Optical Materials EXPRESS

Complex-valued trainable activation function hardware using a TCO/silicon modulator

JUAN NAVARRO-ARENAS,[1,2] (ID) **JORGE PARRA,**[1] (ID) **AND PABLO SANCHIS**[1,*] (ID)

[1]*Nanophotonics Technology Center, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain*
[2]*Instituto de Ciencia de Materiales (ICMUV), Universidad de Valencia, C/Catedrático José Beltrán, 2, E-46980 Paterna, Spain*
*pabsanki@ntc.upv.es

Abstract: Artificial neural network-based electro-optic chipsets constitute a very promising platform because of its remarkable energy efficiency, dense wavelength parallelization possibilities and ultrafast modulation speeds, which can accelerate computation by many orders of magnitude. Furthermore, since the optical field carries information in both amplitude and phase, photonic hardware can be leveraged to naturally implement complex-valued neural networks (CVNNs). Operating with complex numbers may double the internal degrees of freedom as compared with real-valued neural networks, resulting in twice the size of the hardware network and, thus, increased performance in the convergence and stability properties. To this end, the present work revolves on the concept of CVNNs by offering a design, and simulation demonstration, for an electro-optical dual phase and amplitude modulator implemented by integrating a transparent conducting oxide (TCO) in a silicon waveguide structure. The design is powered by the enhancement of the optical-field confinement effect occurring at the epsilon-near-zero (ENZ) condition, which can be tuned electro-optically in TCOs. Operating near the ENZ resonance enables large changes on the real and imaginary parts of the TCO's permittivity. In this way, phase and amplitude (dual) modulation can be achieved in single device. Optimal design rules are discussed in-depth by exploring device's geometry and voltage-dependent effects of carrier accumulation inside the TCO film. The device is proposed as a complex-valued activation function for photonic neural systems and its performance tested by simulating the training of a photonic hardware neural network loaded with our custom activation function.

1. Introduction

A wide-ranging extent of new and emergent nanophotonic technologies rely on signal's control of both, phase and amplitude, or complex-valued (CV) modulation, to enable high performance operation. In the case of quantum communications, for example, CV modulation of a two-photon entangled wave-function allows a much larger variety of transfer functions and, therefore, increases the versatility of quantum information processing setups to a large extent [1]. Another interesting perspective for a CV modulator device is in the development of innovative photonic complex-valued neural network (CVNN) chips. The CVNN allows the input/variables and activation function in neural networks to be complex. Recent work suggests that complex numbers could have a richer representational capacity [2] and robustness against spurious local minima [3], leading to improved stability and convergence properties when compared with their real-valued counterparts. Additionally, CVNNs exhibit superior performance in applications in which the input is naturally complex-valued, such as synthetic aperture radar (SAR) imagery classification [4]. Despite their attractive properties however, the mainstream deep learning architectures are still based on real-valued algebra. Research on the topic of complex-valued deep neural networks have been side-lined due to the lack of well-defined building blocks, as

for instance complex-valued activation functions, and training strategies for such models [5]. Furthermore, since complex arithmetic is only rarely mapped into the processing structures of digital signal processors, processing complex numbers requires a higher number of instructions, therefore introducing slowdowns in the execution of CVNNs [6]. In this regard, since the optical field encodes information in both, phase and amplitude, complex arithmetic can be naturally implemented in photonic hardware offering significantly enhanced computational speed and energy efficiency [7–10].

The implementation of CVNN in photonic hardware requires weighting units and nonlinear activation functions (NLAFs) to perform CV modulation. To achieve phase modulation, most prevalent methods rely either on the silicon thermo-optic effect, requiring long footprint microheaters, or on MOS-like silicon devices that exploit either the plasma dispersion, the quadratic electro-optic or DC Kerr effects of silicon. However, the phase modulation efficiency of each of these technologies is lacking a key performance metric that limits its possibilities (i.e. low modulation speed and energy efficiency in the case of the thermo-optic phase shifters, or high insertion losses in the case of plasma dispersion effect). On top of that, enabling phase and amplitude modulation in the same chip is accomplished by means of a concatenation of a Mach–Zehnder interferometer (MZI) and one of the aforementioned phase shifter technologies. While successful demonstrations of coherent CVNNs have been undertaken with this strategy [9], it also comes at the cost of a large device footprint, reduced component density and, thus, smaller NN capacities. Thus, finding a suitable mechanism that can perform CV modulation in silicon photonics remains a critical task for the implementation CVNNs.

In this context, there has been a surge of interest in epsilon-near-zero (ENZ) materials with vanishing permittivity at NIR-telecom wavelengths. It was recently demonstrated that the ternary composition of Indium Tin Oxide (ITO), a material of the family of transparent conducting oxides (TCOs) and also a CMOS-compatible material, which is currently widely used in microelectronics, exhibits an ultra-large nonlinear response linked to the ENZ effect [11]. Unlike noble metals, which typically have their ENZ wavelengths at the ultraviolet region accompanied with very high imaginary epsilon, TCOs can be considered low-loss materials at telecom wavelengths, enabling the development of new concepts in the field of plasmonics [12]. Indeed, hybrid TCO/silicon devices show an extraordinary enhancement of local electric field [13] and extreme optical nonlinear effects [14–16]. Moreover, TCO thin films can be actively tuned to perform all-optical and electro-optical complex permittivity modulations. Drude model describes the possibility of such modulation mechanisms, either by manipulating the effective electron mass or the electronic concentration of TCOs —given that plasma frequency is proportional these quantities. In current literature, it has been demonstrated that effective electron mass can be altered all-optically by intraband transitions triggered upon the absorption of high-power pulses [17,18]. On the other hand, electron concentration can be tuned electro-optically by means of a metal-oxide-semiconductor (MOS) capacitor structures [19]. Such scheme has been used in the past to design electro-refraction and electro-absorption optical modulators [19–23]. Recently, the integration of TCOs in silicon structures has also been proposed as capable solution to implement absorption-modulator—thus, real-valued—activation functions in all-optical [24] as well as electro-optic schemes [25,26].

