Photonic-accelerated AI for cybersecurity in sustainable 6G networks

Emilio Paolini
and Luca Valcarenghi
Scuola Superiore Sant'Anna
TeCIP Institute
Via Giuseppe Moruzzi, 1, 56124 Pisa
Email: emilio.paolini@santannapisa.it

Luca Maggiani
Sma-RTy Italia SRL
Via dell'Artigianato 2, 20061 Carugate

Nicola Andriolli
CNR-IEIIT
Pisa, 56122, Italy

Abstract—**The sixth generation (6G) of mobile communications, expected to be deployed around the year 2030, is predicted to be characterized by ubiquitous connected intelligence. With Artificial Intelligence (AI) operations being deployed in every aspect of future network infrastructure, network security will also evolve from current solutions to intelligent architectures. To meet the massive amount of operations computed by AI models, photonic hardware can be exploited, delivering higher processing speed and computing density and lower power consumption with respect to electronic counterparts.**

In this paper, we propose a photonic-based Convolutional Neural Network (CNN) solution able to work on real-time traffic, capable of identifying Denial of Service (DoS) Hulk attacks with 99.73 **mean F1-score when exploiting 4 bits. We also compared photonic accelerators with their electronic counterparts, showing limited F1-score degradation, especially in the 4 and 8 bit scenarios.**

I. Introduction

In the foreseeable 6G [1] along with continuous advancements of 5G services, i.e., enhanced Mobile Broadband (eMBB), massive Machine Type Communications (mMTC), and Ultra Reliable Low Latency Communications (URLLC), new 6G services are expected to emerge [2], such as truly immersive extended reality, high-fidelity mobile hologram, and many more. To meet the requirements of new services and new use-cases, 6G networks need to be enhanced in many aspects when compared to 5G, such as data rate and end-to-end latency.

Among the enabling technologies, Artificial Intelligence (AI) techniques will have a relevant role [3]. It is expected that AI systems will be deployed already in 5G networks within the next few years in at least three main scenarios: (i) replacing model-based Layer 1 and Layer 2 algorithms, e.g., channel estimation; (ii) network deployment optimization without human intervention, e.g., finding the optimal number of parameters at the time of deployment; (iii) localization of end devices exploiting learning techniques for improved accuracy [1]. However, the integration of AI models in 6G network architecture will be deeper, switching from AI-enhanced network functions to AI-driven network infrastructure.

AI serving as a foundation for future network architecture will pave the way for a paradigm shift, with a Radio Access Network (RAN)-Core Network (CN) convergence, both in the user-plane and the control-plane [2], [4]. In current 5G networks, although some flexibility can be adopted, the definition of Network Functions (NFs) and network protocols strictly separate the RAN and the CN, limiting the levels of flexibility and efficiency provided by the network. In future 6G networks, a new approach will be exploited providing more flexibility in network deployment, where the RAN and the CN functions can be converged in the same platform and optimized together according to the use-case requirements [2], [4]. Therefore, the entire network architecture should be reconsidered as a collection of intelligent nodes, each one capable of making decisions independently. This paradigm shift will fulfill the so-called collective network intelligence [5], where AI techniques are exploited to provide distributed autonomy and at the same time federate nodes to collaboratively learn and make decisions. To support this 6G vision, network functions must be deployed in a federated manner among base stations. Among the features of 6G networks to enable distributed autonomy, high security, secrecy, and privacy will have primary importance in critical scenarios such as finance and military [6].

However, one main drawback of this network vision is the amount of computations performed by AI techniques, which will exponentially increase due to the ubiquitous presence of intelligent operations in every aspect of network management. Hence, to both support the large quantity of data processed by Machine Learning (ML) models, reduce their computational burden, and comply with 6G vision on green computing [7], [8], new hardware platforms for Deep Learning (DL) inference are investigated [9], [10]. Photonics-based solutions attracted a lot of interest with the promise of outperforming electronic counterparts in speed, power consumption, and computing density [11], [12]. Photonic-based accelerators also comply with energy efficiency requirements of future network generations [7], making them the perfect candidates for DL model deployment.

