PAPER   *Special Section on Recent Advances in Simulation Techniques and Their Applications for Electronics*

A COM Based High Speed Serial Link Optimization Using Machine Learning

Yan WANG[†], *Student Member and* **Qingsheng HU**[†a)], *Nonmember*

SUMMARY   This paper presents a channel operating margin (COM) based high-speed serial link optimization using machine learning (ML). COM that is proposed for evaluating serial link is calculated at first and during the calculation several important equalization parameters corresponding to the best configuration are extracted which can be used for the ML modeling of serial link. Then a deep neural network containing hidden layers are investigated to model a whole serial link equalization including transmitter feed forward equalizer (FFE), receiver continuous time linear equalizer (CTLE) and decision feedback equalizer (DFE). By training, validating and testing a lot of samples that meet the COM specification of 400GAUI-8 C2C, an effective ML model is generated and the maximum relative error is only 0.1 compared with computation results. At last 3 link configurations are discussed from the view of tradeoff between the link performance and cost, illustrating that our COM based ML modeling method can be applied to advanced serial link design for NRZ, PAM4 or even other higher level pulse amplitude modulation signal.

key words: channel operating margin (COM), deep neural networks (DNNs), high-speed link, machine learning, signal integrity

1.   Introduction

With the increased speed of wireline transmission, signal integrity is facing huge challenges nowadays. To improve signal quality in high-speed wireline communication, various optimization methods have been explored not only in the system development, but also in transceiver design. Among them, IBIS-AMI (Input/output Buffer Information Specification-Algorithm Model Interface) [1], which models input/output behavior as well as algorithm for end-to-end high-speed serial link, has become an effective method, attracting much interests since it is proposed. This method, however, requires a large number of bits in simulation to optimize parameters, resulting in a long simulation time. Except for IBIS-AMI, two other methods commonly used are eye-mask [2] and channel operating margin (COM) [3]: although the former can estimate the channel performance intuitively, it often throws channel margin away necessarily [4]. While the latter, *i.e.*, COM based method stands out as a powerful one for system and transceiver designers to explore design space at the early stage of design, as well as to optimize the link parameters, thus overcoming the classic channel performance metrics such as eye diagram and bit error rate (BER).

COM based method along with measurement/simulation has been widely used in industry since it is proposed by the consortium of IEEE 802.3bj. For example, Bradon Gone *et al.* make comparisons between COM and Annex69B for 10Gb/s Ethernet (10GbE) channel evaluation, illustrating that the former outperforms the latter in accuracy [5]. Francesco *et al.* develop and validate a physical model of 400GbE channel depending on COM, moreover they make trade-off among channel length, losses and crosstalk and also select viable channel solutions based on COM [6]. Mike Resso *et al.* present a COM based backplane characterization techniques to alleviate the conflict that old backplane faces the new product generations [7].

Although COM has many advantages in evaluating and optimizing high-speed links, too much iteration should be taken in design space search, making it inflexible for the optimization of a large amount of channels. Fortunately, machine learning (ML) method, an efficient method to predict result quickly, has been applied in design and simulation of high-speed link recently and attracts more and more interests. For example, Bowen Li *et al.* use ML to accelerate the physical verification for high-speed link IC design [8]. In 2017, his team use ML and system identification approach to mimic the behavior model of receiver respectively and the result indicates the ML method has a good performance [9], [10]. Furthermore, they build a powerful ML architecture which can self-evaluate according to its own failure experiences to predict adaptive codes of receiver equalizers [11]. Tianjian Lu *et al.* apply a ML model to characterize the underlying relationships between performance metrics, such as eye height and width, and physical parameters of high-speed link. Their approach requires no complex circuit simulation and the trained model reused conveniently for link design [12]. In 2018, a high-speed channel modeling and optimization technique [13] are proposed by Kim H *et al.*, in which a RLGC (resistance, inductance, conductivity, capacitance) matrices are adapted to characterize the channels and a ML model based on artificial neural network are built for predicting target channels matrices.

These researches, however, either concern COM based link optimization or concern effective ML algorithm in which eye diagram is used as usual for link performance evaluation. In this paper, a serial link optimization method in which COM is taken as the evaluate measure as well as ML is applied predict link parameters is proposed. By taking advantages of COM and ML, our method can offer optimized link configuration with very rapid speed.

2. High-Speed Serial Link and COM

A typical high-speed serial system can be divided into three sections named channel, transmitter (TX) and receiver (RX), shown as Fig. 1. A channel, e.g., backplane channel including print circuit board (PCB) traces, vias, connectors and packages, may be very complicated and always behaves like a low-pass filter. In TX, a typical equalizer such as feed forward equalizer (FFE) can be employed to release the pressure of receiver [14]. In RX, two types of equalizer, *i.e.*, linear and nonlinear one are adopted generally to alleviate the effects of channel loss further. A typical linear equalizer is continuous time linear equalizer (CTLE) which functions as a high-pass filter [15]. While decision feedback equalizer (DFE) commonly used to cancel post-cursor inter symbol interference (ISI) is a nonlinear equalizer [16].