In this work, a TCO/Si modulator that exploits the ENZ effect to perform complex-valued modulations is proposed. The ENZ resonance is tuned electro-optically with a MOS capacitor mounted on top of a Si waveguide. Although we propose such device to implement CV-NLAFs, it can also be used as a coherent weighting unit in non-unitary (unconstrained) neural networks [27–29]. To implement the NLAF, the nonlinearity is built electro-optically in a feedback loop where the voltage, used to bias the MOS capacitor, is generated by a photodetector in proportion to the power of light received, thus converting the optical inputs into electrical currents that tune the effect of the ENZ resonance. Transceiver circuits such as this one have been already adopted

in the field of programmable photonics to artificially implement nonlinearity in the optical neural-networks with phase change materials and thermo-optic phase shifters [30–34]. Indeed, as discussed above, although the tuning of the ENZ resonance can be triggered all-optically, one of the advantages of the opto-electronic approach is the possibility to operate with low power continuous-wave signals. Another important advantage is the possibility to map the input optical power by using any given arbitrary function to scale the optical inputs into electrical outputs. Thus, the resulting nonlinear activation function can be reconfigured during execution. In this way, the parameters of the NLAF can be optimized to improve of the performance in quantification, classification, and reinforcement learning problems, approaching the concept and possibilities of an universal activation function [35] but operating in photonic hardware. Additionally, the reconfigurability of the NLAF can be leveraged to introduce the novel concept of NLAFs with trainable parameters to improve the total response region of the neural network [36]. All in all, the proposed device demonstrates for the first time an ultracompact implementation of a complex-valued activation function for photonic neural networks with trainable parameters.

2. Results and discussion

2.1. Concept of the complex-valued perceptron with reconfigurable activation functions

A perceptron is a simplified model of a biological neuron and constitutes the basic building block for any neural network. It is comprised of four basic layers: input/output layers, a weighted sum layer with a bias, and a reconfigurable non-linear activation function (NLAF) layer. Coherent detection is applied in the final layer to recover the phase and amplitude of the propagated signals. Perceptrons are conceived as modular structures that can be replicated and instantiated to create on-demand network topologies. A representation of such architecture is shown in Fig. 1.
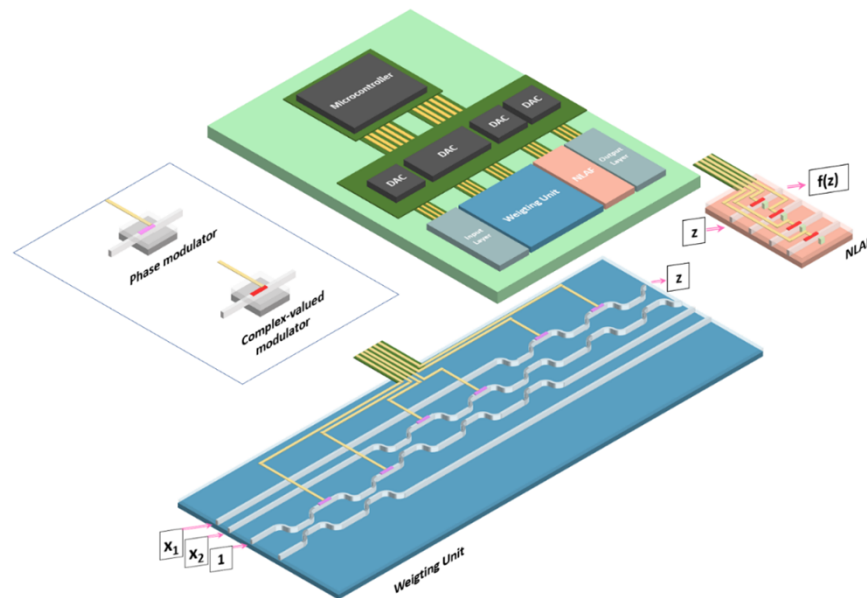
Fig. 1. Schematic representation of the complex-valued perceptron. Phase modulator and complex-valued modulators represent the basic building blocks. The microcontroller powers the reconfigurable optical hardware to perform biased N-dimensional unitary signal transformation (Weighting Unit) and reconfigurable non-linear activation (NLAF).

The weighted sum and bias are performed in the Weighting Unit, signalled in blue in the scheme of Fig. 1, wherein an example of a single output neuron performing the weighting of

signals x_1, x_2 and bias is shown. Trainable weights are applied to the phase shifters (shown in pink in Fig. 1) to perform unitary transformations with the N-channel input. This scheme can be easily implemented with arrayed cascades of MZI containing two 50:50 directional couplers, preceded by a phase shift at one input port [37,38]. Since losses reduce the representation capacity of the unitary space, usually the phase shift is achieved by means of lossless thermo-optic phase shifters [39]. In dissipative photonic circuits, singular value decomposition (SVD) can be applied to a 2D hybrid waveguide mesh to perform any matrix linear operation, including input matrix weighting products [40]. Thereby, a complex matrix could also be implemented solely by using complex-valued modulators.

The outputs of the Weighting Unit are processed in the reconfigurable NLAF, where the activation is applied to the weighted signals. Detail on the anatomy of the proposed NLAF unit is shown in Fig. 2. Such device operates by transducing power of the optical signal (i.e. a complex number z) into a gate voltage to drive the complex-valued modulator. The photodetector implementation has to be non-destructive, which could be achieved in a number of ways (e.g., by using directional coupler to measure a small fraction of the input signal with a photodetector, like in Fig. 2) [41]. The device's activation function manifests the effect of the complex refractive index acting upon an optical field passing through a segment of length L. For a complex number z and its complex-conjugate z^*, the function $f : \mathbb{C} \rightarrow \mathbb{C}$ can be expressed as $f(z) = |z|e^{i\emptyset_z} \times e^{-\alpha L}e^{-i\Delta\emptyset L}$, by defining the absorption coefficient as $\alpha = 4\pi\kappa_{eff}(V_G)/\lambda_0$ and phase change per unit length as $\Delta\emptyset = 2\pi n_{eff}(V_G)/\lambda_0$, where λ_0 is the channel wavelength. Here $\widetilde{n_{eff}} = n_{eff}(V_G) + i\kappa_{eff}(V_G)$, the complex effective refractive index of the modulator, is a function of the gate voltage V_G, which is made proportional to the signal registered by the photodetector (i.e. the modulus squared of the optical field). Such dependence can be expressed linearly as $V_G = a(zz^*) + b.$, where a and b are reconfiguration parameters provided by the electronic controller that drives the modulator. Indeed, the device offers NLAF reconfigurability because the modulator's driving voltage can be externally mapped by gauging the amount of signal from the photodetector that is transduced. This possibility is expressed linearly, although an arbitrary function can be introduced here, in the real-valued coefficients a and b, to smooth and bias the shape of the activation function. Such parameters can be made trainable so that effect of the activation function becomes dynamic and capable to adapt as per the requirements of its neighbouring layers [36].