In this paper, we exploit photonic hardware, i.e., Photonic-Aware Neural Network (PANN) [13], to implement a threat mitigation system at the base station level, performing DL computations in the analog optical domain. This solution, leveraging neuromorphic processors deployed at base stations, complies with the aforementioned 6G visions : (i) the proposed system enables 6G nodes to automatically detect Denial of Service (DoS) attacks , fulfilling node autonomy and intelligence; (ii) by relying on photonic solutions, it can perform DL computations with sub-microsecond latencies, high bandwidth, and low energy consumption satisfying the 6G vision on green computing.

In the following: (i) related works on AI solutions dedicated to intrusion detection systems are reviewed; (ii) the proposed system architecture is described, highlighting the enabling technologies of such a solution; (iii) photonic accelerators are also discussed, showing the advantages with respect to electronic counterparts and the limitations introduced by the underlying hardware and how they affect Neural Network (NN) computations; (iv) experiments to validate the proposed PANN architectures are carried out, discussing their theoretical performance over electronic counterparts. Finally, possible future research directions are highlighted.

II. Related works

Recently, many works try to address threat mitigation aspects leveraging statistical and AI techniques, to enable the deployment of future network infrastructure in critical scenarios.

Towards this direction, authors in [14] propose to use the Deep Neural Networks (DNNs) to detect DoS attacks on network traffic. However, the discussed solution works on statistical features extracted from traffic flows, preventing real-time detection.

In [15] an optimal feature selection algorithm and a Convolutional Neural Network (CNN) model for threat detection in Software Defined Networking (SDN) are developed. The discussed solution suffers from limitations in terms of computational workload since it requires very deep convolutional layers. Even though it is possible to accelerate these layers by exploiting conventional hardware, i.e., Graphics Processing Units (GPU) and Field Programmable Gate Array (FPGA), the power consumption and packet arrival rate would make this solution unamenable for a 6G network.

Another interesting work [16] proposes a Xilinx Zynq Z-7020 FPGA implementation of a Fully-Connected (FC) NN capable of detecting malicious traffic. Although supporting high throughput, this solution suffers from both energy consumption issues, being based on electronics, and limitations concerning real-time suitability since it employs statistics extracted on complete traffic flows.

Finally, in [17] a hierarchical architecture for protecting 5G-enabled Internet of Things (IoT) networks is discussed. The model exploits Markov stochastic process to predict and detect malicious device behavior. Leveraging the Markov process enables this type of solution to be deployed at the edge of the network. However, the system is based on log files and IoT devices, limiting the scalability of the architecture to known device types. On the other hand, a DL technique can automatically extract features and potentially work on all types of traffic packets.

Compared to currently proposed solutions, we are the first to exploit NN algorithms and neuromorphic photonic hardware to address all the issues related to a threat mitigation system amenable to 6G implementations.

