Photonic Kolmogorov Arnold Networks based on Self-Phase Modulation in Nonlinear waveguides

KOSTAS SOZOS,[1*] DIMITRIOS SPANOS,[2] STAVROS DELIGIANNIDIS,[1] GEORGE SARANTOGLOU,[3] NIKOLAOS PASSALIS,[4] NIKOS PLEROS,[5] CHARIS MESARITAKIS,[3] ANASTASIOS TEFAS,[2] ADONIS BOGRIS[1]

[1] *Department of Informatics and Computer Engineering, University of West Attica, Aghiou Spiridonos, 12243, Egaleo, Athens, Greece*
[2] *Computational Intelligence and Deep Learning Group, AIIA Lab, Dept. of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[3] *Department of Biomedical Engineering, University of West Attica, Aghiou Spiridonos, 12243, Egaleo, Athens, Greece*
[4] *Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece*
[5] *Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece, Wireless and Photonic Systems and Networks Group, Center for Interdisciplinary Research and Innovation, Thessaloniki, Greece*
*ksozos@uniwa.gr

Photonic neural networks have attracted intense interest during the past decade, with an extensive number of propositions trying to mimic various forms of the established digital models. This field envisages low-power and low-latency, highly parallelized hardware implementations counting on the vast capacity of photonics. Kolmogorov-Arnold Networks were recently proposed offering improved scalability and interpretability compared to conventional multilayer perceptrons. From this point of view, Kolmogorov Arnold Networks constitute a promising field for implementations which exploit the full pallet of photonic strengths, mitigating the scalability barrier. Here, we propose and experimentally showcase the potential of self-phase modulation in nonlinear waveguides as a versatile generator of arbitrary nonlinear activation functions for the ultra-fast photonic implementation of Kolmogorov-Arnold networks. The efficacy of the proposed approach is demonstrated on both regression and classification tasks, where the photonic implementation matches digital performance and, in specific cases, surpasses the digital baseline used in our comparison.

Artificial intelligence (AI) and machine learning are currently driving in a disruptive manner the technological evolution in the field of high-performance computing in applications spanning from natural language processing to generative biology. Up to date, artificial intelligence relies on neural networks whose main building block is a multi-layer perceptron that performs linear weighting, summation and fixed nonlinear transformation sequentially [1]. Photonic implementations of neural networks have attempted to mimic the same computing sequence mostly focusing on the implementation of high fidelity and scalable linear transformations in the form of matrix-vector or matrix-matrix multiplications [2–4]. However, limitations in scalability and cascadability hinder photonic networks from achieving parity with their digital counterparts. Very recently, the paradigm of Kolmogorov-Arnold networks (KAN) has attracted the interest of AI community [5]. KANs constitute a different approach to universal approximation by leveraging learnable univariate functions on each input dimension of a layer. The use of nonlinear transformations followed by simple summation allows KANs adapt their functional basis dynamically by learning low-dimensional nonlinear transformations, thereby decomposing complex functions into sums of univariate components. This representation grants passive interpretability, as the model's learned functional elements are explicitly accessible, and can enhance efficiency on structured or low-dimensional tasks. [5]. Most importantly, KANs have the potential to achieve accuracy comparable to, or sometimes exceeding, that of MLP-based neural networks, while requiring fewer neurons or shallower architectures thus enabling efficient hardware implementations for computing acceleration. With photonic technologies inherently supporting a strong nonlinearity portfolio, a first simulation-based attempt has been recently successfully presented for exploiting the potential of Ring-assisted Mach-Zehnder interferometers towards serving as parametric generators of nonlinearity in photonic KAN implementations [6]. Another recent analysis unifies k-local Ising optimization and optical Kolmogorov-Arnold network (KAN) learning on a single photonic platform based on computing primitives implemented in spatial light modulators [7]. In this work, we rely on the foundational nonlinear Kerr effect in nonlinear waveguides, leveraging Self-Phase Modulation (SPM) followed by offset filtering for implementing arbitrary nonlinear functions that constitute the core computational primitives of our KAN approach. We repurpose the historical Mamyshev regenerator [8] and experimentally demonstrate the use of SPM induced spectral broadening and offset filtering for generating the necessary pool of nonlinear functions, which are subsequently utilized for the successful demonstration of photonic KANs. Their