A complex-valued representation of the $f(z)$ and the effects of the reconfiguration parameters on its output is shown in Fig. 3 (a)-(c). It can be seen that the effect of the linear parameter a on the function $f(z)$ is to scale the action of the activation on both, phase and amplitude. On the other hand, parameter b has a stronger effect on the phase change. Biasing $f(z)$ comprises the scaling of $\Delta\emptyset$ so that complex numbers with lower magnitudes achieve higher distortions on its phase. There is a physical limit on the maximum achievable phase change of $f(z)$. This limit depends on the design properties of the photonic device that performs the amplitude-phase modulation. Additionally, amplitude and phase changes of $f(z)$ are linked magnitudes and cannot be decoupled. In this scenario, the activation function is invariant with respect to phase rotation, as can be seen in Fig. 3(a) and 3(c). In literature, phase invariance can be used to map complex-valued activation functions such as cardioid function, which scales the input magnitude but retains the input phase. Phase retention has proven useful to process magnetic resonance with the fingerprinting method [42].

However, this invariance can be broken if optical phase detection (e.g. with a coherent scheme) is introduced in the device of Fig. 2 to recover the phase information of the input optical field [43]. In this way, the coefficient a can be made a complex-valued array so that phase dependence can be introduced into $f(z)$ by redefining the electrical signal that biases the device with a dot product, $V_G = a_{Re}Re(z)^2 + a_{Im}Im(z)^2 + b.$. Here, the coefficients break the phase invariancy when $a_{Re} \neq a_{Im}$. In Fig. 3(b) and 3(d), an activation function with a slight asymmetry was plotted.
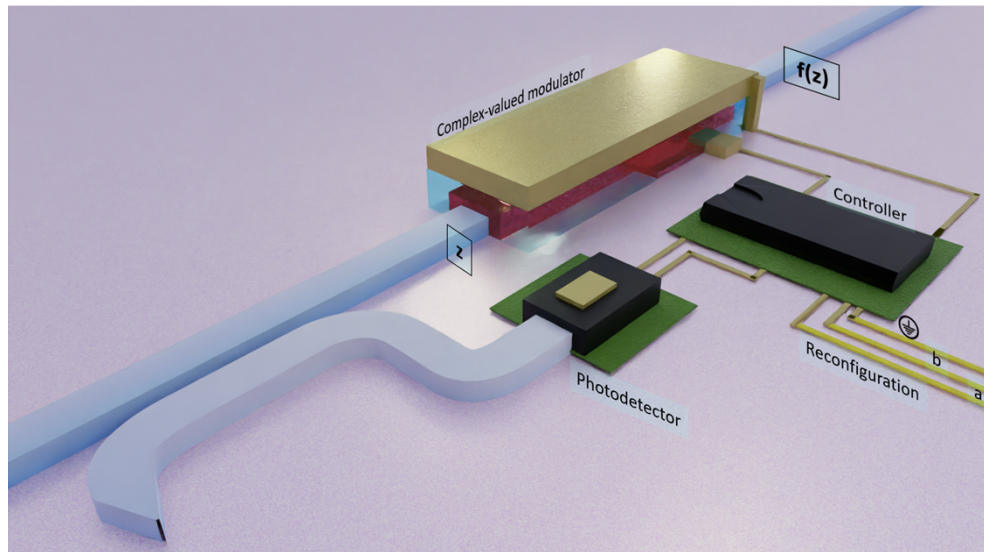
Fig. 2. Schematic representation of the reconfigurable NLAF unit. An optical signal *z* is partially detected with a photodetector and the measured electrical signal is calibrated in the controller with the reconfiguration parameters *a* and *b* to supply the capacitor's gate voltage.

The present activation function offers a high level of reconfigurability due to its various learnable degrees of freedom. Such flexibility enables the function to adapt its shape given the training data. It has been demonstrated that adaptable functions with sufficient degrees of freedom can outperform neural networks with fixed functions [44].

2.2. Description of the complex-valued modulator

The essential building block in the prosed NLAF unit shown in Fig. 2 is the CV modulator. A TCO-based dual modulator is proposed which simultaneously exploits the plasma dispersion effect and the ENZ absorption enhancement effect to introduce phase/amplitude mixed modulation states with an ultra-compact footprint. The plasma dispersion contributes changing the effective refractive index of a hybrid silicon waveguide by manipulating the carrier concentration in the TCO through carrier accumulation mechanisms in a metal-oxide-semiconductor (MOS) capacitor stack. This capacitor is built on top of a silicon waveguide (see schematic of the device in Fig. 4).

In such a hybrid device, only the excitation of the fundamental TM mode is considered (dominant field component in the y-direction) to ensure maximal modal overlap within the MOS stack. When the MOS is biased with a forward voltage in the metal electrode, free carriers from the bulk TCO accumulate at the interface with the insulator creating a charged thin (accumulation) film. Thus, since the electron density has an impact in the optical constants of the TCO material, the effective index of the modal structure that includes the accumulation layer can be modulated electro-optically with an external bias. For such device, the modulation speed is primarily limited by the capacitive delay time of the structure. Calculations with data taken from literature reveal a realistic upper bound in the range of 100 GHz [45]—however, it is important to note that the final device would be equipped with an optoelectronic translation circuit, which would introduce limitations in terms of the device's operational speed [34]. The theoretical model, that describes the change on the optical constants of the TCO when the capacitor is biased, is provided in the S1 Section of the Supplement 1. The relationship between gate voltage (voltage applied between the metal and the TCO) and the change of refractive index near the ENZ resonance is calculated in
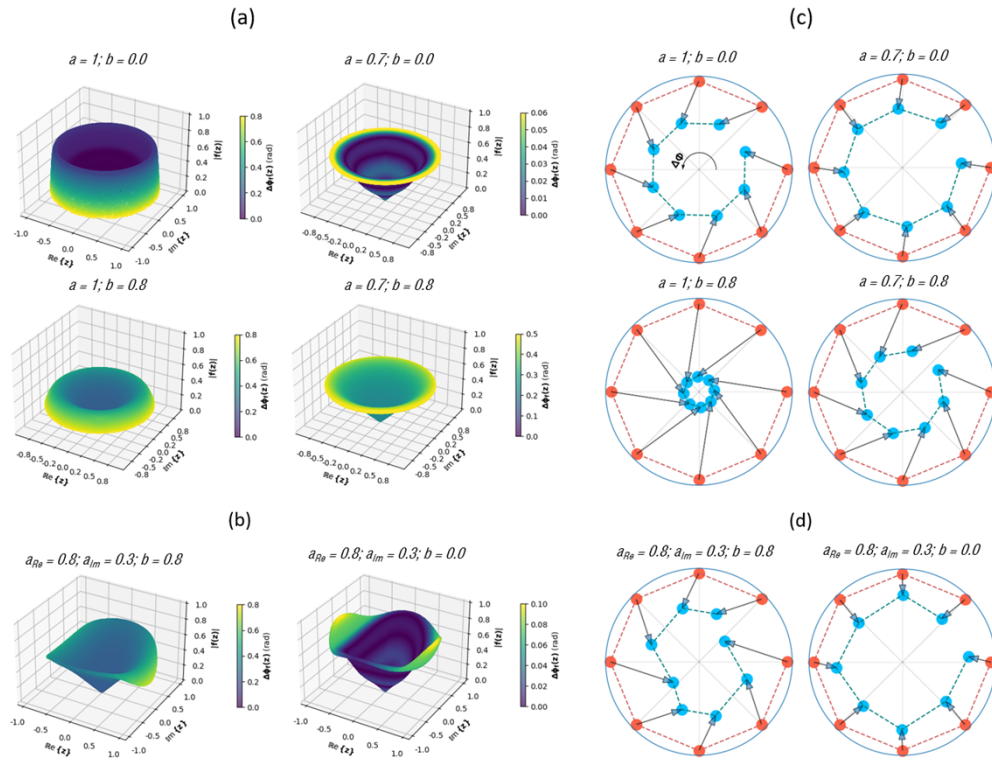
Fig. 3. Surface plot of *f(z)*. The z-axis represents the modulus the activation function and the colour mapping represent the phase change for (a) real- and (b) complex-valued *a* parameter. Representation of *f(z)* action (blue dots) over the input complex numbers *z* (in red) with unitary radius (c) real- and (d) complex-valued *a* parameter. Two device configurations for *b* were chosen to display its effects.

