Early Detection of DDoS Attacks using Photonic Neural Networks

M. Kirtas[1], N. Passalis[1], D. Kalavrouziotis[3], D. Syrivelis[3], P. Bakopoulos[3], N. Pleros[2] and A. Tefas[1]

[1]*Computational Intelligence and Deep Learning Group*
[2]*Wireless and Photonic Systems and Networks Group*
Dept. of Informatics, *Aristotle University of Thessaloniki*, Thessaloniki, Greece
[3]*Nvidia Mellanox Technologies Ltd., Yokneam, Israel*
{eakirtas, passalis}@csd.auth.gr, {dimitriosk, dimitriss, paraskevasb}@nvidia.com, {npleros, tefas}@csd.auth.gr

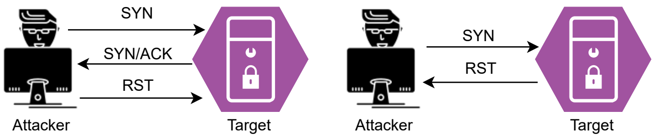
***Abstract*—Deep Learning (DL) has been extensively used in challenging tasks including security applications such as Distributed Denial of Service (DDoS) attacks. However, the high speed requirements of such applications along with the high complexity of DL models restrict the practical use of DL in real systems. Photonic neuromorphic hardware provides several advantages over electronic counterparts since it can operate at very high frequencies with lower power consumption. To this end, in this paper, we propose employing a photonic neuromorphic lookaside accelerator, aiming to perform real-time traffic inspection, enabling us to detect port-scanning attacks, which are indicative of DDoS attacks. We have designed, trained, and evaluated a Photonic Neural Network (PNN) capable of detecting DDoS attacks and operating on such photonic neuromorphic lookaside accelerators. The experimental evaluation is performed on Transport Control Protocol (TCP) traces obtained by simulating a port scanning attack and demonstrates the effectiveness of the proposed approach.**

I. Introduction

DL has been widely applied by both the academic community and industry, leading to state-of-the-art performance [1]. Over the recent years, there is an increasing interest in employing DL on real time intrusion and flooding attack detection, such as detection of DDoS attacks, employing different types of Artificial Neural Networks (ANNs) [2], [3]. DDoS attacks are commonly performed by multiple infected zombies/agents systems that are designed to attack a particular target or network with different types of packets [4]. DDoS attacks have been extensively studied and their detection is prioritized since they can cost organizations and individuals a great amount of time, money and reputation. Although DL seems an attractive approach to early detection and prevention of such attacks, the high complexity of DL models along with the high speed requirements of network applications generally prohibit the application of such DL approaches in most practical settings, since powerful hardware is required that increases both cost and energy consumption [5]. Over the recent years, specialized accelerators have been developed to serve the demanding nature of DL, ranging from Tensor Processing Units (TPUs) [6] to advanced neuromorphic hardware [7], increasing both training and inference speed, while also reducing power and energy consumption. To this end, *photonic* hardware is gaining attention as a very promising approach [8], due to its ability to provide ultra-fast matrix-based operations with very low power consumption [9], [10]. In neuromorphic photonics, signals are encoded using light, instead of electrical quantities, which are then manipulated to provide the neuron's functionality [11], [12]. Such approaches exploit the massive parallelism potential [13] and the ability of photonic components to operate at ultra-high frequencies [14] and form them in purely optical and/or advanced electro-optical devices [15], [16] exceeding their electronic counterparts in terms of bandwidth, speed, and energy consumption. These advantages have fueled the research on developing PNNs that can be efficiently deployed on such platforms.

Indeed, neuromorphic photonics are capable of operating at very high frequencies and can be integrated on a backplane pipeline of a modern high-end switch, which makes them an excellent choice for challenging DDoS detection applications, where high-speed and low-energy inference is required. However, despite these advantages, training ANNs that are oriented to photonic hardware also introduces new challenges to the DL training arising from photonic hardware. More specifically, PNNs rely mostly on sigmoid [16] and sinusoidal [17] based activation functions that are susceptible to early saturation, in contrast to traditionally used functions (such as ReLU [18]). Additionally, current photonic architectures face difficulties in supporting extremely large architectures that are traditionally used in modern DL. Therefore, despite these advantages, PNNs require a significantly different training pipeline in order to ensure that the resulting models will behave similarly to regular DL models that are deployed in software.

