Designing Energy Efficient Neural Networks According to Device Operation Principles

Ergun Simsek
Department of Computer Science and Electrical Engineering
University of Maryland Baltimore County, Baltimore, MD 21250, USA

Abstract— This article compares the accuracy and efficiency of two neural network architectures for estimating the performance metrics of a modified uni-traveling wave carrier (MUTC) photodetector: a fully connected neural network (FCNN) and a recurrent neural network (RNN)-like architecture designed specifically for the unique properties and operation principles of the device. The RNN-like architecture mimics the current flow and wave propagation inside the MUTC PD, with distinct neural network layers used to pair parameters of each PD layer, inform the network about interfaces between layers, mimic current transport between neighboring layers, and teach the network about the order and bi-directional interaction among the paired inputs. Results show that while the RNN-like architecture exhibits slightly higher learning accuracy than the FCNN, the RNN-like architecture consumes less energy and requires less time for training due to using fewer neurons. This study highlights the potential of physics-inspired neural networks in solving problems in electromagnetics and photonics.

2023 Photonics & Electromagnetics Research Symposium (PIERS) | 979-8-3503-1284-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/PIERS59004.2023.10221493

1. INTRODUCTION

The design of photonic and optoelectronic devices is usually done through numerical solvers. These devices display complex and nonlinear behavior, and machine learning algorithms can be used to design them more efficiently. However, to use these algorithms, a large amount of data is required to make accurate predictions. Therefore, a fast numerical solver is needed to create a sufficient dataset. We recently achieved this goal by developing a new drift-diffusion equation solver that is much faster than traditional solvers, allowing for the study of group IV and III-V photodetectors [1]. This significant reduction in simulation time is possible by using non-uniform time-stepping [2] and implementing window functions as broadband modulations [3].

Once we have a moderately large database, we can use a variety of machine learning techniques for both forward and inverse problems, such as predicting the device performance metrics from design parameters and designing a photonic device that performs at a desired level, respectively. However, not all the machine learning algorithms would perform well or be efficient for a given task. For example, fully connected neural networks (FCNNs) can be computationally intensive and require significant energy for training and inference, especially for large models with many neurons and multiple hidden layers. Even though the energy consumption of FCNNs can be reduced through techniques such as weight quantization and pruning, designing neural networks (NNs) according to the physical specifications of the system (or device) under investigation and according to that system's (device's) operation principles can lead to a more significant reduction in energy consumption of the NN. In this work, we compare the accuracy and efficiency (in terms of training time and energy consumption) of two NN architectures: an FCNN vs. a recurrent neural network (RNN)-like architecture designed according to the unique properties of the device under investigation, which is a modified uni-traveling wave carrier (MUTC) photodetector (PD).

2. DATASET

The MUTC PDs that we study include 17 layers of semiconducting materials with varying thicknesses (t_i) and doping levels (d_i) for $i = 1, 2, \ldots, 17$, as shown in Fig. 1(a). The photodetector is reverse-biased ($V_{\text{bias}} = -5\,\text{V}$) and is illuminated by a continuous wave laser operation at 1550 nm that is modulated by an input RF signal. The diameters of the incident beam and photodetector are 40 µm and 28 µm, respectively. The load resistance is $50\,\Omega$. More details of our simulation model may be found in [5]. We use our fast solver [2, 3] in combination with the particle swarm optimization (PSO) [6, 7] algorithm to design an MUTC photodetector with a lower phase noise. In the end, we obtained database with 1611 unique MUTC designs. In Fig. 2, we provide histogram plots of (a) phase noise, (b) quantum efficiency, (c) decay times (ps), and (d) average current of these photodetectors. As it clear from the histogram plots, we have a mildly skewed dataset (negative skewness for quantum efficiency and average current and positive skewness for phase noise and decay time).

