

Analysis of Integration Technologies for High-Speed Analog Neuromorphic Photonics

Lorenzo De Marinis, *Member, IEEE*, Nicola Andriolli, *Senior Member, IEEE*,
and Giampiero Contestabile

Abstract—While the use of graphic processing units fueled the success of artificial intelligence models, their future evolution will likely require overcoming the speed and energy efficiency limitations of current implementations with the use of specialized neuromorphic hardware. In this scenario, neuromorphic photonic processors have recently proved to be a feasible solution.

In this paper, we first discuss basic analog photonic processing elements based on Mach-Zehnder modulators and assess their effective bit resolution. Then we evaluate how different photonic integration technologies affect the performance and the scalability of analog optical processors, in order to provide a clearer path toward real-world implementations of such engines. To this aim, we focus our analysis on the silicon on insulator (SOI), lithium niobate on insulator (LNOI), and indium phosphide (InP) platforms. In particular, we have numerically evaluated the performance of the Photonic Electronic Multiply-Accumulate Neuron (PEMAN) and its tensorial version, both based on Mach-Zehnder modulators, with the three technologies in terms of resolution, energy efficiency, and footprint efficiency.

LNOI modulators achieve the best resolution at high speed, with 4.3 bits at 56 GMAC/s for the single PEMAN and 3.6 bits at 224 GMAC/S for the tensorial version. The energy consumption in InP and LNOI platforms is the lowest, accounting for just 13.2 pJ/MAC and 4.6 pJ/MAC for the single and tensorial PEMAN, respectively. Nonetheless, SOI devices outperform the others in terms of footprint efficiency, reaching 18.6 GMAC/s/mm² in the single-neuron version and 29.6 GMAC/s/mm² in the tensorial version.

Index Terms—photonic analog computing, photonic neural networks, photonic integration technologies.

I. INTRODUCTION

MODERN computers are general-purpose devices that process information encoded in a symbolic form through a chain of serial logic-based instructions. These digital processors are typically based on the Von-Neumann architecture, where a central processing unit (CPU) fetches and executes instructions stored in a separated memory [1]. While this architecture is well suited to run sequential and iterative algorithms with high-precision data representation, recent breakthroughs in deep learning (DL) sprouted the research toward alternative computing paradigms [2].

Deep neural networks (DNNs), the core elements of DL, exploit a parallel computing strategy, with thousands of simple primitives (i.e., artificial neurons) concurrently working in interconnected structures [3]. The distributed architecture of DNNs cannot be efficiently implemented by

means of sequential instructions in a CPU. The use of graphic processing units (GPUs) has allowed a higher degree of parallelism in DNN processing, and thus a significant throughput increase of DL models [4]. However, there are two main issues when using CPUs/GPUs in this context:

The data movement problem: Logic operations in digital processors are energy-efficient, while data loading/offloading from memories constitutes the main source of energy consumption even in very specialized hardware [5]. Forefront GPUs rely on a parallelization strategy, with hundreds or thousands of cores processing concurrently, whose communication relies on complex high-speed and energy-draining interconnect fabrics.

The hardware lottery: Introduced by Hooker [6], this term describes a research idea that is successful thanks to the compatibility with the available hardware and software, not because of its superiority over other research directions. The compatibility with GPUs was key for the current success of DNNs. However, the use of such hardware could hamper the development of novel AI paradigms.

Nowadays, the main strategy adopted to advance DL models is to exploit increasingly fast and interconnected GPUs to support larger and larger DNNs. In this way, models such as OpenAI chatGPT [7], the most advanced dialog and code generator model, or DeepMind AlphaFold [8], which predicts protein structures with unprecedented accuracy, have been developed. However, this “bigger is better” paradigm cannot be indefinitely followed for two main reasons: (i) the energy consumed by DNNs is already hardly sustainable, just training huge models emits as much carbon as five cars in their lifetime [9]; (ii) their performance cannot be indefinitely increased just by increasing the computing capacity [6]. DL models cannot be easily deployed in application scenarios where the footprint and the energy consumption of GPUs can not be easily managed, such as in edge computing [10]. For these reasons, DL is facing a new hardware bottleneck.

In this context, analog computing is experiencing a renaissance related to the development of compact and efficient DL accelerators [4]. Analog devices rely on a physical implementation of artificial neurons to realize processors with a parallel and in-memory compute strategy [11], [12]. Electronic memristor crossbar arrays are a class of such accelerators, where simple resistive tunable elements (memristors) are arranged in square meshes to encode synaptic

L. De Marinis and G. Contestabile are with Scuola Superiore Sant’Anna, Pisa, Italy.

N. Andriolli is with CNR-IEIT, Pisa, Italy

Manuscript received XXX; revised XXX.

connections and perform distributed computations [2]. Among all analog approaches, photonics-based ones have the potential to meet the power consumption, bandwidth, and latency requirements for DNN deployment, being a versatile technology able to develop a notable variety of neural architectures comprising convolutional DNN, spiking neural networks, and reservoir computing [13], [14]. Leveraging the growing maturity of integration platforms, photonic integrated circuits (PICs) are widely used for analog information processing [15] with a growing number of PICs proposed for neuromorphic computing [16].

In this paper, we discuss the impact of photonic technologies on the performance and scalability of analog optical processors. The goal is to delineate a path toward real-world implementations of these processors, whose potentiality has been proven in recent demonstrations. The paper is organized as follows: in Sec. II we outline the role of photonics in neuromorphic computing and the resolution limits of basic photonic processors exploiting either single-output or X-coupled Mach-Zehnder modulators (MZM). Sec. III discusses the most relevant features of several promising integration platforms, and their impact on the development and performance of analog photonic accelerators. In particular, we considered the silicon on insulator (SOI), lithium niobate on insulator (LNOI), and indium phosphide (InP) technologies, as they provide sufficiently mature high-speed modulators. To assess the impact of the integration platform on the performance of neuromorphic accelerator devices, in Sec. IV we report the numerical evaluation of the photonic-electronic multiply-accumulate neuron architecture (PEMAN [17]) and the 4-neuron Tensor-PEMAN[18], designed with the different technologies. Three key performance indicators have been assessed, namely (i) the bit resolution, (ii) the energy per operation, and (iii) the footprint efficiency (FE). Finally, Sec. V concludes the paper.

II. PHOTONIC NEUROMORPHIC ANALOG COMPUTING

The idea of optical computing arose in the past century: already in 1965 Reimann and Kosonocky proposed the use of laser devices to build a digital processor with the aim to overcome the interconnection problems of electrical circuitry [19]. Since then, several schemes and demonstrations of optical logic gates with bulk and integrated implementations have been developed, as well as photonic memories [20]. However, the benefits arising from the miniaturization of electronic components outpaced the ones promised by optics, and the idea of a photonic digital processor faded [13].

In the meanwhile, photonics has proven to be *the* technology for long-distance communications. The advances in photonic technologies have significantly increased the transmission capacity of optical transceivers while reducing their cost and power consumption. Today optics is used also in short-reach intra-datacenter communications, exchanging information at an energy cost comparable to a DRAM-based CPU communication [21].

