

IEICE **TRANSACTIONS**

on Communications

DOI:10.23919/transcom.2023EBP3121

This advance publication article will be replaced by the finalized version after proofreading.

A PUBLICATION OF THE COMMUNICATIONS SOCIETY



The Institute of Electronics, Information and Communication Engineers
Kikai-Shinko-Kaikan Bldg., 5-8, Shibakoen 3chome, Minato-ku, TOKYO, 105-0011 JAPAN

Overfitting Problem of ANN- and VSTF-based Nonlinear Equalizers Trained on Repeated Random Bit Sequences

Kai IKUTA^{†a}, Jinya NAKAMURA^{††}, *Student Members*, and Moriya NAKAMURA[‡], *Member*

SUMMARY In this paper, we investigated the overfitting characteristics of nonlinear equalizers based on an artificial neural network (ANN) and the Volterra series transfer function (VSTF), which were designed to compensate for optical nonlinear waveform distortion in optical fiber communication systems. Linear waveform distortion caused by, e.g., chromatic dispersion (CD) is commonly compensated by linear equalizers using digital signal processing (DSP) in digital coherent receivers. However, mitigation of nonlinear waveform distortion is considered to be one of the next important issues. An ANN-based nonlinear equalizer is one possible candidate for solving this problem. However, the risk of overfitting of ANNs is one obstacle in using the technology in practical applications. We evaluated and compared the overfitting of ANN- and conventional VSTF-based nonlinear equalizers used to compensate for optical nonlinear distortion. The equalizers were trained on repeated random bit sequences (RRBSs), while varying the length of the bit sequences. When the number of hidden-layer units of the ANN was as large as 100 or 1000, the overfitting characteristics were comparable to those of the VSTF. However, when the number of hidden-layer units was 10, which is usually enough to compensate for optical nonlinear distortion, the overfitting was weaker than that of the VSTF. Furthermore, we confirmed that even commonly used finite impulse response (FIR) filters showed overfitting to the RRBS when the length of the RRBS was equal to or shorter than the length of the tapped delay line of the filters. Conversely, when the RRBS used for the training was sufficiently longer than the tapped delay line, the overfitting could be suppressed, even when using an ANN-based nonlinear equalizer with 10 hidden-layer units.

key words: optical nonlinear compensation, nonlinear equalizers, artificial neural network, Volterra series transfer function, overfitting

1. Introduction

Data traffic through communication systems has been continuing to grow exponentially with the technological development of cloud computing and fifth-generation (5G) mobile communications. Increasing the capacity further will require optical fiber communications technology that supports these services. To meet this demand, multi-level modulation, including quadrature amplitude modulation (QAM), is an important technology that can increase the spectral efficiency in the limited optical bandwidth. However, a QAM signal has a large peak-to-average power ratio (PAPR) and is susceptible to nonlinear waveform distortion caused by optical nonlinear effects such as self-phase modulation (SPM) and cross-phase modulation (XPM). Techniques to compensate for the nonlinear

waveform distortion using digital signal processing (DSP), digital backpropagation (DBP) and nonlinear equalizers based on the Volterra series transfer function (VSTF) have been studied [1-4]. However, the significant computational complexity of these methods poses a technical barrier to their practical implementation. On the other hand, nonlinear equalizers based on artificial neural networks (ANNs) are attracting attention as another possible candidate. ANN-based nonlinear equalizers have been experimentally demonstrated with various modulation formats, including intensity modulation and direct detection (IM/DD), QAM, and orthogonal frequency-division multiplexing (OFDM) [5-7]. The effectiveness of the equalizers has been verified not only in laboratory experiments but also with an 11,000-km live-traffic carrying submarine cable [8]. Recently, several field-programmable gate array (FPGA) implementations of ANN-based nonlinear equalizers have been demonstrated [9,10]. One implementation realized both the equalization and training stages within the same FPGA simultaneously [11]. In our research group, we demonstrated complex-valued ANN-based nonlinear equalizers, which showed improved learning speed and reduced computational complexity compared to a conventional real-valued ANN [12]. Furthermore, we clarified the necessary number of ANN units for compensating for chromatic dispersion (CD) and SPM [13]. We also reported that an ANN can effectively compensate for nonlinearities using significantly less computational effort compared to DBP and the VSTF [14,15].

An issue that has been pointed out with the ANN-based nonlinear equalizers is overfitting. In particular, when a pseudo-random binary sequence (PRBS) is used in the training, the ANN configures a logic circuit that is optimized for the specific PRBS [16-18]. Consequently, the ANN predicts the incoming PRBS signals, resulting in overestimation of the compensation performance. Conversely, when the compensation performance is evaluated using a PRBS different from the one used in the training, the compensation performance is underestimated. Some reports investigated the dependence of the tap length of the ANN and the length of the PRBS on the overfitting characteristics [19,20]. It is also reported that the overfitting becomes stronger when the number of hidden-layer of the ANN is increased from three to four [21]. We evaluated the overfitting characteristics of VSTF-based nonlinear

[†]The authors are with Meiji University, Kawasaki-shi, 214-8571 Japan.

^{††}The author is with Fujikura Ltd., Koto-ku, 135-8512 Japan.
a)E-mail: k_ikuta@meiji.ac.jp

equalizers using the same method that has been employed to evaluate the overfitting of ANN-based nonlinear equalizers. As a result, we revealed that the overfitting of the ANN- and VSTF-based nonlinear equalizers occurs under the same conditions when PRBSs are used in the training [22]. This is because the VSTF has a high function representation capability and thus acquires the logic circuit of the PRBS as well as the ANN. We should consider that the overfitting is not a problem that is unique to ANN-based nonlinear equalizers but possibly occurs with any equalizers using learning algorithms.

