Zero-Multiplier Sparse DNN Equalization for Fiber-Optic QAM Systems with Probabilistic Amplitude Shaping

Toshiaki Koike-Akino[(1)], Ye Wang[(1)], Keisuke Kojima[(1)], Kieran Parsons[(1)], Tsuyoshi Yoshida[(2)]

[(1)] Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA 02139, USA. koike@merl.com
[(2)] Information Technology R&D Center, Mitsubishi Electric Corporation, 5-1-1 Ofuna, 247-8501, Japan.

Abstract *We propose a multiplier-less deep neural network (DNN) to mitigate fiber-nonlinear distortion of shaped constellations. Our DNN achieves an excellent performance-complexity trade-off with progressive lottery ticket hypothesis (LHT) weight pruning and additive powers-of-two (APoT) quantization.*

Introduction

Deep neural network (DNN) has been recently investigated for the next-generation fiber-optic communications, e.g., for nonlinear compensation[1]–[5] and end-to-end design[6]–[11]. Although a high potential of DNN has been successfully demonstrated in literature, DNN generally requires high computational complexity and high power operation for high-speed real-time processing. In this paper, we propose a hardware-friendly DNN framework for nonlinear equalization. Our DNN realizes multiplier-less operation based on powers-of-two quantization[12]–[14]. We demonstrate that quantization-aware training (QAT) for additive powers-of-two (APoT) weights can fully eliminate multipliers without causing any performance loss (but slight improvement). In addition, we introduce weight pruning based on the lottery ticket hypothesis (LTH)[15]–[17] to sparsify the over-parameterized DNN weights for further complexity reduction. We verify that progressive LTH pruning can prune more than 99% of the weights, yielding power-efficient equalization applicable to high-throughput communications.

DNN Equalization for Probabilistic Amplitude Shaping (PAS) QAM Systems

The optical communications system under consideration is depicted in Fig. 1. Eleven-channel dual-polarization quadrature-amplitude modulations (DP-QAM) for 34 GBaud and 35 GHz spacing are sent over fiber plants towards coherent receivers. We consider N spans of dispersion unmanaged links with 80 km standard single-mode fiber (SSMF). The SSMF has a dispersion parameter of $D = 17$ ps/nm/km, a nonlinear factor of $\gamma = 1.2$ /W/km, and an attenuation of 0.2 dB/km. The span loss is compensated by Erbium-doped fiber amplifiers (EDFA) with a noise figure of 5 dB, where total noise is added before the receiver for simplicity. We use digital root-raised cosine filters with 2% rolloff at both transmitter and receiver. The receiver employs standard digital signal processing with symbol synchronization, carrier-phase recovery, dispersion compensation, and polarization recovery with 61-tap linear equalization (LE).

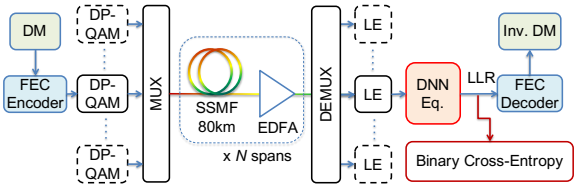


Fig. 1: Optical fiber transmission with shaped QAM.

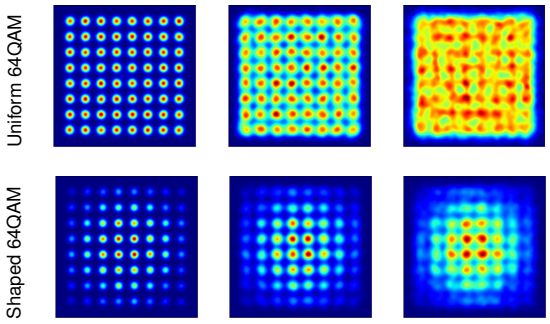


Fig. 2: Nonlinear distortion of DP-64QAM.