The main contribution of this work is the development of a DDoS detection PNN model that can be appropriately trained after taking into account the hardware limitations that arise from photonic neuromorphic accelerators. More precisely, we propose to prevent DDoS attacks during the reconnaissance attack (RA) phase, when the attacker tries to determine critical information about the target's configuration. Before deploying a DDoS attack, a port scanning procedure is compiled to track open ports on a target machine, as shown in Figures 1a and 1b. During this procedure, port scanning tools, such as Nmap [19], create synthetic traffic that can be captured and analyzed by the proposed PNN. We evaluated the proposed architecture on synthetically obtained data simulating port-scanning and collecting the transport control protocol (TCP) traces on the targeted machine. Furthermore, to demonstrate the generality of the proposed method, we also employed two different photonic activation functions that correspond to two



(a) SYN scan of an opened port (b) SYN scan of a closed port

Fig. 1: Port scanning using SYN scans

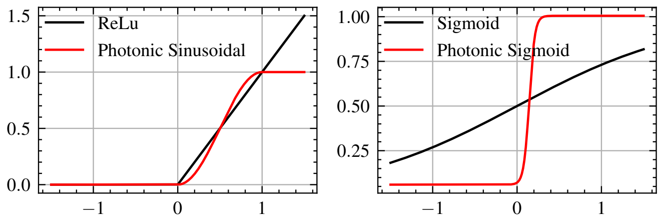


Fig. 2: Photonic activation functions in reference to traditionally used activation functions

different photonic neuromorphic configurations.

The rest of the paper is structured as follows. In Section II the proposed method is presented where the proposed hardware pipeline is presented along with the proposed architecture. Then, we experimentally evaluate the proposed setup in Section III. Finally, in Section IV conclusions are drawn.

II. Proposed Method

A. DDoS attacks and neuromorphic lookaside accelerators

In this paper we build on the concept of a neuromorphic lookaside accelerator, targeting to perform real-time traffic inspection, searching for DDoS attack patterns. However, further evaluation of this architecture pointed towards the identification of RA patterns rather than DDoS attack patterns, allowing for more timely and efficient recognition of a cyberattack.

RA is a common way to prepare a cyberattack on communication infrastructure, such as distributed DDoS. The technique involves the collection of probe responses that target discovering available services. An RA is typically a carefully engineered stealth attack that does not raise an alarm at the target site. An RA usually involves sending synthetic control traffic that generates random probes in an effort to collect the aforementioned probe responses, and thus discover available services. In other words, an RA typically sends adversarial control packets in order to characterize the infrastructure to be attacked. The very low traffic volume and rate of adversarial control packets, as compared to data traffic volume/rate, makes it practically impossible to detect the RA in real time using existing techniques.

RAs can be implemented with various algorithms. For example, in the case of TCP communications, a possible approach is to leverage TCP control messages for connection establishment and tear down, while tricking the remote server into sending responses that reveal whether a port is open or closed and what service it provides. In case of user datagram protocol (UDP) communications, probes may be sent to known service ports with known service requests in anticipation of receiving a response.

Typically, during an RA a small number of seemingly trustworthy control packets, such as TCP packets, UDP packets, or internet control message protocol (ICMP) packets, are sent to a particular point, such as to an N-port switch. Trying to hide in normal (data) traffic, the adversary network protocol traffic during RA is usually very slow and hence spans a relatively long-time window (e.g., 0.1 milliseconds) in order not to trigger high-traffic anomaly alarms that can be easily captured with simpler observation of statistics. Without an advance warning, the subsequently cyber attacked infrastructure may be damaged with the possibility of severe consequences. Timely detection of RAs, such as port scans, allows taking proactive measures to mitigate an imminent cyberattack like DDoS, for example by slowing down port connection and filtering of malicious IPs, and thereby to protect against severe socio economic consequences of such cyberattacks.