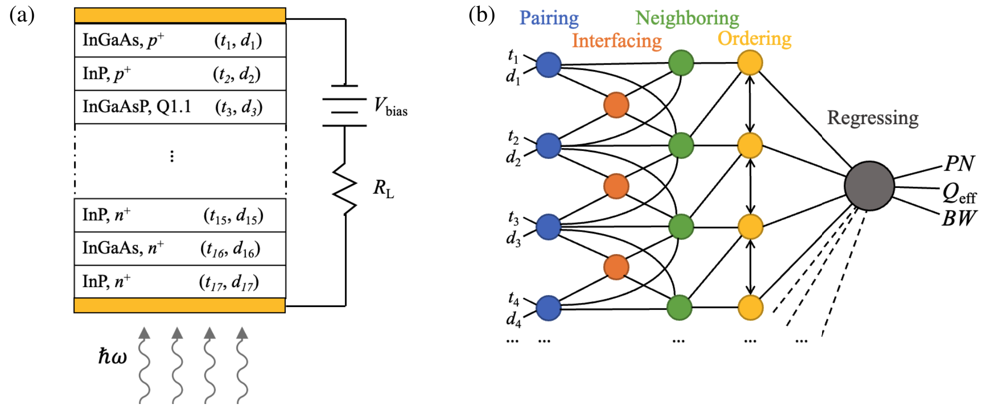

Figure 1: (a) A schematic of the MUTC photodetector with varying layer thicknesses and doping concentrations, t_i and d_i, respectively, for $i = 1, 2, 3, \ldots, 17$, and (b) an illustration of the neural network specifically designed to learn from device parameters to predict performance metrics such as the phase noise (PN), quantum efficiency (Q_{eff}), bandwidth (BW), etc.

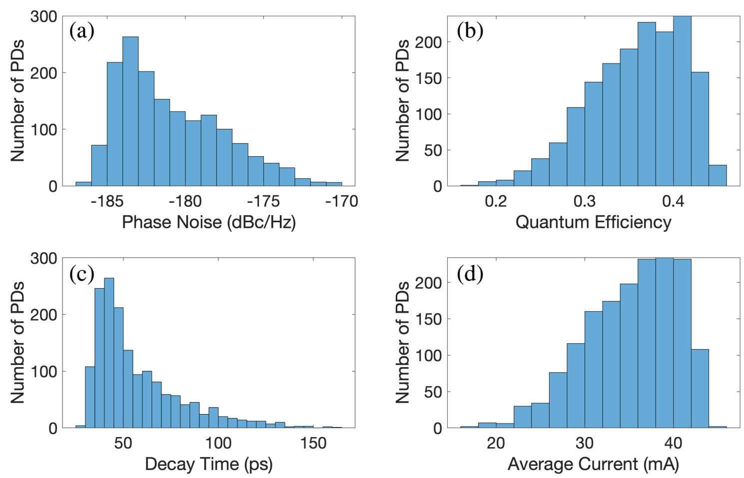

Figure 2: Histogram plots of (a) phase noise, (b) quantum efficiency, (c) decay times (ps), and (d) average current of the 1661 unique MUTC photodetectors generated during the optimization study.

3. ENERGY EFFICIENT NEURAL NETWORKS

The goal is to estimate the photodetector performance metrics, namely the phase noise (PN) and quantum efficiency Q_{eff}, from t_i's and d_i's. When we use traditional machine learning methods, such as linear regression, k-nearest neighbor, random forests, and fully-connected neural networks, we can achieve predicting out metrics with mean absolute errors of 9.9%, 11.3%, 7.1%, and 5.04%, respectively.

To increase the accuracy and to decrease the energy consumption, we design a recurrent neural network (RNN)-like architecture in a unique way to mimic the current flow and wave propagation occurring inside the device. To achieve this, we first utilize a neural network (NN) layer to pair the parameters of each PD layer as depicted in Fig. 1(b). By doing so, we inform the network that these are not just 34 uncorrelated numbers and there are two distinct groups. Then we use another NN layer to inform our network about the interfaces separating PD layers. The third NN layer is designed to mimic the current transport among neighboring PD layers. The fourth group aims to teach the NN that there is a certain order among these 17 paired inputs and the interaction is bi-directional, similar to reflections of electromagnetic waves back and forth between the physical layers. In the end, we have additional NN layers to do the regression of the device performance metrics. Fig. 3(a) compares the ground truth vs. prediction of the quantum efficiency. Fig. 3(b) shows how loss decays as a function of epoch number during the training. With this novel architecture, we achieve to reduce the mean absolute error to 2%. More importantly, due to using a fever number of neurons in our physics-inspired neural network, we achieve to reduce the energy consumed for computing and the period of time spent during training.

