End-to-end Deep Learning for VCSEL’s Nonlinear Digital Pre-Distortion

(Article begins on next page)

End-to-end Deep Learning for VCSEL's Nonlinear Digital Pre-Distortion

Leonardo Minelli
Politecnico di Torino (DET)
Torino, Italy
leonardo.minelli@polito.it

Fabrizio Forghieri
CISCO Photonics Italy S.r.l.
Vimercate, Italy

Roberto Gaudino
Politecnico di Torino (DET)
Turin, Italy

***Abstract*—We propose a novel optimization method for a Neural Network based Digital Pre-Distorter (DPD), applied in Intensity Modulation-Direct Detection transmission systems leveraging Multi-Modal Fiber and Vertical-Cavity Surface-Emitting Laser. We train the DPD using End-to-end Deep Learning of the optical link, together with a Direct Learning Approach leveraging experimental measurements for modeling the transmission channel. The optimization considers VCSEL amplitude constraints, the use of an FFE at the receiver side, and the presence of a receiver non-flat Colored Gaussian Noise (CGN). We verify our optimized DPD on an experimental setup transmitting a 92 Gbps PAM-4 modulated signal. We achieve, for BER=0.01, a performance gain of more than 1 dB in terms of Optical Path Loss with respect to the best performing non-pre-distorted scenario.**

***Index Terms*—Deep learning applications on communication systems, nonlinear equalization, VCSEL, optical PAM-4 IM-DD systems, Data Center Interconnects**

I. Introduction

Emerging applications such as the Internet of Things, Cloud Computing and ultra-high-definition multimedia streaming require communication networks with very high data rates. Specifically, in Data Centers Interconnects (DCI) the short-term goal is to increase their link capacity beyond 100 $Gbps/\lambda$ using the same hardware that now allows them to transmit up to 28 Gbps per lane. Current solutions leverage Intensity Modulation and Direct Detection (IM-DD) optical links, using binary On-Off Keying (OOK) modulation around 100 meters of optical fiber. Around 50% of optical links are using Multi-Modal Fibers (MMF) [1], mostly paired to Vertical-Cavity Surface-Emitting Lasers (VCSEL). These optical sources have a low cost per chip production, high power efficiency, and long device lifetime. However, pushing these MMF-VCSEL links to the extreme of their data rate capabilities leads to linear and nonlinear distortions (i.e., bandwidth limitations and VCSEL nonlinearities), that severely affect the transmitted signals. To thus counteract these impairments, the use of higher-order modulation formats such as Pulse Amplitude Modulation (PAM) and Digital Signal Processing (DSP) techniques seems to be the most feasible solution. In particular, nonlinear equalizers have been investigated for VCSEL-MMF IM-DD systems, with a focus on PAM-4 modulated signals [2]. These DSP devices can be adopted at the receiver (RX) as post-equalizers, implemented in several forms such as Volterra-series nonlinear equalizers (VNLE) [3], Support Vector Machines (SVM) [4] and Artificial Neural Networks (ANN) [5]. Nonlinear equalization can be moreover applied at the transmitter (TX), where the devices are called Digital Pre-Distorters (DPD). In short-reach IM-DD optical links, the usage of DPDs is favored with respect to nonlinear post-equalization, as DSP algorithms are more easily implementable at TX with respect to RX. Several technologies have been proposed for nonlinear DPD, such as VNLE [6] [7], Look-up Tables (LUT) [8] [9] and ANN [10] [11] [12] [13]. Recent attention has been posed on determining the best approach to optimize these devices: solutions based on Indirect Learning (ILA) [6] and Direct Learning Architectures (DLA) [7] [13] [11] are the mostly adopted (having the latter proved to give better performances [13]), but recently have also been proposed techniques based on reinforcement learning [12] and End-to-end (E2E) deep learning of the optical transmission system [14] [15]. The latter is supposed to be the optimal approach since it involves the joint training of a DPD at TX together with a post-equalizer at RX: in this way it should be possible to achieve the best absolute performances over an optical communication link. However, optimization of nonlinear DPD for VCSEL-MMF IM-DD systems requires considering amplitude constraints at the Digital-to-Analog Converter (DAC) output and VCSEL input, as well as noise with a non-flat power spectral response at the receiver. In this paper, we propose a novel E2E neural network, based on FIR-based Neural Networks (FIRNN), that is able to optimize a nonlinear DPD natively taking into account a VCSEL input amplitude constraint, an RX non-white Gaussian noise, and a linear Feed-Forward Equalizer at RX. We evaluate the performances of our DPD on an experimental setup, transmitting a 92 Gbps PAM-4 signal transmitted with $\bar{P}_{TX} = 5$ mW, and using a VCSEL with $B_{3dB} = 20$ GHz. Finally, we compare the results with respect to using only FFE at RX, assessing the achieved gain.