In addition to PRBSs, the overfitting characteristics of the ANN-based nonlinear equalizers have also been investigated in a case where finite-length repeated random bit sequences (RRBSs) were used in the training [16-18]. As the number of input and hidden layer units in the ANN is increased, the ANN-based nonlinear equalizers have a higher function representation capability to memorize the random bit sequence, resulting in overfitting. However, it is known that the overfitting of ANN-based nonlinear equalizers with an RRBS is weaker than that with a PRBS. On the other hand, the overfitting characteristics of VSTF-based nonlinear equalizers with an RRBS have not been investigated, to the authors' best knowledge. Therefore, it remains unclear whether the overfitting of ANN-based nonlinear equalizers with an RRBS is larger than that of the VSTF. This paper focuses on comparing the overfitting characteristics of the ANN- and VSTF-based nonlinear equalizers trained on RRBSs, in contrast to the characteristics of the ANN trained on PRBS, which were investigated in [19,20].

In this study, we evaluated and compared the overfitting characteristics of nonlinear equalizers based on the ANN and VSTF which were trained on a finite-length RRBS. We clarified that the overfitting characteristics of the ANN-based nonlinear equalizer were comparable to those of the VSTF when the number of hidden-layer units of the ANN was as large as 100 or 1000. However, when the number of hidden-layer units was 10, which is usually enough to compensate for optical nonlinear distortion, the overfitting was weaker than that of the VSTF.

The remainder of this paper is organized as follows: Section 2 summarizes the theory and computational complexity of ANN- and VSTF-based nonlinear equalizers. In Section 3, we explain the system setup for evaluating overfitting characteristics. Section 4 offers a comparison between the overfitting of the ANN and that of the VSTF. Finally, Section 5 provides the conclusion of this paper.

2. ANN- and VSTF-based Nonlinear Equalizers and Computational Complexity

2.1 ANN-based Nonlinear Equalizer

Figure 1(a) shows the construction of the ANN-based nonlinear equalizer used for optical nonlinear compensation

[12]. The ANN consists of three layers: an input layer, a hidden layer, and an output layer. Input signal $x(n)$ is fed to the input layer through a feedforward tapped delay line, where n represents the time index of the sampled signal with a sampling interval of T . $L = 2N+1$ expresses the tap length of the tapped delay line. $y(n)$ is the output signal of the ANN-based nonlinear equalizer. $x(n)$ and $y(n)$ are real values, while complex values are employed in [12]. This is because binary signals are used in this investigation of the overfitting. Therefore, we employed a real-valued ANN. Input-layer units simply distribute the input signal to the hidden-layer units. Figure 1(b) shows a hidden-layer unit used in the ANN. The inner potential of the j -th hidden-layer unit, $u_j(n)$, is described as

$$u_j(n) = \sum_{i=-N}^N w_{ji}^{(1)} x(n+i) + b_j^{(1)}, \quad (1)$$

where $w_{ji}^{(1)}$ is the weight between the i -th input-layer unit and the j -th hidden-layer unit, and $b_j^{(1)}$ is the bias. The units in the hidden layer have a sigmoid function expressed as

$$z_j(n) = \frac{1}{1 + e^{-u_j(n)}}, \quad (2)$$

where $z_j(n)$ is the output of the hidden-layer unit. The units in the output layer have a linear function. The output of the ANN-based nonlinear equalizer, $y(n)$, is described as

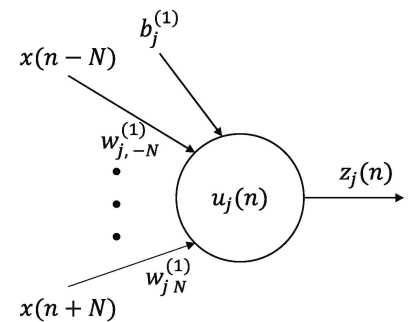
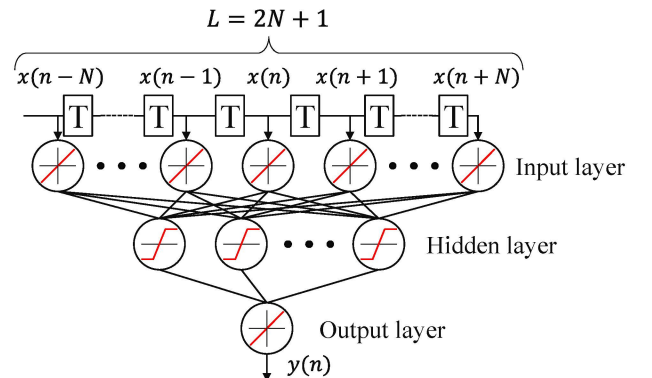


Fig. 1 ANN-based nonlinear equalizer and hidden-layer unit.

$$y(n) = \sum_{j=1}^M w_j^{(2)} z_j(n) + b^{(2)}, \quad (3)$$

where $w_j^{(2)}$ is the weight between the j -th hidden-layer unit and the output-layer unit, and $b^{(2)}$ is the bias. M is the number of hidden-layer units. We trained the ANN by using the error backpropagation (EBP) method, a type of least mean square (LMS) algorithm. We trained the ANN sample by sample. We did not use batches or minibatches. The error function is described as

$$e(n) = |y(n) - t(n)|^2, \quad (4)$$

where $t(n)$ is the ideal signal point at the time index n , namely a supervised signal. The error, $e(n)$, is minimized by updating the weights using the equation described as

$$\mathbf{w}(n+1) = \mathbf{w}(n) - \mu \frac{\partial e(n)}{\partial \mathbf{w}}, \quad (5)$$

where μ is the step size parameter which decides the learning speed and its stability. \mathbf{w} represents all the weights in the ANN. The number of hidden layer units required to compensate for SPM is about ten or less [13]. The required number of input layer units, which is equal to the number of taps of the tapped delay line, is decided by the amount of CD [13].