Section 3. However, the expected variation of the optical phase and amplitude has also been illustrated in Fig. 4.

To evaluate the performance of the proposed device we can study its behaviour under conditions that either maximize phase change or conditions that maximize losses. Attending to the scheme of Fig. 4, the proposed modulator can be configured in the *steady state (OFF)*, the modulator has low propagation losses, the transmission is maximized, and the phase is a constant value. A second *amplitude modulation state (ON$^{\Delta\kappa}$)* tunes response of the device at the centre of the ENZ resonance so that optical absorption is maximized. Another *phase modulation state (ON$^{\Delta n}$)* can be tuned neighbouring the ENZ resonance so that a phase difference is extracted with minimal losses. Therefore, the two dominant modulation mechanisms can be deployed within a shared device architecture by slightly adjusting the gate voltage. For the activation function, the device can be configured by tuning the voltage as a continuous variable to obtain mixed phase-absorption states.
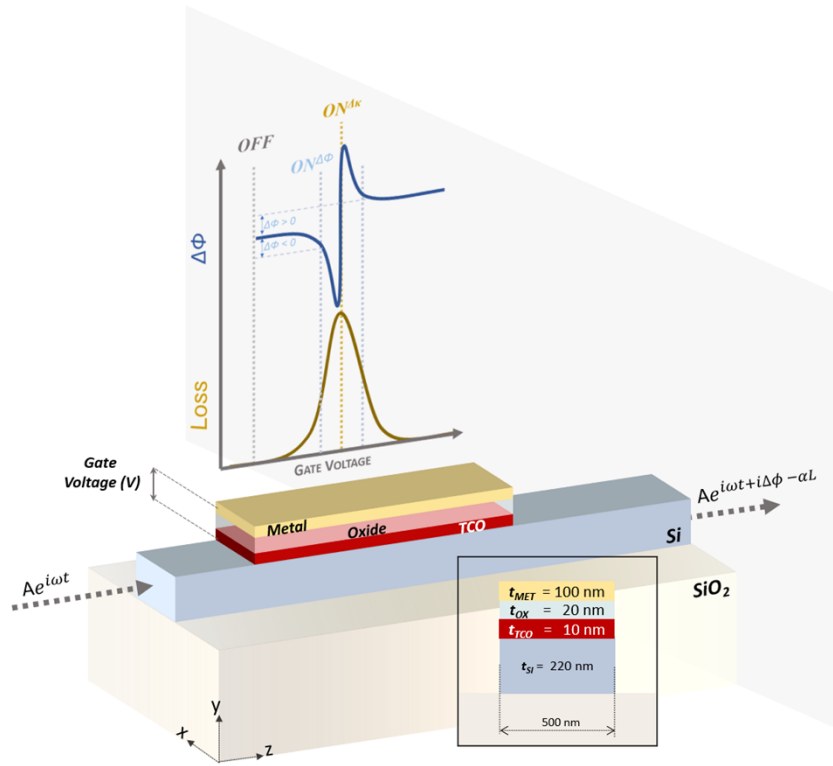
Fig. 4. Schematic (not to scale) of the complex-valued modulator and cross-section of the hybrid waveguide. The variation of the optical phase and amplitude in the hybrid waveguide as a function of gate voltage is also illustrated.

3. Performance analysis and optimization of the complex-valued modulator

The performance analysis and optimization of CV modulator has been carried out by the finite element method implemented in the RSoft CAD environment from Synopsys. The permittivity of the TCO film was parametrized in two regions: the 1 nm accumulation layer, whose permittivity depends on the gate voltage, which is described in Section S1 of the Supplement 1, and the bulk layer, whose permittivity is described by a Drude model loaded with the background electron concentration given by the optimal doping level of $N_D = 1.0 \times 10^{19}\,\text{cm}^{-3}$, which was optimized in Section S2 of the Supplement 1.

Two figures of merit (FOMs) can be defined to establish the most optimal configuration to operate the device in the $ON^{\Delta\kappa}$ and $ON^{\Delta n}$ states. The FOM that describes best the absorption state scenario is $FOM^{\Delta\kappa} = |\Delta\kappa/(\kappa_{OFF}E_S)|$, in which $\Delta\kappa = \kappa_{ON} - \kappa_{OFF}$ is the increment of the extinction coefficient (imaginary part of the effective index) between the *ON* and *OFF* gate voltage states. This metric ensures a compromise solution between extinction ratio (ER) with minimal insertion losses (IL) at the lowest switching energy. On the other hand, the FOM used to describe the phase switching states is $FOM^{\Delta n} = |\Delta n/((\Delta\kappa + \kappa_{OFF})\,E_S)|$, in which Δn is the increment of the real part of the effective index between two gate voltage states. Placing $\Delta\kappa$ in the denominator ensures that the $ON^{\Delta n}$ state losses are minimized.