Due to fiber nonlinearity, residual distortion after LE will limit the achievable information rates. Fig. 2 shows a sample of distorted DP-64QAM constellation after least-squares LE for 1-/10-/20-span transmissions. Here, we compare uniform QAM and shaped QAM, which uses distribution matcher (DM) for probabilistic amplitude shaping (PAS) following Maxwell–Boltzmann distribution; $\Pr(x_i) \propto \exp(-\lambda |x_i|^2)$ with $\lambda = 2$. We can observe that the shaped constellation is more distinguishable as the Euclidean distance is increased with a reduced entropy (11.51 b/s/4D symbol).

978-0-7381-4679-9/21/$31.00 ©2021 IEEE

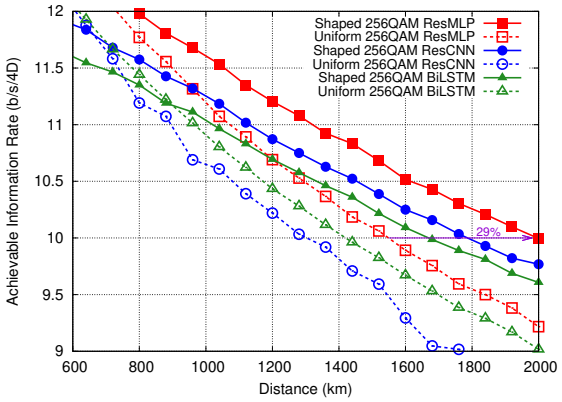
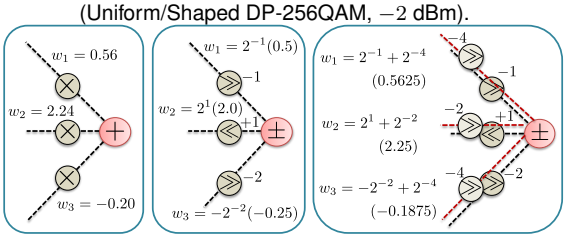


Fig. 3: Rate vs. distance for various DNN equalizers (Uniform/Shaped DP-256QAM, -2 dBm).



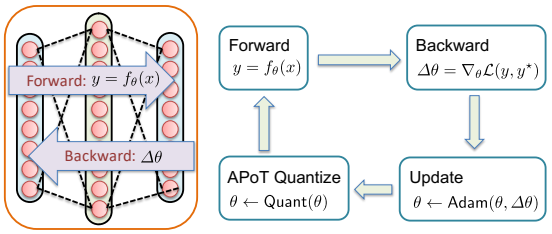
(a) FP: Multiply & Add (b) PoT: Shift & Add (c) APoT: Dual Shift & Add

Fig. 4: DNN weights with floating-point precision, PoT, and APoT quantizations. PoT/APoT can realize multiplier-less implementation. APoT provides more accurate quantization.

To compensate for the residual nonlinear distortion, we use DNN-based equalizers, which directly generate bit-wise soft-decision log-likelihood ratios (LLRs) for the decoder. Fig. 3 shows the achievable rate of DP-256QAM across SSMF spans for various DNN equalizers; residual multi-layer perceptron (6-layer 100-node ResMLP), residual convolutional neural network (4-layer kernel-3 ResCNN), and bidirectional long short-term memory (2-layer 100-memory BiLSTM). Binary cross entropy loss is minimized via Adam with a learning rate of 0.001 for $2{,}000$ epochs over 2^{16} training symbols to evaluate 2^{14} distinct testing symbols. For ResMLP, it is seen that the constellation shaping can achieve a reach extension by 29% for a target rate of 10 b/s. We found that the use of more hidden layers for CNN and LSTM architectures will further improve the training performance, while degrading testing performance due to over-fitting. It suggests that even larger training data size is required to offer better performance for deeper models.