COM is defined as the ratio of available signal amplitude to statistical noise in dB:

$$COM = 20 \times \log_{10}\left(\frac{A_s}{A_{ni}}\right) \tag{1}$$

where A_s is the signal amplitude and A_{ni} represents the amplitude of all noise and interference. From Eq. (1), we know that the greater the COM, the better the link performance. In other words, sufficient COM means a design can work robustly. In a typical case, 3dB is defined as a channel compliance threshold [4]. It is worth to note that the calculation of COM allows us to achieve critical data that can be used for the validating and margining of transceiver.

Figure 2 gives a typical link model for COM calculation. In TX, FFE and transmitter filter are included, modeled as $H_{ffe}(f)$ and $H_t(f)$, respectively. In RX, many filters such as receive filter $H_r(f)$, continuous time filter $H_{ctf}(f)$ and

DFE are contained, meanwhile the contributions of noise, interference and jitter are taken into account as well. For victim channel and aggressor channels which model the interference, S-parameter based models $S^{(0)}$ and $S^{(k)}$ are used, respectively, where k stands for the kth aggressive channel. Additionally, device package $S^{(tp)}$, $S^{(rp)}$ and path termination R_d for TX/RX are also included in order to emulate practical link as precisely as possible.

The link optimization flow based on COM is shown as Fig. 3. After achieving channel S-parameters by measurement or other methods, a channel transfer function $H_{ch}(f)$, representing channel and package & termination can be obtained at first. Then one set of TX+RX equalization configuration exclude DFE is chosen. Thus, we can express the transfer function as:

$$H(f) = H_{ffe}(f)H_t(f)H_{ch}(f)H_r(f)H_{ctf}(f) \tag{2}$$

where $H_{ffe}(f)$, $H_t(f)$, $H_r(f)$, and $H_{ctf}(f)$ are the transfer functions of TX FFE, TX transition time filter [17], RX noise filter [18] and CTLE given by Eq. (3) [4].

$$H_{ffe}(f) = \sum_{i=1} c_i \exp(j2\pi(i+1)(f/f_b))$$

$$H_t(f) = \exp 2\left(\frac{\pi f T_r}{1.6832}\right)^2$$

$$H_r(f) = \frac{1}{1 - 3.4(f/f_b)^2 + (f/f_b)^4 + j2.6((f/f_b) - (f/f_b)^3)}$$

$$H_{ctf}(f) = \frac{(10^{\frac{g_{DC}}{20}} + j(f/f_z))(10^{\frac{g_{DC2}}{20}} + j(f/f_{LF}))}{(1 + j(f/f_{p1}))(1 + j(f/f_{p2}))(1 + j(f/f_{LF}))}$$

$$\tag{3}$$

where c_i is the coefficient of the i_{th} tap of FFE, $f_b = 1/T_b$, where T_b is the time duration of a symbol, T_r is the transition
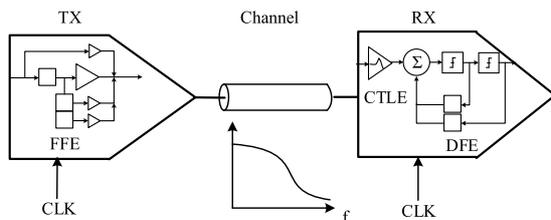

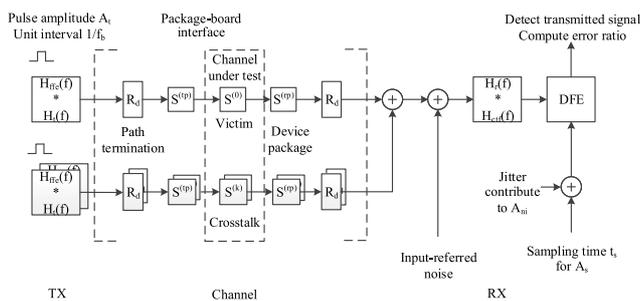**Fig. 1** A typical high speed serial link

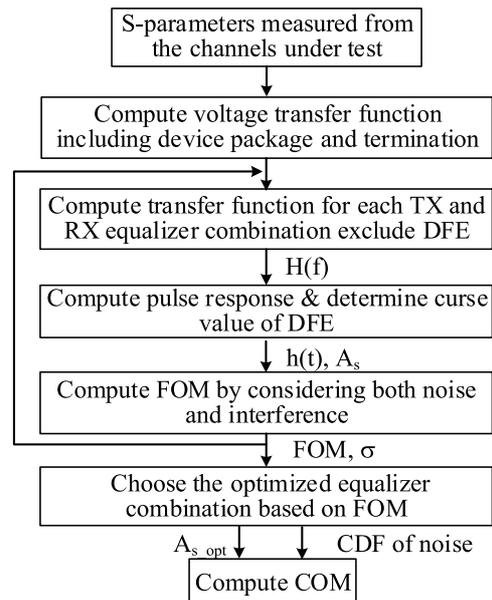**Fig. 2** A typical link model for COM calculation

Fig. 3 COM based link optimization flow

time of received signal and f_r is 3dB bandwidth of TX filter. While g_{DC} and g_{DC2}, f_{p1}, f_{p2}, f_z and f_{LF} are DC gains, poles, zero and low frequency pole/zero of CTLE, respectively.