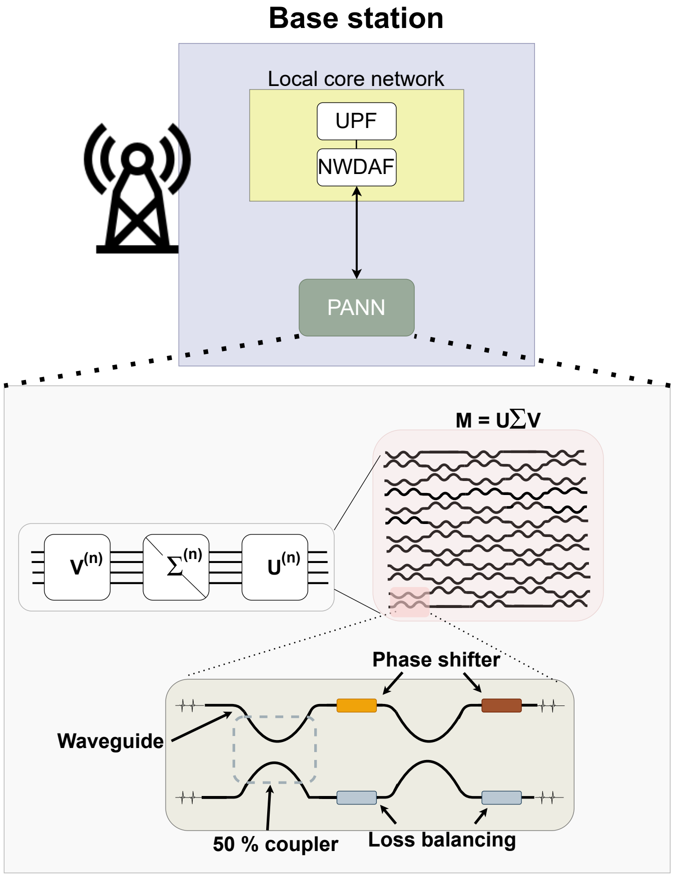

Fig. 1. Proposed architecture. 6G base station is deployed along with a local CN and PANN accelerator. For the sake of clarity, only the two NFs discussed are depicted.

III. Proposed architecture

In this section, we detail how the proposed architecture can be implemented in a future 6G base station. As discussed above, 6G network cloudification will provide less strict separation between RAN and CN, ultimately leading to the convergence of the two. With each 6G base station equipped with functionalities coming from both CN and RAN, it will be possible to deploy a local CN on top of each node.

Among the novel NFs, the Network Data Analytics Function (NWDAF) will become more important in 6G networks, with its use becoming more prominent for delivering network intelligence. The NWDAF can be deployed in each base station, providing data analytics upon request from any other NFs [18].

With security becoming the pillar of 6G networks, NWDAF can be exploited to perform intelligent threat mitigation on user data. Hence, the NWDAF can collect User Plane Function (UPF) data coming from the various User Equipments (UEs) and use them as input to the DL system to identify malicious traffic. With this architecture, a possible threat can be directly identified at the base station level , without forwarding it to the rest of the network. Moving security functionalities as close as possible to where possible threats can be generated is essential. Furthermore, real-time detection is a fundamental aspect of this scenario. Indeed, it is estimated that the average cost of service disruption can reach 5600$ per minute [19]. Hence, a solution placed at the base station level that can perform real-time threat mitigation on user data is of paramount importance.

To perform traffic classification and identify malicious packets, the solution proposed in this paper exploits CNN. In particular, packets belonging to the same flow are collected to create input matrices [20] . However, DL architectures are characterized by a heavy computational burden (especially due to the large number of required Multiply-Accumulate (MAC) operations). Indeed, NNs are resource-intensive algorithms, which might increase the latency and thus the response time of the node. To have a solution compliant with the massive amount of traffic expected in 6G networks, novel technological solutions need to be explored.

In particular, in this paper, we resort to photonic accelerators to speed up NN computations. Photonic hardware can outperform electronic counterparts in energy ($> 10^2$), speed (10^3), and compute density (10^2) [21]. To further unleash the potential of a drastic reduction in power consumption, several photonic solutions also leverage passive components to implement NN weights, i.e., not requiring energy besides input generation and output acquisition [13]. Hence, the underlying photonic hardware can improve the system's energy efficiency and realize sustainable green networks, one of the main goals of 6G vision [22].

The proposed system architecture is depicted in Fig. 1. For the sake of brevity, we report only the NFs in the CN that we use for this solution, i.e., UPF and NWDAF. The developed method leverages photonic accelerators that receive packets collected by the NWDAF from the UPF and perform classification at a very high rate and with time-of-flight latency.

The proposed solution is based on a mesh of Mach-Zender Interferometer (MZI) performing the typical NN computations, i.e., matrix-vector multiplications [23]. The photonic hardware relies on a parallel architecture with data stored and processed in the same place, i.e., within the MZI elements. The MZI mesh physically implements the matrix of interest by singular value decomposition. Once in the optical domain, the time needed to perform MAC operations corresponds to the time-of-flight of the photonic chip. Thus, the throughput of PANN architectures is mainly limited by the driving electronics in the input layer, reaching tens of GHz or even more than one hundred with the lithium-niobate on insulator platform [24]. However, analog photonic hardware introduces some limitations, typically not encountered in electronics, that must be faced in both the training and inference phases. We present them in Sec. III-A, discussing the methods to make these accelerators compliant with NN computations.