Zero-Multiplier DNN with Additive Powers-of-Two (APoT) Quantization

In order to reduce the computational complexity of DNN equalizers for real-time optical communications, we integrate APoT quantization[14] into a DeepShift framework[12]. In the original DeepShift, DNN weights are quantized into a signed PoT as $w = \pm 2^u$, where u is an integer



(a) Multiplier-Less DNN (b) Quantization-Aware Training

Fig. 5: Straight-through rounding QAT for multiplier-less DNN with trainable APoT weights θ.

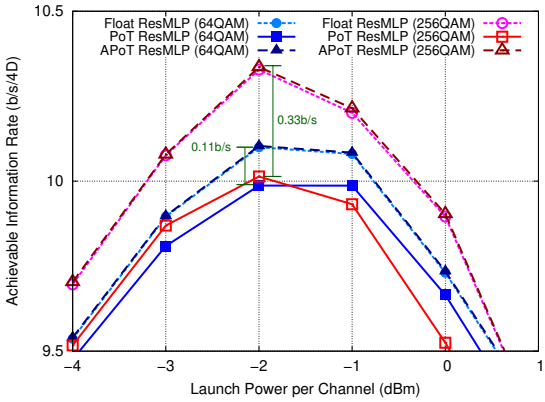


Fig. 6: Rate vs. launch power for zero-multiplier DNN equalizer (Shaped DP-64/256QAM, SSMF 22 spans).

to train. Note that the PoT weights can fully eliminate multiplier operations from DNN equalizers as it can be realized with bit shifting for fixed-point precision or addition operation for floating-point (FP) precision. It is illustrated in Fig. 4. We further improve the DeepShift by using APoT weights, i.e., $w = \pm(2^u + 2^v)$, where we use another trainable integer $v < u$. It requires an additional summation, but no multiplication likewise PoT. Using APoT weights, we can significantly decrease the residual quantization error of the conventional PoT as depicted in Fig. 4. Note that the original APoT[14] uses a deterministic non-trainable look-up table, whereas our paper extends it as an improved DeepShift with trainable APoT weights through the use of QAT. Fig. 5 shows our QAT updating, where we use a straight-through rounding to find dual bit-shift integers u and v after each epoch iteration. We also use a pre-training phase before the QAT fine-tuning in order to stabilize the DNN learning.

Fig. 6 shows the achievable rate across launch power at the 22nd span for 6-layer 100-node ResMLP. PoT quantization has a small degradation of 0.11 b/s compared to FP precision DNN equalizer for shaped DP-64QAM, whereas a considerable loss of 0.33 b/s is seen for shaped DP-256QAM. Notably, our multiplier-less DNN with APoT quantization has no degradation (but slight improvement) from FP precision. This is a great

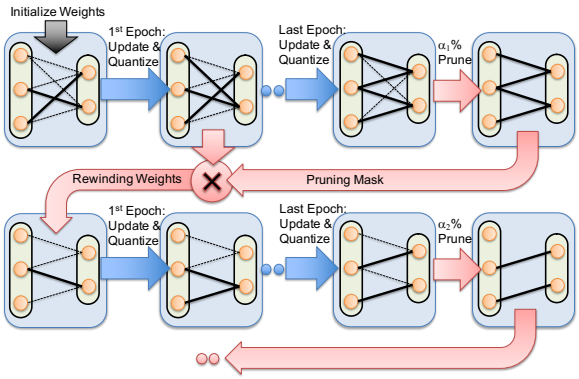


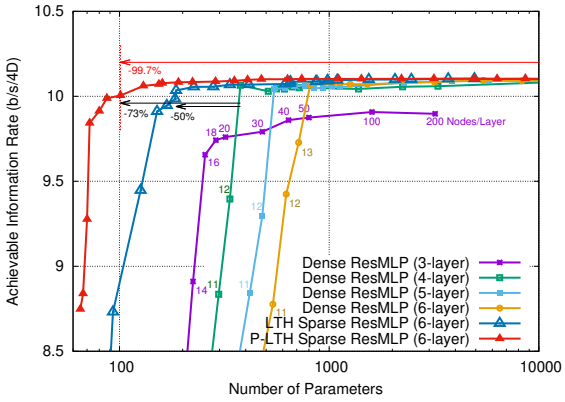
Fig. 7: Progressive LTH pruning for sparse DNN: 1) Initialize DNN weights; 2) QAT updates over multiple epochs; 3) Determine pruning mask by weak weights at the last epoch; 4) Rewind weights to the one at early epoch and prune the weights according to the mask; 5) Repeat steps 2–4 by gradually increasing the pruning percentages.

advantage in practice for real-time fiber nonlinearity compensation as there is no performance loss yet no multipliers are required.

