# NN-based PCS distribution optimization for practical channels

Xueyang Li[(1)], Zhenping Xing[(1)], Md Samiul Alam[(1)], Maxime Jacques[(1)], Stéphane Lessard[(2)], David V. Plant[(1)]

[(1)] Department of Electrical and Computer Engineering, McGill University, Montréal, Québec H3A 0E9, Canada, xueyang.li@mail.mcgill.ca
[(2)] Ericsson Canada, Montréal, Québec H4S 0B6, Canada

**Abstract** *We propose an optimization scheme for the distribution of probabilistically shaped signals in practical non-AWGN channels based on a neural network and genetic algorithm. The optimized input distribution enables 23.5% higher throughput than the Maxwell-Boltzmann distribution in a short reach channel with a SiP transmitter.*

## Introduction

Probabilistic constellation shaping (PCS) enables an adaptable information rate (IR) with a fixed rate forward error correction (FEC) code and can provide an ultimate shaping gain of 1.53 dB in additive white Gaussian noise (AWGN) channels. For this, PS has been introduced into intensity-modulated direct detection (IMDD) systems in recent years for an improved throughput[1,2]. The Maxwell-Boltzmann (MB) distribution is widely used in PCS since it is a close to optimal distribution in AWGN channels. However, the AWGN channel approximation often deviates from real-life channels over a short reach where the signal is impacted by the device nonlinearity and non-AWGN noise sources including the laser flicker noise and clock leakage frequencies. Thus, in this scenario the MB distribution is suboptimal, and it is desirable to optimize the input distribution with respect to the actual channel characteristics in a practical way. Previous papers have adopted either exhaustive search[3] or numerical optimization techniques[4,5] based on the prior knowledge of an analytical channel model for nonlinear channels. However, a comprehensive analytical model that accurately characterizes practical channels is hard to construct and requires precise pre-measurement of multiple channel parameters. Neural networks have a strong nonlinear approximation capability and can be used to accurately model the characteristics of practical channels so that the performance metric can be precisely predicted with respect to the input distribution.

In this paper, we propose an optimization scheme for the input distribution of practical channels using a neural network (NN) and the genetic algorithm (GA). The NN is used to predict the generalized mutual information (GMI) with respect to the input distributions, whereas the GA is used to optimize the input distributions at desired spectral efficiencies (SE) based on the trained NN. We show numerically that the NNGA-optimized input distribution outperforms the MB distribution in an IMDD system with device nonlinearity. We also experimentally validate the efficiency of the NNGA scheme in optimizing the input distribution of an O-band IMDD system over 2-km of single-mode fibre (SMF) with a silicon photonic (SiP) traveling wave Mach-Zehnder modulator (TWMZM). It is found that the NNGA-optimized distribution achieves 23.5 % higher throughput relative to the MB distribution assuming a practical 8/9 fixed-rate low-density parity-check (LDPC) code with an NGMI threshold of 0.91 at 80 Gbaud, which corresponds to a net data rate of 197.3 Gb/s.

## NNGA-assisted distribution optimization

The AWGN channel assumption for short reach channels are often not valid due to (i) the device nonlinearity resulting from the phase shifter, the modulator and the trans-impedance amplifier (TIA); (ii) the coloured noise spectrum after equalization in a bandlimited channel; (iii) non-AWGN sources such as the laser flicker noise and the clock leakage frequencies. Fig. 1. depicts a schematic of the NNGA-assisted distribution optimization scheme for a non-AWGN channel.

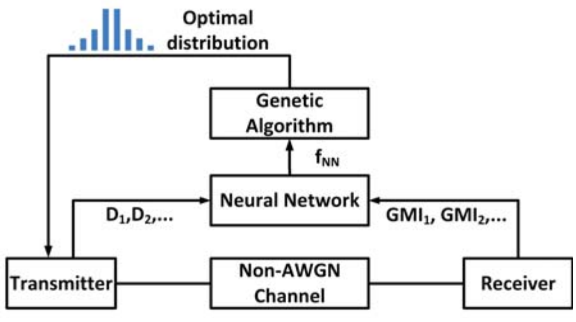


**Fig. 1:** Schematic of the distribution optimization using NNGA.

In this scheme, we use a neural network (NN) to approximate the nonlinear multivariable function that maps the input distribution D at the transmitter to the GMI at the receiver, which is taken as the performance metric in a probabilistic amplitude shaping (PAS) scheme[6]. Note that shaped signals of different distributions $D_1$, $D_2$… are sent at the transmitter to probe the channel so that the NN can effectively learn the channel characteristic. The trained NN as a multivariable function $f_{NN}$ is fed to a GA module which optimizes the input distribution at desired SEs, i.e. entropies. Note that this combined NNGA scheme is also applicable to the distribution optimization of long-reach channels with low dispersion.