The recent advances in artificial intelligence (AI) reignited the interest in photonic computing in the analog domain, as

discussed in Sec. I. Compared to analog electronics, optics can support large structures with long interconnections as it is not affected by skin and inductance effects introducing distortions and limiting the operation frequency. Analog photonics based on passive elements allows to build distributed computing structures without dynamic power consumption beyond input generation (transmitter) and output collection (receiver). In these parallelized structures, the computation time is constant with respect to the number of inputs n , corresponding to the light time of flight. In the recent past, many photonic neuromorphic devices have been reported in the literature [13] implementing a great variety of architectures, such as convolutional, feed-forward, and spiking neural networks, and reservoir computing [22], [23]. Nonetheless, the increased throughput and power efficiency brought by photonic accelerators come at a cost of some limitations in the neural network model design [24]. The number of inputs per neuron can vary from a few in all-optical approaches to several hundreds in electro-optic solutions [17]. Also, the depth of the architectures (i.e., the number of neural layers) and the implementable nonlinearities vary significantly [22].

In typical optical neuromorphic architectures, the most relevant limiting factor is noise, which accumulates and produces unwanted fluctuations reducing the appreciable variations in signals. Together with noise, also distortions limit the accuracy at which values can be distinguished as a consequence of the operations performed in an analog processor. The number of different values that can be resolved defines the analog system resolution. The latter is usually quantified by the Effective Number of Bits (ENOB), which denotes the number of bits required to digitally store the value. Hence, DL models designed to be run on analog accelerators cannot exploit the common floating point number representation. Instead, trained models should exploit integer parameters, with a bitwidth resembling the ENOB of the analog hardware. Considering these hardware constraints, a proper design of DL models is pivotal. The recent literature reports some practical strategies to this end [24], [25], [26].

The most critical aspect of photonic processors concerns their high-speed stages, essential to fully leverage the bandwidth and latency advantages of optical approaches. For this reason, at least one stage for high-speed input generation is necessary, which ultimately sets the limits of resolution, speed, and power consumption of the whole system. Despite this pivotal role, fast input generation is often not adequately discussed in the literature.

In the following, we discuss in depth how noise, distortion, and bandwidth affect the bit resolution of optical signals when generated by high-speed modulators as inputs for analog neuromorphic processors. In particular, we consider both (i) single-output and (ii) X-coupled MZM acting on the intensity of a continuous-wave laser source, and terminated on a single and balanced photodetector, respectively.

A. Single-Output Mach-Zehnder Modulator

An effective metric to account for both noise and distortions in analog devices is the Spurious-Free Dynamic Range

(SFDR). This represents the ratio between the signal power and the power of distortions when the latter equals the power of noise [27]. Hence, the SFDR can be regarded as the maximum achievable signal-to-noise and distortion ratio, and can be thus related to the maximum ENOB through the equation [28]:

$$\text{ENOB} = \frac{\text{SFDR}_{\text{dB}} - 1.76}{6.02} \quad (1)$$

where the SFDR value is represented in decibels. To quantify the SFDR, a two-tone test on the modulating device can be performed to assess the intermodulation distortions, of which the third order one is the most critical as its components usually fall within the bandwidth of the signal [27]. Noise is usually independent of the modulating device and three sources of noise can be distinguished: (i) thermal noise, (ii) shot noise, and (iii) laser relative intensity noise (RIN). The first one derives from the thermal agitation of electrons in conductors at thermal equilibrium, arising at the transmitting and receiving devices in photonic circuits. The shot noise derives from the quantum nature of light, reflecting the random arrival of photons at the PD. The spontaneous emission in the laser source causes the RIN, which we consider frequency- and temperature-independent for simplicity. Their powers per unit bandwidth [W/Hz] can be expressed as:

$$\begin{aligned} p_{th} &= k_B T \\ p_{sh} &= \frac{1}{2} q I_0 R \\ p_{rin} &= 10^{\frac{\text{RIN}}{10}} I_0^2 R \end{aligned} \quad (2)$$

where p_{th} , p_{sh} and p_{rin} are the powers of thermal, shot, and RIN noises, respectively, k_B is the Boltzmann constant, R is the load resistance at the PD side, q is the electron charge, I_0 is the average power flowing in the load resistance, and RIN is the relative laser noise. I_0 depends directly from the input laser power with the relation $I_0 = r_{PD} P_i / 2L$, where P_i is the laser power, r_{PD} is the PD responsivity and L are the losses between laser and PD. Considering the noise sources of Eq. (2), the SFDR is derived through the power intercept point between the fundamental tone and the intermodulation distortion of interest [27], i.e., the point in which the power of the fundamental and the distortion components are equal. The SFDR when using a single-output (SO) MZM is then:

$$\text{SFDR}_{\text{SO}} [\text{dB Hz}^{\frac{2}{3}}] = \frac{2}{3} \left[10 \log_{10} (I_0^2 R) - 10 \log_{10} \left((1+g)k_B T + \frac{qI_0 R}{2} + 10^{\frac{\text{RIN}}{10}} \frac{I_0^2 R}{4} \right) \right] \quad (3)$$

where the factor g accounts for the thermal noise of the transmitter part, $g \leq 1$ in absence of amplification and is usually negligible. The second term between square brackets limits the SFDR growth and, as the largest term dominates in the logarithm of a sum, three noise regimes can be identified. The first one is the thermal noise regime, which is in the low laser power range. Here, increasing the laser power gives a quadratic growth of the SFDR. The second regime is

established when the shot noise component becomes dominant. The SFDR still increases with the pump power but in a linear fashion, as also the shot noise depends on the average PD current I_0 . Eventually, the system enters the RIN regime, where the SFDR saturates.

To quantify how the analog resolution varies with the laser power, we consider a realistic scenario with the following parameters: $R = 50 \Omega$, $T = 290 \text{ K}$, $L = 10$, $r_{PD} = 0.8 \text{ A/W}$, and $\text{RIN} = -160 \text{ dB/Hz}$. Fig. 1 shows the SFDR of a single-output MZM configuration as a function of input laser power with a dashed line, together with the contribution of thermal noise (red line), shot noise (blue line) and RIN (green line). The red line slope is double the blue line slope, reflecting the quadratic increase of the thermal noise contribution related to the laser pump power while for shot noise the increase is linear. The green line is horizontal, as RIN is independent of input laser power, and gives an upper bound. The right vertical axis of the figure reports the bit resolution at 10 GHz, which is evaluated through Eq. 1 and the formula $\text{SFDR}(B \text{ Hz}) = \text{SFDR}(1 \text{ Hz}) - \frac{2}{3} 10 \log(B)$, where B is the bandwidth [27].

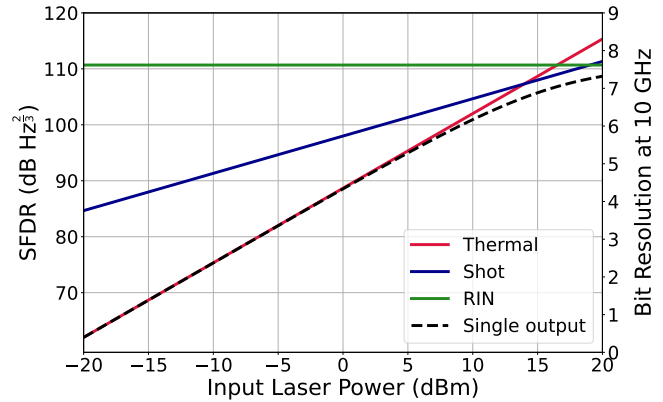


Fig. 1: SFDR of a single-output MZM as a function of input laser power (dashed line). Thermal, shot and RIN components are depicted with a red, blue and green line, respectively.

