PHOTONICS Research

Photonic extreme learning machine by free-space optical propagation

DAVIDE PIERANGELI,[1,2,3,*] **GIULIA MARCUCCI,**[4] (iD) **AND CLAUDIO CONTI**[1,2,3] (iD)

[1]*Institute for Complex System, National Research Council (ISC-CNR), 00185 Rome, Italy*
[2]*Physics Department, Sapienza University of Rome, 00185 Rome, Italy*
[3]*Centro Ricerche Enrico Fermi (CREF), 00184 Rome, Italy*
[4]*Department of Physics, University of Ottawa, Ottawa, Ontario K1N 6N6, Canada*
*Corresponding author: davide.pierangeli@roma1.infn.it

Photonic brain-inspired platforms are emerging as novel analog computing devices, enabling fast and energy-efficient operations for machine learning. These artificial neural networks generally require tailored optical elements, such as integrated photonic circuits, engineered diffractive layers, nanophotonic materials, or time-delay schemes, which are challenging to train or stabilize. Here, we present a neuromorphic photonic scheme, i.e., the photonic extreme learning machine, which can be implemented simply by using an optical encoder and coherent wave propagation in free space. We realize the concept through spatial light modulation of a laser beam, with the far field acting as a feature mapping space. We experimentally demonstrate learning from data on various classification and regression tasks, achieving accuracies comparable with digital kernel machines and deep photonic networks. Our findings point out an optical machine learning device that is easy to train, energetically efficient, scalable, and fabrication-constraint free. The scheme can be generalized to a plethora of photonic systems, opening the route to real-time neuromorphic processing of optical data.   © 2021 Chinese Laser Press

https://doi.org/10.1364/PRJ.423531

1. INTRODUCTION

Artificial neural networks excel in learning from data to perform classification, regression, and prediction tasks of vital importance [1]. As data information content increases, simulating these models becomes computationally expensive on conventional computers, making specialized signal processors crucial for intelligent systems. Training large networks is also costly in terms of energy consumption and carbon footprint [2]. Photonics provides a valuable alternative toward sustainable computing technologies. For this reason, machine learning through photonic components is gaining enormous interest [3]. In fact, many mathematical functions, which enable complex feature extraction from data, find a native implementation on optical platforms. Pioneering attempts using photosensitive masks [4] and volume holograms [5,6] have been recently developed into coherent optical neural networks that promise fast and energy-efficient optical computing [7–16]. These schemes exploit optical units such as nanophotonic circuits [7], on-chip frequency combs [14–16], and spatial light modulators to perform matrix multiplications in parallel [17] or to carry out convolutional layers [18–20]. Training consists in adjusting the optical response of each physical node [21], also by adopting external optical signals [22], which is demanding [23]. Moreover, photonic neural networks based on nanofabrication

still have a considerable energy impact. A general and promising idea to overcome the issue is to adapt machine-learning paradigms and make them inclined to optical platforms. In this paper, we pursue this approach by constructing an easy-to-train optical architecture that requires only free-space optical propagation.

Photonic architectures that do not need control of the entire network are particularly attractive. A remarkable method for their design and training is reservoir computing [24–26], in which the nonlinear dynamics of a recurrent system processes data, and training occurs only on the readout layer. Optical reservoir computing has demonstrated striking performance on dynamical series prediction by using delay-based setups [27–30], laser networks [31], multimode waveguides [32], and random media [33]. Since several interesting complex systems can be exploited as physical reservoirs, the inverse approach is also appealing, i.e., the scheme can be trained for learning dynamical properties of the reservoir itself [34].

Extreme learning machines (ELMs) [35,36], or closely related schemes based on random neural networks [37,38], support vector machines [39] and kernel methods [40], are a powerful paradigm in which only a subset of connections is trained. ELMs perform comparably with basic deep and recurrent neural networks on the majority of learning tasks [36]. The basic mechanism is to use the network to establish a non-

linear mapping between the data set and a high-dimensional feature space, where a properly trained classifier performs the separation. Training occurs at the readout layer; however, at variance with reservoir computing, ELMs have no recurrences in the connection topology and do not possess dynamical memory. In optics, an interesting case of this general approach has been implemented by using speckle patterns emerging from multiple scattering [41] and multimode fibers [42] as a feature mapping. Although in principle many optical phenomena can form the building block of the architecture [43,44], the general potential of the ELM framework for photonic computing remains largely unexplored.