A. Photonic-Aware Neural Networks

Although being one of the most promising NN accelerators, photonic hardware still presents several limitations [9]. The most important aspect regards the training phase: exploiting the optical domain to perform training is very cumbersome. Hence, photonic solutions are being investigated with the aim of carrying out just the inference phase [25]. Furthermore, photonic accelerators are analog engines, working with analog values that can in principle vary in a continuous set. Nonetheless, the resolution is constrained by noise and distortions. A particular aspect is that noise does not depend on the represented values [26]. Hence, photonic accelerators have constant noise intervals. For this reason, photonic devices can distinguish only a finite number of different equally-spaced levels. This behavior justifies the exploitation of the Effective Number Of Bits (ENOB) to relate the number of distinguishable values to the corresponding number of bits needed for digital storage. In the context of PANNs, the typical bit resolution is very limited i.e., $\leq$ 8 bits [11], [27]. This is in contrast with the typical bit resolution, i.e., 32-bit floating point, used in the digital electronic implementation of NNs. Thus equally-spaced types, i.e., fixed-point format, on a low number of bits are suitable for PANN computations. In addition to this, such accelerators suffer from a trade-off between attainable operation speed and supported resolution, as well as power consumption [27].

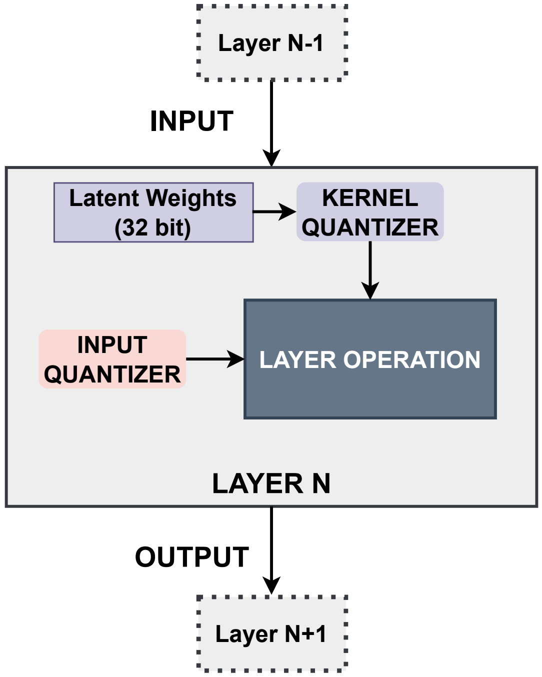

Fig. 2. PANN training strategy. A layer is defined along with an input quantizer and a kernel quantizer.

To overcome this issue, as we already proposed in [13], in this paper, we exploit reduced-precision fixed-point type for PANN inference and present a suited training approach. Indeed, the direct quantization of the weights computed using floats significantly reduces the accuracy of the obtained NN [28].

The PANN training process is shown in Fig. 2. Leveraging this method, we can alleviate the issue of accuracy loss due to quantization. A layer is defined along with an input quantizer and a kernel quantizer, which describe the strategy to quantize the incoming inputs and weights, respectively. Hence, a quantized layer computes the activation y as:

$$y = \sigma(f(q_{\text{kernel}}(w), q_{\text{input}}(x)) + b)$$

with full precision weights w, arbitrary precision input x, layer operation f, activation function σ and bias b.

An essential aspect regards the latent weights, which take into account the quantization loss at training time. Indeed, latent weights are full-precision (i.e., 32-bit floating point) copies of the quantized weights, which are used to store the gradient updates computed during the training phase. In the forward pass, a quantized version of these weights is used. Finally, when a model has finished the training phase, only the quantized weights are kept and used in the inference phase. Concerning the quantizers exploited for the PANN training-to-inference strategy, DoReFa quantizers [29] are chosen since they allow to flexibly define the bitwidths on both inputs and weights.

Additionally, photonic hardware presents constraints concerning the NN inputs. Indeed, inputs are required to be positive-valued, since they are encoded in the intensity of optical signals. This imposes to have both normalization and activation functions with positive values. ReLU and sigmoid functions are suitable candidates for PANN architectures.