In the considered case, with a rather low RIN value of -160 dB/Hz , the maximum resolution at 10 GHz is close to 8 bits. However, this requires a large laser power, which contrasts with the target of energy efficiency that underpins the use of photonic neuromorphic processors. In practice, the input laser power varies in the range 0-10 dBm, where the resolution at 10 GHz goes from 4 to slightly above 6 bits. This trend outlines an inherent trade-off between resolution, power, and bandwidth in photonic analog processors, where a higher resolution can be achieved by increasing the power consumption or decreasing the bandwidth, i.e., the computing speed.

B. X-Coupled Mach-Zehnder Modulator

A dual-output (or X-coupled) MZM followed by balanced photodetectors can be used to reduce the impact of RIN, with reported results of 24 dB RIN suppression [29]. Indeed, the

common mode rejection deriving from the balanced detection suppresses the RIN associated with the DC component if the MZM is biased in quadrature. The RIN is not entirely suppressed since the component related to the modulating signal remains. Moreover, balanced detection allows for the representation of negative numbers and gives a theoretical 6 dB output power advantage compared to the single PD case, as the received current doubles with the same received power. However, the shot noise doubles as well, due to the contribution of both detectors. The total SFDR becomes [16]:

$$\text{SFDR}_X[\text{dB Hz}^{\frac{2}{3}}] = \frac{2}{3} \left[10 \log_{10} (4I_0^2 R) - 10 \log_{10} \left((1+g)k_B T + qI_0 R + 10^{\frac{\text{RIN}_X}{10}} I_0^2 R \right) \right] \quad (4)$$

where I_0 is the average photocurrent for the single-output MZM case and RIN_X is the suppressed RIN value in logarithmic scale. Fig. 2 reports the SFDR and the bit resolution at 10 GHz for this configuration (dashed line), considering the following parameters: $R = 50 \Omega$, $T = 290 \text{ K}$, $L = 10$, $r_{PD} = 0.8 \text{ A/W}$, $\text{RIN} = -150 \text{ dB/Hz}$ and a RIN suppression of 20 dB (i.e., $\text{RIN}_X = -170 \text{ dB/Hz}$). The three noise contributions are depicted with a red (thermal), a blue (shot), and a green (RIN) line. The bit resolution at 10 GHz exceeds 8 bits at high power, while in the typical 0-10 dBm input laser power range, the bit resolution goes from ~ 4.5 to almost 7 bits. The graph also reports with a black solid line the total SFDR for the same parameters in the case of a single-output MZM. The X-coupled device provides a 6-dB gain in SFDR up to 10 dBm, where the single-output scenario reaches the RIN regime and the advantage increases further.

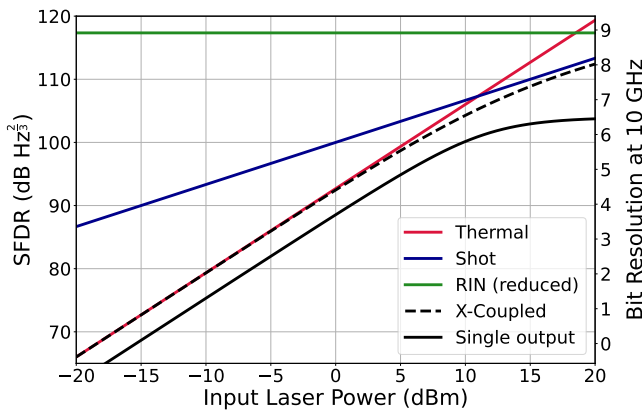


Fig. 2: SFDR of a X-coupled MZM as a function of input laser power (dashed line). Thermal, shot, and RIN components are depicted with a red, blue, and green line, respectively. The black solid line represents the SFDR of a single-output MZM with the same parameters.

III. PHOTONIC INTEGRATION TECHNOLOGIES

The technological platform used to implement the neuromorphic photonic devices affects the performance of

the overall processor. The previous section has discussed the impact of the analog photonic processor parameters on the achievable bit resolution. This metric is of paramount importance as is directly related to the accuracy of DL models running on the photonic hardware [24]. However, several other aspects play an important role in the practical implementation of DL photonic accelerators. In this section, we discuss the main features of SOI, LNOI, and InP photonic platforms affecting the performance and the scalability of neuromorphic accelerators. These platforms have been chosen as they provide quite mature high-speed modulating devices. We evaluate the following key metrics: available optical components, required optical I/O, footprint usage, CMOS compatibility, and losses in waveguides. Additionally, we report four important metrics regarding high-speed modulators since they are critical for the power efficiency and the bit resolution of analog processors. The metrics related to the MZM are insertion loss (IL), bandwidth, half-wave voltage, and dynamic ER.

A. Silicon on Insulator

The silicon photonics industry has seen a formidable development in the last two decades, pushed by the datacom and telecom fields [30]. The SOI technology is inherently compatible with the CMOS process as it exploits the standard silicon fabrication techniques. Currently, the SOI platform is a high-yield, robust, and mature process. A relevant aspect of the development of novel neuromorphic PICs is the availability of design and simulation software for SOI, from computer-aided design tools to optical simulators. Silicon photonics manufacturers usually provide process design kits (PDK) with compact models of their devices, allowing circuit simulation and design of novel architectures in a fast and predictable way [31].

The high refractive index of silicon allows photonic structures to be very compact [30] and connected with tight bends (radii of curvature even below $10 \mu\text{m}$), while propagation losses are typically in the range 2-3 dB/cm. Fast modulators can be developed relying on the plasma dispersion effect of PN junctions, usually with depletion modulation of reverse PN junctions, whose 3-dB bandwidth hardly exceeds 50 GHz. High-speed modulators exhibit rather high losses and a rather low dynamic extinction ratio (ER), both reducing resolution. The MZM length is usually a couple of millimeters, resulting in rather compact devices [32]. The linearity of SOI integrated modulators is moderate, characterized by an SFDR more than 20 dB below the one of bulk LNOI modulators. The linearity can be improved at a cost of more complex electronic and photonic hardware [33]. The quite large half-wave voltage (V_π) requires RF drivers, which impacts the circuitry complexity and the energy consumption [31]. High-speed PDs are embedded in the waveguides thanks to germanium processing [34]. Silicon is an indirect band-gap semiconductor, which precludes the development of practical lasers, a fundamental building block also for neuromorphic applications. SOI PICs require external light sources, with the consequent need for optical I/O. Here, the high refractive index of silicon is a disadvantage, causing a large optical mode

TABLE I: Comparison between SOI, LNOI, and InP photonic integration technologies.

Platform	Optical Components	Optical I/O	Footprint	CMOS Compatibility	Waveguide Loss	MZM IL	MZM Bandwidth	MZM V_π	MZM Dynamic ER
Silicon on Insulator	Passives, PD and phase modulators	Input laser	Lowest	Yes	2-3 dB/cm	~ 6 dB/MZM	~ 50 GHz	3-6 V	< 10 dB
Lithium Niobate on Insulator	Passives and phase modulators	Input laser output PD	Highest	No	< 1 dB/cm	~ 1 dB/MZM	> 100 GHz	< 1.6 V	> 20 dB
Indium Phosphide	Passives and all actives	Not required	Moderate	No	1-2 dB/cm	~ 6 dB/MZM	> 70 GHz	< 1 V	< 10 dB

and refractive index mismatch with optical fibers. Despite the advances in focusing grating and edge couplers, this remains a critical aspect for the losses [31]. To reduce input losses, a valid alternative is the hybrid integration of III-V lasers on silicon PICs, which has seen notable progress in terms of reliability and robustness [35].