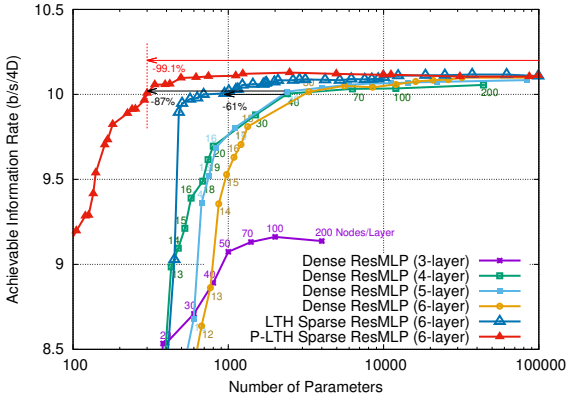
Lottery Ticket Hypothesis (LTH) Pruning for Sparse DNN

Even though our DNN equalizer does not require any multipliers, it still needs a relatively large number of addition operations due to the over-parameterized DNN architecture having huge number of weights. We introduce a progressive version of the LTH pruning method[15]–[17] to realize low-power sparse DNN implementation. It is known that an over-parameterized DNN can be significantly sparsified without losing performance and that sparsified DNN can often outperform the dense DNN. The progressive LTH pruning is illustrated in Fig. 7. We first train the dense DNN via QAT for APoT quantization, starting from random initial weights. We then prune a small percentage of the edges based on the trained weights. We re-train the pruned DNN after rewinding the weights to the early-epoch weights for non-pruned edges. Rewinding, QAT updating, and pruning are repeated with a progressive increase of the pruning percentages. We use late rewinding[16] of the first-epoch weights.

Fig. 8 shows a trade-off between the achievable rate and the number of non-zero weights. For dense DNNs, more hidden nodes and more hidden layers can improve the performance in general, at the cost of computational complexity. In consequence, a moderate depth such as 4-layer DNN can be best in the Pareto sense of the performance-complexity trade-off in low-complexity regimes as shown in Fig. 8. The LTH pruning can significantly improve the trade-off, i.e., the sparse DNNs can achieve more than



(a) Shaped DP-64QAM (SSMF 22 spans)



(b) Shaped DP-256QAM (SSMF 24 spans)

Fig. 8: Performance-complexity trade-off of dense/sparse DNN equalizers (-2 dBm).

50% complexity reduction over the dense DNNs to achieve a target rate of 10 b/s. Using progressive pruning, we can prune more than 99% of the weights of 6-layer 100-node ResMLP to maintain 10 b/s. Consequently, the sparse DNNs can be significantly lower-complex than the best dense DNNs by 73% and 87% for shaped 64QAM and 256QAM, respectively.

Conclusion

We compared various DNN equalizers for nonlinear compensation in optical fiber communications employing probabilistic amplitude shaping. We then proposed a zero-multiplier sparse DNN equalizer based on state-of-the-art APoT quantization and LTH pruning techniques. We showed that APoT quantization can achieve floating-point arithmetic performance without using any multipliers, whereas the conventional PoT quantization suffers from a severe penalty. We also demonstrated that the progressive LTH pruning can eliminate 99% of the weights, enabling highly power-efficient implementation of DNN equalization for real-time fiber-optic systems.