2.2 VSTF-based Nonlinear Equalizer

Figure 2 shows the VSTF-based nonlinear equalizer. Here, the Volterra kernels for the nonlinear compensation are acquired using the LMS algorithm. Optical nonlinearity of the optical fibers can be approximated by using only first- and third-order Volterra kernels [3,4]. We omitted second-order Volterra kernels, because it is known that the second-order terms of the VSTF are not effective in equalizing the optical-fiber nonlinearity. The output of the VSTF is expressed as

$$y(n) = \sum_{m_1=-N}^N h_{m_1} x(n-m_1) + \sum_{m_1=-N}^N \sum_{m_2=m_1}^N \sum_{m_3=-N}^N h_{m_1 m_2 m_3} x(n-m_1) x(n-m_2) x^*(n-m_3), \quad (6)$$

where $x(n)$ and $y(n)$ are the real-valued input and real-valued output of the VSTF at time index, n , respectively, h_{m_1} and $h_{m_1 m_2 m_3}$ are the first- and third-order Volterra kernels, respectively, and $L = 2N+1$ expresses the number of taps of the tapped delay line. If we use only first-order Volterra kernels, omitting third-order terms in Eq. (6), the equalizer is equivalent to an FIR filter.

2.3 Computational Complexity of ANN- and VSTF-based Nonlinear Equalizers

The number of multiplications required for the ANN-based nonlinear equalizer to compensate for a symbol is expressed as

$$M_{\text{ANN}} = L \times S_{\text{hidden}} + S_{\text{hidden}}, \quad (7)$$

where M_{ANN} is the number of real-valued multiplications, L is the number of taps of the tapped delay line, and S_{hidden} is the number of hidden-layer units [14,15]. Here, we neglect the calculations for the sigmoid functions of the hidden-layer units, assuming that a lookup table is employed. The number of real-valued multiplications required for a first-order VSTF (equivalent to an FIR filter) is expressed as

$$M_{\text{VSTF(1st, order)}} = L. \quad (8)$$

The number of real-valued multiplications per symbol of first- and third-order VSTF-based nonlinear equalizers can be expressed as

$$M_{\text{VSTF(1st,3rd order)}} = L + 3L^2(L+1)/2 = \frac{3}{2}L^3 + \frac{3}{2}L^2 + L, \quad (9)$$

where we eliminated the redundant terms, taking into account the symmetry of the Volterra kernels [14,15]. Figure 3 shows the number of multiplications of the equalizers

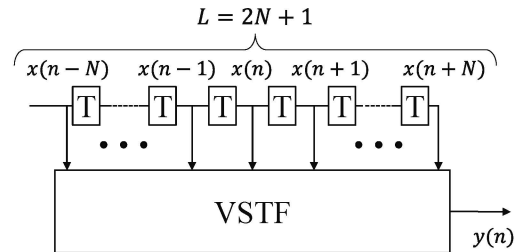


Fig. 2 VSTF-based nonlinear equalizer.

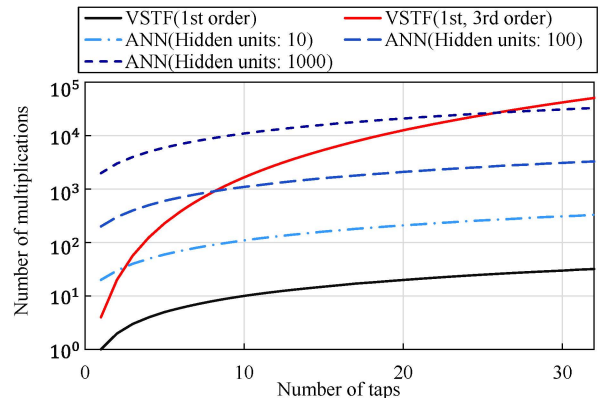


Fig. 3 Required number of multiplications versus the number of taps.

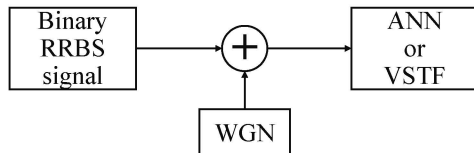


Fig. 4 Additive WGN channel with RRBS.

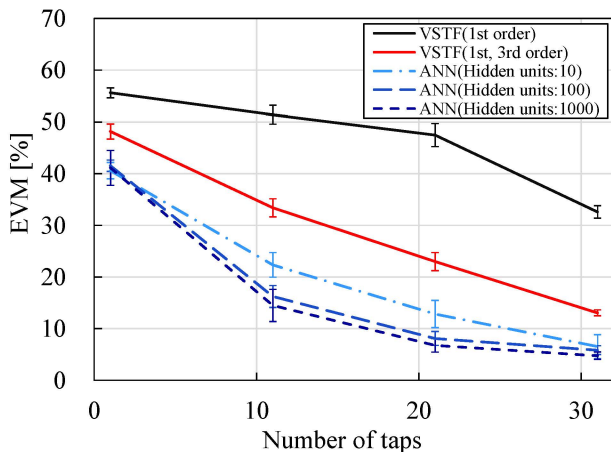


Fig. 5 EVM versus the number of taps. (Trained on 15-bit RRBS)

versus the number of taps. The number of multiplications in the ANN-based nonlinear equalizer increases linearly with the number of taps and hidden layer units. The number of multiplications in the first-order VSTF also increases linearly. On the other hand, for the first- and third-order VSTF, the number of multiplications increases in proportion to the cube of the number of taps. Therefore, if we need a long tapped delay line, the VSTF-based nonlinear equalizer will require significantly more multiplications than the ANN-based nonlinear equalizer.