B. Lithium Niobate on Insulator

Lithium niobate has been sometimes referred to as the ‘‘Silicon of Photonics’’, to remark its importance in photonics equivalent to silicon in electronics. Lithium niobate combines several features that make this platform particularly suitable for photonic analog computing [18]. For once, propagation losses have been demonstrated to be lower than 0.3 dB/cm consistently across 6-in. thin film wafers, showing a high process maturity [36]. Remarkably low losses have been demonstrated also in traveling-wave (TW) MZMs (< 1 dB/MZM) with an electro-optic bandwidth exceeding the hundred of GHz [37]. The strong Pockels effect in LNOI enables very low half-wave voltages ($V_\pi < 1.6$ V), compatible with conventional CMOS electronics voltages. Indeed, high-speed DACs can directly drive LNOI modulators, avoiding the use of cumbersome, power-hungry and potentially distorting RF amplifiers. Dynamic ERs above 20 dB have been demonstrated in LNOI MZMs [36].

The LNOI platform has also some disadvantages, stemming from a rather large footprint of devices, with MZM length falling in the tens of mm range. For this reason, complex circuits developed with LNOI are not very cost-effective. Additionally, a light source is missing in LNOI, leading to similar considerations as for SOI. In this platform also photodetectors are still missing, thus practically requiring optical I/O handling both for input and output. Nonetheless, demonstrations of hybrid integration of InGaAs/InP high-speed PDs [38], on-chip LNOI photodetectors [39], and lasers [40], [41] are promising to fill this gap in the foreseeable future.

The SFDR in devices demonstrated so far is not optimal, around $100 \text{ dB Hz}^{2/3}$, but slightly larger compared to bulk modulators [36]. As this technology is still at an early stage of development, improvements are expected in terms of SFDR, bandwidth, ER, and V_π [36]. This does not hold for the modulator length, which is likely to prevent the implementation of very complex all-optical neuromorphic circuits in this platform [16].

C. Indium Phosphide

Despite not being CMOS compatible, the InP platform is a common choice when dealing with complex PIC architectures [42]. Indium phosphide is the only established photonic platform enabling the monolithic integration of all passive elements (i.e., waveguides, couplers, filters...) and all active components, such as lasers, semiconductor optical amplifiers (SOA), PDs, and high-speed modulators [43]. This monolithic integration significantly reduces packaging costs, avoiding the need for optical I/O and the related coupling losses, while improving the overall system reliability [44].

InP high-speed modulators have been reported with bandwidths exceeding 70 GHz and low half-wave voltages, compatible with direct DAC driving [45]. A low V_π below 1 V has been demonstrated with a bandwidth of 67 GHz already in 2014 [46]. Thus, CMOS DACs can be directly used to drive the modulators avoiding bulky and power-hungry RF drivers, as in the LNOI case. With this technology, optical losses fall in the 2 dB/cm range, comparable with SOI, but one order of magnitude larger than LNOI. Lower propagation losses have been reported exploiting specific process steps, with a notable 0.4 dB/cm [47]. Additionally, SOAs make this platform able to recover losses and, in neuromorphic chips, to implement diagonal matrices with gain [16]. Moreover, the non-linear behavior of SOAs can be exploited to both recover power and apply an activation function in all-optical neural networks, as demonstrated in [48].

Still, this platform presents some drawbacks. For instance, the bending radius of InP waveguides is more than ten times larger than SOI waveguides. The transition between active and passive regions requires proper converters, the modulator lengths fall in the mm range, longer than SOI modulators, but shorter than LNOI ones. The dynamic ER (similar to SOI) is moderate, with values below 10 dB in the tens of GHz, limiting the dynamic range and thus the resolution. The RIN of InP integrated lasers usually falls near -140 dB/Hz , also limiting the maximum ENOB, as discussed in Sec. II, while SOAs have noise factors above 4 and usually require additional optical filters.

D. SOI, LNOI, and InP Comparison

Table I reports a comparison among SOI, LNOI, and InP photonic platforms, discussing the main features of interest for the design and fabrication of photonic neuromorphic devices. They encompass the available optical components, the required optical I/O, the chip footprint, the platform CMOS

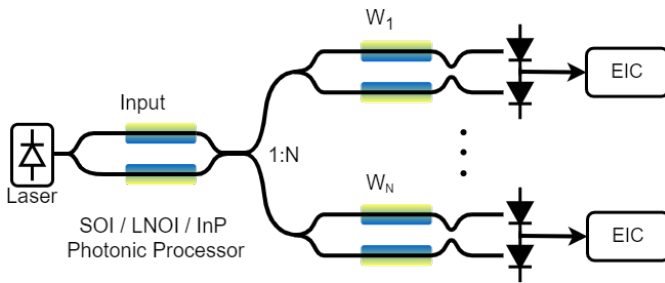


Fig. 3: The Tensor-PEMAN architecture, composed of a single-output MZM that broadcasts the input signal to N weighting X-coupled MZMs connected to as many PDs and electronic front-ends.

compatibility, and the waveguide losses. A focus is given to the metrics related to high-speed MZMs, as these are fundamental elements for high-speed analog signal generation and processing. In particular, the MZM insertion loss (IL), bandwidth, half-wave voltage (V_{π}), and dynamic ER have been considered.

The InP platform stands out for the available optical elements, as it allows a monolithic integration from the laser to PD. For this reason, InP PICs do not require optical I/O, and can thus avoid the associated optical losses while reducing the complexity of chip packaging. The SOI platform lacks SOAs and lasers, hence it requires at least managing a laser input. Nonetheless, the advancements in hybrid-integration of III-V lasers on silicon PICs give practical solutions to simplify packaging and reduce insertion losses [35]. The LNOI platform lacks both light sources and integrated PDs, thus requiring the handling of optical I/O for both inputs and outputs. Despite being the most recent platform, the successful hybrid integration of both lasers and PDs in LNOI has been recently reported [40], [41].

The footprint of photonic devices impacts directly both the degree of architectural complexity that can be managed in a PIC and the CAPEX for chip fabrication. Combining compactness, CMOS compatibility, and substrate availability and cost, the SOI process dominates when considering a large-scale production of complex integrated chips.

The lithium niobate technology is a leap forward concerning the modulator metrics. LNOI waveguides have the lowest propagation losses, with demonstrated values below 0.1 dB/cm [40]. This allows LNOI modulators, whose length falls in the tens of mm range, to have a loss as low as 1 dB per element, four times lower compared to InP and SOI modulators. Microstructured LNOI electrodes guarantee an excellent matching between the effective index of the electrical signal and the group index of the optical signal, significantly increasing the bandwidth of modulators, in excess of 100 GHz [37]. The low half-wave voltage, the high linearity, and the dynamic ER of LNOI modulators make them the best option for low-power and high-speed analog signal generation.