References

[1] Rios-Müller, Rafael, J. M. Estarán, and J. Renaudier, "Experimental estimation of optical nonlinear memory channel conditional distribution using deep neural networks", in *Optical Fiber Communication Conference (OFC)*, 2017, W2A–51.

[2] V. Kamalov, L. Jovanovski, V. Vusirikala, *et al.*, "Evolution from 8QAM live traffic to PS 64-QAM with neural-network based nonlinearity compensation on 11000 km open subsea cable", in *Optical Fiber Communication Conference (OFC)*, 2018, Th4D–5.

[3] P. Li, L. Yi, L. Xue, and W. Hu, "56 Gbps IM/DD PON based on 10G-class optical devices with 29 dB loss budget enabled by machine learning", in *Optical Fiber Communications Conference (OFC)*, 2018.

[4] C.-Y. Chuang, C.-C. Wei, T.-C. Lin, *et al.*, "Employing deep neural network for high speed 4-PAM optical interconnect", in *European Conference on Optical Communication (ECOC)*, 2017.

[5] T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "Neural turbo equalization: Deep learning for fiber-optic nonlinearity compensation", *Journal of Lightwave Technology (JLT)*, vol. 38, no. 11, pp. 3059–3066, 2020.

[6] C. Ye, D. Zhang, X. Hu, X. Huang, H. Feng, and K. Zhang, "Recurrent neural network (RNN) based end-to-end nonlinear management for symmetrical 50Gbps NRZ PON with 29dB+ loss budget", in *European Conference on Optical Communication (ECOC)*, 2018.

[7] B. Karanov, M. Chagnon, F. Thouin, *et al.*, "End-to-end deep learning of optical fiber communications", *Journal of Lightwave Technology (JLT)*, vol. 36, no. 20, pp. 4843–4855, 2018.

[8] S. Li, C. Häger, N. Garcia, and H. Wymeersch, "Achievable information rates for nonlinear fiber communication via end-to-end autoencoder learning", in *European Conference on Optical Communication (ECOC)*, 2018.

[9] R. T. Jones, T. A. Eriksson, M. P. Yankov, and D. Zibar, "Deep learning of geometric constellation shaping including fiber nonlinearities", in *European Conference on Optical Communication (ECOC)*, 2018.

[10] M. Chagnon, B. Karanov, and L. Schmalen, "Experimental demonstration of a dispersion tolerant end-to-end deep learning-based IM-DD transmission system", in *European Conference on Optical Communication (ECOC)*, 2018.

[11] V. Talreja, T. Koike-Akino, Y. Wang, D. S. Millar, K. Kojima, and K. Parsons, "End-to-end deep learning for phase noise-robust multi-dimensional geometric shaping", in *European Conference on Optical Communications (ECOC)*, 2020.

[12] M. Elhoushi, Z. Chen, F. Shafiq, Y. H. Tian, and J. Y. Li, "DeepShift: Towards multiplication-less neural networks", *arXiv preprint arXiv:1905.13298*, 2019.

[13] B. McDanel, S. Q. Zhang, H. Kung, and X. Dong, "Full-stack optimization for accelerating CNNs using powers-of-two weights with FPGA validation", in *ACM International Conference on Supercomputing (ICS)*, 2019, pp. 449–460.

[14] Y. Li, X. Dong, and W. Wang, "Additive powers-of-two quantization: An efficient non-uniform discretization for neural networks", *arXiv preprint arXiv:1909.13144*, 2019.

[15] J. Frankle and M. Carbin, "The lottery ticket hypothesis: Finding sparse, trainable neural networks", in *International Conference on Learning Representations (ICLR)*, 2018.

[16] H. Zhou, J. Lan, R. Liu, and J. Yosinski, "Deconstructing lottery tickets: Zeros, signs, and the supermask", *arXiv preprint arXiv:1905.01067*, 2019.

[17] H. You, C. Li, P. Xu, *et al.*, "Drawing early-bird tickets: Toward more efficient training of deep networks", in *International Conference on Learning Representations (ICLR)*, 2020.