3. System Setup for Evaluating Overfitting

Figure 4 shows the system setup used to evaluate the overfitting, which had been employed in previous studies on the overfitting evaluation of ANN-based nonlinear equalizers [16-18]. By employing this setup, we can simplify the evaluation to focus on the essential characteristics of the overfitting, eliminating the effects of the transmission parameters such as CD, SPM, pulse shape, and modulation formats. Even in actual transmission systems, the effects of the transmission parameters can be compensated by the equalizers, theoretically. Therefore, the essential characteristics of the overfitting are also applicable in actual transmission systems. A binary RRBS was generated by the Mersenne Twister (MT) algorithm. White Gaussian noise (WGN) was added to this binary baseband signal so that the signal-to-noise ratio (SNR) was adjusted to 4 dB. The bit lengths were changed from 15 to 31, 127, and 511 bits. The nonlinear equalizers were trained to try to “compensate” for the noise. The signal quality after the “compensation” was evaluated using the error vector magnitude (EVM). Essentially the noise cannot be

compensated for using the equalizers. When the overfitting occurs, however, the equalizers predict the next incoming signals, resulting in an improvement of the apparent EVM values. The numbers of hidden-layer units of the ANN were 10, 100, and 1000. As noted in Section 2.1, only about ten or fewer hidden layer units are enough to compensate for the fiber nonlinearity [13]. Nevertheless, we attempted to use as many as 100 or 1000 hidden layer units to evaluate the overfitting characteristics of the ANN-based nonlinear equalizers with a computational complexity comparable to that of the VSTF. We employed the first-order VSTFs and the first- and third-order VSTFs. In the training of the ANN and VSTF, we did not employ the techniques such as batch normalization, a dropout layer, and an early stopping algorithm. This approach was chosen to compare the overfitting characteristics of ANN and VSTF in the simplest condition. This simplicity of the training algorithm is important in high-speed optical communication systems. We trained the equalizers over 100,000 epochs, which we confirmed to be a sufficient number of epochs. Each epoch involved the training and test samples with different noise generated using different seeds. We used the same RRBS generated using one seed through the training over 100,000 epochs to observe the overfitting to the RRBS. The numbers of the training and test samples correspond to the bit length of the RRBS used. The learning rate was adjusted to minimize the average learning error for each combination of the number of taps, the number of hidden units, and RRBS length.

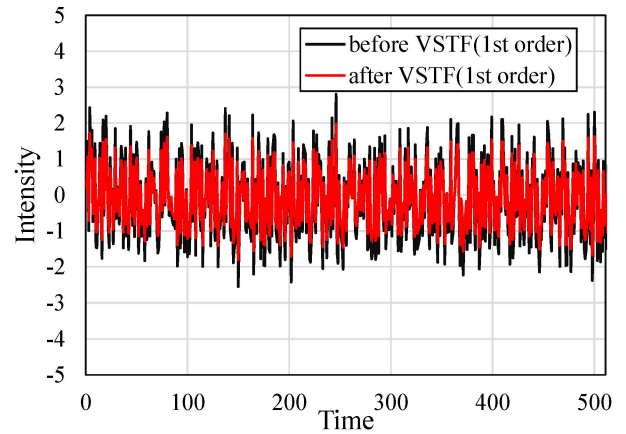
4. Results and Discussion

First, we evaluated the overfitting with a short RRBS of 15 bits, which is comparable to or shorter than the number of taps of the tapped delay line of the nonlinear equalizers. 15 bits is impractically short, and it is easily expected that strong overfitting is prone to occur. However, we performed this investigation using the short RRBS to evaluate the overfitting of the first-order VSTFs (equivalent to FIR filters). Figure 5 shows the EVM versus the number of taps of the first-order VSTF-based nonlinear equalizer when trained on the 15-bit RRBS. In the figure, the characteristics of the first- and third-order VSTF and ANN are also presented for comparison. We plotted the averages of ten trials of the training, with the error bars representing the standard deviation at each tap length of the equalizers. The RRBSs for the ten trials were generated using different seeds. In the case of the first-order VSTF with one tap, the equalizer simply multiplies the input signal by a Volterra kernel. Therefore, the equalizer does not change the EVM of the input signal with WGN, and the value was about 55%. It should be noted that the EVM was decreased by overfitting when we increased the number of taps of the first-order VSTF. When the number of taps was as large as 31, the EVM was decreased by about 23%. In the case of the first- and third-order VSTFs and ANNs, the EVM values were decreased to about 48% and 41%, respectively, even when

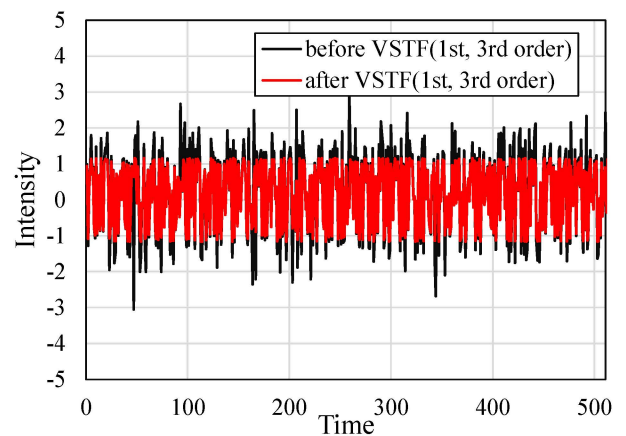
the number of taps was one. This is not due to the overfitting, but due to the clipping of WGN caused by the nonlinearity of the third-order terms of the VSTFs and the sigmoid functions of the ANNs.