The detection technique performs real-time monitoring of ingress and egress traffic of control message packets. A processor analyzes in real-time metadata that is indicative of a temporal pattern of control messages communicated via one or more ports of a network device and identifies an RA on the network device by applying a DL-based DDoS detection algorithm to the temporal pattern of the control message traffic only. The rationale is that when a port scan (e.g., a TCP port scan) is in progress, the ingress traffic control messages and egress traffic control messages (e.g., TCP messages) that are exchanged in a given time window exhibit a detectable pattern anomaly as compared to normal control message traffic. For example, port scans may probe many dead ports, resulting in different flows of control message responses from hosts that affect the normal control message traffic pattern as it evolves in time.

To address this, a system-level design that includes coupling a processor optimized for AI, such as a neuromorphic coprocessor, to a backplane pipeline of a modern high-end switch can be employed. The architecture includes a processing circuitry that pre-processes and transforms specific switch telemetry data in real time, for direct feed to an AI processor, such as photonic neuromorphic coprocessor. The coprocessor is configured to instantly detect RAs happening within a small time frame of a few milliseconds over an entire set of ports of a network device. Therefore, for instance, a security scan of several tens of ports of a switch would take a few milliseconds to detect an RA, making it possible for the system to monitor hundreds of ports in real time (e.g., at a rate of 100 Hz).

B. Neuromorphic DL Training for DDoS detection

Fully connected PNNs similarly to software implemented ANNs are based on perceptron with their ultimate goal to approximate a function f^*. The training process is applied on software and then the PNNs parameters are deployed to the photonic hardware. More precisely, the PNN is trained iteratively using a training dataset that contains samples composed of the input signal that is denoted as $\boldsymbol{x} \in \mathbb{R}^M$, where M represents the number of features. Every sample in the train dataset is annotated using a binary label vector $\boldsymbol{t} \in \mathbb{R}^2$, which is equal to $[1, 0]^T$ for samples that do not belong to DDoS attack periods, indicating that the first output neuron of the model should activate (benign traffic). For samples that do belong to DDoS attack periods this vector is set to to $[0, 1]^T$, indicating that the second output neuron of the PNN should be activated (malicious traffic).

Multilayer perceptron (MLP) approximate f^* by using more than one layer, i.e., $f_n(...(f_2(f_1(\boldsymbol{x}; \boldsymbol{\theta}_1); \boldsymbol{\theta}_2); \boldsymbol{\theta}_n)$ and learn the parameters $\boldsymbol{\theta}_i$, where $0 \leq i \leq n$ with $\boldsymbol{\theta}_i$ consisting of weights $\boldsymbol{w}_i \in \mathbb{R}^{N_i \times M_i}$ and biases $\mathbf{b}_i \in \mathbb{R}^{N_i}$. N_i and M_i denote the output and input dimensionality of each layer respectively, while n is the total number of layers. Subsequently, each layer's output is denoted as:

$$\boldsymbol{z}_i = \boldsymbol{w}_i \boldsymbol{y}_{i-1} + \boldsymbol{b}_i. \tag{1}$$

In turn, the output of the linear part of each neuron is fed to a non-linear function $g(\cdot)$, named activation function, to form the final output of the layer:

$$\boldsymbol{y}_i = f_i(\boldsymbol{y}_{i-1}) = g(\boldsymbol{z}_i). \tag{2}$$

The training process is guided by the loss function $J(\boldsymbol{y}, \boldsymbol{t})$, where $\boldsymbol{t}$ represents the training labels and $\mathbf{y}$ the output of the network, which defines the correction step of the trainable parameters for every training step calculated by the backpropagation algorithm [20]. The cross-entropy loss is typically used in classification defined as:

$$J(\boldsymbol{y}, \boldsymbol{t}) = -\sum_{c=1}^{N} t_c \log y_c, \tag{3}$$

where N is the number of classes / output neurons of the network. The weight and biases of each layer are updated according to the propagated loss given by:

$$\Delta \boldsymbol{b}_i = -\eta \frac{\partial J}{\partial \boldsymbol{b}_i}, \text{ and } \Delta \boldsymbol{W}_i = -\eta \frac{\partial J}{\partial \boldsymbol{W}_i}. \tag{4}$$

After the training procedure, the weights and biases are deployed to the neuromophic photonic hardware for the inference phase. However, currently available hardware does not support traditionally used activation functions due to the unique nature of analog computing and the typically used photonic hardware configurations. To this end, we take into account the actual transfer function of the optical activation function during the training procedure. In this case of study two photonic activation functions are used, as shown in Fig. 2. The first one is the photonic sigmoid activation function [21], defined as:

$$g(z) = A_2 + \frac{A_1 - A_2}{1 + e^{(z - z_0)/d}}, \tag{5}$$

where the parameters $A_1 = 0.060, A_2 = 1.005, z_0 = 0.154$ and $d = 0.033$ are tuned to fit on the experimental observations as implemented on real hardware devices [21].