Once $H(f)$ is obtained, the main cursor A_s of the pulse response $h(t)$ corresponding to $H(f)$ can be determined by Eq. (4):

$$A_s = h(t_s) \tag{4}$$

where t_s is the sampling time which is one UI after the pre-zero-crossing nearest to the peak of $h(t)$ [22].

After that, a figure of merit (FOM) which evaluates the effect of various equalization configurations should be calculated as:

$$FOM = 10 \log_{10} \frac{A_s^2}{\left(\sigma_{TX}^2 + \sigma_{ISI}^2 + \sigma_J^2 + \sigma_{XTK}^2 + \sigma_N^2 \right)} \tag{5}$$

where the denominator is the sum of variances of all noise and interference. Among them σ_{TX}^2 represents the variance of noise from TX end. σ_{ISI}^2 represents the amplitude variance of ISI [19]. σ_J^2 refers to the error coming from timing jitters. σ_{XTK}^2 is the sum of the maximum amplitude variance for all crosstalk paths, and σ_N^2 is that of Gaussian noise of receiver, respectively.

Finally, we traversal for all combinations of TX and RX equalizers parameter to find the optimal equalizers set, COM can be computed based on A_s and A_{ni} which correspond to the maximum FOM.

It is worth to point out that the total noise and interference amplitude A_{ni} in Eq. (1) can be obtained from the cumulative distribution function (CDF) $P(y)$ defined by Eq. (6):

$$P(y) = \int_{-\infty}^{y} p(y) dy \tag{6}$$

where $p(y)$ is the combined probability density function (PDF) yielded by convoluting all noise and interference amplitude distribution function [4].

Last but not least, A_{ni} is defined by the magnitude of y_0 that satisfies $P(y_0) = DER$, where DER is the target detector error ratio.

3. A COM Based Machine Learning Method

It is well known that machine learning can be applied in parameter optimization as well as result prediction in many fields with the advantages of high accuracy and effectiveness [20]. After learning from a large amount of data, the training model may build some logic of its own. Once the model is ready to be applied, it allows the system to generate decision results at a very fast speed.

Figure 4 gives our COM based ML modeling for serial link optimization, in which two steps are included: the first is data pre-processing and the second is ML modeling. At first, amounts of training channels presented in S-parameters [21] are inputted to be pre-processed. Then after
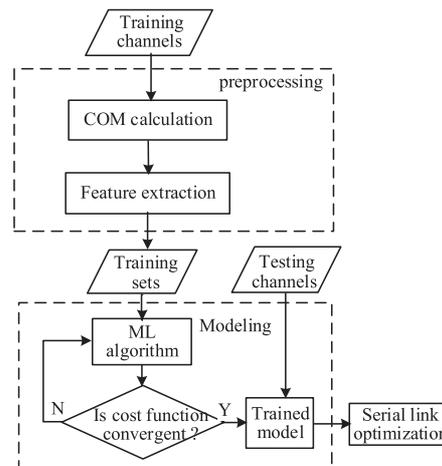

Fig. 4　COM based ML method for link optimization

completing the COM calculation, several important features are extracted which will be grouped as a training set, validating set and testing set for ML modeling.

In the second step, *i.e.*, ML modeling, we use the training set obtained in the first step to train ML model until the cost function is converged and then we use the validation set to optimize model parameters. After that, the testing set is use to validate the model by comparing predicted result with practical one.

Once the modeling process is finished, the model can be employed for serial link design such as link performance evaluation and equalizer configuration optimization.

3.1　Pre-Processing

The data pre-processing is responsible for generating the useful data to construct training and testing set for ML algorithm. First COM is calculated for training channels. In this flow, on the one hand, the optimized equalization configuration such as TX FFE coefficients, RX CTLE DC gain and DFE taps are recorded, on the other hand, zeros and poles of CTLE and some important time domain waveform such as single bit response (SBR) as well as long pulse response (LPR) are extracted. These features will be used to construct training set for ML based modeling. For example, for M training channels each with X TX FFE taps, if N cursors amplitude are extracted from SBR of each channel, then a $M \times X$ coefficient matrix and a $M \times N$ feature matrix can be obtained for training TX FFE.

For RX, more features can be measured on receiver input waveform, including high-frequency/low-frequency spectrum density for SBR and LPR, as well as some features of CTLE. If L features are recorded, then a $M \times L$ and a $M \times Y$ matrix can be generated for modeling RX CTLE, where Y is the coefficient of CTLE in frequency domain. Because of the SBR data can show channel loss. The LPR would provide DC gain information [11]. Table 1 gives main features used in modeling and the prediction targets.