Figure 6(a) shows the waveforms of the RRBSs with WGN before and after the first-order VSTF-based nonlinear equalizer with only one tap. As noted above, the equalizer simply multiplies the input signal by a Volterra kernel. Therefore, a linear relationship exists between the input and output waveforms. Figure 6(b) shows the waveforms before and after the first- and third-order VSTFs with one tap. In this case, we can observe that the amplitude of the WGN was clipped by the nonlinearity of the third-order terms of the VSTF. When the overfitting is evaluated by using the EVM, we have to take into account the effect of the clipping caused by the nonlinearity of the equalizers. Figure 6(c) shows the waveforms before and after the ANN with ten hidden-layer units and one tap. The saturation curve of the sigmoid functions of the hidden-layer units causes stronger clipping than the VSTF. Figure 6(d) shows the principle of the clipping caused by the nonlinearity of the equalizers. When the transfer function of the equalizer is nonlinear, the large amplitude of the input signal is clipped to some extent, according to the nonlinear curve of the function. The first- and third-order VSTF-based nonlinear equalizers caused this clipping due to the nonlinear operation in the second term of Eq. (6), whereas the ANN-based nonlinear equalizers caused the clipping due to the nonlinearity of the activation function. These clippings decreased the apparent EVM, as shown in Fig.5 and Fig.6 (b) and (c).

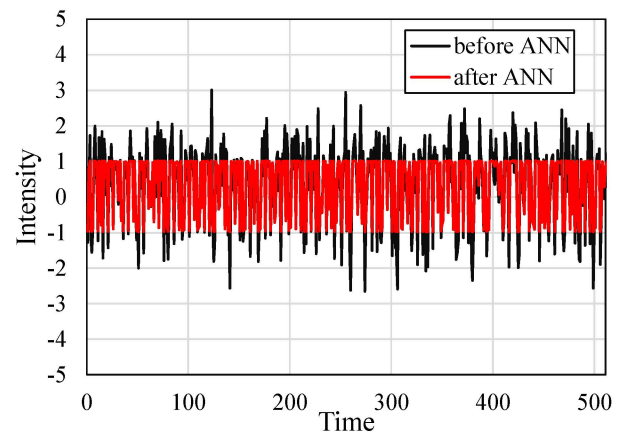
To eliminate the effects of the clipping, we plotted the variations in EVM, ΔEVM , from the value that was evaluated with one tap. Figure 7(a) is the replotted version of Fig. 5, showing the variations, ΔEVM , versus the number of taps of the VSTF- and the ANN-based nonlinear equalizers when trained on the 15-bit RRBS. In the case of the first-order VSTF, the EVM decreased by about 23% when the number of taps was 31, as mentioned above. When we used the first- and third-order VSTFs, the EVM decreased by about 35% with 31 taps, which shows larger overfitting than that which occurred in the case of the first-order VSTF. When we used the ANNs with 10, 100, and 1000 hidden-layer units, we observed stronger overfitting than observed with the VSTF. This result implies the high function representation capability of the ANN-based equalizers. However, when the number of taps was 31, the EVM decreased by about 35%, which was approximately equal to that of the first- and third-order VSTFs. This is due to the lower limit of the EVM, as shown in Fig. 5. Figure 7(b) shows ΔEVM versus the number of taps of the equalizers when trained on 31-bit RRBS. In the case of the first-order VSTF, the EVM decreased by 7% when the number of taps was 31. When we used the first- and third-order VSTFs, the EVM decreased by 27% with 31 taps. When we used the ANN with 10 hidden-layer units, the overfitting characteristics were comparable to those of the



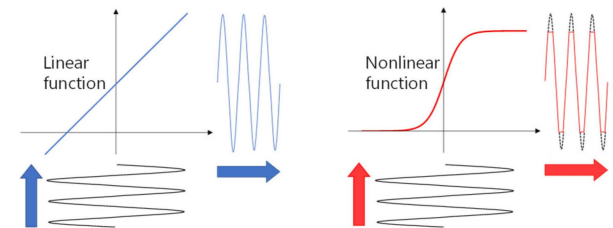
(a) First-order VSTF with one tap.



(b) First- and third-order VSTFs with one tap.



(c) ANN with one tap.



(d) Principle of clipping by nonlinearity.

Fig. 6 Clipping of WGN by nonlinearity of equalizers.