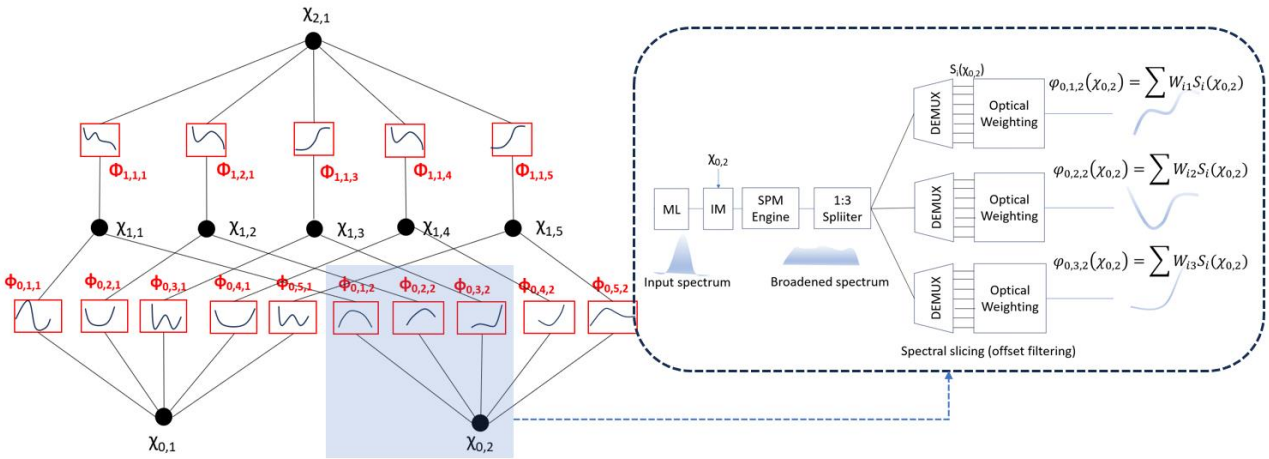

Fig. 1: The conceptual scheme of a photonic KAN operating with the use of SPM modules followed by spectral slicing and weighting that offers programmable nonlinearity. With the use of a splitter, SPM output can be broadcasted to multiple slicing-weighting modules for tailoring the univariate nonlinearity on the edges.

evaluation in both classification (IRIS dataset) and regression (heartrate prediction) tasks reveals a performance that equals the performance of respective digital KAN implementations if the SPM based univariate functions are used to approximate those generated by the digital models. When the KAN is trained using SPM-based univariate basis functions, the regression and classification results show a clear improvement with respect to the considered digital solutions, meaning that photonic KANs based on Mamyshev-inspired nonlinear activation functions have a significant potential in offering universal approximation at very high bandwidths leveraging the ultra-fast response of Kerr effect.

The universal nonlinear function generator based on SPM is depicted in fig. 1. A mode locked pulsed laser passes through an intensity modulator (IM) where the signal $\chi_{0,i}$ is applied and enters a highly nonlinear waveguide. The modulated signal experiences self-phase modulation which leads to spectral broadening. This is what we call SPM engine in fig. 1. If power and wavelength are properly selected, it may also experience modulation instability in the anomalous dispersion regime that results in stronger amplitude modulation [9]. Using offset filtering at different frequency detunings from the central wavelength one can obtain differentiated outputs which are nonlinear representations of the input signal $\chi_{0,i}$. By using simple optical weighting one can finally attain to tailor the nonlinearity so as to approximate any arbitrary nonlinear activation required by KANs in regression or classification tasks. The filtering-weighting unit can be implemented in the optical domain with the use of demultiplexers and weight banks [10]. Here, we use a wavelength selective switch (WSS) in order to implement both demultiplexing and optical weighting. Depending on the fan-out (one to many), each SPM engine can be re-used with a 1:N broadcasting to different filtering-weighting units so as to generate multiple learnable nonlinear functions $\varphi(\chi)$ on edges as a linear combination of the SPM generated functions as shown in fig. 1. Therefore, each output of the DEMUX depicted in fig. 1 ($S_i(\chi)$) acts as a basis function in our SPM-based programmable nonlinear function generator similarly to splines or radial basis functions commonly used in KANs [5]. The summation of different $\varphi_{0,i,j}$ depicted in fig. 1 can be accomplished with the use of photodetectors and their output will modulate the intensity of mode-locked lasers in order to repeat the process of SPM, filtering and optical weighting. The process is fully analog avoiding digitization bottleneck.