Different from the coefficients of TX FFE and RX

Table 1   Features and prediction targets of proposed method

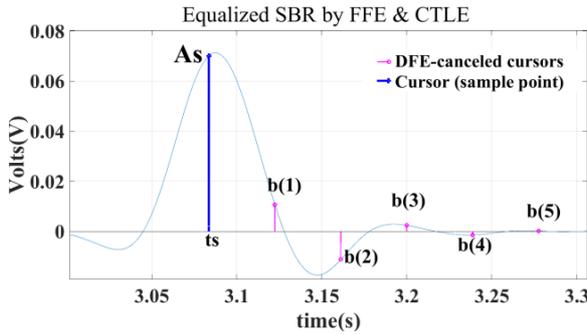|	TX FFE	RX CTLE
Features	Signal amplitude of SBR	Signal amplitude of SBR and LPR
		Frequency spectrum density of SBR and LPR
		Zeros, Poles
Prediction targets	FFE coefficients	CTLE DC gain

Fig. 5   h(t) used to determine DFE coefficients

CTLE which are chosen by searching whole design space of FOM computation, DFE coefficients are determined according to $h(t)$ which is corresponding to the best FOM. Figure 5 illustrates how to determine DFE coefficients. Suppose n is the index of DFE tap, and $n = 1 \sim N_b$ where N_b is number of DFE length. Then the coefficients of the nth DFE tap $b(n)$ can be determined as following:

$$b(n) = \begin{cases} -b_{\max}(n) & h(t_s + nT_b)/h(t_s) < -b_{\max}(n) \\ b_{\max}(n) & h(t_s + nT_b)/h(t_s) > b_{\max}(n) \\ h(t_s + nT_b)/h(t_s) & otherwise \end{cases}$$

$$(7)$$

where $b_{max}(n)$ is the normalized coefficient magnitude limit for tap n.

Figure 5 shows a typical $h(t)$ after applying TX FFE and RX CTLE, in which A_s is the amplitude of $h(t)$ at sample time t_s and $b(1) \sim b(5)$ are the post-cursors which should be canceled by DFE.

3.2 DNN Modeling

A deep neural network (DNN) [23] which mimics the learning process of human brain neurons is explored in this paper. Figure 6 (a) shows the basic unit of DNN, *i.e.* M-P neuron [24] which is created by McCulloch and Pitts through a mathematical model to simulate its "excited" behavior when it receives a signal beyond a certain threshold. In Fig. 6 (a), $x = \{x_1, x_2, \dots, x_n\}$ represents N input features of neuron, coming from input layer or output stretched from other neurons. These input signals are then transmitted to the neuron through linearly weighted connections which can be represented by a weight vector $w = \{w_1, w_2, \dots, w_n\}$. Thus, the linear output z can be expressed as:
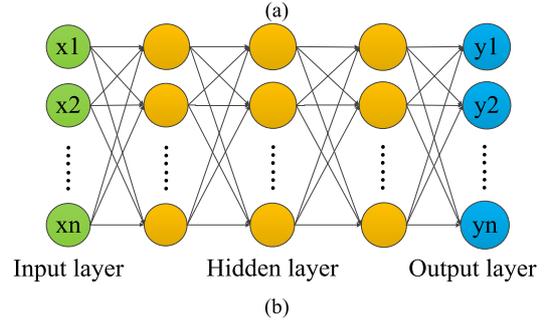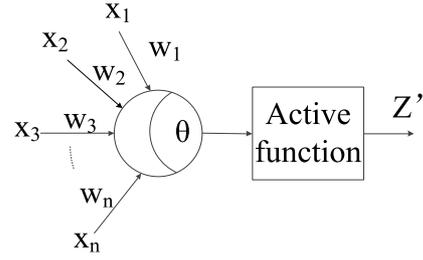

(a)

(b)

Fig. 6   DNN structure

$$z = x \cdot w \tag{8}$$

After that, z', the output of neurons can be generated by active function as following:

$$z' = f_a(z - \theta) \tag{9}$$

where $f_a(\cdot)$ stands for the active function and θ is threshold value.

Compared with perceptron [25] which is a simplest structure only consisting of input and output layer, DNN increases hidden layers between input and output layer, shown as Fig. 6 (b). We can see that the input features received by input layer are passed onto the hidden layers one by one until output layer and each neuron in hidden layer has an active function used to standardize the output.

The active function used in this paper is hyperbolic tangent function shown as following:

$$z' = \tanh(z - \theta) \tag{10}$$

where z' is the output of hidden layer, it is the input of next neuron as well. At the end of network, a linear regression output layer [26] is constructed to generate a predicted result.

To evaluate the gap between predicted value and real value, a cost function should be employed. Here, mean square error (MSE) is used as the cost function:

$$J = \frac{1}{m} \sum_{i=1}^{m} (y_i - \hat{y}_i) \tag{11}$$

where y_i and $\hat{y}_i$ are the real and predicted value respectively and m is the number of data set.

To minimize the cost function, the scaled conjugate gradient (SCG) [23] method is used in our work. Additionally, back propagation (BP) [27] and Bayesian optimization [28] schemes are applied in training process to improve

the performance of machine learning. In practice, we train the DNN model iteratively until J is below 0.03, then we apply the features extracted from the test channels to the trained DNN to validate an optimization the model. After that, we can obtain the predicted targets, *i.e.*, TX FFE coefficients and RX CTLE parameters from the output of DNN. Once FFE and CTLE have been modeled successively, $h(t)$ can be acquired easily, and the DFE tap coefficients can be determined by Eq. (7).