The second photonic activation function considered in this paper is a sinusoidal activation function, which corresponds to photonic activation layout that employs a Mach-Zender Modulator device (MZM) [22] to convert the data into optical signal along with a photodiode [23]. Note the similarity between this activation function and ReLu between 0 and 1, as shown in Fig. 2. The transfer function of this photonic activation is:

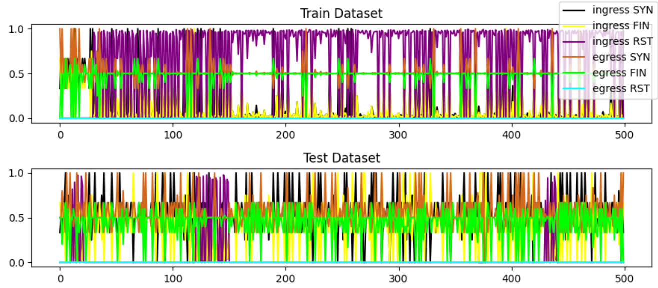


Fig. 3: TCP traces used for the training and evaluation datasets

$$g(z) = \begin{cases} 0, & \text{if } z < 0. \\ \sin \frac{\pi^2}{2} z, & \text{if } 0 < z < 1. \\ 1, & \text{if } z > 1. \end{cases} \tag{6}$$

It should be noted that the narrow activation range of the input domain of these photonic activations causes additional training difficulties, since networks tend to be easily saturated, leading to slower convergence, while in some cases this can even halt the training process [24].

In this work we employ a single hidden layer MLP, that can be implemented using the currently available photonic hardware taking as input 6 features that correspond to the statistics obtained from the TCP layer. More precisely, the PNN takes as input the percentage of each TCP flag (SYN, FIN, and RST) for the total ingress packages and the percentage of each TCP flag for the total egress packages, resulting in 6 values for every selected time window. After the hidden layer, a photonic activation function is applied. The classification layer consists of two neurons and it classifies a trace within the time window either as benign or malicious.

III. Experimental Evaluation

We compiled a dataset by monitoring a node of the network in which we perform SYN port scanning to collect positive samples. Periods of benign traffic also exist to provide the negative/benign samples. Three traces were generated: a) a malicious trace, b) a benign trace, and c) a mixed trace. Each TCP package is labeled according to the timestamp that is transmitted, enabling us to compile labeled datasets to perform supervised learning. We created two datasets, one for the training and one for the evaluation, including six features that correspond to the percentage of the TCP flags involved during 3 ms. The training dataset is formed by appropriately concatenating the malicious and plain traces, resulting in 23,473 records of which 30.97% are malicious. The evaluation dataset includes 12,604 records, from which 7.56% are malicious. A slice of the resulting time-series for both the train and evaluation datasets is depicted in Figure 3. Please note that we assume that every sample in the training dataset refers to a specific time-window that is labeled as malicious if inside this window a port scanning is performed. Otherwise, this sample is labeled as benign.

We demonstrate the experimental results of a lightweight fully connected PNN, which consists of 6 neurons on the

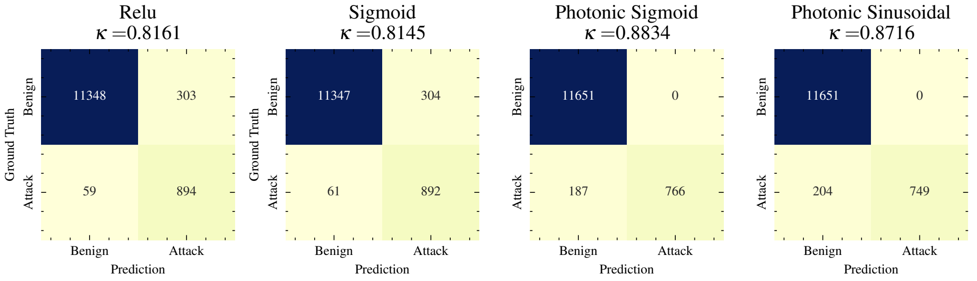


Fig. 4: Confusion matrices presenting the best evaluation run of each employed architecture