Fig. 2 shows the experimental setup for the measurement of the independent SPM generated functions $S_i(\chi)$. As the SPM source, we have used a 50 fs mode-locked laser with a repetition rate of 100 MHz. We properly attenuate its power and then slice its optical spectrum using an optical band pass filter in order to keep almost 0.2 nm which results in a pulse broadening to ~50 ps. The pulses are then sent to a Mach-Zehnder modulator that is driven with a linear ramp signal at 5 MHz resulting in a dynamic range of almost 20 dB in the input power (χ in fig. 1). The signal is then amplified before being launched into a 1 km long highly nonlinear fiber (HNLF) with λ_0=1542 nm, γ=12 W^{-1}km^{-1} and dispersion slope equal to $dD/d\lambda$ = 0.03 ps nm^{-2} km^{-1}. At the output we use the WSS as single or multi-10 GHz selective switch in order to mimic the operation of the DEMUX depicted in fig. 1. We select and record different timetraces at different spectral positions. In total we select 24 positions (12 per sideband) for an input power of 6 dBm with a frequency separation of 10 GHz. We scan the transfer characteristics of different spectral regions by detuning the WSS in terms of the input central frequency by 10 GHz from 120 GHz at the blue shifted sideband and up to the 120 GHz in the red shifted sideband. In this respect we take 24 unique power transfer characteristics at the output of the optical nonlinear processor. In fig. 2b, the spectrally broadened spectra are depicted for different input power levels scaling from 0 to 8 dBm. These spectra are characteristic of SPM broadening and have appeared in many publications in the past and utilized for regeneration and oscillation using Mamyshev's pioneering idea of offset filtering [11].

The recorded $S_i(\chi)$ functions at twelve spectral positions (blue shifted components) are depicted in fig. 2c when we

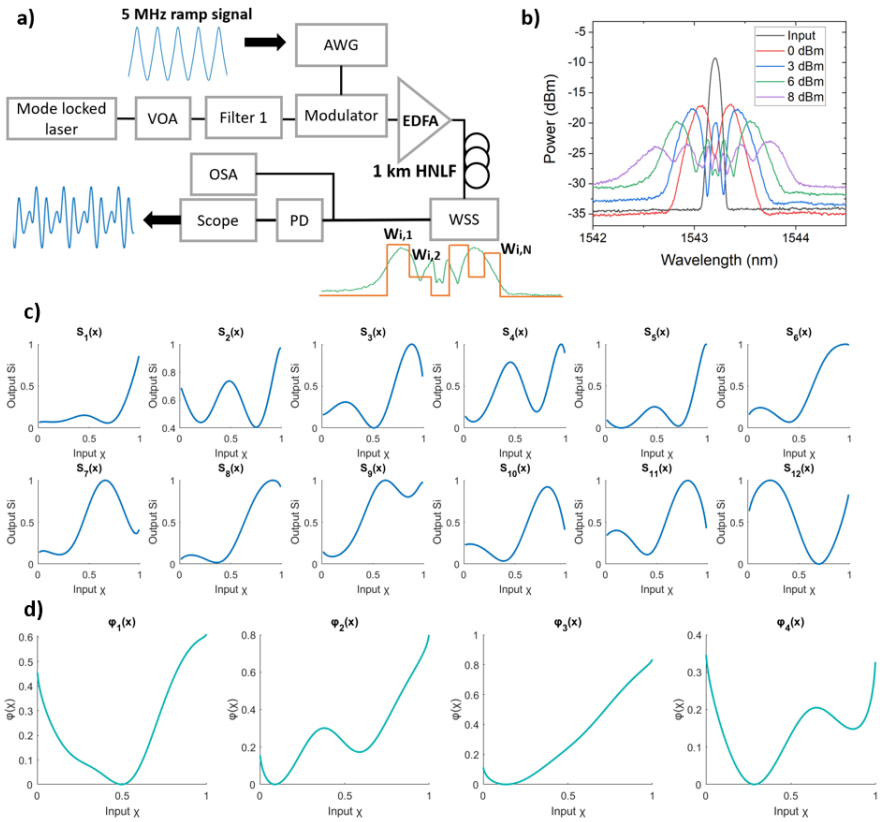

Fig. 2: a. Experimental Setup. WS: Waveshaper, AWG: Arbitrary waveform generator, MZM: Mach-Zehnder Modulator, EDFA: Erbium Doped Fiber Amplifier, TOBPF: tunable optical bandpass filter, VOA: variable optical attenuator, OSA: Optical Spectrum Analyzer. b. The broadened optical spectra obtained vs. power. c. Si functions (normalized values) after blue shifted offset filtering (20 GHz to 130 GHz) – 6 dBm input power. d. φ(χ) functions produced by optical weighting of 4 from the 12 Si traces in the waveshaper.