In Figs. 5(a) and 5(b), the FOMs were calculated as a function of the gate voltage, which also depends on the thickness of the insulator layer, considering a MOS stack made by a high mobility ($\mu = 200\,\text{cm}^2\text{V}^{-1}\text{s}^{-1}$) TCO, cadmium oxide, and a HfO$_2$ oxide film. The gate voltage was then

used to evaluate the energy consumption. The details on how to perform such calculations are discussed in Section S2 of the Supplement 1. The oxide thickness, t_{ox}, impacts the energy consumption of the complete modulation. Additionally, it also affects the effective index of the hybrid waveguide. The index contrast in the TCO layer is changed which influences the intensity of the ENZ effect. Therefore, the oxide thickness is a critical design parameter for the optimization of the device. After an optimization search, a thickness of $t_{ox} = 15$ nm was chosen for our design because it offers offer best index contrast with the lowest energy consumption ratio.
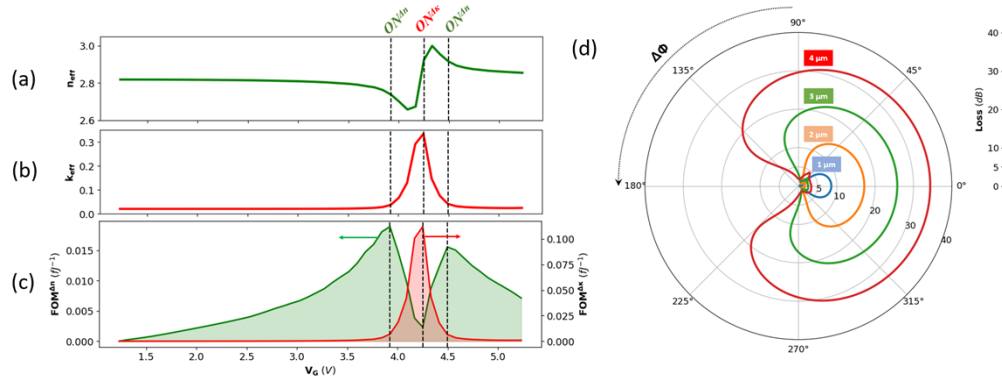
Fig. 5. (a) Real and (b) imaginary parts of the effective refractive index as a function of the applied gate voltage. (c) $FOM^{\Delta\kappa}$ and $FOM^{\Delta n}$ showing the best configuration for amplitude and phase switching as a function of the gate voltage. (d) Representation of all the possible mixed phase-amplitude states of the modulator. The change in amplitude (loss in dB) is defined as the radius and the change in phase as the angle of the polar plot. In such representation possible states with different lengths are also shown as closed loops with different losses.

A comparison of the two figures reveals that both FOMs are mutually exclusive. In Fig. 5(c) the optimal gate voltage that maximizes $FOM^{\Delta\kappa}$ occurs at 4.25 *V* while the one required to enable phase switching ($FOM^{\Delta n}$ maxima) is around 3.8 *V*. It is worth mentioning that due to the symmetric shape of the ENZ resonance, phase switching could also be attained with a higher voltage, specifically at 4.5 V. However, this configuration is discarded because it requires a higher switching energy. Table 1 summarizes the performance metrics for the designed complex-valued modulator. It's important to note here that the implementation of the proposed activation function would necessitate the incorporation of additional control electronics to provide voltage to the MOS capacitor (see Fig. 2). While the impact of these control electronics on the device's energy consumption was not accounted for in the results presented in Table 1, it should be taken into consideration for the practical application.

Table 1. Performance metrics for the designed complex-valued modulator.

	$ON^{\Delta n}$	$ON^{\Delta\kappa}$
OFF Loss (dB µm^{-1})	0.7	0.7
ON Loss (dB µm^{-1})	1.2	11.8
$\Delta\Phi$ (rad µm^{-1})	0.32	-
V_G (V)	3.8	4.25
E_S (fJ µm^{-1})	122	145

The performance metrics are comparable with previously published high mobility cadmium-based modulators [19]. Here we report a device that can achieve $\pi/2$ phase shift in a compact length of 5 microns with a $V_\pi L_\pi = 0.04$ V/cm. On the other hand, a very high extinction ratio (>50 dB in just 5 microns) can be attained with the device configured for amplitude modulation.

Regarding losses, the device shows relatively low insertion losses, of only 0.7 dB μm^{-1}, when operated in the OFF state, and losses of just 1.2 dB μm^{-1} at the point of optimal phase contrast. Coupling between the hybrid waveguide and the bare Si waveguide was also analyzed with 3D-FDTD simulations. Results can be found in the Supplement 1 Section S3 and reveal coupling losses of around 1.5 dB.

As a final remark, Fig. 5(d) shows in a polar plot the compendium of all possible mixed phase-amplitude configurations so that any possible configuration within the loop can be achieved. In such plot, the radial coordinate represents the amplitude losses while the angle component quantifies the total phase change. The shape of such projection is a closed loop, with a radius that depends only on the active length of the device.

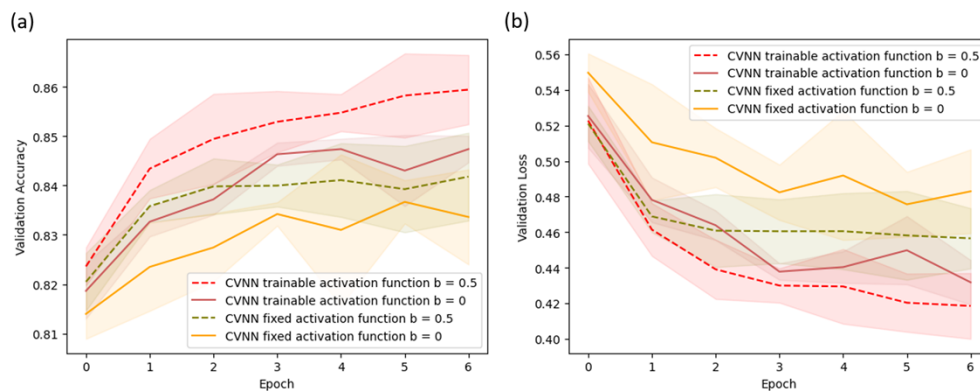## 4.   Complex-valued activation functions and photonic neural networks

In this section, we implement CVNN with standard fully-connected layers followed by pointwise nonlinearity denoted by $X_0 \xrightarrow{t_1} X_1 \xrightarrow{t_2} \ldots X_{d-1} \xrightarrow{t_d} X_d$. In this simulation, the phase and amplitude of the optical field inputted at the beginning of the optical circuit are represented by the complex variable X_0. Following the scheme introduced in Section 1, the weighting unit is abstracted as an arbitrary complex-valued matrix; the activation applies the t_d nonlinear transformation defined after the modeling results in Section 3 of the device with an active length of 1 micron. Here the proposed complex-valued trainable non-linear activation function is evaluated by means of simulations for different tasks of the problem of multi-class classification in the complex domain. Comparison with a real-valued neural networks (RVNNs) was also computed in similar conditions.