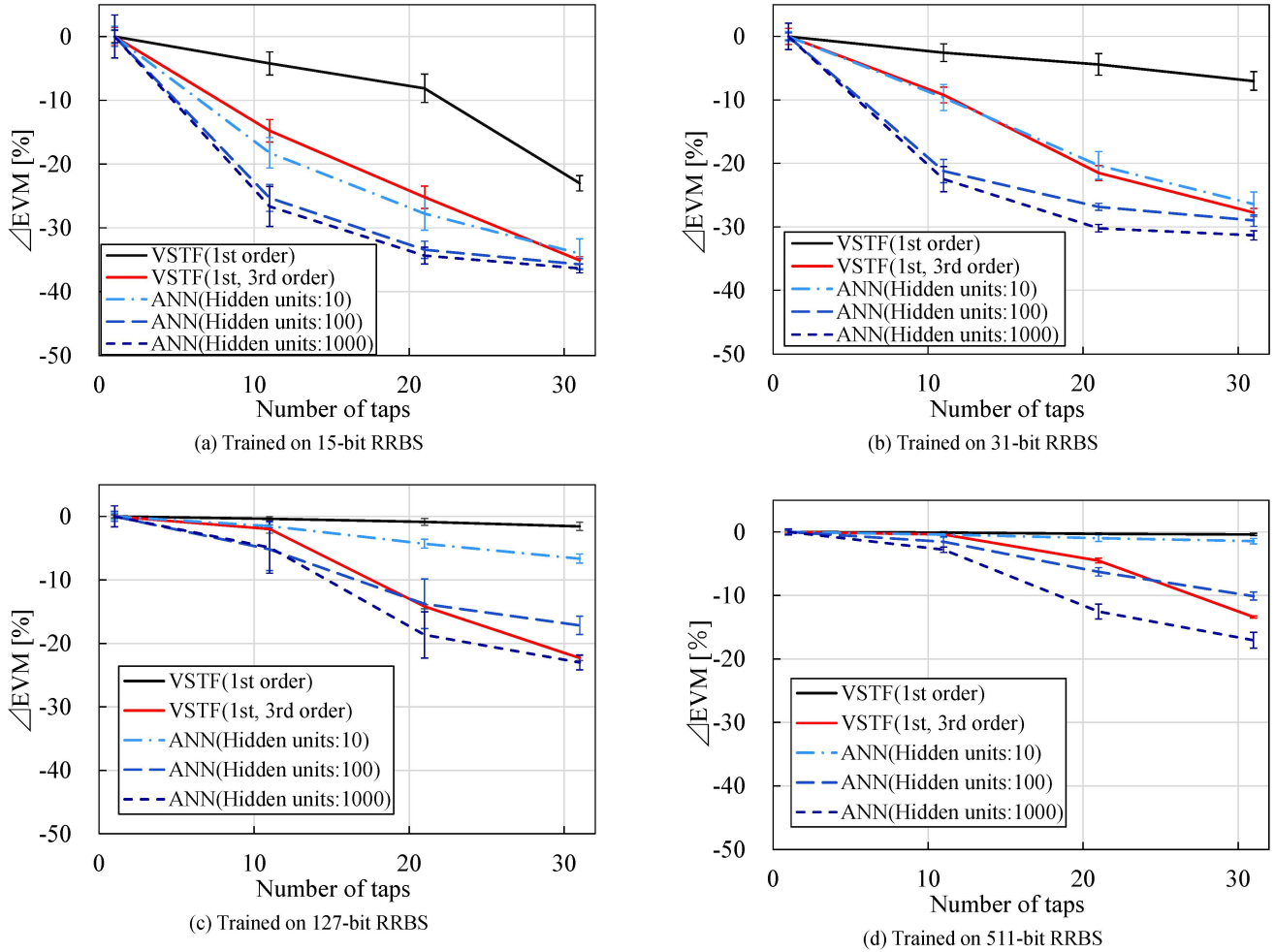


Fig. 7 Δ EVM versus the number of taps.

first- and third-order VSTFs. When we used the ANNs with 100 and 1000 hidden-layer units, we observed stronger overfitting than observed with the VSTF. This result shows the tendency toward weaker overfitting with an increase in the length of the RRBS used for the training. In order to investigate the overfitting characteristics with longer RRBS than the number of taps, we set the length to 127 bits. Figure 7(c) shows Δ EVM versus the number of taps of the equalizers which was trained on 127-bit RRBS. In the case of the first-order VSTF, EVM decreased by only 2% when the number of taps was 31, indicating the weak overfitting. When we used the first- and third-order VSTFs, the EVM decreased by 22% when the number of taps was 31. When we used the ANN with 10 hidden layer units, however, the EVM decreased by 7%, which is much smaller than that of the first- and third-order VSTFs. When we used the ANNs with 100 and 1000 hidden-layer units, the overfitting characteristics were comparable to those of the first- and third-order VSTFs. Figure 7(d) shows Δ EVM versus the number of taps when a 511-bit RRBS was employed for the training. In the case of the first-order VSTF, the EVM variation was about 0%, even when the number of taps was

as large as 31. When we used the first- and third-order VSTFs, the EVM decreased by 13%, when the number of taps was 31. On the other hand, when we used the ANN with 10 hidden-layer units, Δ EVM was only about 1%, even when the number of taps was as large as 31. In this case, the overfitting was suppressed enough, although we employed the ANN-based nonlinear equalizer. However, when we used the ANN and the number of hidden-layer units was as many as 100 and 1000, the overfitting characteristics were comparable to that of the first- and third-order VSTFs.

Figures 8(a) and (b) show the variations Δ EVM versus the bit length of the RRBS used for the training under the condition where the number of taps of the nonlinear equalizers was 31. First, we should note that the first-order VSTF, which is equivalent to an FIR filter, showed strong overfitting when the RRBS was as short as 31 or less. However, when the RRBS was longer than 127, the overfitting was sufficiently suppressed. In the case of the first- and third-order VSTFs, we observed strong overfitting, even when the RRBS was as long as 511. This result indicates that the first- and third-order VSTFs have a high function representation capability, and the VSTF-based

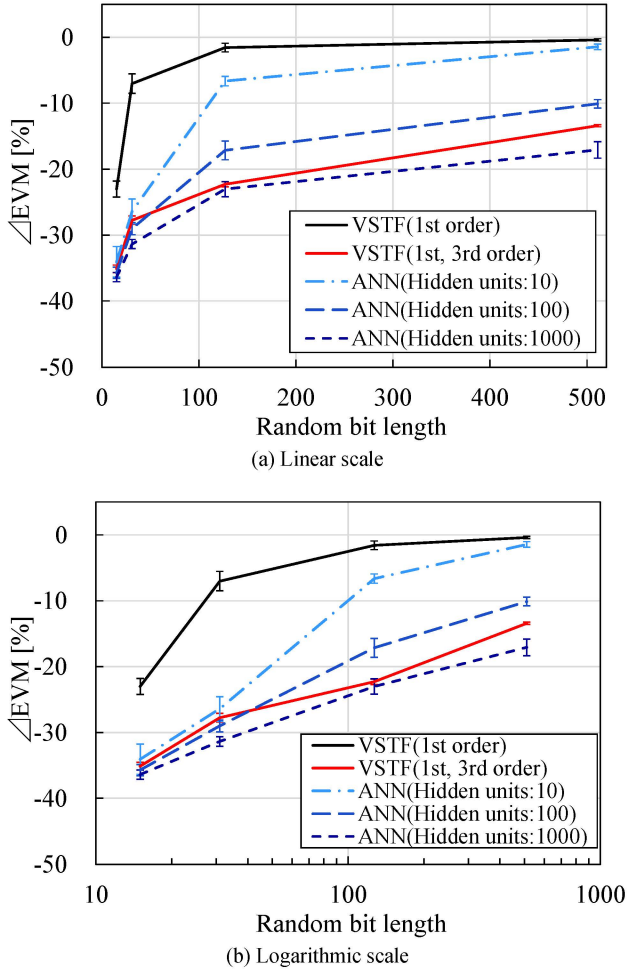


Fig. 8 Δ EVM versus bit-length of RRBS.