Here, we propose a photonic extreme learning machine (PELM) for performing classification and regression tasks using coherent light. We find that a minimal optical setting composed by an input element, which encodes embedded input data into the optical field, a wave layer, corresponding to propagation in free space, and a nonlinear detector, enables simple and efficient learning. The encoding method plays a crucial role in determining the learning ability of the machine. The architecture is experimentally implemented via phase modulation of a visible laser beam by a spatial light modulator (SLM). We benchmark the realized device on problems of different classes, achieving performance comparable with digital ELMs. These include a classification accuracy exceeding 92% on the well-known MNIST database, which overcomes fabricated diffractive neural networks [8]; further, it is comparable with convolutional artificial networks that employ photonic accelerators [15,16,19]. Given the massive parallelism provided by spatial optics and the ease of training, our approach is ideal for big data, i.e., extensive data sets with large dimension samples. Our PELM is intrinsically stable and adaptable over time, as it does not require engineered or sensitive components. It can potentially operate in dynamic environments as an intelligent device performing on-the-fly optical signal processing.

2. PELM ARCHITECTURE

An ELM is a feed-forward neural network in which a set of input signals is connected to at least one output node by a large structure of generalized artificial neurons [35]. An outline of the architecture is illustrated in Fig. 1(a). The hidden neurons form the computing reservoir. Unlike for neurons in a deep neural network, they do not require any training, and their individual responses can be unknown to the user [36]. Given an input data set with N samples $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_N]$, the reservoir furnishes an hidden-layer output matrix $\mathbf{H} = [\mathbf{g}(\mathbf{X}_1), \ldots, \mathbf{g}(\mathbf{X}_N)]$, where $\mathbf{g}(\mathbf{x})$ is a nonlinear feature mapping, $\mathbf{H}$ is linearly combined with a set of weights β to give the output $\mathbf{Y} = \mathbf{Y}(\mathbf{H}; \beta)$ performing the classification, and $\mathbf{Y} = \mathbf{H}\beta$. To train an ELM classifier, the optimal output weights β_i are the sole parameters that need to be determined, a problem that can be solved via ridge regression. Details on this technique are reported in Appendix A.

We transfer the ELM principle into the optical domain by considering the three-layer structure illustrated in Fig. 1(b). In the encoding layer, the input vectors $\mathbf{X}_i$ are embedded into the phase and/or amplitude of the field by an optical modulator. The reservoir consists of linear optical propagation and nonlin-
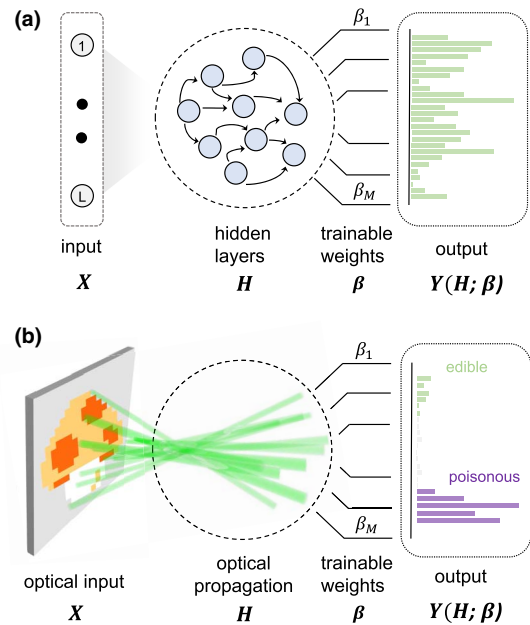

Fig. 1. Schematic architecture of the photonic extreme learning machine (PELM). (a) General ELM scheme with the input data set $\mathbf{X}$, which is fed into a reservoir and gives out the hidden-layer output matrix $\mathbf{H}$. The trainable readout weights β_i determine the network output $\mathbf{Y} = \mathbf{Y}(\mathbf{H}; \beta)$. (b) In the optical case, the input (a mushroom in the example) is encoded on the optical field, and hidden neurons have been replaced by modes that interact during propagation. Training of the photonic classifier is enabled by M detection channels.