The first tasks was built with the well-known IRIS dataset [46] bundled with the scikit-learn library [47]. The data set contains 3 classes of 50 instances each. Two of the classes are not linearly separable. Each class refers to a type of iris plant, each entry with four real-valued inputs: length and width of the petals and length and width of the sepals. Inputs are codified into the complex plane by assigning the petal parameters to the real part and sepal parameters to the imaginary. In this way, the dimensionality of the input is halved. For the evaluation, a fully connected two-hidden-layer neural network, with small layer sizes of 8 and 16 neurons, respectively, was implemented. At the output of the classifier, softmax activation was computed on the real and imaginary parts and then averaged. The networks were trained by minimizing categorical cross-entropy with the Adam optimizer ($\beta_1 = 0.9$, $\beta_2 = 0.999$ and learning rate of 0.001). All networks were trained for 50 epochs with a batch size of 7. The proposed models were implemented in Python using Tensorflow as back-end with the aid of an open-source library that adds support to complex-number data types [48,49]. In section S4 of the Supplement 1 it is shown how the trainable activation function layer was defined in Keras. Code for this section can be accessed here [50].

For this experiment, the network model described above was trained with the activations of real ReLU and Parametric-ReLU (PReLU)—where the activation slope is a learned array [51] —to demonstrate the performance of a real-valued neural networks (RVNNs). The results were compared with a CVNN with fixed and learnable parameters. Simulations in Fig. 6 show that the neural network with the proposed NLAF with trainable parameters performs better than the CVNN case with fixed parameters, obtaining very high accuracy for the complex valued CVNN, with a median of over 95%, and the lowest validation loss at the end of the training. The trainable activation function also improved the stability of the CVNN, as the dispersion of the loss function was greatly reduced as can be seen Fig. 6(b). Both trainable and non-trainable complex NLAFs
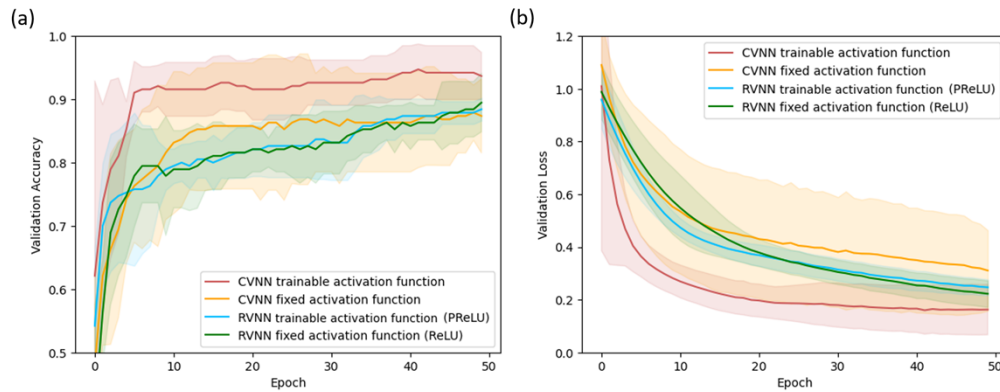
Fig. 6. (a) Mean validation accuracy and (b) validation loss function evolution for the complex-valued IRIS dataset task implemented with a CVNN with fixed NLAF (orange solid line) and with trainable NLAF (learnable a parameter and fixed b = 0, red), compared with RVNN with fixed (green) and learnable (blue) ReLU. Training was repeated 10 times to calculate 95% confidence interval (shadowed region). The bias parameter b had little impact in the training and was kept to zero.

Fig. 7. (a)Mean validation accuracy and (b) validation loss function evolution for the complex-valued Fashion MNIST dataset task implemented with a CVNN equipped with a trainable activation function on the a parameter and fixed b parameter

show faster convergence rates than the real-valued NLAFs. On the other hand, ReLU-based fixed and learnable RVNN show a similar accuracy, both below 90%.

In the next exercise, a more complex task was implemented; the Fashion MNIST dataset which contains 70,000 grayscale low resolution (28 by 28 pixel) clothing images in 10 categories, was tested with the CVNN [47]. To cast real-valued data from the images into the complex plane, phase information was generated proportional to the magnitude of each pixel. This kind of pre-processing adds no extra information and has been previously used in related publications [52]. We considered networks with two hidden layers having 32 and 128 neurons each with softmax activation at the output layer. The networks were trained by minimizing categorical cross-entropy with the Adam optimizer. All networks were trained for 7 epochs with a batch size of 128. Results of Fig. 7 show that the bias parameter b can be used to improve the sensitivity of the neural network on more demanding tasks. In this example, by tuning the bias the accuracy of the CVNN with learnable parameters was improved up to over 86% while it also provided stability to the loss function. Training of the same dataset with RVNN using ReLU and PReLU

rendered lower validation accuracies, of over 84%. It is worth mentioning that this architecture was not optimized, and it was meant only as a demonstration. Thus, several improvements are possible, such as including b as a trainable parameter.

The performance improvements introduced by the complex field are most likely due to the superior capacity in terms of information storage offered by the phase-amplitude pair and the adaptability offered by the reconfiguration parameters. The complex network can be decoupled into two sub-networks, ending up with neural network with doubled width when compared to traditional real-valued networks [9]. This characteristic of CVNN allows to reduce the quantity of integrated devices required to build dense networks. Additionally, the shape of the function evolves to adapt better into the network's topology. This versatility allows the CVNN to achieve high performance in different deep-learning problems with shared hardware.