input layer, 10 neurons on the hidden layer and 2 neurons on the output layer that is capable of detecting DDoS attacks precisely in a high frequent manner (3 milliseconds) holding the credentials that can be implemented in photonic hardware. More specifically, we evaluate four different architectures: a) a baseline architecture that is implemented with ReLU activation function, b) a second baseline architecture employing sigmoid activation function, c) a sigmoid photonic architecture based on parameters arisen from experimental observation on real hardware implementation and d) a sinusoidal photonic architecture. During the training, we sampled data with replacement from the malicious class to overcome limitations arising from using a highly imbalanced dataset. The Adam optimizer is used for 15 training epochs with a learning rate equal to 0.0001, while batches of 32 samples are used.

We report the average Cohen's κ score [25] and the variance over 10 evaluation runs in Table I. As it is shown, the employed lightweight network can lead to sufficient classification performance. Both photonic architectures achieve better kappa scores in the average case than the traditional activation functions, the accuracy of the photonic models is sufficient enough for such a challenging task, while keeping all the advantages of the photonic substrate. It is worth noting that using the ReLU activation leads to significantly lower DDoS detection performance. This can be attributed to the intrinsic property of the ReLU activation [18], which is completely deactivated for half of its input range (negative values). This usually does not negatively impact the performance of networks, when enough neurons are available. However, when ReLU is used in lightweight architectures with a small number of neurons, it can lead to significant information loss, since we can expect that for normally distributed inputs, half of the neurons will be deactivated. Indeed, for the employed lightweight architecture this can lead to deactivating half of the neurons of the hidden layer, reducing the effective number of active neurons to 5. As a result, using ReLU essentially reduces the effective capacity of the network, compared to sigmoid activations. A similar effect is also observed for the sinusoidal function.

By comparing the confusion matrix of the best run of each different architecture, depicted in Figure 4, we observe that traditionally used activation functions have a significantly high false positive rate in contrast to photonic activation functions. In such security applications, a high false positive rate increases the sensitivity of a system (higher recall) but also increases the possibility that the prediction could be false positive (lower precision) resulting in systems that are difficult to be used in practical scenarios. On the other hand, photonic activation results in higher precision models eliminating the false positive rate to zero. Thus, when the model outputs that there is malicious traffic, it indicates that indeed a port-scanning is occurring and a further examination should be contacted by the network's administrators. Furthermore, ReLU activation has significantly unstable performance highlighted by the fact that even though on the average case is significantly worse than the sigmoid one, in the best case achieves slightly better performance. This confirms the aforementioned weakness of ReLU activation in lightweight architectures. Even when a variance-preserving initialization scheme is used [26], leading to a reduction in the number of active neurons and significantly affecting the classification layer outputs. The employed photonic architectures result in models with higher performance in contrast to traditionally used activations with the photonic sigmoid to achieve the best performance among all architectures.

TABLE I: Average kappa score and variance for each architecture

Activation Function	**Kappa score**
Baseline Architectures	
Relu	0.7785 ± 0.0760
Sigmoid	0.8067 ± 0.0167
Photonic Architectures	
Photonic Sinusoidal	0.8535 ± 0.0423
Photonic Sigmoid	$\mathbf{0.8788 \pm 0.0218}$

IV. Conclusions

In this work, we examined a method for preventing DDoS attacks using PNNs. More precisely, we proposed an ANN architecture based on photonic activation functions that can be integrated on a backplane pipeline of a modern high-end switch and that can effectively detect port scanning in a target machine using TCP traces. We evaluated the proposed method on synthetically obtained data, which are simulating port-scanning attacks, demonstrating that they can lead to adequate performance.

Acknowledgment: This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871391 (PlasmoniAC). This publication reflects the authors' views only. The European Commission is not responsible for any use that may be made of the information it contains.

References

[1] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] A. Saied, R. E. Overill, and T. Radzik, “Detection of known and unknown ddos attacks using artificial neural networks,” *Neurocomputing*, vol. 172, pp. 385–393, 2016.