launch 6 dBm of average power in the HNLF. Twelve, clearly different, sigmoid-like or ReLU like functions with some of them exhibiting non-monotonic behaviour constitute the S_i functions. This non-monotonic S-shaped profile is theoretically predicted [12] and is associated with the S-shaped transfer characteristics of the Mamyshev regenerator [8], modulated by the periodic spectral lobes of SPM-induced spectral broadening whose position varies as a function of power (fig. 2b). Applying offset filtering at varying positions induces basis functions with maxima and minima distributed across different regions of the input domain, thereby enhancing the diversity of local responses. This diversity is fundamental to achieving the expressive capacity required for universal function approximation. Similar nonlinear properties of the diverse activation functions have been theoretically estimated based on FWM [13]. In Fig. 3d., we provide four optically weighted nonlinear functions $\varphi(\chi)$, produced by optically multiplying the 4 of the 12 S_i with random weights in the waveshaper.

We validate the SPM-based KAN in well-known classification and regression tasks. For the classification task, we employ the IRIS dataset [14]. We trained a digital KAN model with one learnable nonlinearity of the form $\varphi(x) = \sum_{i=1}^{8} c_i B_i$ where B_i is a B-spline per neuron [5]. Each nonlinearity uses 8 basis functions and incorporates a residual connection with the SiLU activation function. The model architecture consists of 4 input neurons and 3 output neurons. Training is conducted over 100 epochs using a batch size of 32. We use the Adam optimizer with a learning rate of 0.01. For the time series regression task, we utilize a heart rate prediction dataset comprising 1800 evenly spaced measurements of instantaneous heart rate (in beats per minute), recorded at 0.5-second intervals over a 15-minute period from a single subject engaged in continuous activity [15]. The first 1200 samples (approx. 67%) were used to train the model. The remaining 600 samples (approx. 33%) were reserved for evaluation. The model predicts the next timestep's heart rate based on a sequence of the preceding 12 measurements. The KAN used in this setup has the same form of nonlinearity per neuron as in the classification task. Training is performed for 300 epochs with a batch size of 16, using the Adam optimizer and a fixed learning rate of 10^{-3}.

For the evaluation of the SPM-KAN scheme we follow two approaches. Firstly, the nonlinear functions produced by the digital model at the edges of each neuron are approximated by weighting the basis functions of the SPM-based engine. Using the 10 out of the 12 acquired basis functions from the blue shifted slices, we have managed to

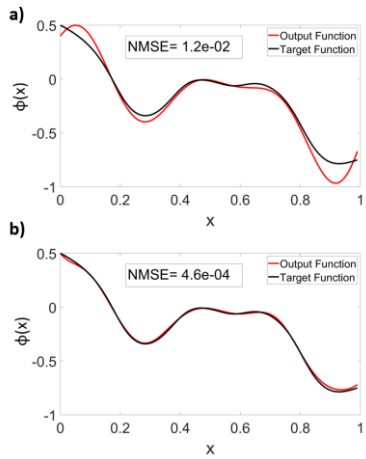

Fig. 3: SPM-based generator approximation to digitally produced nonlinear activations from IRIS task. 10 basis functions provide almost ideal approximation with NMSE < 0.001. 6 basis functions result to NMSE~0.012.