II. The proposed End-to-end system

The optimization of the ANN-based nonlinear DPD leverages the use of the End-to-end system illustrated in Fig. 1. This architecture is a neural network composed of the cascade of an ANN TX (i.e., the DPD), an ANN model for the optical transmission channel (CH), and an ANN RX (i.e., an FFE). The E2E system, working at a sample-per-symbol

Identify applicable funding agency here. If none, delete this.

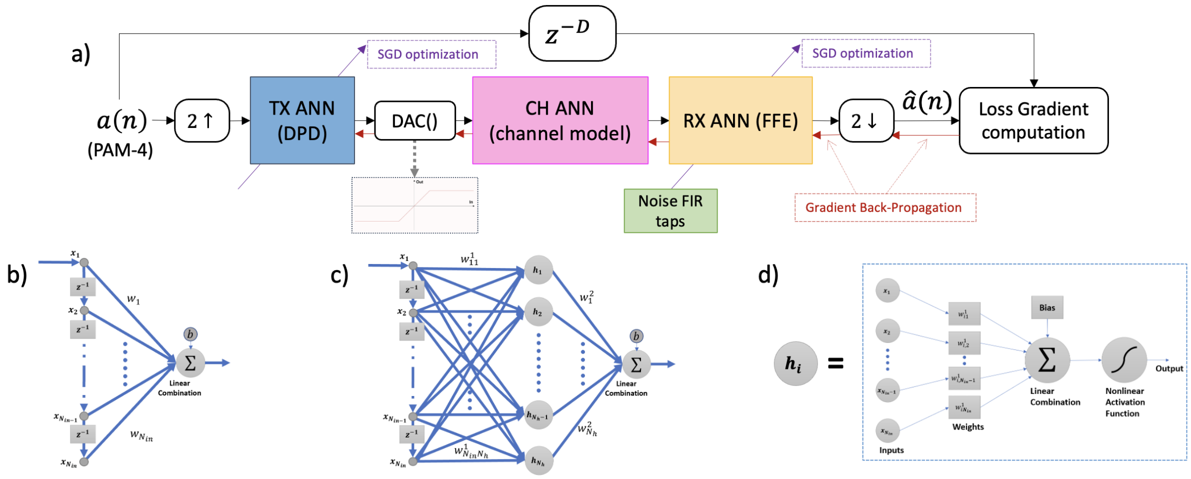

Fig. 1. a) The End-to-end system optimization scheme; b) structure of the RX ANN (i.e., a FFE); c) structure of the TX ANN and the CH ANN; d) structure of an hidden nonlinear neuron (present in TX and CH). TX and CH can be equivalently be viewed as FIR-based Neural Networks, with synapses extended to FIR filters only in their first layer.

(sps) ratio equal to 2, emulates the behavior of the considered communication link in the form of an autoencoder [16]. The E2E optimization consists thus of training the whole E2E neural network, keeping fixed the ANN CH coefficients, to jointly optimize the DPD and the post-equalizer. To properly take into account the memory of the channel, the DPD, as well as the ANN CH, is modeled as a FIRNN: this kind of ANN is characterized to generalize their "synapses" (i.e., the branches from a node to another) to be FIR filters. In our case, for the DPD and the channel model, the branches are FIR-based only in the first layer: this is equivalent to modeling them as a Feed-Forward Neural Network whose inputs are the successive samples stored in a digital delay line (Fig. 1.c). Likewise, the FFE can be viewed as a simplified FIRNN without nonlinearities. As a consequence, the whole E2E system is a FIRNN itself, whose loss gradients can be computed by exploiting the temporal backpropagation algorithm proposed in [17]. Moreover, to impose an amplitude peak-to-peak constraint relative to the VCSEL input and DAC output dynamics, we model the DAC as a hard-limiter function, applied as output nonlinear activation of the DPD. In addition, to include the effects of a Colored Gaussian noise (CGN) at RX, we assume introducing at the FFE input the outcome of a White Gaussian Noise (WGN) filtered by an FIR filter, whose taps are modeled according to the desired noise spectral behavior.