4. Numerical Examples

In our work, two numerical examples are presented to evaluate the effectiveness of applying our ML method in serial link optimization. Table 2 gives the summary of our data set for ML modeling. Totally 214 channels including 140 training sets, 40 validation sets and 44 test sets (10 from validation set) are contained listed in Table 2. Among the test channels, 40 test channels dedicated for illustrating the effectiveness of ML are grouped as example A. While the rest 4 channels illustrating how to use COM to optimize link configuration belong to example B.

4.1 Link Parameter Prediction

In this example, the serial link used should meet the physical layer specification 400GAUI-8 C2C [4], in which part of important parameters are shown as Table 3. We can see that the symbol rate is 26.5625GBd and modulation way

is PAM4 [29]. In TX, 3 taps FFE is used and the coefficient range for pre/post-cursor are specified as −0.15~0 and −0.25~0 with a step of 0.01, respectively. For CTLE, the poles and zeros are fixed and g_{DC} and $g_{DC}2$ can be optimized in a range of −20~0 and −10~0 dB with a step of 0.2 and 0.1, respectively. Similarly, the DFE including 10 taps and their magnitude limit are also specified.

Concretely say, for TX, two fully connect neurons layers with 60 nodes are used and the training set contains a 140×14 feature matrix (where 14 is the number of sampled cursors) and a 140×3 coefficient matrix. For RX, a three-hidden layer with 100, 80, and 60 nodes model is used. The feature and coefficient matrix are 140×225 (where 225 is the number of extracted feature) and 140×2, respectively. In addition, the learning rate and batch size are set as 0.005 and 20, respectively in our work.

Figure 7 shows the inset losses (IL) of some training channels. It is worth to illustrate that 40 test channels are divided into two groups according to their IL: one is for low IL named group1 normally with COM more than 3dB. The other is for high IL named group2 and may be failed in channel compliance check.

Figure 8 shows the examples of ML modeling results for a test channel with 12.24dB IL. We can see that the parameters of three types equalizer achieved by ML method are very close to that extracted during COM calculation.

Figure 9 gives the relative errors between ML and COM calculation for major link parameters, such as FFE main tap, DC gain and DFE tap1 coefficient. Among 40 samples of test set, No.1~No.20 belong to group1 and No.21~No.40 belong to group2. We can see that the fitting result for FFE is the best. Since g_{DC1} and g_{DC2} error affects the determination of DFE coefficients, it can be observed obviously that the DFE curve fluctuates with g_{DC1}. For group2, *i.e.*, No.21~No.40, however, DFE relative error seems disappear. This is because that the insert loss is so high that the coefficients of CTLE and DFE reach the maximum value.

Figure 10 (a) and Fig. 10 (b) show the comparison of COM between calculation and ML method as well as the relative error for group1 and group2, respectively. We can see that for both groups the results by ML match that by calculation well and the both of maximum relative error is
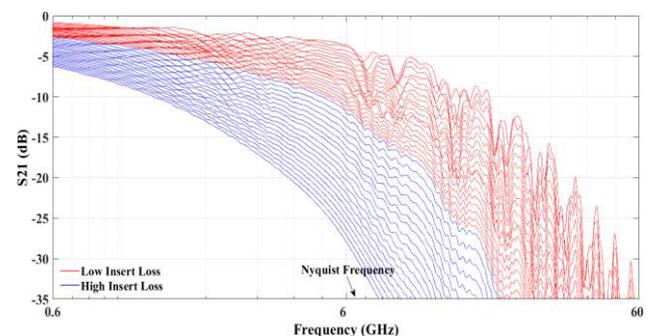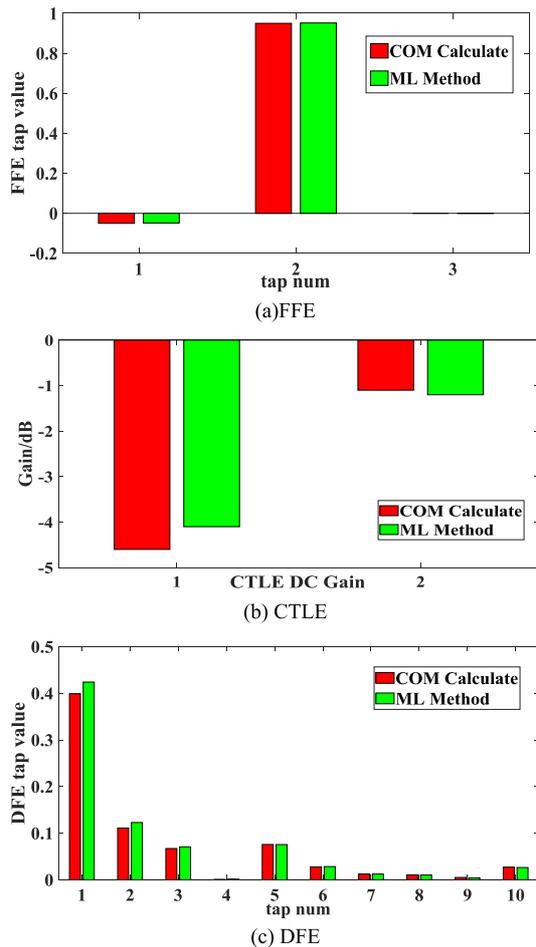
Table 2 Data set for numerical examples