IV. PERFORMANCE ON PHOTONIC NEUROMORPHIC PROCESSORS

Many photonic neuromorphic architectures can be realized stemming from the basic high-speed MZM discussed in Sec. II. The recent literature reports various photonic configurations able to implement different DNNs, most of them focusing either on fully-connected or convolutional layers [3], [13], [14]. The two main strategies to physically realize these layers rely either on coherent architectures, such as tunable interferometer meshes, or WDM devices, which encode inputs in multiple wavelengths and use filters to act selectively on them. These architectures usually exploit slowly-varying thermally-tunable elements to leverage the parallelism of photonics and perform the weighing part of several neurons in parallel on fast-varying inputs, which are produced by high-speed MZMs. Nonetheless, these approaches have a limited scalability imposed by the circuit complexity and the associated losses, which grow quadratically with the number of inputs [49]. In practical devices, the number of inputs per layer does not exceed ten, while projections on future devices are limited to 64 ports per chip [17].

To quantify the impact of different integration technologies on photonic neuromorphic devices, we have numerically evaluated the performance of a specific device: the Photonic Electronic Multiply-Accumulate Neuron (PEMAN) [17]. We have chosen this device as its processing core relies on two cascaded high-speed MZMs. Moreover, it allows to process a number of inputs per neuron ranging from hundreds to a thousand, thanks to its electro-optic strategy.

The PEMAN is a precision-scalable multiply-accumulate (MAC) architecture, able to perform in the analog domain all the operations required by an artificial neuron, i.e., multiplications, accumulations, and application of a nonlinearity. Fig. 3 represents the PEMAN in the case where $N = 1$, i.e., with one output branch. This optoelectronic neuron performs multiplications at high speed and low power exploiting a single-output and an X-coupled MZM that respectively impress neuron inputs and weights in the amplitude of a continuous lightwave with a time serial strategy. The X-coupled MZM, able to impress both positive and negative weights, is connected to a balanced photodetector. The output photocurrent encodes the multiplication results, which are accumulated within the front-end analog Electronic Integrated Circuit (EIC). A final nonlinearity is imposed within the Analog-to-Digital Converter (ADC), designed with a nonlinear characteristic. The obtained output can be readily re-used by the same structure, allowing to compute multiple DNN layers without cascading issues.

The structure can be scaled to perform the computations of N neurons in parallel by broadcasting the analog signal generated by the input modulator to N weighting units. Fig. 3 depicts this version of the device, referred to as Tensor-PEMAN [50]. The Tensor-PEMAN has an N -time advantage in the MAC rate: when operating at a given frequency, N MAC in parallel are executed. Moreover, the energy consumption per MAC operation is reduced, as the laser and the input

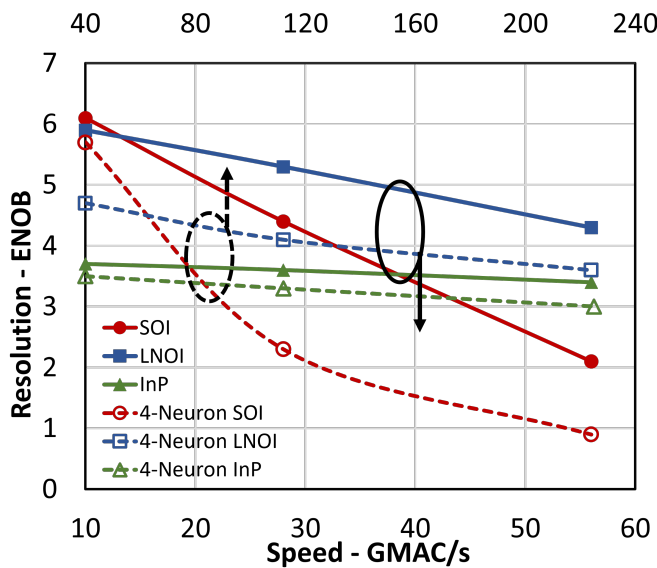


Fig. 4: Bit resolution as a function of speed for different technologies and parallel PEMAN neurons. Solid/dashed lines refer to the lower/upper axis.

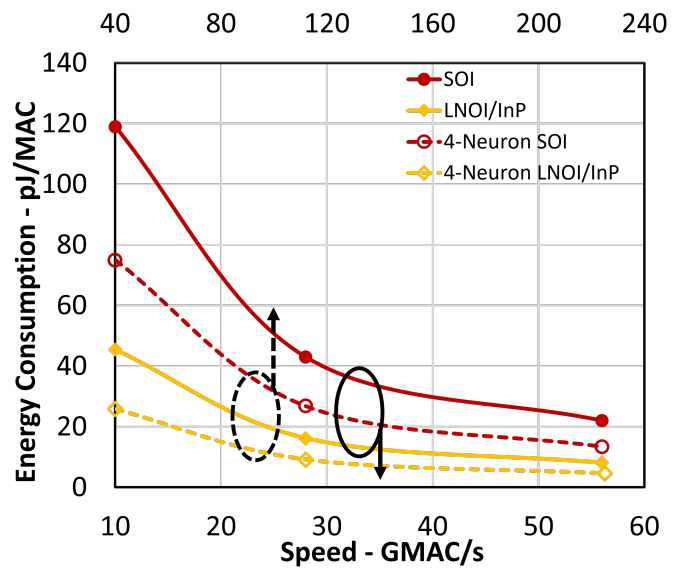


Fig. 5: Energy consumption as a function of speed for different technologies and parallel PEMAN neurons. Solid/dashed lines refer to the lower/upper axis.

modulator are shared among the N weighting elements. This comes at the cost of a more complex device and a lower optical power budget due to the power splitter, which translates into a lower resolution, as discussed in Sec. II. The Tensor-PEMAN computes the output of N neurons at a time, storing the N outputs on a local buffer to be used as inputs for the successive layer. In this way, the inputs for the computations within a layer do not need to be loaded at every cycle.

We have numerically evaluated the performance of the single PEMAN and the 4-neuron Tensor-PEMAN ($N = 4$) realized in different photonic platforms through the Lumerical INTERCONNECT environment. To this aim, we have performed frequency-dependant simulations evaluating three metrics: (i) the output signal resolution, (ii) the energy consumption (in pJ/MAC), and (iii) the footprint efficiency (in GMAC/s/mm²). The results are summarized in Fig. 4, 5, and 6, respectively. In the plots, the lower horizontal axis refers to the MAC rate of the single case (equivalent to the operating frequency of both the single-neuron and the 4-neuron instances), while the upper axis refers to the total MAC rate of the scaled structure. The maximum speed has been chosen to be 56/224 GMAC/s for the single-neuron and the 4-neuron instances, respectively, a value matching the bandwidth limitations of the SOI devices, whose output signals drown in noise beyond this point. The SOI devices have been simulated with the Imec iSiPP50G process [51], for the LNOI devices the parameters reported in [37] have been considered, while the InP implementations were simulated according to [45].

Fig. 4 shows the bit resolution as a function of the MAC rate. Time-domain simulations have been performed considering a dataset of 1024 input-weight pair multiplications, taking the photocurrent waveform as an output. The time traces have been analyzed to obtain the standard deviation of the multiplication error σ , then used to

derive the bit resolution with a 6σ metric. In all platforms, the resolution decreases with the MAC rate as the PEMAN trades off speed with resolution. LNOI devices achieve the best performance especially at high speed, with 4.3 bits at 56 GMAC/s for $N = 1$ and 3.6 bits at 224 GMAC/s for $N = 4$. The InP technology provides a rather stable ENOB, between 3 and 4 for both structures and all considered operating speeds. The relatively low resolution at lower frequencies is due to the low dynamic ER of the MZM and the high RIN of the laser, respectively. However, the InP device is mildly affected by frequency effects due to the large 3-dB bandwidth of the InP modulator, thus generating a quite stable resolution. Conversely, the finite bandwidth of the SOI devices particularly limits their resolution at higher speed. For the single neuron case, the ENOB drops from 6.1 bits at 10 GMAC/s to 2.1 at 56 GMAC/s. In the 4-neuron case, the SOI signal drowns in noise, i.e., goes below 1 ENOB, for the highest rate of 224 GMAC/s.