III. Experimental optimization of the DPD

We applied experimentally our DPD optimization on a setup whose schematics are illustrated in Fig.2. The procedure consists of the following steps:

1) Acquisition of the RX signal: On the experimental transmission setup, we acquire more than 1000 periods of a 46 GBaud PAM-4 modulated PRBS sequence $a[n]$ (period 2^{15}). After TX shaping using a Gaussian filter with order 2 and $B_{3dB} = 0.75 \cdot R_s$ (where R_s is the Baud rate), the resulting signal $x[n]$ is transmitted from an Arbitrary

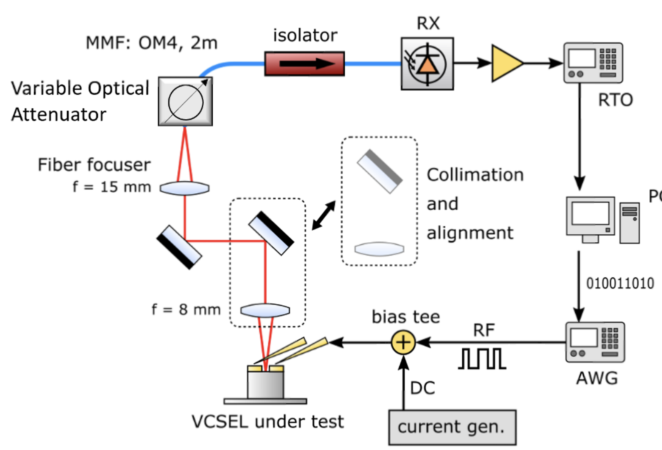

Fig. 2. The experimental VCSEL-MMF IM-DD transmission channel

Waveform Generator (AWG) with sampling rate equal to 92 GSample/s. The AWG is driven by imposing a certain peak-to-peak modulation voltage to control the VCSEL input swing. After bias current addition, the signal is injected into an 850 nm VCSEL, which emits light in free space. The optical waveform is guided into an OM4 fiber through a free-space lens setup. The signal is then attenuated through an Optical Variable Attenuator, and after passing through 2 meters of OM4 fiber and an isolator (to avoid backward reflections) it is converted into the electric domain through a PIN Photodiode RX. Finally, the sequence is digitally acquired through a Real-Time Oscilloscope with a sampling frequency equal to 200 GSample/s.

2) RX signal denoising and experimental noise retrieval: After the acquisition, the RX signal $y[n]$ is averaged over the repetitions of the PRBS period, minimizing the noise impairments. The resulting noiseless signal $\bar{y}[n]$ is then subtracted from the original acquired signal, obtaining a signal