Another constraint imposed by the underlying photonic hardware concerns the maximum number of inputs, i.e., fan-in, to each neuron. MZI meshes, on which the proposed solution is based, can perform block operations at the expense of several electro-optical conversions: the maximum number of inputs is about 200 for these implementations.

A final limitation due to the underlying hardware concerns the maximum kernel size in convolutional layers. Given the current experimentally validated photonic convolutional kernel implementations [23], the maximum kernel size is equal to 3×3 elements.

IV. Experiments

The experiments are carried out using CIC-IDS 2017 network traffic dataset collected over several days, which includes DoS attacks [30]. Both raw packets, i.e., PCAP files, and the results of the network traffic analysis in the form of flow statistics, i.e., CSV files, are available.

The baseline model is an NN architecture working with features extracted from network traffic. This approach, afterward called offline, can be considered as a theoretical benchmark for the proposed system: this solution does not provide real-time response and thus it is not amenable for a 6G environment. Instead, the proposed system architecture, working on matrices of raw packets directly collected at the base station, i.e., real-time, can be seamlessly implemented in 6G nodes. In particular, this approach works on matrices of shape $N \times M$, where N is the number of packets for each flow and M represents the number of features extracted from packets. In particular, during our experiments we set $N = 10$ to keep the length of the matrices small and $M = 11$, selecting the features that intuitively help the model generalize better, i.e., discarding those features that are too deterministic such as source/destination address/port.

Concerning the NN models, two main architectures have been developed: (i) FC-NN used in both offline, i.e., working on statistics, and real-time FC scenarios, i.e., working on flattened matrices. The number of input features of the two models is changed accordingly: 40 for the offline model and 110 for the real-time architecture; (ii) CNN used for the real-time convolutional experiment. Regarding the first model, 5 FC layers are deployed, with $200, 192, 128, 64, 32$ and 1 neurons, respectively. Dropout layers are deployed among FC layers with 0.1 rate. The second model is composed of 5 convolutional layers with $2, 4, 6, 8, 16, 32$ filters, respectively, each one leveraging 3×3 kernel. In both the models, ReLU has been adopted as activation function, except for the last layer that uses Sigmoid to perform binary classification. Both models comply with PANN constraints in terms of input features, kernel size, and activation function. Each training phase has been composed of 30 epochs, using Adam as optimizer and a batch size of 128.

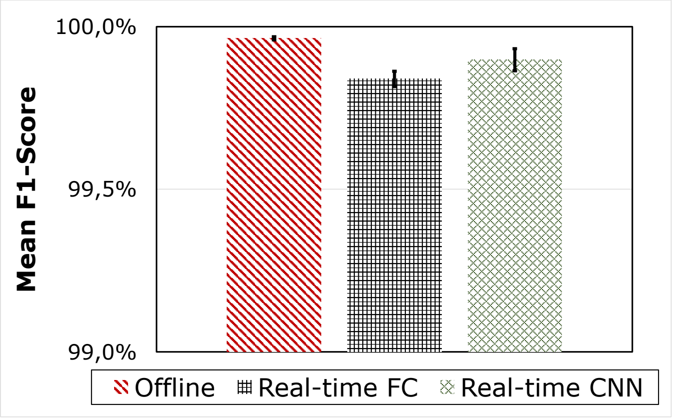

Fig. 3. Comparison among the theoretical baseline, i.e., Offline, and two real-time models, i.e., FC and CNN.

To evaluate our proposed architectures, we considered DoS Hulk attack types, that generate unique requests bypassing caching engines. Concerning the training set, we resort to random oversampling techniques to have the same data distribution among benign and malicious classes. On the other hand, the test set must be as close as possible to real-world scenarios, hence its distribution has been kept imbalanced. In particular, we test our models in a scenario where a DoS attack is currently being performed to the network, resulting in a $14\% - 86\%$ balancing among normal and attack packets. Regarding the metrics, we resort to F1-score since it can give a better understanding of the classification performance when dealing with imbalanced data.

To identify the candidate for photonic implementation, a first comparison among full-precision FC and CNN models has been carried out, with the offline model providing an upper bound to the detection performance. Results are shown in Fig. 3 in terms of average F1 score and confidence intervals at 95% confidence level computed with 10 training phases.