The energy consumption (in pJ/MAC) is depicted in Fig. 5 as a function of the speed. It has been evaluated considering 81 mW for the laser, 13 mW per EIC, 180 mW per high-speed DAC, and 400 mW per RF amplifier [17]. The RF amplifiers dissipate the highest share of power, which justifies the largest energy consumption for SOI devices. Conversely, both LNOI and InP modulators are compatible with direct DAC driving, thus they avoid the use of RF drivers and achieve the same energy consumption. All curves show a decreasing behavior for increasing speed, due to the more efficient spread of static energy dissipation. Indeed, the lowest energy consumption is achieved at a working frequency of 56 GHz (224 GMAC/s for the 4-neuron instance), where the single LNOI/InP device consumes just 8.1 pJ/MAC, a nearly three-times advantage compared to SOI, which requires 22 pJ/MAC. The Tensor-PEMAN shares laser and input MZM

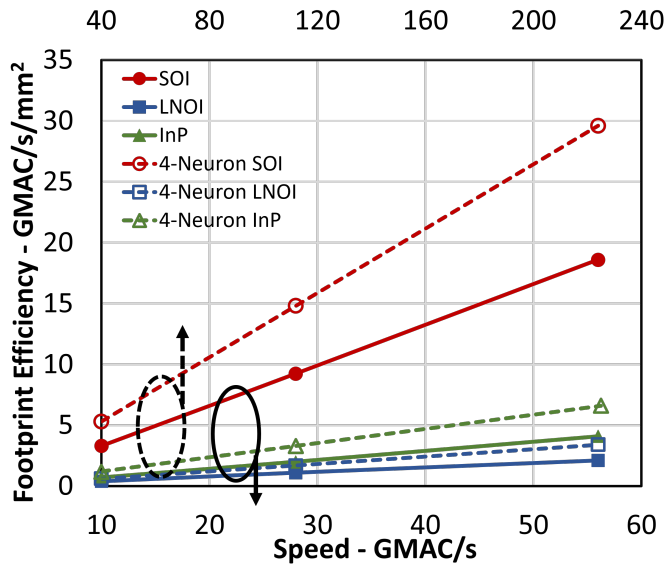


Fig. 6: Footprint efficiency as a function of speed for different technologies and parallel PEMAN neurons. Solid/dashed lines refer to the lower/upper axis.

among the 4 weighting structures, hence the power per MAC operation can be further reduced. This configuration has a power consumption of 13.2 pJ/MAC and 4.6 pJ/MAC for the SOI and LNOI/InP devices, respectively.

Results in terms of footprint efficiency as a function of MAC rate are shown in Fig. 6. This metric gives a deeper insight into the neuromorphic device performance compared to plain area usage since it quantifies how efficiently the chip is used to produce MAC operations per second [52]. The reported footprint efficiency values have been derived as the number of MAC/s per unit area of the photonic chip. For each design, this value linearly grows with the MAC rate. All Tensor-PEMAN implementations achieve a higher footprint efficiency with respect to the single case since the area occupied by the laser and the input MZM is shared by more neurons. At the highest working frequency of 56 GHz, the LNOI, InP, and SOI PEMAN achieve 2.1, 4.1, and 18.6 GMAC/s/mm², respectively. The much larger footprint efficiency of the SOI platform is mainly due to the larger index contrast and compactness of the MZM. Regarding the 4-neuron Tensor-PEMAN implementation, the LNOI, InP, and SOI devices reach the values of 3.4, 6.6, and 29.6 GMAC/s/mm², respectively. Also in this case the SOI platform stands out, with a 4× advantage with respect to InP and an almost 9× increase compared to LNOI.

V. CONCLUSIONS

The recent advances and demonstrations in photonic neuromorphic computing support the speed, latency, and power consumption advantages of optical AI accelerators. Moreover, such novel photonic hardware has the potential to enable non-conventional AI paradigms. Having proven the working principle, the current challenge is to provide a path toward real-world implementations of photonic

analog processors. In this paper, we have discussed how different photonic technologies affect the performance and the scalability of neuromorphic processors, in order to gain a deeper understanding of the most suited enabling technologies.

Our analysis focused on SOI, LNOI, and InP photonic integration platforms, as they provide sufficiently mature building blocks and in particular high-speed modulators, which are essential elements to exploit the bandwidth and latency advantages of photonic neuromorphic approaches. We motivated the use of analog photonics in this context and reported a theoretical analysis on the achievable resolution of MZM-based basic elements, depending on noise sources and distortions. We have then discussed the strengths and weaknesses of the three integration platforms. LNOI modulators overcome the others for each aspect except the footprint. A monolithic integration of all optical components can be achieved only by the InP technology, while the SOI platform has the most compact photonic structures and is the only CMOS compatible.

To assess how the different photonic platforms affect the performance of neuromorphic devices, we have numerically evaluated the operation of the PEMAN and the 4-neuron Tensor-PEMAN devices when designed with a specific technology. We have considered three performance metrics, i.e., bit resolution, energy consumption, and footprint efficiency as a function of the device MAC rate. Notably, all the Tensor-PEMAN implementations achieve a speed, power consumption, and footprint advantage compared to the single-neuron instance leveraging the optical parallelism, with an unavoidable resolution reduction due to the tighter power budget.

The best resolution has been achieved by the LNOI platform, with 4.3 bits at 56 GMAC/s for the single PEMAN and 3.6 bits at 224 GMAC/s for the 4-neuron instance. For energy consumption, both InP and LNOI perform best, requiring only 13.2 pJ/MAC and 4.6 pJ/MAC at the highest speed for the single and 4-neuron implementations, respectively. The SOI reduced energy efficiency arises from the need for RF amplifiers to drive the modulators, while LNOI and InP MZMs allow for direct DAC driving. Nonetheless, the SOI stands out for footprint efficiency, achieving 18.6 and 29.6 GMAC/s/mm² at the highest speed for the single and 4-neuron instances, respectively. The SOI achieves a 4-times advantage compared to InP, and nearly an order of magnitude over the LNOI.

These results highlight the features of each platform, with no technology clearly prevailing in the overall picture. Up to date, only the priority deriving from a specific application can delineate which platform to use. The LNOI will be the technology of choice if high-speed and high-resolution neuromorphic processing is required. InP provides the unique feature of monolithic integration, which avoids the need for optical I/O resulting in a simplified PIC design. However, in the roadmap for a wide deployment of photonic neuromorphic processors, SOI is the only technology that currently supports cost-effective and large-scale production. We expect that the advances in the performance, maturity, and robustness of the various platforms provide a clearer vision concerning the most

suitable photonic analog computing devices in the near future.

ACKNOWLEDGMENT

This work was supported in part by PNRR MUR project PE0000023-NQSTI, and in part by the Italian Ministry of Foreign Affairs and International Cooperation, grant number IN22GR06.