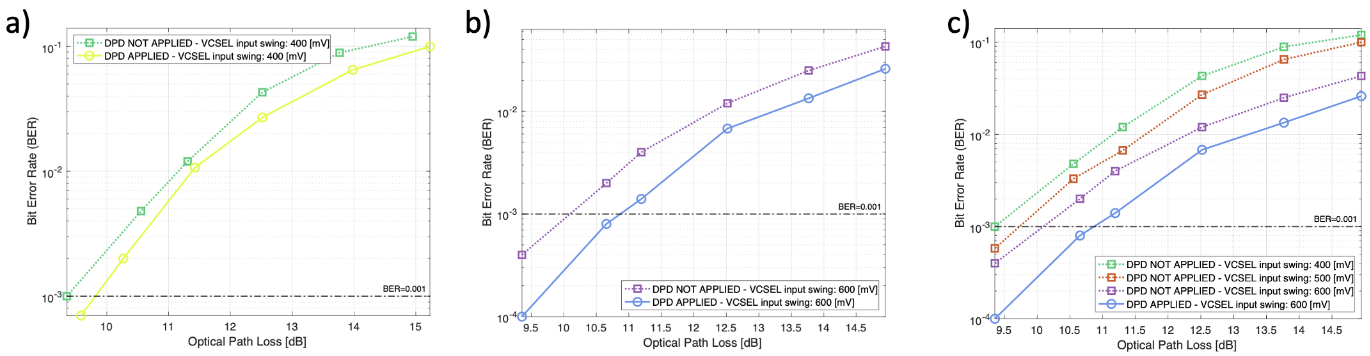

Fig. 3. BER vs OPL performances on the experimental setup, with and without applying a ANN-based Digital Pre-Distorter, transmitting PAM-4 symbols at 92 Gbps: a) VCSEL driven in linear condition; b) VCSEL driven in nonlinear condition; c) overall comparison between best DPD performances and non predistorted scenarios with different VCSEL input swings.

$r[n] = y - \bar{y}[n]$ which represents a stochastic realization of the experimental noise.

3) ANN channel and RX noise modelization: The experimental setup from the AWG to the RTO is directly modeled by the ANN channel model, which is trained according to the DLA illustrated in [13]: the sequence $x[n]$ is used for the input training examples, and the sequence $y[n]$ for the output labels. Since the ANN is working at a 2 sample-per-symbol ratio, the two sequences are resampled accordingly. The experimental RX noise instead is modeled as a White Gaussian random process, passed through an FIR filter whose taps are modeled so that the magnitude response fits the Powers Spectral Density of the signal $r[n]$.

4) End-to-end optimization: After obtaining the ANN channel model and the RX noise FIR filter, we fix the retrieved coefficients in the E2E system, to jointly train the DPD and the FFE. The E2E optimization consists of a Stochastic Gradient Descend algorithm to minimize the Mean Square Error (MSE) between a random PAM-4 sequence $a[n]$ (different from the one used during experimental acquisition to avoid overfitting issues) given in input to the E2E system and the produced output. The MSE is computed after delaying $a[n]$ by D symbols, being D the delay introduced by the E2E system. In our method, the RX Colored Gaussian Noise (CGN) is not directly injected in the time domain at the RX FFE input, but its effect is introduced analytically as an additive regularization term in the FFE SGD update [18]: we verified that this reduces the variance of the back-propagated gradient, facilitating the convergence of the system. Moreover, as the amplitude constraints tend to penalize the PAM-4 outer levels (i.e., overshoots for counteracting bandwidth limitations cannot exceed the dynamics), in the MSE loss computation more weight is given to the external symbols. Therefore, the loss is computed as follows:

$$Loss = \begin{cases} (a[n-D] - \hat{a}[n])^2 & |a[n-D]| = 1 \\ h \cdot (a[n-D] - \hat{a}[n])^2 & |a[n-D]| = 3 \end{cases} \tag{1}$$

where h is an gain factor that has been heuristically found to be optimal when set to 50.

Concerning the optimization hyperparameters, for simplicity in our experiments every FIR filter in the E2E system (i.e., belonging to TX and CH ANN, FFE, CGN FIR) has been set to have 31 taps. The same number has been used for the number of hidden neurons in the ANN, where to keep low the DPD complexity the number of hidden layers has been set to 1 (with ReLU activation function). Moreover, 1e5 samples with a learning rate (lr) equal to 0.001 have been used in the channel modelization, while 3e5 samples and lr=0.0001 have been used during the E2E optimization.

IV. Experimental Results

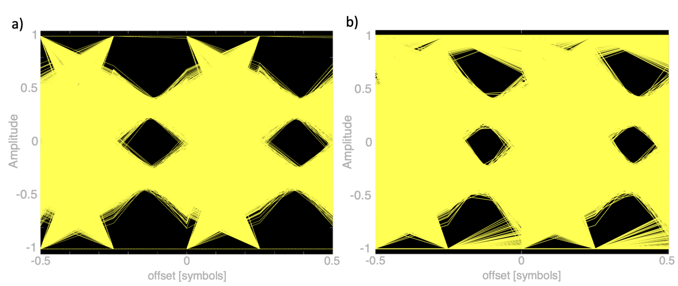

Fig. 4. Digitally Pre-Distorted transmitted PAM-4 signal a) Driving the laser in a linear condition (VCSEL input swing = 400 mV) b) Driving the laser in a nonlinear condition (VCSEL input swing = 600 mV)