	Number of channels	
Training Set	140	
Validation Set	40	
Test Set	Example A	Example B
	40 (10 from validation set)	4
Total	214	

Table 3 COM setting for 400GAUI-8 C2C

	Parameters	Symbol	Value			Units
General	symbol rate	f_b	26.5625			GBd
	number of signal levels	L	4			-
	COM threshold	th	3			dB
	Target detector error ratio	DER	10^{-5}			
			Minimum	Maximum	Step	
FFE	main cursor	$c(0)$	0.6			-
	pre-cursor	$c(-1)$	-0.15	0	0.01	-
	post-cursor	$c(1)$	-0.25	0	0.01	-
CTLE	DC gain 1	g_{DC}	-20	0	0.2	dB
	DC gain 2	g_{DC2}	-10	0	0.1	dB
	zero for g_{DC} =0	f_z	$f_b/2.5$			GHz
	pole1	f_{p1}	$f_b/2.5$			GHz
	pole2	f_{p2}	$2 \times f_b$			GHz
	low frequency pole/zero	f_{LF}	$f_b/40$			GHz
DFE	length	N_b	10			UI
	normalized tap magnitude limit	$n=1$	$b_{max}(n)$	0.5		-
		$n=2\sim10$		0.2		

Fig. 7 IL of some training channels

(a)FFE

(b) CTLE

(c) DFE

Fig. 8   Examples of modeling result for 3 types of equalizer
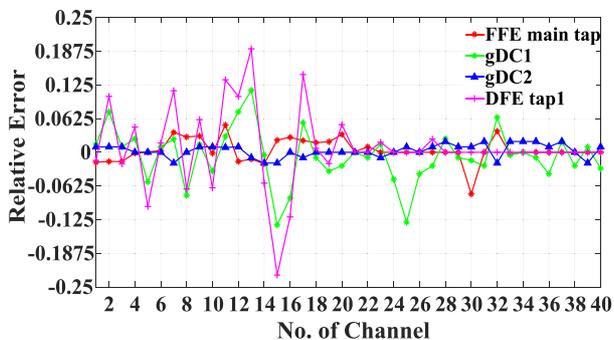
Fig. 9   Relative error of critical coefficients

(a) Channels with low IL

(b) Channels with high IL

Fig. 10   COM results by calculating and ML modeling

Table 4   Performance of predicted FFE CTLE and COM value

	FFE			CTLE		COM
	pre	main	post	$g_{DC}1$	$g_{DC}2$	
Convergence of J	0.008			0.021		
Maximum relative error(%)	5.7	8.1	3.6	12.6	3.7	7
RMSE	0.016	0.015	0.011	0.895	0.27	0.067

RMSE are below 1. As for the prediction accuracy of COM, which should be most concerned, it shows that its RMSE and maximum relative error are 0.067 and 7%, respectively. The results above illustrate that our method is effective and has high accuracy.

4.2   Link Optimization

The aim of example B is to illustrate how to choose optimized equalize configuration according to COM. In this example, 4 types of channels, which stand for different return loss (RL) and IL situation, *i.e.*, low IL+low RL (Ch.1), low IL+high RL(Ch.2), high IL+low RL(Ch.3) and high IL+high RL (Ch.4) are investigated, shown as Fig. 11. Totally three equalization combinations are available: TX FFE + RX CTLE+DFE, RX CTLE+DFE and TX FFE + RX DFE. Table 5 gives the COM values of 4 channels and the corresponding histogram is shown in Fig. 12.

We can see that for Ch.1 which has low IL and RL, COM is great than 3dB for all combinations and the maximum COM can be up to 6.66dB. For Ch.2 and Ch.3, their COM values also meet the requirement of 3dB although the

below 0.1, illustrating that our method works correctly for all cases, including that of negative COM for No.36~No.40 as shown in Fig. 10 (b).