5. Conclusions

A way to implement a complex-valued activation function with trainable parameters has been proposed, offering improved training accuracy and convergence speed due to the superior (virtually twice) size of the data encoded in phase and amplitude, and adaptability provided by the learnable parameters of the activations. Such activation function is powered by an electro-optical TCO/Si modulator with reconfigurable phase-amplitude response. The modulator design offers novelty on itself because, while phase-amplitude modulation is usually accomplished by means of a concatenation of Mach–Zehnder interferometers preceded by an individual phase shifter (this approach comes at expense of having a large device footprints that limits the scalability of integration), the proposed TCO/Si waveguide device, however, comprises the possibility of CV-modulation in a single device with ultracompact footprint, achieving a $V_\pi L_\pi = 0.04$ V/cm and ER over 11.8 dB/µm with a switching energy lower than 500 fJ—notice that energy consumption should be higher considering the control electronics required to operate the activation function. These results pave the road towards future and more efficient post-von-Neuman architectures based on photonics, an emerging technology with disruptive potential.

Disclosures. The authors declare no conflicts of interest.

Data availability. The code used to obtain the results of section 4 are available in Ref. [50].

Supplemental document. See Supplement 1 for supporting content.

References

1. F. Zäh, M. Halder, and T. Feurer, "Amplitude and phase modulation of time-energy entangled two-photon states," Opt. Express **16**(21), 16452–16458 (2008).
2. A. Caragea, D. G. Lee, J. Maly, G. Pfander, and F. Voigtlaender, "Quantitative approximation results for complex-valued neural networks," SIAM J. Math. Data Sci. **4**(2), 553–580 (2022).
3. X. Liu, "Neural networks with complex-valued weights have no spurious local minima," arXiv, arXiv:2103.07287 (2021).
4. T. Scarnati and B. Lewis, "Complex-valued neural networks for synthetic aperture radar image classification," in *2021 IEEE Radar Conference (RadarConf21)* (2021), pp. 1–6.
5. C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal, "Deep complex networks," arXiv, arXiv.1705.09792 (2017).
6. A. Hirose, *Complex-Valued Neural Networks* (Springer Berlin Heidelberg, 2012).
7. H. H. Zhu, J. Zou, H. Zhang, Y. Z. Shi, S. B. Luo, N. Wang, H. Cai, L. X. Wan, B. Wang, X. D. Jiang, J. Thompson, X. S. Luo, X. H. Zhou, L. M. Xiao, W. Huang, L. Patrick, M. Gu, L. C. Kwek, and A. Q. Liu, "Space-efficient optical computing with an integrated chip diffractive neural network," Nat. Commun. **13**(1), 1044 (2022).
8. R. Wang, P. Wang, C. Lyu, G. Luo, H. Yu, X. Zhou, Y. Zhang, and J. Pan, "Multicore photonic complex-valued neural network with transformation layer," Photonics **9**(6), 384 (2022).

9. H. Zhang, M. Gu, X. D. Jiang, J. Thompson, H. Cai, S. Paesani, R. Santagati, A. Laing, Y. Zhang, M. H. Yung, Y. Z. Shi, F. K. Muhammad, G. Q. Lo, X. S. Luo, B. Dong, D. L. Kwong, L. C. Kwek, and A. Q. Liu, "An optical neural chip for implementing complex-valued neural network," Nat. Commun. **12**(1), 457 (2021).
10. G. Mourgias-Alexandris, A. Totović, A. Tsakyridis, N. Passalis, K. Vyrsokinos, A. Tefas, and N. Pleros, "Neuromorphic Photonics With Coherent Linear Neurons Using Dual-IQ Modulation Cells," J. Lightwave Technol. **38**(4), 811–819 (2020).
11. M. Z. Alam, I. De Leon, and R. W. Boyd, "Large optical nonlinearity of indium tin oxide in its epsilon-near-zero region," Science **352**(6287), 795–797 (2016).
12. B. C. Yildiz and H. Caglayan, "Epsilon-near-zero media coupled with localized surface plasmon modes," Phys. Rev. B **102**(16), 165303 (2020).
13. I. V. A. K. Reddy, J. M. Jornet, A. Baev, and P. N. Prasad, "Extreme local field enhancement by hybrid epsilon-near-zero–plasmon mode in thin films of transparent conductive oxides," Opt. Lett. **45**(20), 5744–5747 (2020).
14. J. Wu, Z. T. Xie, Y. Sha, H. Y. Fu, and Q. Li, "Epsilon-near-zero photonics: infinite potentials," Photonics Res. **9**(8), 1616–1644 (2021).
15. N. Kinsey, C. DeVault, A. Boltasseva, and V. M. Shalaev, "Near-zero-index materials for photonics," Nat. Rev. Mater. **4**(12), 742–760 (2019).
16. O. Reshef, I. De Leon, M. Z. Alam, and R. W. Boyd, "Nonlinear optical effects in epsilon-near-zero media," Nat. Rev. Mater. **4**(8), 535–551 (2019).
17. J. Navarro-Arenas, J. Parra, and P. Sanchis, "Ultrafast all-optical phase switching enabled by epsilon-near-zero materials in silicon," Opt. Express **30**(9), 14518–14529 (2022).
18. E. Li and A. X. Wang, "Femto-joule all-optical switching using epsilon-near-zero high-mobility conductive oxide," IEEE J. Sel. Top. Quantum Electron. **27**(2), 1–9 (2021).
19. I. C. Reines, M. G. Wood, T. S. Luk, D. K. Serkland, and S. Campione, "Compact epsilon-near-zero silicon photonic phase modulators," Opt. Express **26**(17), 21594–21605 (2018).
20. M. G. Wood, S. Campione, S. Parameswaran, T. S. Luk, J. R. Wendt, D. K. Serkland, and G. A. Keeler, "Gigahertz speed operation of epsilon-near-zero silicon photonic modulators," Optica **5**(3), 233–236 (2018).
21. A. P. Vasudev, J.-H. Kang, J. Park, X. Liu, and M. L. Brongersma, "Electro-optical modulation of a silicon waveguide with an ``epsilon-near-zero'' material," Opt. Express **21**(22), 26387–26397 (2013).
22. M. Ayata, Y. Nakano, and T. Tanemura, "Silicon rib waveguide electro-absorption optical modulator using transparent conductive oxide bilayer," Jpn. J. Appl. Phys. **55**(4), 042201 (2016).
23. S. Mohammadi-Pouyan, M. Miri, and M. H. Sheikhi, "Efficient binary and QAM optical modulation in ultra-compact MZI structures utilizing indium-tin-oxide," Sci. Rep. **12**(1), 8129 (2022).
24. J. Gosciniak, Z. Hu, M. Thomaschewski, V. J. Sorger, and J. B. Khurgin, "Bistable All-Optical Devices Based on Nonlinear Epsilon-Near-Zero (ENZ) Materials," Laser Photonics Rev. **17**(4), 2200723 (2023).
25. R. Amin, J. K. George, S. Sun, T. Ferreira de Lima, A. N. Tait, J. B. Khurgin, M. Miscuglio, B. J. Shastri, P. R. Prucnal, T. El-Ghazawi, and V. J. Sorger, "ITO-based electro-absorption modulator for photonic neural activation function," APL Mater. **7**(8), 81112 (2019).
26. R. Amin, J. K. George, H. Wang, R. Maiti, Z. Ma, H. Dalir, J. B. Khurgin, and V. J. Sorger, "An ITO–graphene heterojunction integrated absorption modulator on Si-photonics for neuromorphic nonlinear activation," APL Photonics **6**(12), 120801 (2021).
27. H.-Y. Chang and K. L. Wang, "Deep unitary convolutional neural networks," arXiv, arXiv.2102.11855 (2021).
28. P. Xu and Z. Zhou, "Silicon-based optoelectronics for general-purpose matrix computation: a review," Adv. Photonics **4**(04), 044001 (2022).
29. H. Zhou, J. Dong, J. Cheng, W. Dong, C. Huang, Y. Shen, Q. Zhang, M. Gu, C. Qian, H. Chen, Z. Ruan, and X. Zhang, "Photonic matrix multiplication lights up photonic accelerator and beyond," Light: Sci. Appl. **11**(1), 30 (2022).
30. T. Y. Teo, X. Ma, E. Pastor, H. Wang, J. K. George, J. K. W. Yang, S. Wall, M. Miscuglio, R. E. Simpson, and V. J. Sorger, "Programmable chalcogenide-based all-optical deep neural networks," Nanophotonics **11**(17), 4073–4088 (2022).
31. A. Jha, C. Huang, and P. R. Prucnal, "Reconfigurable all-optical nonlinear activation functions for neuromorphic photonics," Opt. Lett. **45**(17), 4819–4822 (2020).
32. C. Pappas, S. Kovaios, M. Moralis-Pegios, A. Tsakyridis, G. Giamougiannis, M. Kirtas, J. Van Kerrebrouck, G. Coudyzer, X. Yin, N. Passalis, A. Tefas, and N. Pleros, "Programmable Tanh-, ELU-, Sigmoid-, and Sin-based nonlinear activation functions for neuromorphic photonics," IEEE J. Sel. Top. Quantum Electron. **29**(6: Photonic Signal Processing), 1–10 (2023).
33. Z. Xu, B. Tang, X. Zhang, J. F. Leong, J. Pan, S. Hooda, E. Zamburg, and A. V.-Y. Thean, "Reconfigurable nonlinear photonic activation function for photonic neural network based on non-volatile opto-resistive RAM switch," Light: Sci. Appl. **11**(1), 288 (2022).
34. I. A. D. Williamson, T. W. Hughes, M. Minkov, B. Bartlett, S. Pai, and S. Fan, "Reprogrammable electro-optic nonlinear activation functions for optical neural networks," IEEE J. Sel. Top. Quantum Electron. **26**(1), 1–12 (2020).
35. B. Yuen, M. T. Hoang, X. Dong, and T. Lu, "Universal activation function for machine learning," Sci. Rep. **11**(1), 18757 (2021).