[3] A. Chonka, J. Singh, and W. Zhou, “Chaos theory based detection against network mimicking ddos attacks,” *IEEE Communications Letters*, vol. 13, no. 9, pp. 717–719, 2009.

[4] E. Alomari, S. Manickam, B. B. Gupta, S. Karuppayah, and R. Alfaris, “Botnet-based distributed denial of service (ddos) attacks on web servers: classification and art,” *arXiv preprint arXiv:1208.0403*, 2012.

[5] P. J. Freire, Y. Osadchuk, B. Spinnler, A. Napoli, W. Schairer, N. Costa, J. E. Prilepsky, and S. K. Turitsyn, “Performance versus complexity study of neural network equalizers in coherent optical systems,” *Journal of Lightwave Technology*, vol. 39, no. 19, p. 6085–6096, Oct 2021.

[6] N. P. Jouppi, C. Young, N. Patil, D. Patterson, G. Agrawal, R. Bajwa, S. Bates, S. Bhatia, N. Boden, A. Borchers, R. Boyle, P. Cantin, C. Chao, C. Clark, J. Coriell, M. Daley, M. Dau, J. Dean, B. Gelb, T. V. Ghaemmaghami, R. Gottipati, W. Gulland, R. Hagmann, C. R. Ho, D. Hogberg, J. Hu, R. Hundt, D. Hurt, J. Ibarz, A. Jaffey, A. Jaworski, A. Kaplan, H. Khaitan, D. Killebrew, A. Koch, N. Kumar, S. Lacy, J. Laudon, J. Law, D. Le, C. Leary, Z. Liu, K. Lucke, A. Lundin, G. MacKean, A. Maggiore, M. Mahony, K. Miller, R. Nagarajan, R. Narayanaswami, R. Ni, K. Nix, T. Norrie, M. Omernick, N. Penukonda, A. Phelps, J. Ross, M. Ross, A. Salek, E. Samadiani, C. Severn, G. Sizikov, M. Snelham, J. Souter, D. Steinberg, A. Swing, M. Tan, G. Thorson, B. Tian, H. Toma, E. Tuttle, V. Vasudevan, R. Walter, W. Wang, E. Wilcox, and D. H. Yoon, “In-datacenter performance analysis of a tensor processing unit,” in *2017 ACM/IEEE 44th Annual International Symposium on Computer Architecture (ISCA)*, 2017, pp. 1–12.

[7] G. Indiveri, B. Linares-Barranco, T. Hamilton, A. van Schaik, R. Etienne-Cummings, T. Delbruck, S.-C. Liu, P. Dudek, P. Häfliger, S. Renaud, J. Schemmel, G. Cauwenberghs, J. Arthur, K. Hynna, F. Folowosele, S. SAÏGHI, T. Serrano-Gotarredona, J. Wijekoon, Y. Wang, and K. Boahen, “Neuromorphic silicon neuron circuits,” *Frontiers in Neuroscience*, vol. 5, p. 73, 2011.

[8] G. Dabos, G. Mourgias-Alexandris, A. Totovic, M. Kirtas, N. Passalis, A. Tefas, and N. Pleros, “End-to-end deep learning with neuromorphic photonics,” in *Integrated Optics: Devices, Materials, and Technologies XXV*, vol. 11689. International Society for Optics and Photonics, 2021, p. 116890I.

[9] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund *et al.*, “Deep learning with coherent nanophotonic circuits,” *Nature Photonics*, vol. 11, no. 7, p. 441, 2017.

[10] G. Mourgias-Alexandris, M. Moralis-Pegios, A. Tsakyridis, N. Passalis, M. Kirtas, A. Tefas, T. Rutirawut, F. Y. Gardes, and N. Pleros, “Channel response-aware photonic neural network accelerators for high-speed inference through bandwidth-limited optics,” *Opt. Express*, vol. 30, no. 7, pp. 10 664–10 671, Mar 2022. [Online]. Available: http://opg.optica.org/oe/abstract.cfm?URI=oe-30-7-10664

[11] N. Pleros, M. Moralis-Pegios, A. Totovic, G. Dabos, A. Tsakyridis, G. Giamougiannis, G. Mourgias-Alexandris, N. Passalis, M. Kirtas, and A. Tefas, “Compute with light: Architectures, technologies and training models for neuromorphic photonic circuits,” in *Proceedings of the European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.