In the numerical study, we optimize the distribution of probabilistically shaped (PS) PAM-8 signals in a short reach IMDD system with device nonlinearity. As the distribution of PS-PAM-8 symbols is symmetric following the PAS scheme, we reserve the probability mass function (PMF) of the positive half PS-PAM-4 symbols to reduce the number of inputs to the NN. The PMF is formulated as D = [$P_1$ $P_2$ $P_3$ $P_4$] and each PMF is created by first generating a 1×4 vector of uniformly distributed random numbers within [0,1], whose sum is normalized to 1 for a valid PMF. Constant composition distribution matching (CCDM) is used to approach the desired PMFs[7]. Fig. 2 shows a schematic of the optimized NN with two hidden layers that contain 16 and 32 hidden units, respectively. The PMF is taken as the input of the NN, whereas the GMI is taken as the output of the NN. We use the rectified linear unit (ReLU) function as the activation function at the hidden layers and minimize the mean squared error (MSE) between the NN output and the GMI from the training set. We use a mini-batch size of 128 and the Adam algorithm to optimize the NN weights. The training set, validation set, and testing set contain 1000, 250, and 250 PMF-GMI pairs, respectively. The NN is trained over 800 epochs.

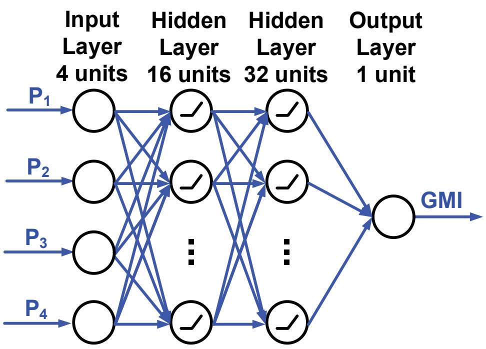


**Fig. 2:** Schematic of the NN with two hidden layers.

Fig.3 plots a simplified model for the short reach IMDD system. The sinusoidal function simulates the modulator nonlinearity, where k is a parameter within [0,1] that controls the swing of a normalized signal x. The maximum swing, i.e. from -π/2 to π/2 is obtained at k=1, which corresponds to a maximum modulator nonlinear effect. A 20 dB AWGN source is cascaded after the nonlinear module to simulate the overall

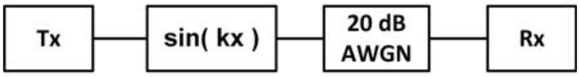


**Fig. 3:** A simplified model for short reach IMDD systems.

system noise. Note that this simplified model is used for the purpose of validation of the NNGA scheme for a short reach channel with nonlinearity. In actual short reach systems, the material nonlinearity of the phase shifter, the saturation of the TIA also contributes to the device nonlinearity.

After training, the NN is seen as a multivariable function $f_{NN}$ and finding the PMF that maximizes the function output, i.e. GMI at a certain entropy is equivalent to an optimization problem formulated as follows

$$
\begin{aligned}
&\underset{D}{\mathit{Minimize}} \;\; -f_{NN}(D) \\
&\mathit{subject\ to} \; \sum P_i = 1, \\
&\qquad\qquad \sum -P_i \log_2(P_i) = \mathit{Entropy} - 1, \\
&\qquad\qquad 0 \le P_i \le 1, \\
&\mathit{where}\;\; i = 1,2,3,4 \;\mathit{for\ all\ constraints}.
\end{aligned}
$$

Since this is a nonconvex nonlinear minimization problem, we use the genetic algorithm to search for approximated solutions by setting $-f_{NN}$ as the fitness function and performing mutation, crossover, and selection of the encoded input distributions. Fig. 4 (a) plots the GMI and NGMI as a function of k for both the MB and the NNGA-optimized distributions at entropy of 2.6.