REFERENCES

- [1] D. Shin and H.-J. Yoo, "The heterogeneous deep neural network processor with a non-von Neumann architecture," *Proceedings of the IEEE*, vol. 108, no. 8, pp. 1245–1260, 2019.
- [2] W. Zhang, B. Gao, J. Tang, P. Yao, S. Yu, M.-F. Chang, H.-J. Yoo, H. Qian, and H. Wu, "Neuro-inspired computing chips," *Nature electronics*, vol. 3, no. 7, pp. 371–382, 2020.
- [3] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.
- [4] W. Haensch, T. Gokmen, and R. Puri, "The next generation of deep learning hardware: Analog computing," *Proceedings of the IEEE*, vol. 107, no. 1, pp. 108–122, 2018.
- [5] D. A. Miller, "Attojoule optoelectronics for low-energy information processing and communications," *Journal of Lightwave Technology*, vol. 35, no. 3, pp. 346–396, 2017.
- [6] S. Hooker, "The hardware lottery," *Communications of the ACM*, vol. 64, no. 12, pp. 58–65, 2021.
- [7] "ChatGPT: Optimizing Language Models for Dialogue," <https://openai.com/blog/chatgpt/>, 2022, accessed: 2023-02-03.
- [8] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko *et al.*, "Highly accurate protein structure prediction with AlphaFold," *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.
- [9] K. Hao, "Training a single AI model can emit as much carbon as five cars in their lifetimes," *MIT technology Review*, vol. 75, p. 103, 2019.
- [10] X. Wang, Y. Han, V. C. Leung, D. Niyato, X. Yan, and X. Chen, "Convergence of edge computing and deep learning: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 869–904, 2020.
- [11] T. P. Xiao, C. H. Bennett, B. Feinberg, S. Agarwal, and M. J. Marinella, "Analog architectures for neural network acceleration based on non-volatile memory," *Applied Physics Reviews*, vol. 7, no. 3, p. 031301, 2020.
- [12] J. Kaiser, S. Billaudelle, E. Müller, C. Tetzlaff, J. Schemmel, and S. Schmitt, "Emulating dendritic computing paradigms on analog neuromorphic hardware," *Neuroscience*, vol. 489, pp. 290–300, 2022.
- [13] 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.
- [14] M. Moralis-Pegios, A. Totovic, A. Tsakyridis, G. Giamougiannis, G. Mourgias-Alexandris, G. Dabos, N. Passalis, M. Kirtas, A. Tefas, and N. Pleros, "Photonic neuromorphic computing: Architectures, technologies, and training models," in *2022 Optical Fiber Communications Conference and Exhibition (OFC)*. IEEE, 2022, pp. 01–03.
- [15] W. Bogaerts, D. Pérez, J. Capmany, D. A. Miller, J. Poon, D. Englund, F. Morichetti, and A. Melloni, "Programmable photonic circuits," *Nature*, vol. 586, no. 7828, pp. 207–216, 2020.
- [16] L. De Marinis, "Integrated Photonic Neuromorphic Computing," Ph.D. dissertation, Scuola Superiore Sant'Anna, 2022.
- [17] 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*, vol. 58, no. 5, pp. 1–10, 2022.
- [18] L. De Marinis, E. Paolini, G. Contestabile, and N. Andriolli, "Leveraging Lithium Niobate on Insulator Technology for Photonic Analog Computing," in *2022 Italian Conference on Optics and Photonics (ICOP)*. IEEE, 2022, pp. 1–4.
- [19] O. A. Reimann and W. F. Kosonocky, "Progress in optical computer research," *IEEE spectrum*, vol. 2, no. 3, pp. 181–195, 1965.
- [20] T. Alexoudi, G. T. Kanellos, and N. Pleros, "Optical RAM and integrated optical memories: a survey," *Light: Science & Applications*, vol. 9, no. 1, p. 91, 2020.
- [21] E. Agrell, M. Karlsson, A. Chraplyvy, D. J. Richardson, P. M. Krummrich, P. Winzer, K. Roberts, J. K. Fischer, S. J. Savory, B. J. Eggleton *et al.*, "Roadmap of optical communications," *Journal of Optics*, vol. 18, no. 6, p. 063002, 2016.
- [22] 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.
- [23] G. Van der Sande, D. Brunner, and M. C. Soriano, "Advances in photonic reservoir computing," *Nanophotonics*, vol. 6, no. 3, pp. 561–576, 2017.
- [24] E. Paolini, L. De Marinis, M. Cococcioni, L. Valcarengi, L. Maggiani, and N. Andriolli, "Photonic-aware neural networks," *Neural Computing and Applications*, vol. 34, no. 18, pp. 15 589–15 601, 2022.
- [25] G. Giamougiannis, A. Tsakyridis, M. Moralis-Pegios, C. Pappas, M. Kirtas, N. Passalis, D. Lazovsky, A. Tefas, and N. Pleros, "Analog nanophotonic computing going practical: silicon photonic deep learning engines for tiled optical matrix multiplication with dynamic precision," *Nanophotonics*, 2023.
- [26] S. Garg, J. Lou, A. Jain, Z. Guo, B. J. Shastri, and M. Nahmias, "Dynamic precision analog computing for neural networks," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 29, no. 2: Optical Computing, pp. 1–12, 2022.
- [27] C. H. Cox, *Analog optical links: theory and practice*. Cambridge University Press, 2006.
- [28] A. N. Tait, "Quantifying power in silicon photonic neural networks," *Physical Review Applied*, vol. 17, no. 5, p. 054029, 2022.
- [29] D. A. I. Marpaung, "High dynamic range analog photonic links: design and implementation," *University of Twente*, 2009.
- [30] N. Margalit, C. Xiang, S. M. Bowers, A. Bjorlin, R. Blum, and J. E. Bowers, "Perspective on the future of silicon photonics and electronics," *Applied Physics Letters*, vol. 118, no. 22, p. 220501, 2021.
- [31] S. Y. Siew, B. Li, F. Gao, H. Y. Zheng, W. Zhang, P. Guo, S. W. Xie, A. Song, B. Dong, L. W. Luo *et al.*, "Review of silicon photonics technology and platform development," *Journal of Lightwave Technology*, vol. 39, no. 13, pp. 4374–4389, 2021.
- [32] A. Rahim, A. Hermans, B. Wohlfeil, D. Petousi, B. Kuyken, D. Van Thourhout, and R. Baets, "Taking silicon photonics modulators to a higher performance level: state-of-the-art and a review of new technologies," *Advanced Photonics*, vol. 3, no. 2, pp. 024 003–024 003, 2021.
- [33] N. Hosseinzadeh, A. Jain, K. Ning, R. Helkey, and J. F. Buckwalter, "A 0.5–20 ghz rf silicon photonic receiver with 120 db•hz 2/3 sfdR using broadband distributed im3 injection linearization," in *2019 IEEE Radio Frequency Integrated Circuits Symposium (RFIC)*. IEEE, 2019, pp. 99–102.
- [34] M. Piels and J. E. Bowers, "Photodetectors for silicon photonic integrated circuits," *Photodetectors*, pp. 419–436, 2023.
- [35] M. Tang, J.-S. Park, Z. Wang, S. Chen, P. Jurczak, A. Seeds, and H. Liu, "Integration of III-V lasers on Si for Si photonics," *Progress in Quantum Electronics*, vol. 66, pp. 1–18, 2019.
- [36] M. Zhang, C. Wang, P. Kharel, D. Zhu, and M. Lončar, "Integrated lithium niobate electro-optic modulators: when performance meets scalability," *Optica*, vol. 8, no. 5, pp. 652–667, 2021.
- [37] P. Kharel, C. Reimer, K. Luke, L. He, and M. Zhang, "Breaking voltage-bandwidth limits in integrated lithium niobate modulators using micro-structured electrodes," *Optica*, vol. 8, no. 3, pp. 357–363, 2021.
- [38] X. Guo, L. Shao, L. He, K. Luke, J. Morgan, K. Sun, J. Gao, T.-C. Tzu, Y. Shen, D. Chen *et al.*, "High-performance modified uni-traveling carrier photodiode integrated on a thin-film lithium niobate platform," *Photonics Research*, vol. 10, no. 6, pp. 1338–1343, 2022.
- [39] X. Sun, Y. Sheng, X. Gao, Y. Liu, F. Ren, Y. Tan, Z. Yang, Y. Jia, and F. Chen, "Self-Powered Lithium Niobate Thin-Film Photodetectors," *Small*, vol. 18, no. 35, p. 2203532, 2022.
- [40] D. Zhu, L. Shao, M. Yu, R. Cheng, B. Desiatov, C. Xin, Y. Hu, J. Holzgrafe, S. Ghosh, A. Shams-Ansari *et al.*, "Integrated photonics on thin-film lithium niobate," *Advances in Optics and Photonics*, vol. 13, no. 2, pp. 242–352, 2021.
- [41] A. Shams-Ansari, D. Renaud, R. Cheng, L. Shao, L. He, D. Zhu, M. Yu, H. R. Grant, L. Johansson, M. Zhang *et al.*, "Electrically pumped laser transmitter integrated on thin-film lithium niobate," *Optica*, vol. 9, no. 4, pp. 408–411, 2022.
- [42] J. Klamkin, H. Zhao, B. Song, Y. Liu, B. Isaac, S. Pinna, F. Sang, and L. Coldren, "Indium phosphide photonic integrated circuits: Technology and applications," in *2018 IEEE BiCMOS and Compound Semiconductor Integrated Circuits and Technology Symposium (BCICTS)*. IEEE, 2018, pp. 8–13.