nonlinear equalizer memorized the trained RRBS. Consequently, the equalizer predicted the incoming RRBS, and the EVM decreased. The ANN-based nonlinear equalizers have a high function representation capability as good as one based on the VSTF. However, when the number of hidden-layer units was as small as 10, the Δ EVM was only about 1%, and the overfitting was sufficiently suppressed against the 511-bit RRBS, whereas the first- and third-order VSTF showed strong overfitting in the same condition. As mentioned in Section 2.1, only about ten or fewer hidden layer units are sufficient to compensate for the fiber nonlinearity [13]. It should be noted that the computational complexity of the ANN-based nonlinear equalizer is much smaller than that of the VSTF, as shown in Fig. 3. However, when we increased the number of hidden-layer units to more than required, namely, 100 or 1000, we observed strong overfitting similar to the case of the VSTF. The results indicate that we need to carefully consider the overfitting and the required number of hidden-layer units of ANN-based nonlinear equalizers. In [22], the overfitting characteristics of the ANN- and VSTF-based nonlinear equalizers were compared using PRBSs. In this case, both equalizers showed stronger overfitting than what

was observed in this study using RRBSs. This is because the ANN and VSTF can learn the simple generation rule of the PRBSs and consequently predict the received pattern. The overfittings of the nonlinear equalizers with RRBSs were weaker than that with PRBSs. In particular, when the number of the hidden-layer units of the ANN was as small as 10, the overfitting of the ANN was weaker than that of VSTF in the case of RRBSs.

5. Conclusion

We investigated the overfitting of ANN- and VSTF-based nonlinear equalizers trained on a finite-length RRBS. The results show that the VSTF used for nonlinear compensation in optical communication causes stronger overfitting than the ANN, depending on the conditions, in particular, the length of the RRBS and the number of taps. Nevertheless, it should be noted that we have to take care in deciding the number of hidden-layer units of the ANN. If we use more hidden-layer units than necessary, this will result in stronger overfitting. The problem of overfitting occurs not only with ANN-based nonlinear equalizers but also with general equalizers using learning algorithms. Depending on the conditions, the overfitting can occur even when we use a simple FIR filter.

Acknowledgments

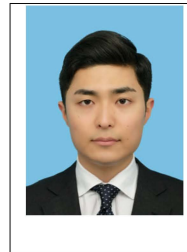
This work was supported by JSPS KAKENHI Grant Number 20K05367.

References

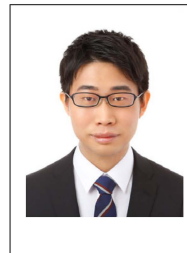
- [1] E. Ip and J. M. Kahn, "Compensation of dispersion and nonlinear impairments using digital backpropagation," *J. Lightw. Technol.*, vol. 26, no. 20, pp. 3416-3425, Oct. 2008. DOI: 10.1109/JLT.2008.927791
- [2] L. Zhu, F. Yaman, and G. Li, "Experimental demonstration of XPM compensation for WDM fibre transmission," *IET Electron. Lett.*, vol. 46, no. 16, pp. 1140-1141, Aug. 2010. DOI: 10.1049/el.2010.1444
- [3] Y. Gao, F. Zhang, L. Dou, Z. Chen, and A. Xu, "Intra-channel nonlinearities mitigation in pseudo-linear coherent QPSK transmission systems via nonlinear electrical equalizer," *Optics Commun.*, vol. 282, no. 12, pp. 2421-2425, Jun. 2009. DOI: 10.1016/j.optcom.2009.03.002
- [4] L. Liu, L. Li, Y. Huang, K. Cui, Q. Xiong, F. N. Hauske, C. Xie, and Y. Cai, "Intrachannel nonlinearity compensation by inverse Volterra series transfer function," *J. Lightw. Technol.*, vol. 30, no. 3, pp. 310-316, Feb. 2012. DOI: 10.1109/JLT.2011.2182038
- [5] J. Estarán, R. Rios-Müller, M. A. Mestre, F. Jorge, H. Mardoyan, A. Konczykowska, J.-Y. Dupuy, and S. Bigo, "Artificial neural networks for linear and non-linear impairment mitigation in high-baudrate IM/DD systems," ECOC2016, M.2.B.2, Sep. 2016.
- [6] S. Owaki, Y. Fukumoto, T. Sakamoto, N. Yamamoto, and M. Nakamura, "Experimental demonstration of SPM compensation based on digital signal processing using a three-layer neural-network for 40-Gbit/s optical 16QAM signal," *IEICE Commun. Express.*, vol. 7, no. 1, pp. 13-18, Jan. 2018. DOI: 10.1587/comex.2017XBL0148
- [7] M. A. Jarajreh, E. Giacomidis, I. Aldaya, S. T. Le, A. Tsokanos, Z. Ghassemlooy, and N. J. Doran, "Artificial neural network nonlinear equalizer for coherent optical OFDM," *IEEE Photon. Technol. Lett.*, vol.