Table 4 summarizes the performance of the proposed method measured by convergence, maximum relative error and root-mean-square error (RMSE). We can see that the convergences of both cost functions are below 0.03. The maximum relative errors for FFE pre/main/post coefficient are 5.7%, 8.1% and 3.6%, respectively. The worst error presents at $g_{DC}1$, which is 12.6%. We can also see that all
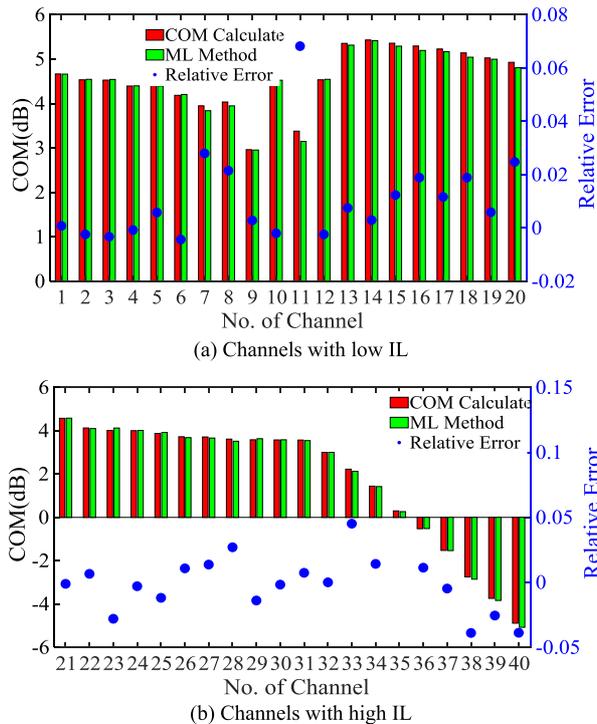
Fig. 11   RL and IL of 4 channels

Table 5   COM values for 4 channels of example B

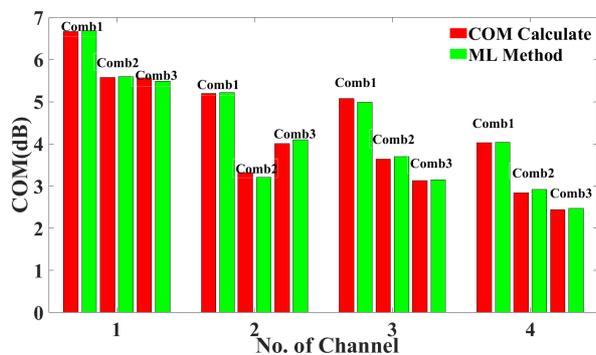| Combination\\\No. of channels	Comb1 (FFE+CTLE+DFE)	Comb2 (CTLE+DFE)	Comb3 (FFE+DFE)
Ch.1(Low IL, Low RL)	6.66	5.57	5.57
Ch.2(Low IL, High RL)	5.19	3.32	4.10
Ch.3(High IL, Low RL)	5.07	3.63	3.13
Ch.4(High IL High RL)	4.03	2.84	2.44

Fig. 12   Histogram of COM for different equalization combinations

maximum value decreases from 6.66dB to 5.07dB due to either high IL or high RL. Additionally, it can be inferred that FFE plays an important role in canceling RL while CTLE has no effect on it since the COM for Comb.3 is higher than Comb.2 about 0.8dB. It is worth to note that, for saving area or power, the combination 2 and 3 are acceptable with Ch.1~Ch.3 and the result of our method also can match well. Unfortunately, for Ch.4 only one combination, *i.e.* FFE+CTLE+DFE can pass compliance check and all other cases may be failed due to the heavy IL and RL.

5. Conclusions

In this work, a machine learning method has been proposed for high speed link optimization, in which both TX FFE and RX CTLE are modeled by using DNN method combined with COM threshold to obtain effective link configuration. Through lots of samples that meet the COM specification of 400GAUI-8 C2C, two optimized DNN model with maximum relative error of 0.08 and 0.12 is generated to predict the link parameters. Furthermore, 3 different equalization combinations for 4 channels with different IL and RL are

evaluated in detail based on the proposed ML method. It can be observed that for the worst channel which has heavy IL and RL only TX FFE+ RX CTLE+DFE can pass compliance check of COM, while for the rest 3 channels all combinations can meet the threshold of 3dB and the maximum COM can be up to 6.7dB. For these cases further optimization should be performed in order to get a good tradeoff between the link performance and the cost such as area and power consumption.

In this work, the machine learning method can be utilized in the COM based link parameters prediction with high accuracy and high speed. Besides, the ML model can change another type to broaden the range of application such as simple linear regression or support vector regression or RNN [30].

References

[1] Y. Zhan, Q. Hu, and Y. Zhang, "IBIS-AMI based PAM4 signaling and FEC technique for 25Gb/s serial link," Proc. International Conference on Wired/wireless Internet Communication, pp.271–281, Sept. 2017.

[2] C. Wu, "Eye-pattern templates help evaluate SerDes performance," Edn Europe, no.9, pp.50–56, May 2008.

[3] Y. Sun and Q. Hu, "COM based Channel-Compliance analysis and its comparison with BER-Way for 25 Gbps NRZ serial link," Proc. 2019 3rd International Conference on Circuits, System and Simulation (ICCSS), pp.172–177, Aug. 2019.

[4] IEEE 802.3-2018, "IEEE Standard for Ethernet," 2018.

[5] B. Gore and R. Mellitz, "An excerise in applying channel operating margin (COM) for 10GBASE-KR channel design," IEEE, no.6, pp.648–653, 2014.

[6] F. de Paulis, T. Wang-Lee, R. Mellitz, M. Resso, R. Rabinovich, and O.J. Danzy, "Backplane channel design exploration at 112 gbps using channel operating margin (COM)," Proc. 2020 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI), Feb. 2020.