To assess the performances of our DPD, we trained the structure for several values of Optical Path Loss (OPL) adopting 2 different VCSEL input swing constraints. First, we imposed through the AWG a VCSEL input swing equal to 400 mV (i.e., to drive the laser in a linear condition). Secondly, we augmented the dynamics to 600 mV, to operate with the VCSEL driven in a nonlinear condition. In Fig. 4 an eye-diagram related to the pre-distorted signal obtained in the two conditions (for OPL$\sim$ 9.5 dB) is illustrated. It can be observed that when the DPD is trained to bound the output in a linear region (Fig.4.a), it pre-equalizes the internal PAM-4 levels with time-domain overshoots (i.e., to compensate for the low-pass system response): in the outer levels instead are modulated with almost square pulses. When predistortion is applied in a nonlinear laser condition (Fig.4.b), the DPD seems still to produce overshoots in the inner levels, but the predistorted eye is visibly asymmetric: in this condition, the nonlinear skew caused by the VCSEL is indeed pre-compensated by the ANN.

In both situations, we applied the DPD on a PAM-4 modulated PRBS sequence different from the one used in the training acquisition (period 2^{16}), and after transmission through the experimental setup we post-processed the acquired signal using an FFE and evaluated the BER. For comparison we performed the experiment also shaping the sequence with a Gaussian filter (the same used for training), keeping the same amplitude constraint for both the pre-distorted and non-predistorted signals. The BER vs OPL experimental curves are illustrated in Fig. 3.a and 3.b. It can be observed how applying nonlinear DPD gives a performance gain ($\sim$0.3 dB of OPL for BER=0.001) even in a linear lasing condition (Fig. 3.a). The performance gain is however enhanced when the laser is driven with a higher input swing, as the DPD compensates for the nonlinear distortions (Fig. 3.b). Finally, in Fig. 3.c a comparison is illustrated between the best DPD and the non-pre-distorted scenario for different VCSEL input swings. It can be noticed that the performances using the nonlinear DPD are better than any non-pre-distorted scenario, with an improvement of at least 1 dB of OPL for BER=0.01 and 0.7 dB for BER=0.001. Without DPD, as the input swing is increased the OMA is higher (i.e., +1.76 dB moving from 400 mV to 600 mV if the VCSEL was linear), and so is the signal-to-noise ratio. However, for BER=0.001 the gain in OPL doesn't increase accordingly (0.7 dB from 400 to 600 mV) since the nonlinear distortions become more relevant: our proposed DPD can be thus applied to overcome this limitation.

V. Conclusion and discussion

In this paper, we showed a new End-to-end system to optimize a nonlinear DPD for a VCSEL-MMF IM-DD link. The proposed solution is able to natively satisfy the VCSEL and DAC amplitude constraints, together with assessing the effects of a non-flat Gaussian noise at the receiver input as an additive regularization term in the FFE optimization. The obtained DPD provides a consistent gain in terms of BER vs OPL performances, even if compared with a non-pre-distorted signal having the same amplitude peak-to-peak swing. The illustrated method is however properly applicable only when the DPD works with an sps ratio equal to 2, limiting the Baud Rate to half of the DAC sampling frequency: further study is thus required to extend this technique at different bit rates and sps ratios.

Acknowledgment

This work was carried out under a research contract with Cisco Photonics. We also acknowledge the PhotoNext initiative at Politecnico di Torino (http://www.photonext.polito.it/) and its laboratory, where all experiments have been performed.

References

[1] J. King, "In Support of 200G MMF Ethernet PMDs," IEEE 802.3 Next-generation 200 Gb/s and 400 Gb/s MMF PHYs Study Group. [Online]. Available: //www.ieee802.org/3/NGMMF/public/Jan18/young_NGMMF_01a_jan18.pdf

[2] H. Zhou, Y. Li, Y. Liu, L. Yue, C. Gao, W. Li, J. Qiu, H. Guo, X. Hong, Y. Zuo, and J. Wu, "Recent Advances in Equalization Technologies for Short-Reach Optical Links Based on PAM4 Modulation: A Review," *Applied Sciences*, vol. 9, no. 11, 2019. [Online]. Available: https://www.mdpi.com/2076-3417/9/11/2342

[3] Y. Gao, C. Yang, J. Wang, X. Qin, H. Guo, X. Zhang, C.-C. Shen, H. Li, Z. Chen, and C. J. Chang-Hasnain, "288 Gb/s 850 nm VCSEL-based Interconnect over 100 m MMF based on Feature-enhanced Recurrent Neural Network," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, 2022, pp. 01–03.