36. S. Balaji, T. Kavya, and N. Sebastian, "Learn-Able Parameter Guided Activation Functions," in *Intelligent Systems and Applications*, K. Arai, S. Kapoor, and R. Bhatia, eds. (Springer International Publishing, 2021), pp. 583–597.

37. W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, "Optimal design for universal multiport interferometers," Optica **3**(12), 1460–1465 (2016).

38. S. A. Fldzhyan, M. Y. Saygin, and S. P. Kulik, "Optimal design of error-tolerant reprogrammable multiport interferometers," Opt. Lett. **45**(9), 2632–2635 (2020).

39. S. Bandyopadhyay, R. Hamerly, and D. Englund, "Hardware error correction for programmable photonics," Optica **8**(10), 1247–1255 (2021).

40. D. A. B. Miller, "Self-configuring universal linear optical component," Photonics Res. **1**(1), 1–15 (2013).

41. R. Tian, X. Gan, C. Li, X. Chen, S. Hu, L. Gu, D. Van Thourhout, A. Castellanos-Gomez, Z. Sun, and J. Zhao, "Chip-integrated van der Waals PN heterojunction photodetector with low dark current and high responsivity," Light: Sci. Appl. **11**(1), 101 (2022).

42. E. Cole, J. Cheng, J. Pauly, and S. Vasanawala, "Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications," Magn. Reson. Med. **86**(2), 1093–1109 (2021).

43. C. Bruynsteen, M. Vanhoecke, J. Bauwelinck, and X. Yin, "Integrated balanced homodyne photonic–electronic detector for beyond 20 GHz shot-noise-limited measurements," Optica **8**(9), 1146–1152 (2021).

44. S. Scardapane, S. Van Vaerenbergh, A. Hussain, and A. Uncini, "Complex-valued neural networks with nonparametric activation functions," IEEE Trans. Emerg. Top. Comput. Intell. **4**(2), 140–150 (2020).

45. V. J. Sorger, N. D. Lanzillotti-Kimura, R.-M. Ma, and X. Zhang, "Ultra-compact silicon nanophotonic modulator with broadband response," Nanophotonics **1**(1), 17–22 (2012).

46. R. A. Fisher, "The use of multiple measurements in taxonomic problems," Ann. Eugen. **7**(2), 179–188 (1936).

47. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning* (2013), pp. 108–122.

48. J. A. Barrachina, C. Ren, G. Vieillard, C. Morisseau, and J.-P. Ovarlez, "Theory and implementation of complex-valued neural networks," arXiv, arXiv:2302.08286 (2023).

49. J. A. Barrachina, "Library to help implement a complex-valued neural network (CVNN) using tensorflow as back-end," (2021).

50. J. Navarro-Arenas, "Complex valued trainable activation," Zenodo, 2023, https://zenodo.org/record/8328002.

51. K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on ImageNet classification," arXiv, arXiv:1502.01852 [cs.CV] (2015).

52. R. Savitha, S. Suresh, and N. Sundararajan, "A Fast Learning Complex-valued Neural Classifier for real-valued classification problems," in *The 2011 International Joint Conference on Neural Networks* (2011), pp. 2243–2249.