nowadays, data centers are very sensitive to energy consumption. Therefore, models with lower computational complexity are more suitable for data centers to reduce the energy consumption.

### 3.2 Computational complexity

DBN-HMM based equalizer is trained offline during the training process, and the parameters remain unchanged at the equalizing stage. Therefore, we mainly focus on the computational complexity during the equalizing process. The equalizing process mainly contains two steps: the first step is to calculate the probability density of received symbols by means of trained DBNs, and the second step is the decision stage by Viterbi algorithm according to the transition matrix and the probability densities calculated in the first step.

In the first step, computing the natural exponential function in the neural network is costly. Fortunately, the natural exponential function in DBNs can be calculated by coordinate rotation digital computer (CORDIC) method, which means that the calculating process of exponential function has a lot of optimization space in the implementation of hardware [22]. Therefore, the computational complexity of the natural exponential function is ignored in this paper [15]. So equalizing one symbol requires $(L+1)\Gamma R+4R$ multiplications and $(L+1)\Gamma R+4R$ additions.

For the second step, the computational complexity mainly comes from Viterbi algorithm. When the state number $N$ is determined, for every received symbol, $(L+1)N$ multiplications and $(L+3)N$ additions are required. Therefore, totally equalizing every received symbol requires $(L+1)\Gamma R+4R+(L+1)N$ multiplications and $(L+1)\Gamma R+4R+(L+3)N$ additions for DBN-HMM based equalizer.

## 4 Conclusion

In this paper, in order to mitigate the nonlinear distortions, we propose DBN-HMM based equalizer for short-range optical interconnect and also experimentally study the performance of DBN-HMM based equalizer. DBN is used to fit the true distribution of received symbols while HMM is used for considering the sample's relevant information to reduce the computational complexity. The experimental results show that the BER performance can be greatly improved by introducing DBN-HMM compared with MLSE. Moreover, the computational complexity of DBN based equalizer can be about 41% lower than that of ANN based method with a similar BER performance. Therefore, the conclusion can be reached that DBN-HMM has a significant advantage in dealing with nonlinear distortion, making it quite suitable in high-speed optical interconnect systems.

**Acknowledgements** This work was supported by National Key R&D Program of China (Grant No. 2019YFB1802904) and Joint Fund of the Ministry of Education (Grant No. 6141A02033347).

**References**

1 Xie C. Datacenter optical interconnects: requirements and challenges. In: Proceedings of IEEE Optical Interconnects Conference, Santa Fe, 2017. 37–38
2 Tatum J, Landry G, Gazula D, et al. VCSEL-based optical transceivers for future data center applications. In: Proceedings of Optical Fiber Communication Conference, San Diego, 2018. M3F6
3 Meloni G, Malacarne A, Fresi F, et al. 6.27 bit/s/Hz spectral efficiency VCSEL-based coherent communication over 800 km of SMF. In: Proceedings of Optical Fiber Communication Conference, Los Angeles, 2015. Th2A.30
4 Rubsamen M, Winzer P J, Essiambre R-J. MLSE receivers for narrow-band optical filtering. In: Proceedings of Optical Fiber Communication Conference, Anaheim, 2006
5 Kottke C, Caspar C, Jungnickel V, et al. High speed 160 Gb/s DMT VCSEL transmission using pre-equalization. In: Proceedings of Optical Fiber Communications Conference and Exhibition, Los Angeles, 2017. 1–3
6 Karinou F, Prodaniuc C, Stojanovic N, et al. Experimental performance evaluation of equalization techniques for 56 Gb/s PAM-4 VCSEL-based optical interconnects. In: Proceedings of the 41th European Conference on Optical Communication, Valencia, 2015. 1–3
7 Tan Z, Yang C, Zhu Y, et al. High speed band-limited 850-nm VCSEL link based on time-domain interference elimination. IEEE Photon Technol Lett, 2017, 29: 751–754
8 Szczerba K, Lengyel T, Karlsson M, et al. 94-Gb/s 4-PAM using an 850-nm VCSEL, pre-emphasis, and receiver equalization. IEEE Photon Technol Lett, 2016, 28: 2519–2521
9 Tan Z W, Yang C C, Zhu Y X, et al. A 70 Gbps NRZ optical link based on 850 nm band-limited VCSEL for data-center intra-connects. Sci China Inf Sci, 2018, 61: 080406
10 Lavrencik J, Kota Pavan S, Thomas V A, et al. Noise in VCSEL-based links: direct measurement of VCSEL transverse mode correlations and implications for MPN and RIN. J Lightw Technol, 2017, 35: 698–705
11 Estaran J, Rios-Müller R, Mestre M A, et al. Artificial neural networks for linear and non-linear impairment mitigation in high-baudrate IM/DD systems. In: Proceedings of the 42th European Conference on Optical Communication, Dusseldorf, 2016. M.2.B.2
12 Chuang C, Wei C, Lin T, et al. Employing deep neural network for high speed 4-PAM optical interconnect. In: Proceedings of the 43th European Conference on Optical Communication, Gothenburg, 2017. W.2.D.2
13 Goodfellow I, Bengio Y, Courville A. Deep Learning. Cambridge: The MIT Press, 2016
14 Dahl G E, Yu D, Deng L, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Audio Speech Lang Process, 2012, 20: 30–42
15 Liang A, Yang C, Zhang C, et al. Experimental study of support vector machine based nonlinear equalizer for VCSEL based optical interconnect. Opt Commun, 2018, 427: 641–647
16 Rabiner L R, Juang B H. An introduction to hidden Markov models. IEEE Acoust, Speech, Signal Process Mag, 1986, 3: 4–16
17 Badino L, Canevari C, Fadiga L, et al. Deep-level acoustic-to-articulatory mapping for DBN-HMM based phone recognition. In: Proceedings of IEEE Spoken Language Technology, 2012
18 Dahl G, Yu D, Deng L, et al. Large vocabulary continuous speech recognition with context-dependent DBN-HMMS. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011
19 Mohamed A, Sainath T N, Dahl G, et al. Deep belief networks using discriminative features for phone recognition. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2011
20 Sheri A M, Rafique A, Pedrycz W, et al. Contrastive divergence for memristor-based restricted Boltzmann machine. Eng Appl Artif Intell, 2015, 37: 336–342
21 Rabiner L R. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE, 1989, 77: 257–286
22 Garrido M, Kallstrom P, Kumm M, et al. CORDIC II: a new improved cordic algorithm. IEEE Trans Circ Syst II, 2016, 63: 186–190