- [43] H. Zhao, S. Pinna, F. Sang, B. Song, S. T. Š. Brunelli, L. A. Coldren, and J. Klamkin, "High-power indium phosphide photonic integrated circuits," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 25, no. 6, pp. 1–10, 2019.
- [44] M. Smit, K. Williams, and J. van der Tol, "Past, present, and future of InP-based photonic integration," *APL Photonics*, vol. 4, no. 5, p. 050901, 2019.
- [45] Y. Ogiso, J. Ozaki, Y. Ueda, H. Wakita, M. Nagatani, H. Yamazaki, M. Nakamura, T. Kobayashi, S. Kanazawa, Y. Hashizume *et al.*, "80-GHz bandwidth and 1.5-V V_{π} InP-based IQ modulator," *Journal of Lightwave Technology*, vol. 38, no. 2, pp. 249–255, 2019.
- [46] S. Dogru and N. Dagli, "0.77-V drive voltage electro-optic modulator with bandwidth exceeding 67 GHz," *Optics letters*, vol. 39, no. 20, pp. 6074–6077, 2014.
- [47] D. D'Agostino, G. Carnicella, C. Ciminelli, P. Thijs, P. J. Veldhoven, H. Ambrosius, and M. Smit, "Low-loss passive waveguides in a generic InP foundry process via local diffusion of zinc," *Optics express*, vol. 23, no. 19, pp. 25 143–25 157, 2015.
- [48] B. Shi, N. Calabretta, and R. Stabile, "InP photonic integrated multi-layer neural networks: Architecture and performance analysis," *APL Photonics*, vol. 7, no. 1, p. 010801, 2022.
- [49] 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.
- [50] L. De Marinis, G. Contestabile, and N. Andriolli, "A Lithium Niobate on Insulator Based Photonic Neural Network," in *2022 27th OptoElectronics and Communications Conference (OECC) and 2022 International Conference on Photonics in Switching and Computing (PSC)*. IEEE, 2022, pp. 1–3.
- [51] S. Tanaka, T. Simoyama, T. Aoki, T. Mori, S. Sekiguchi, S.-H. Jeong, T. Usuki, Y. Tanaka, and K. Morito, "Ultralow-power (1.59 mW/Gbps), 56-Gbps PAM4 operation of Si photonic transmitter integrating segmented PIN Mach-Zehnder modulator and 28-nm CMOS driver," *Journal of Lightwave Technology*, vol. 36, no. 5, pp. 1275–1280, 2018.
- [52] A. R. Totović, G. Dabos, N. Passalis, A. Tefas, and N. Pleros, "Femtojoule per MAC neuromorphic photonics: An energy and technology roadmap," *IEEE Journal of selected topics in Quantum Electronics*, vol. 26, no. 5, pp. 1–15, 2020.

Lorenzo De Marinis Lorenzo De Marinis received the B.S. and M.S. degrees (cum laude) in electronic engineering from the University of Pisa, in 2017 and 2019, respectively. He received the Ph.D. (cum Laude) in 2022 from Scuola Superiore Sant'Anna (SSSA), Pisa, Italy. From November 2021 to April 2022, he was a visiting scholar at the Aristotle university of Thessaloniki, Thessaloniki, Greece, at the WinPhos laboratory. He was a research fellow in quantum and neuromorphic photonics from September 2022 to February 2023 at SSSA, where he is currently an Assistant Professor. His main research concerns the conceptualization and design of photonic integrated circuits for quantum and neuromorphic applications, analog computing, photonic-electronic codesign and machine learning for networking scenarios.

Nicola Andriolli Nicola Andriolli received the Laurea degree in telecommunications engineering from the University of Pisa in 2002, and the Diploma and Ph.D. degrees from Scuola Superiore Sant'Anna, Pisa, in 2003 and 2006, respectively. He was a Visiting Student at DTU, Copenhagen, Denmark and a Guest Researcher at NICT, Tokyo, Japan. Since 2007 he has been an Assistant Professor at Scuola Superiore Sant'Anna, since 2019 he is with the National Research Council of Italy. He is currently a Senior Researcher at the Institute of Electronics, Information Engineering and Telecommunications (CNR-IEIT). He has been working in the field of optical processing and optical interconnection network architectures and scheduling, and more recently has been investigating integrated transceivers, comb sources, and architectures and subsystems for photonic neural networks. He has authored more than 200 publications in international journals and conferences, contributed to one IETF RFC, and filed 11 patents. He has been a TPC Member in several international conferences (ICC, GLOBECOM, ECOC, EuCNC, IPR, PSC), he served as an Associate Editor of IEEE Access between 2019 and 2022, and he is an Associate Editor of IEEE Photonics Journal.

Giampiero Contestabile Giampiero Contestabile received the Laurea degree in physics from "La Sapienza" University of Rome, Rome, Italy, in 1998 and the Ph.D. degree in electrical engineering and telecommunications from "Tor Vergata" University of Rome in 2001. From 1996 to 2000, he was with the Semiconductor Devices Group of "Fondazione Ugo Bordoni," Rome. In 2001, he was with Optospeed Italia. He is an Associate Professor with Scuola Superiore Sant'Anna, Pisa, Italy. He has coauthored more than 200 papers published in international peer-reviewed journals and presented in leading international conferences. His research interests include photonic integrated circuits, advanced optical systems, access networks, and semiconductor optical amplifiers and lasers.