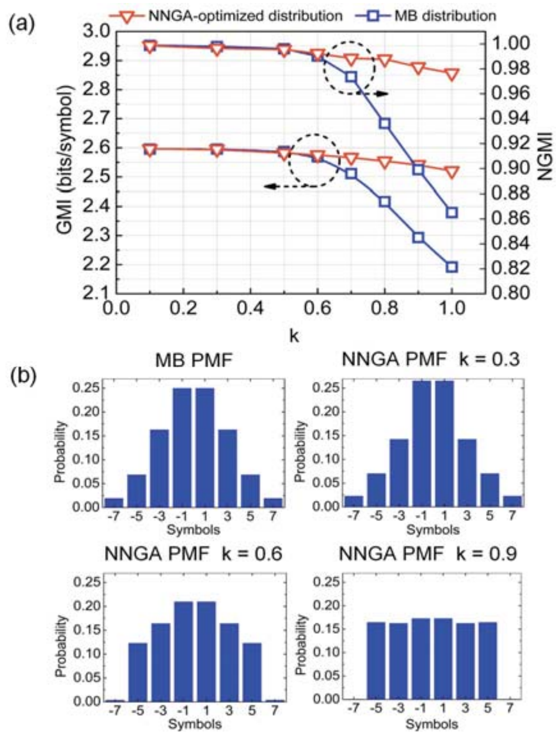


**Fig. 4:** (a) GMI and NGMI as a function of k for the MB and NNGA-optimized distributions. (b) PMF histograms of the MB distribution and NNGA-optimized distributions at different k.

The figure shows that the GMI of the MB distribution decreases sharply when k is greater than 0.6, whereas the GMI of the NNGA-optimized distribution decreases in a much slower rate. The NGMIs also show a similar trend for the MB and the NNGA-optimized distributions. As rate-adaptive FECs are costly, a fixed-rate FEC with PCS is more desirable in practical systems to enable the rate adaptation, and the NNGA-optimized distribution with a higher NGMI possesses a higher bound of the allowed code rate of practical FECs with k greater than 0.6. Fig. 4 (b) shows the corresponding PMF histograms of the MB distribution and the NNGA-optimized distribution at k of 0.3, 0.6, 0.9 at entropy of 2.6. It is observed that as k increases, i.e. with a stronger modulator nonlinear effect, the NNGA scheme determines that the probability of the outermost symbol level needs to approach zero as an appropriate distribution to mitigate the impact of the MZM nonlinearity.

## Experimental setup and results

We evaluate the performance of the NNGA-optimized shaping in an IMDD system in the O-band. Fig. 7 depicts the experimental setup. The carrier at 1302 nm is coupled into a SiP-TWMZM with a 3-dB E-O bandwidth of 45 GHz. The TWMZM modulates the carrier with an amplified RF signal generated by a 120 GSa/s arbitrary waveform generator (AWG). After propagation over 2 km of SMF, the signal is detected by a photodiode (PD)+TIA with a 3-dB bandwidth of 45 GHz and sampled by a 62 GHz real-time oscilloscope (RTO) for the post-processing. The transmitter and receiver DSP are also shown in the figure.

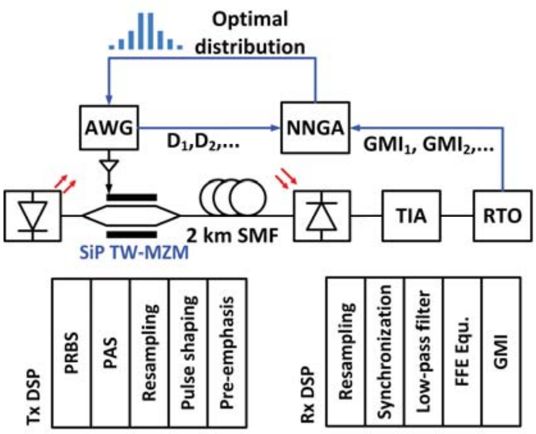


**Fig. 5:** Experimental setup and DSP.

The NN is trained following the procedure as described earlier. We transmit 80 Gbaud PS-PAM 8 signals to assess the performance of the NNGA-optimized shaping in this setup. Fig. 6 shows the PMF histograms and eye diagrams of the MB distribution and NNGA-optimized distribution, respectively, at entropy of 2.6. For the MB distribution, it is seen that the eye-openings close to the outermost levels are almost indiscernible due to the system nonlinearity. By contrast, the NNGA distribution has much more distinguishable eye openings and thus a 1.2-bits/symbol higher GMI.