[12] G. Giamougiannis, A. Tsakyridis, G. Mourgias-Alexandris, M. Moralis-Pegios, A. Totovic, G. Dabos, N. Passalis, M. Kirtas, N. Bamiedakis, A. Tefas, D. Lazovsky, and N. Pleros, “Silicon-integrated coherent neurons with 32gmac/sec/axon compute line-rates using eam-based input and weighting cells,” in *Proceedings of the European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.

[13] G. Mourgias-Alexandris, M. Moralis-Pegios, S. Simos, G. Dabos, N. Passalis, M. Kirtas, T. Rutirawut, F. Y. Gardes, A. Tefas, and N. Pleros, “A silicon photonic coherent neuron with 10gmac/sec processing line-rate,” in *Proceedings of the Optical Fiber Communications Conference and Exhibition (OFC)*, 2021, pp. 1–3.

[14] G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, M. Kirtas, A. Tefas, T. Rutirawut, F. Y. Gardes, N. Pleros, and M. Moralis-Pegios, “25gmac/sec/axon photonic neural networks with 7ghz bandwidth optics through channel response-aware training,” in *Proceedings of the European Conference on Optical Communication (ECOC)*, 2021, pp. 1–4.

[15] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, “All-optical machine learning using diffractive deep neural networks,” *Science*, vol. 361, no. 6406, pp. 1004–1008, 2018.

[16] G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, A. Tefas, K. Vyrsokinos, and N. Pleros, “An all-optical neuron with sigmoid activation function,” *Opt. Express*, vol. 27, no. 7, pp. 9620–9630, Apr 2019.

[17] N. Passalis, G. Mourgias-Alexandris, A. Tsakyridis, N. Pleros, and A. Tefas, “Training deep photonic convolutional neural networks with sinusoidal activations,” *IEEE Trans. Emerging Topics in Computational Intelligence*, pp. 1–10, 2019.

[18] X. Glorot, A. Bordes, and Y. Bengio, “Deep sparse rectifier neural networks,” in *Proceedings of the International Conference on Artificial Intelligence and Statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

[19] G. F. Lyon, *Nmap network scanning: The official Nmap project guide to network discovery and security scanning*. Insecure. Com LLC (US), 2008.

[20] H. J. KELLEY, “Gradient theory of optimal flight paths,” *ARS Journal*, vol. 30, no. 10, pp. 947–954, 1960.

[21] G. Mourgias-Alexandris, A. Tsakyridis, N. Passalis, A. Tefas, K. Vyrsokinos, and N. Pleros, “An all-optical neuron with sigmoid activation function,” *Opt. Express*, vol. 27, no. 7, pp. 9620–9630, 2019.

[22] S. Pitris, C. Mitsolidou, T. Alexoudi, D. Pérez-Galacho, L. Vivien, C. Baudot, P. D. Heyn, J. V. Campenhout, D. Marris-Morini, and N. Pleros, “O-band energy-efficient broadcast-friendly interconnection scheme with sipho mach-zehnder modulator (mzm) & arrayed waveguide grating router (awgr),” in *Proceedings of the Optical Fiber Communication Conf.* Optical Society of America, 2018.

[23] L. Danial, N. Wainstein, S. Kraus, and S. Kvatinsky, “Breaking through the speed-power-accuracy tradeoff in adcs using a memristive neuromorphic architecture,” *IEEE Trans. on Emerging Topics in Computational Intelligence*, vol. 2, no. 5, pp. 396–409, 2018.

[24] N. Passalis, M. Kirtas, G. Mourgias-Alexandris, G. Dabos, N. Pleros, and A. Tefas, “Training noise-resilient recurrent photonic networks for financial time series analysis,” in *Proceedings of the European Signal Processing Conference (EUSIPCO)*, 2021, pp. 1556–1560.

[25] M. L. McHugh, “Interrater reliability: the kappa statistic,” *Biochemia medica: Biochemia medica*, vol. 22, no. 3, pp. 276–282, 2012.

[26] K. He, X. Zhang, S. Ren, and J. Sun, “Delving deep into rectifiers: Surpassing human-level performance on imagenet classification,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.