[4] A. Liang, C. Yang, C. Zhang, Y. Liu, F. Zhang, Z. Zhang, and H. Li, "Experimental study of support vector machine based nonlinear equalizer for VCSEL based optical interconnect," *Optics Communications*, vol. 427, pp. 641–647, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030401818305844

[5] Q. Zhou, C. Yang, A. Liang, X. Zheng, and Z. Chen, "Low computationally complex recurrent neural network for high speed optical fiber transmission," *Optics Communications*, vol. 441, pp. 121–126, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0030401819301452

[6] V. Shivashankar, C. Kottke, V. Jungnickel, and R. Freund, "Investigation of Linear and Nonlinear Pre-Equalization of VCSEL," in *Broadband Coverage in Germany; 11. ITG-Symposium*, 2017, pp. 1–5.

[7] V. Bajaj, M. Chagnon, S. Wahls, and V. Aref, "Efficient Training of Volterra Series-Based Pre-distortion Filter Using Neural Networks," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*, 2022, pp. 1–3.

[8] J. Zhang, P. Gou, M. Kong, K. Fang, J. Xiao, Q. Zhang, X. Xin, and J. Yu, "PAM-8 IM/DD Transmission Based on Modified Lookup Table Nonlinear Predistortion," *IEEE Photonics Journal*, vol. 10, no. 3, pp. 1–9, 2018.

[9] Z. He, K. Vijayan, M. Mazur, M. Karlsson, and J. Schröder, "Look-up Table based Pre-distortion for Transmitters Employing High-Spectral-Efficiency Modulation Formats," in *2020 European Conference on Optical Communications (ECOC)*, 2020, pp. 1–4.

[10] Q. Zhang, Z. Wang, S. Duan, N. Jiang, B. Cao, and Y. Wu, "An Improved End-to-end Optical Transmission System Based On Deep Learning," in *2021 19th International Conference on Optical Communications and Networks (ICOCN)*, 2021, pp. 1–3.

[11] V. Bajaj, F. Buchali, M. Chagnon, S. Wahls, and V. Aref, "Deep Neural Network-Based Digital Pre-Distortion for High Baudrate Optical Coherent Transmission," *Journal of Lightwave Technology*, vol. 40, no. 3, pp. 597–606, 2022.

[12] J. Song, Z. He, C. Häger, M. Karlsson, A. G. i. Amat, H. Wymeersch, and J. Schröder, "Over-the-fiber Digital Predistortion Using Reinforcement Learning," in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.

[13] G. Paryanti, H. Faig, L. Rokach, and D. Sadot, "A Direct Learning Approach for Neural Network Based Pre-Distortion for Coherent Nonlinear Optical Transmitter," *Journal of Lightwave Technology*, vol. 38, no. 15, pp. 3883–3896, 2020.

[14] B. Karanov, M. Chagnon, F. Thouin, T. A. Eriksson, H. Bülow, D. Lavery, P. Bayvel, and L. Schmalen, "End-to-End Deep Learning of Optical Fiber Communications," *Journal of Lightwave Technology*, vol. 36, no. 20, pp. 4843–4855, 2018.

[15] V. Neskorniuk, A. Carnio, V. Bajaj, D. Marsella, S. K. Turitsyn, J. E. Prilepsky, and V. Aref, "End-to-End Deep Learning of Long-Haul Coherent Optical Fiber Communications via Regular Perturbation Model," in *2021 European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.

[16] T. O'Shea and J. Hoydis, "An Introduction to Deep Learning for the Physical Layer," *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 563–575, 2017.

[17] E. Wan, "Temporal backpropagation for FIR neural networks," in *1990 IJCNN International Joint Conference on Neural Networks*, 1990, pp. 575–580 vol.1.

[18] C. M. Bishop, "Training with Noise is Equivalent to Tikhonov Regularization," *Neural Computation*, vol. 7, no. 1, pp. 108–116, 1995.