The model with the highest F1 is the offline model, i.e., 99.96%. This approach, working on statistics, slightly outperforms both the FC and CNN approaches by 0.12% and 0.06%, respectively. This behavior can be traced back to the fact that working on statistics extracted from complete traffic flows gives the AI model deeper insights on the traffic. Instead, both the full-precision real-time models, working on 10 packets each time, have a less comprehensive view of the traffic flows. The CNN achieves a slightly higher F1 score, i.e., 0.06% with respect to its FC counterparts: the exploitation of spatial information helps identify between benign and malicious traffic. This validates the convolutional model as the best candidate for a photonic implementation, losing only 0.06% with respect

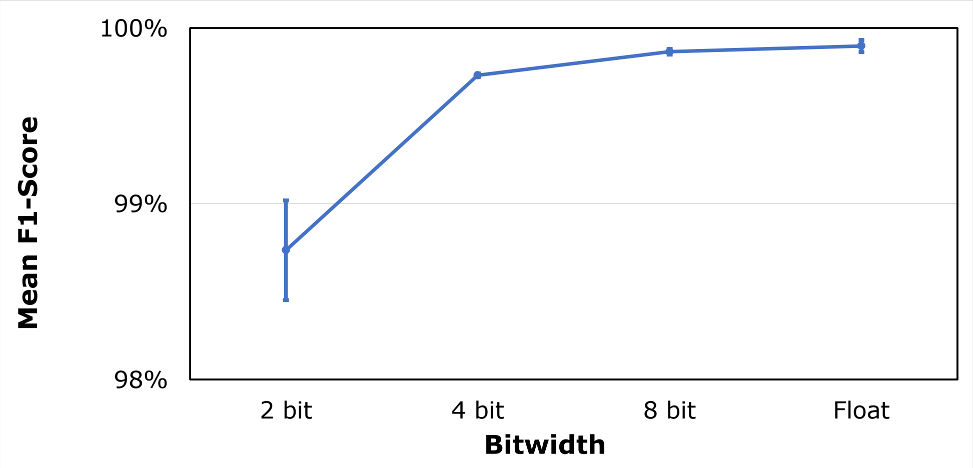

Fig. 4. PANN implementation of the CNN model. The floating-point baseline, i.e., electronic implementation, is also reported.

to the offline approach.

Based on these results, we focused on the theoretical performance of a PANN architecture leveraging the same architecture of the CNN model. As aforementioned, PANN architectures have a focus on the inference phase leveraging fixed-point with limited resolution. Thus we report the results for varying bitwidths compliant with photonic resolution constraint, i.e., $2, 4, 8$, in Fig. 4.

As expected, results show an F1-score increase for increasing bitwidth, even though the differences among the different resolutions are minimal. The highest value is obtained by exploiting 8 bits, i.e., 99.87%, with a 0.03% decrease with respect to the floating-point baseline. Exploiting a higher number of bits can indeed alleviate the loss due to the quantization. Further evidence for this is provided by the 2-bit scenario that reaches the lowest F1-score, i.e., 98.74%.When comparing the 4 and 8-bit experiments with each other, the F1-score decrease is only 0.14%.

Finally, photonic architectures suffer from a trade-off between computing speed and resolution: a higher operation speed leads to a lower bit resolution [27]. Thus, the 4-bit configurations can leverage all the advantages of photonic accelerators, without experiencing a high drop of speed [31] , while achieving good F1-score values.

V. Conclusion

In this paper, to embrace the 6G vision of distributed autonomy we proposed AI-based solution to distinguish between normal and malicious packets at the base station level. This architecture, while fulfilling intelligent security, is compatible with the RAN-CN convergence that will be deployed in 6G networks. Moreover, thanks to the exploitation of photonic hardware, the neuromorphic computations will satisfy 6G green computing requirements, outperforming electronic counterparts in terms of energy, speed, and computing density.

Experiments exploiting FC architectures and CNN architectures reach good F1-score values, with the latter slightly outperforming the former. Thus, we have proposed a PANN architecture based on convolutions, highlighting the trade-off between bitwidths and performance, i.e., both in terms of speed and accuracy. In this scenario, the best compromise is given by the 4-bit experiment.