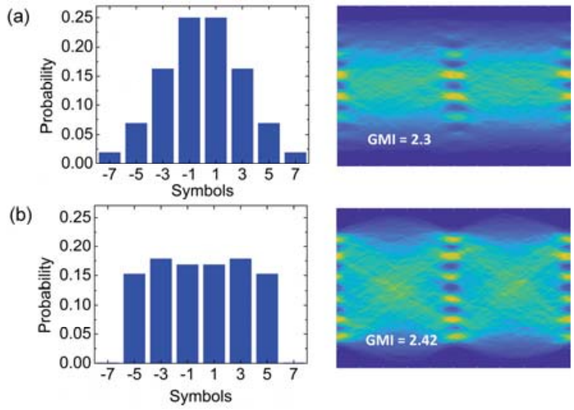


**Fig. 6:** PMF histograms and eye diagrams of (a) the MB distribution, and (b) the NNGA distribution at entropy of 2.6.

Fig. 7 plots the GMI and NGMI versus a varied entropy for both the MB distribution and NNGA-optimized distribution. The figure shows that the NNGA-optimized distribution leads to a higher GMI than the MB distribution at the same NGMI, i.e. the code rate of an ideal fixed-rate FEC. Assuming a practical 8/9 LDPC code with an NGMI threshold of 0.91[8], the NNGA-optimized distribution allows a 0.47 bit/symbol higher information rate. This translates to 23.5 % higher throughput at 80 Gbaud and corresponds to a net data rate of 197.3 Gb/s.

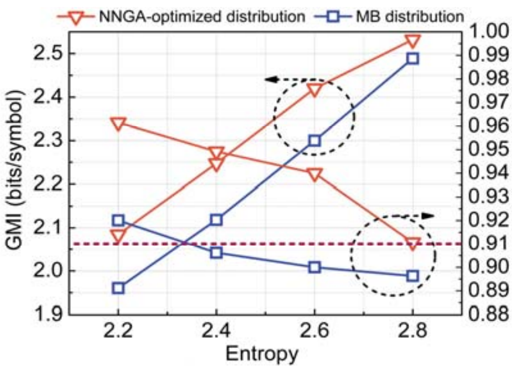


**Fig. 7:** GMI and NGMI as a function of the entropy.

## Conclusion

We demonstrate that the NNGA-optimized distribution outperforms the MB distribution in practical short reach channels with nonlinearity both numerically and experimentally. We show that 23.5 % higher throughput can be achieved using the NNGA-optimized distribution compared to the MB distribution in an O-band IMDD system with a SiP transmitter over 2 km assuming the 8/9 fixed rate LDPC code.

## References


[1] X. Chen *et al*., “Single-wavelength and single-photodiode entropy-loaded 554-Gb/s transmission over 22-km SMF”, in Optical Fiber Communication Conference Postdeadline Papers 2019, paper Th4B.5.

[2] X. Li *et al*., “Cost-minimizing distribution matching supporting net 800 Gbit/s PS-PAM transmission over 2 km using a 4-λ EML TOSA”, Opt. Lett. vol. 45, no. 17, pp. 4718-4721, 2020.

[3] C. Pan *et al.*, “Probabilistic 16-QAM shaping in WDM systems”, in J. Lightwave Technol., vol. 34, no. 18, pp. 4285-4292, 2016.

[4] E. Sillekens *et al*., “A simple nonlinearity-tailored probabilistic shaping distribution for square QAM”, in Optical Fiber Communication Conference 2018, paper M3C.4.

[5] M. P. Yankov *et al*., “Constellation shaping for fiber-optic channels with QAM and high spectral efficiency”, in IEEE Photon. Technol. Lett., vol. 26, no. 23, pp. 2407-2410, 2014.

[6] G. Böcherer *et al*., “Bandwidth efficient and rate-matched low-density parity-check coded modulation”, IEEE Trans. Commun., vol. 63, no. 12, pp. 4651-4665, 2015.

[7] P. Schulte *et al*., “Constant composition distribution matching”, IEEE Trans. Inf. Theory, vol. 62, no. 1, pp. 430-434, 2015.

[8] A. Alvarado *et al*., "Replacing the Soft-Decision FEC Limit Paradigm in the Design of Optical Communication Systems", in J. Lightwave Technol., vol. 33, no. 20, pp. 4338-4352, 2015.