approximate the digitally generated functions using *B*-spline basis functions in all studied tasks with normalized mean square error below 10^{-3}. Fig. 3 shows an example of the approximation capabilities of SPM-based nonlinear generator when ten basis functions or six of them are employed. Then, we evaluate the performance of the SPM-generator using the digitally trained KAN models as baseline by substituting the digitally generated learnable nonlinearities with the photonic approximations and present the performance in Table 1. In IRIS task the baseline model achieved 93.33% accuracy, with the same performance being achieved with the SPM approximator with 8 basis functions. In the regression task, the baseline achieved normalized root MSE (NRMSE) equal to 0.1462 in average heart rates ranging from 80-90 bpm. The photonic approximator achieved RMSE=0.148, only slightly worse than the digital counterpart. As a second approach, we directly train the models with the SPM produced basis functions using gradient descent and back-propagation. In this case, the training with the acquired SPM based splines evidently outperforms the digital algorithm in both tasks resulting in 97.33% accuracy for the IRIS problem and NRMSE=0.0856 for the heart rate problem. The observed superiority can be attributed to the higher representational expressiveness of the SPM-based nonlinear activation functions compared to the specific digital implementation, enabling extensive and diverse mapping across the full input field space as shown in fig. 2c.

In conclusion, we experimentally demonstrated a versatile and programmable nonlinear approximator based on SPM in nonlinear waveguides and evaluated its capabilities in photonic KANs leveraging the ultra-fast and expressive Kerr effect in regression and classification tasks. The approach shows very promising classification and regression performance which is equivalent or even better than that offered by digitally implemented KAN models. Recent works on silicon nitride nonlinear waveguides [16] showcase that the nonlinear efficiency of optical fibers in terms of $L\gamma$ product

Table 1: SPM-based KAN vs KAN based on splines [5]

Task	KAN [5]	SPM – KAN Spline Approximation	SPM – KAN Direct Training with SPM basis functions
IRIS	Accuracy= 93.33%	Accuracy = 93.33%	Accuracy = 97.33%
Heart rate	NRMSE = 0.1462	NRMSE = 0.148	NRMSE = 0.0856

can be adequately approached paving the way for monolithic integration of the approach. Co-integrating a picosecond mode-locked laser and a fast modulator with the SiN waveguides on a single chip could provide a pathway to further miniaturization of the overall system. Future work will focus on training the KAN network and implementing it in real time on the photonic platform.

Funding. This work has been supported by Horizon Europe project PROMETHEUS under grant agreement 101070195. The authors also acknowledge support by the EC via the Horizon Europe programme (via Chips JU) HAETAE (101194393).

Disclosures.
The authors declare no conflicts of interest.
Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. K. Hornik, M. Stinchcombe, and H. White, Neural Networks **2**, 359 (1989).
2. N. Peserico, B. J. Shastri, and V. J. Sorger, J. Lightwave Technol. **41**, 3704 (2023).
3. 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, Light Sci Appl **11**, (2022).
4. A. Tsakyridis, M. Moralis-Pegios, G. Giamougiannis, M. Kirtas, N. Passalis, A. Tefas, and N. Pleros, APL Photonics **9**, (2024).
5. Z. Liu, Y. Wang, S. Vaidya, F. Ruehle, J. Halverson, M. Soljačić, T. Y. Hou, and M. Tegmark, (2025).
6. Y. Peng, S. Hooten, X. Yu, T. V. Vaerenbergh, Y. Yuan, X. Xiao, B. Tossoun, S. Cheung, M. Fiorentino, and R. Beausoleil, (2024).
7. N. Stroev and N. G. Berloff, (2025).
8. P. V. Mamyshev, in *24th European Conference on Optical Communication. ECOC '98 (IEEE Cat. No.98TH8398)* (Telefonica, n.d.), **1**, pp. 475–476.
9. M. J. Potasek and G. P. Agrawal, Phys. Rev. A **36**, 3862 (1987).
10. A. N. Tait, A. X. Wu, T. F. De Lima, E. Zhou, B. J. Shastri, M. A. Nahmias, and P. R. Prucnal, IEEE J. Select. Topics Quantum Electron. **22**, 312 (2016).
11. C. Finot and M. Rochette, Nanophotonics **14**, 2835 (2025).
12. C. Finot, F. Chaussard, and S. Boscolo, J. Opt. Soc. Am. B **35**, 3143 (2018).
13. K. Sozos, S. Deligiannidis, C. Mesaritakis, and A. Bogris, IEEE J. Quantum Electron. **60**, 1 (2024).
14. R. A. Fisher, Annals of Eugenics **7**, 179 (1936).
15. G. B. Moody and R. G. Mark, IEEE Eng. Med. Biol. Mag. **20**, 45 (2001).
16. P. Zhao, M. Karlsson, and P. A. Andrekson, J. Lightwave Technol. **40**, 128 (2022).