Finally, this work is a first demonstration of how two key enabling technologies, i.e., photonics and AI, can be leveraged to fulfill 6G vision. To really have 6G-proof base stations, other aspects should be examined. Among the others, creating a federation of nodes that can collaboratively learn will be of primary importance for a sustainable and intelligent 6G network.

Acknowledgments

This research is partly supported by the project CLEVER (project number 101097560). The project is supported by the Key Digital Technologies Joint Undertaking and its members (including top-up funding by the Italian Ministry of Research and University (MUR)). This work is also partially supported by the European Union under the Italian National Recovery and Resilience Plan (NRRP) of NextGenerationEU, partnership on “Telecommunications of the Future” (PE00000001 - program “RESTART”). This work is partially supported by the project Smart Computing and Communication at the Edge (SMARTCOM) funded by Sma-RTy and CNR.

References

[1] H. Viswanathan and P. E. Mogensen, “Communications in the 6G era,” *IEEE Access*, vol. 8, pp. 57 063–57 074, 2020.

[2] J. Cha, Y. Moon, S. Cho, D. Kim, J. Choi, J. Jung, J. Lee, and S. Choi, “RAN-CN Converged User-Plane for 6G Cellular Networks,” in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 2843–2848.

[3] Q. Xu, Z. Su, and R. Li, “Security and Privacy in Artificial Intelligence-Enabled 6G,” *IEEE Network*, vol. 36, no. 5, pp. 188–196, 2022.

[4] J. Choi, N. Sharma, S. S. Gantha, V. Mandawaria, J. Cha, D. Kim, J. Jung, J. Lee, and S. Choi, “RAN-CN Converged Control-Plane for 6G Cellular Networks,” in *GLOBECOM 2022-2022 IEEE Global Communications Conference*. IEEE, 2022, pp. 1253–1258.

[5] W. Saad, M. Bennis, and M. Chen, “A vision of 6G wireless systems: Applications, trends, technologies, and open research problems,” *IEEE network*, vol. 34, no. 3, pp. 134–142, 2019.

[6] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, “What should 6G be?” *Nature Electronics*, vol. 3, no. 1, pp. 20–29, 2020.

[7] M. Božanić, S. Sinha, M. Božanić, and S. Sinha, "6G: The green network," *Mobile Communication Networks: 5G and a Vision of 6G*, pp. 189–220, 2021.

[8] B. Mao, F. Tang, Y. Kawamoto, and N. Kato, "AI models for green communications towards 6G," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 210–247, 2021.

[9] L. De Marinis, M. Cococcioni, P. Castoldi, and N. Andriolli, "Photonic neural networks: A survey," *IEEE Access*, vol. 7, pp. 175 827–175 841, 2019.

[10] G. Giamougiannis, A. Tsakyridis, M. Moralis-Pegios, G. Mourgias-Alexandris, A. R. Totovic, G. Dabos, M. Kirtas, N. Passalis, A. Tefas, D. Kalavrouziotis *et al.*, "Neuromorphic silicon photonics with 50 GHz tiled matrix multiplication for deep-learning applications," *Advanced Photonics*, vol. 5, no. 1, p. 016004, 2023.

[11] M. A. Nahmias, T. F. de Lima, A. N. Tait, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, "Photonic Multiply-Accumulate Operations for Neural Networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–18, 2020.

[12] A. Cem, S. Yan, Y. Ding, D. Zibar, and F. Da Ros, "Data-driven Modeling of Mach-Zehnder Interferometer-based Optical Matrix Multipliers," *arXiv preprint arXiv:2210.09171*, 2022.

[13] E. Paolini, L. De Marinis, M. Cococcioni, L. Valcarenghi, L. Maggiani, and N. Andriolli, "Photonic-aware neural networks," *Neural Computing and Applications*, vol. 34, no. 18, pp. 15 589–15 601, 2022.

[14] A. E. Cil, K. Yildiz, and A. Buldu, "Detection of DDoS attacks with feed forward based deep neural network model," *Expert Systems with Applications*, vol. 169, p. 114520, 2021.