[7] M. Resso, "Characterization techniques of a terabit backplane using new figures of merit," Proc. 2017 IEEE 21st Workshop on Signal and Power Integrity (SPI), Oct. 2017.

[8] B. Li and P.D. Franzon, "Machine learning in physical design," Proc. Electrical Performance of Electronic Packaging & Systems, 2016.

[9] Y. Choi and C. Cheng, "High-speed link analysis with system identification approach," Proc. 2017 IEEE International Symposium on Electromagnetic Compatibility & Signal/Power Integrity (EMCSI), April 2017.

[10] B. Li, P. Franzen, Y. Choi, and C. Cheng, "Receiver Behavior Modeling Based on System Identification," Proc. 2018 IEEE 27th Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), Sept. 2018.

[11] B. Li, B. Jiao, C.-H. Chou, R. Mayder, and P. Franzon, "Self-Evolution cascade deep learning model for High-Speed receiver adaptation," IEEE Transactions on Components, Packaging, and Manufacturing Technology, vol.10, no.6, pp.1043–1053, 2020.

[12] T. Lu, J. Sun, K. Wu, and Z. Yang, "High-Speed channel modeling with machine learning methods for signal integrity analysis," IEEE Trans. Electromagn. Compat., vol.60, no.6, pp.1957–1964, 2018.

[13] H. Kim, C. Sui, K. Cai, B. Sen, and J. Fan, "Fast and Precise High-speed Channel Modeling and Optimization Technique Based on Machine Learning," IEEE Trans. Electromagn. Compat., vol.60, no.6, pp.2049–2052, 2018.

[14] S. Mukherjee, A. Das, S. Seth, and S. Saxena, "An Energy-Efficient 3 gb/s PAM4 Full-Duplex transmitter with 2-Tap feed forward equalizer," IEEE Trans. Circuits Syst. II, vol.67, no.5, pp.916–920, 2020.

[15] J.F. Bulzacchelli, P.A. Francese, Y. Liu, and T.H. Toifl, "Continuous-time linear equalizer for high-speed receiving unit," US Patent, 09467313B2, 2016.

[16] H. Kim and C. Yoo, "A 6-Gb/s wireline receiver with intrapair skew compensation and Three-Tap Decision-Feedback equalizer in 28-nm CMOS," IEEE Trans. Very Lagre Scale Integr. (VLSI) Syst., vol.28, no.5, pp.1107–1117, 2020.

[17] T. Kalker, "On optimal boundary and transition filters in time-varying filter banks," Image Processing, 1996. Proceedings., International Conference on, Lausanne, Switzerland, pp.625–628, IEEE, Aug. 1996.

[18] I. Bilal, A. Meijerink, and M.J. Bentum, "Optimum receiver filter for a noise-based frequency-offset modulation system," Proc. IEEE International Symposium on Personal, 2016.

[19] S.Y. Mui, "Method for intersymbol interference compensation," US Patent, 6690739B1, 2004.

[20] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, The MIT Press, US, 2016.

[21] J.P. Peter, S-parameters for Signal Integrity, Cambridge University Press, 2020.

[22] K. Mueller and M. Muller, "Timing recovery in digital synchronous data receivers," IEEE Trans. Commun., vol.24, no.5, pp.516–531, 1976.

[23] S. Raschka, Python Machine Learning, Packt Publishing, US, 2014.

[24] L. Zhang and B. Zhang, "A geometrical representation of McCulloch-Pitts neural model and its applications," IEEE Trans. Neural Netw., vol.10, no.4, pp.925–929, 1999.

[25] B. Widrow and M.A. Lehr, "30 years of adaptive neural networks: Perceptron, Madaline, and backpropagation," Proc. IEEE, vol.78, no.9, pp.1415–1442, 1990.

[26] R.G.D. Steel and J.H. Torrie, "Principles and Procedures of Statistics: A Biometrical Approach," Biometrics, vol.37, no.4, pp.859–860, 1981.

[27] P.J. Werbos, "Backpropagation through time: What it does and how to do it," Proc IEEE, vol.78, no.10, pp.1550–1560, Oct. 1990.

[28] J. Mockus, Bayesian Approach to Global Optimization: Theory and Applications, Kluwer Academic Publisher, Berlin, 1989.

[29] intel, "AN835: PAM4 Signaling Fundamentals," 2019.

[30] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting – a novel pooling deep RNN," IEEE Trans. Smart Grid, vol.9, no.5, pp.5271–5280, 2018.

Qingsheng Hu received the circuits and systems doctoral degree from Shanghai Jiao Tong University, Shanghai, China in 1997. Then she engaged in postdoctoral research in communications and electronics department of China Science and Technology University. In 1999, she joined Nanjing Institute of ZTE, where she engaged in research and development of communications products. In 2002, she engaged in postdoctoral research in Electrical and electronic engineering and computer engineering department of Hong Kong Science and Technology University. In 2003, she joined at the school of Information Science and Engineering of Southeast University.

Yan Wang was born in 1992 at Jiangsu Province China. He is currently working toward the Ph.D. degree in Southeast University. He received his M.S. degree from Nanjing University of Technology in 2015. His research interests focus on the communication reliability for high speed SerDes.