- 27, no. 4, pp. 387-390, Feb. 2015. DOI: 10.1109/LPT.2014.2375960
- [8] V. Kamalov, L. Jovanovski, V. Vusirikala, S. Zhang, F. Yaman, K. Nakamura, T. Inoue, E. Mateo, and Y. Inada, "Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable," OFC2018, Th4D.5, Mar. 2018.
- [9] N. Kaneda, Z. Zhu, C.-Y. Chuang, A. Mahadevan, B. Farah, K. Bergman, D. V. Veen, and V. Houtsmas, "FPGA implementation of deep neural network based equalizers for high-speed PON," OFC2020, T4D.2, Mar. 2020.
- [10] P. J. Freire, M. Anderson, B. Spinnler, T. Bex, J. E. Prilepsky, T. A. Eriksson, N. Costa, W. Schairer, M. Blott, A. Napoli, S. K. Turitsyn, "Towards FPGA implementation of neural network-based nonlinearity mitigation equalizers in coherent optical transmission systems," ECOC2022, We1C.2, Sep. 2022.
- [11] K. Liu, E. Börjeson, C. Häger, and P. Larsson-Edefors, "FPGA implementation of multi-layer machine learning equalizer with on-chip training," OFC2023, M1F.4, Mar. 2023.
- [12] M. Nakamura, Y. Fukumoto, S. Owaki, T. Sakamoto, and N. Yamamoto, "Experimental demonstration of SPM compensation using a complex-valued neural network for 40-Gbit/s optical 16QAM signals," *IEICE Commun. Express*, vol. 8, no. 8, pp. 281-286, Aug. 2019. DOI: 10.1587/comex.2019XBL0043
- [13] M. Nakamura, Y. Fukumoto, and S. Owaki, "Size of an artificial neural-network for simultaneous compensation of linear and nonlinear optical waveform distortion," *IEICE Commun. Express*, vol. 8, no. 7, pp. 269-274, Jul. 2019. DOI: 10.1587/comex.2019XBL0049
- [14] Y. Otsuka, Y. Fukumoto, S. Owaki, and M. Nakamura, "Computational-complexity comparison of artificial neural network and Volterra series transfer function for optical nonlinearity compensation," OECC2018, P1-25, Jul. 2018.
- [15] T. Kyono, Y. Otsuka, Y. Fukumoto, S. Owaki, and M. Nakamura, "Computational-complexity comparison of artificial neural network and Volterra series transfer function for optical nonlinearity compensation with time- and frequency-domain dispersion equalization," ECOC2018, Th2.28, Sep. 2018.
- [16] T. A. Eriksson, H. Bülow, and A. Leven, "Applying neural networks in optical communication systems: possible pitfalls," *IEEE Photon. Technol. Lett.*, vol. 29, no. 23, pp. 2091-2094, Dec. 2017. DOI: 10.1109/LPT.2017.2755663
- [17] L. Shu, J. Li, Z. Wan, W. Zhang, S. Fu, and K. Xu, "Overestimation trap of artificial neural network: learning the rule of PRBS," ECOC2018, Tu4F.1, Sep. 2018.
- [18] C.-Y. Chuang, L.-C. Liu, C.-C. Wei, J.-J. Liu, L. Henrickson, C.-L. Wang, Y.-K. Chen, and J. Chen, "Study of training patterns for Employing deep neural networks in optical communication systems," ECOC2018, Tu4F.2, Sep. 2018.
- [19] J. Kim and H. Kim "Length of pseudorandom binary sequence required to train artificial neural network without overfitting," *IEEE Access*, vol. 9, pp. 125358-125365, Sep. 2021. DOI: 10.1109/ACCESS.2021.3111092
- [20] P. J. Freire, A. Napoli, B. Spinnler, N. Costa, S. K. Turitsyn, and J. E. Prilepsky, "Neural networks-based equalizers for coherent optical transmission: caveats and pitfalls," *IEEE J. Sel. Top. Quantum Electron.*, vol. 28, no. 4, art. seq. no. 7600223, Jul./Aug. 2022. DOI: 10.1109/JSTQE.2022.3174268
- [21] J. Nakamura, K. Ikuta, and M. Nakamura, "Overfitting characteristics of four-layer-deep-neural-network-based nonlinear equalizer for optical communication systems," *IEICE Commun. Express*, vol. 11, no. 7, pp. 368-373, Jul. 2022. DOI: 10.1587/comex.2022XBL0035
- [22] K. Ikuta, Y. Otsuka, Y. Fukumoto, and M. Nakamura, "Overestimation problem with ANN and VSTF in optical communication systems," *IET*

Electron Lett., vol. 55, no. 19, pp. 1051-1053, Sep. 2019. DOI: 10.1049/el.2019.2008



Kai Ikuta received the B.S. and M.S. degrees in Electrical Engineering from Meiji University, Kanagawa, Japan, in 2019 and 2021, respectively. In 2022, he became Research Associate with Meiji University, and is pursuing the D.E. degree. His research interests include optical communications and machine learning.



Jinya Nakamura received the B.S. and M.S. degrees in Electrical Engineering from Meiji University, Kanagawa, Japan, in 2021 and 2023, respectively. His research interests include optical communications and machine learning.



Moriya Nakamura received the B.E. and M.E. degrees in Electrical Engineering from Tokyo University of Science, Japan, in 1991 and 1993, respectively. He received the Ph. D. degree in Electrical Engineering from Osaka University, Japan in 2003. In 1993, he joined the Communications Research Laboratory, Ministry of Posts and Telecommunication (currently National Institute of Information and Communications Technology (NICT)). From 2013, he has been an Associate Professor with School of Science and Technology, Meiji University. He has been engaged in research and development of DSP-based linear and nonlinear equalization, coherent optical transmission, optical multi-level modulation, fiber-optic CDMA, and optical interconnection. He has served as a Program Officer of Basic Civilian Technologies in Telecommunications Advancement Organization (TAO), and as a Program Officer of SCOPE in the Ministry of Internal Affairs and Communications. He is a member of IEICE and IEEE.