[15] S. Singh and S. Jayakumar, "DDoS attack detection in SDN: optimized deep convolutional neural network with optimal feature set," *Wireless Personal Communications*, vol. 125, no. 3, pp. 2781–2797, 2022.

[16] L. Ioannou and S. A. Fahmy, "Network intrusion detection using neural networks on FPGA SoCs," in *2019 29th International Conference on Field Programmable Logic and Applications (FPL)*. IEEE, 2019, pp. 232–238.

[17] H. Moudoud, L. Khoukhi, and S. Cherkaoui, "Prediction and detection of fdia and ddos attacks in 5g enabled iot," *IEEE Network*, vol. 35, no. 2, pp. 194–201, 2020.

[18] S. Sevgican, M. Turan, K. Gökarslan, H. B. Yilmaz, and T. Tugcu, "Intelligent network data analytics function in 5G cellular networks using machine learning," *Journal of Communications and Networks*, vol. 22, no. 3, pp. 269–280, 2020.

[19] A. B. de Neira, B. Kantarci, and M. Nogueira, "Distributed denial of service attack prediction: Challenges, open issues and opportunities," *Computer Networks*, p. 109553, 2023.

[20] R. Doriguzzi-Corin, S. Millar, S. Scott-Hayward, J. Martinez-del Rincon, and D. Siracusa, "LUCID: A practical, lightweight deep learning solution for DDoS attack detection," *IEEE Transactions on Network and Service Management*, vol. 17, no. 2, pp. 876–889, 2020.

[21] M. A. Nahmias, T. F. De Lima, A. N. Tait, H.-T. Peng, B. J. Shastri, and P. R. Prucnal, "Photonic multiply-accumulate operations for neural networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–18, 2019.

[22] T. Huang, W. Yang, J. Wu, J. Ma, X. Zhang, and D. Zhang, "A survey on green 6G network: Architecture and technologies," *IEEE access*, vol. 7, pp. 175 758–175 768, 2019.

[23] L. De Marinis, M. Cococcioni, O. Liboiron-Ladouceur, G. Contestabile, P. Castoldi, and N. Andriolli, "Photonic integrated reconfigurable linear processors as neural network accelerators," *Applied Sciences*, vol. 11, no. 13, p. 6232, 2021.

[24] B. J. Shastri, A. N. Tait, T. Ferreira de Lima, W. H. Pernice, H. Bhaskaran, C. D. Wright, and P. R. Prucnal, "Photonics for artificial intelligence and neuromorphic computing," *Nature Photonics*, vol. 15, no. 2, pp. 102–114, 2021.

[25] G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. Miller, and D. Psaltis, "Inference in artificial intelligence with deep optics and photonics," *Nature*, vol. 588, no. 7836, pp. 39–47, 2020.

[26] T. F. de Lima, A. N. Tait, H. Saeidi, M. A. Nahmias, H.-T. Peng, S. Abbaslou, B. J. Shastri, and P. R. Prucnal, "Noise Analysis of Photonic Modulator Neurons," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 26, no. 1, pp. 1–9, 2019.

[27] L. De Marinis, A. Catania, P. Castoldi, G. Contestabile, P. Bruschi, M. Piotto, and N. Andriolli, "A Codesigned Integrated Photonic Electronic Neuron," *IEEE Journal of Quantum Electronics*, pp. 1–1, 2022.

[28] M. Wang, S. Rasoulinezhad, P. H. Leong, and H. K.-H. So, "NITI: Training Integer Neural Networks Using Integer-only Arithmetic," *IEEE Transactions on Parallel and Distributed Systems*, pp. 1–1, 2022.

[29] S. Zhou, Z. Ni, X. Zhou, H. Wen, Y. Wu, and Y. Zou, "DoReFa-Net: Training Low Bitwidth Convolutional Neural Networks with Low Bitwidth Gradients," *CoRR*, vol. abs/1606.06160, 2016. [Online]. Available: http://arxiv.org/abs/1606.06160

[30] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSp*, vol. 1, pp. 108–116, 2018.

[31] M. Miscuglio and V. J. Sorger, "Photonic tensor cores for machine learning," *Applied Physics Reviews*, vol. 7, no. 3, p. 031404, 2020.