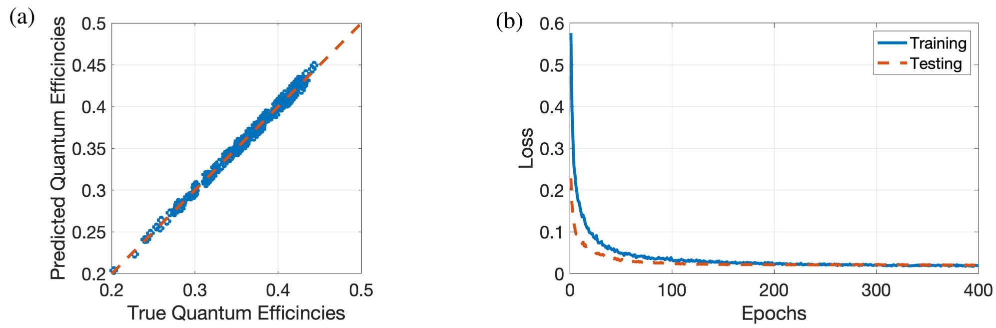

Figure 3: ((a) Ground truth vs. predicted quantum efficiencies, and (b) training loss (blue curve) and validation loss (red dashed curve) vs. epoch number.

4. CONCLUSION

In conclusion, this paper has demonstrated the potential of physics-inspired neural networks in solving problems in electromagnetics and photonics, specifically in the design of photonic and optoelectronic devices. The study compared the accuracy and efficiency of two neural network architectures for estimating the performance metrics of a modified uni-traveling wave carrier photodetector. While FCNNs are commonly used, the study showed that designing an RNN-like architecture according to the unique properties of the device under investigation can lead to a more significant reduction in energy consumption of the NN. The RNN-like architecture, which mimics the current flow and wave propagation occurring inside the device, consumed less energy and required less time for training due to using fewer neurons, while exhibiting similar learning accuracy. The results suggest that physics-inspired neural networks can be a promising approach to efficiently design and optimize complex photonic devices.

REFERENCES

1. Donati, S., *Photodetectors: Devices, Circuits and Applications*, Wiley, 2021.
2. Simsek, E., I. Md Anjum, T. F. Carruthers, and C. R. Menyuk, "Non-uniform time-stepping for fast simulation of photodetectors under high-peak-power, ultra-short optical pulses," *22nd Int. Conf. on Num. Sim. of Optoelec. Dev.*, Sept. 12–16, 2022.
3. Simsek, E., I. M. Anjum, T. F. Carruthers, D. A. Tulchinsky , K. J. Williams, J. C. Campbell, and C. R. Menyuk, "Efficient and accurate calculation of photodetector RF output power," *2022 IEEE Photonics Conference*, Vancouver, BC, Canada, Nov. 13–17, 2022.
4. Simsek, E., I. Md Anjum, T. F. Carruthers, and C. R. Menyuk, "Designing photodetectors with machine learning," *Optica Advanced Photonics Congress 2022*, paper NoW5C.2, 2022.
5. Mahabadi, S. E. J., S. Wang, T. F. Carruthers, C. R. Menyuk, F. J. Quinlan, M. N. Hutchinson, J. D. McKinney, and K. J. Williams, "Calculation of the impulse response and phase noise of a high-current photodetector using the drift-diffusion equations," *Opt. Express*, Vol. 27, No. 3, 3717–3730, 2019.
6. Kennedy, J. and R. Eberhart, "Particle swarm optimization," *Proceedings of the IEEE International Conference on Neural Networks*, 1942–1945, 1995.
7. Pedersen, M. E., "Good parameters for particle swarm optimization," Hvass Laboratories, Luxembourg, 2010.