ear detection of the final state. The output is recovered in the readout layer, where weights β_i are applied to M measured channels. The β set is trained by solving the regression problem via digital hardware. For an extensive training set, with size larger than the number of channels ($N \gg M$), an effective solution reads as [36]

$$\beta = (\mathbf{H}^T\mathbf{H} + c\mathbf{I})^{-1}\mathbf{H}^T\mathbf{T}, \tag{1}$$

where $\mathbf{T}$ indicates the training targets, $\mathbf{I}$ the identity matrix, and c is a regularizing constant. $\mathbf{H}$ contains projections of input samples in the feature space, where data are separated according to their features. This mapping is the key ingredient for learning and occurs entirely by means of optical propagation. Mapping functions can be constructed starting from the elementary components of the optical setup. For a general PELM based on the scheme in Fig. 1(b), we construct the feature mapping

$$H_{ji} = \mathbf{g}_i(\mathbf{X}_j) = G[\mathbf{M} \cdot p(\mathbf{X}_j) \cdot q(\mathbf{W})]_i, \tag{2}$$

which describes linear propagation of input optical data and their nonlinear detection. In Eq. (2), G is a detection function, $\mathbf{M}$ is the complex transfer matrix modelling field transmission, p and q are two encoding functions, and $\mathbf{W}$ is a fixed character of the encoder that we term embedding matrix. $\mathbf{W}$ has no direct equivalent in the basic ELM model and serves taking into account that any laser beam has spatial amplitude modulations and phase inhomogeneities. The elements of this matrix can be understood as biases or as entries of a fan-in matrix connecting the input signal to the reservoir. We remark that, at variance with the ELM

model, in the optical architecture, input and output data, network connections, and processing functions are all complex quantities. This fact can add an important degree of freedom to the network operation [45].

In the PELM architecture, each operation is defined by the corresponding optical elements. Therefore, Eq. (2) can represent the feature space of different optical settings. For instance, in the optical classifier demonstrated by Saade *et al.* [41], in which amplitude modulated data propagate through a scattering medium, $p(\mathbf{X}_j) = \mathbf{X}_j$ (amplitude encoding) and $\mathbf{M}$ is a random complex Gaussian matrix. We validate and experimentally realize a three-component system composed by a phase-only SLM, free space, and a camera. In our case, Eq. (2) can thus be specified as follows. Phase encoding of the input data by spatial light modulation corresponds to $p(\mathbf{x}) = \exp(i\mathbf{x})$, and, since a single encoder is employed, $p = q$, and $\mathbf{H} = G[\mathbf{M} \cdot \exp i(\mathbf{X} + \mathbf{W})]$. The nonlinear function G models the detection of the transmitted field. Using the saturation effect of the camera pixels, we have $G(I) \simeq I/(I + I_s)$, with $I = |A|^2$, optical amplitude A, and saturation intensity I_s. For free-space optical propagation, i.e., the light distribution in the far field or in the focal plane of a lens, $\mathbf{M}$ corresponds to the discrete Fourier transform [46].

We first validate the free-space PELM architecture and assess the condition for learning via numerical simulations. We consider digit recognition and train our classifier on the MNIST handwritten digit database. Figure 2 reports the learning properties for two representative phase-encoding methods: (i) noise embedding and (ii) Fourier embedding. In (i), $\mathbf{W}$ is a disordered matrix modelling a distortion on the encoder (see Appendix A for details). It remains unchanged during both training and testing. Each input is encoded by phase modulation in $[0, \pi]$, and the embedding signal is encoded in the same phase interval. The effect of the noise embedding is illustrated in Fig. 2(a), where a typical digit is shown as a phase mask. Figure 2(b) shows the classification performance of the PELM with $M = 1600$ channels when varying the mean noise amplitude. The machine always reaches accuracies close to that allowed by training. The classification error shows a sharp decrease as noise increases from zero, and it converges to a plateau already for small perturbations. Results varying the noise correlation length at fixed noise level are in Fig. 2(c). This behavior indicates that optimal learning occurs for small-scale noise, i.e., when $\mathbf{W}$ has its maximum information content (rank L). The relevance of $\mathbf{W}$ for learning extends to embedding matrices without randomness. In (ii), $\mathbf{W}$ is a modulated carrier signal [Fig. 2(d)]. We are superimposing the input data on a carrier pulse; feature mapping corresponds to analyzing the intensity of the entire spectrum after a nonlinear amplifier. In this case, the learning process can be interpretable: the learned features point out resonances between the input signal and the carrier pulse. As shown in Fig. 2(e), a few frequencies in the carrier are sufficient for the learning transition. The PELM is much more accurate in the classification as the frequency content of the embedding signal is larger. These results reveal that the embedding matrix plays a key role in enabling data-set learning. Although, in our experiments, we control this matrix via the encoder, we note that it can be
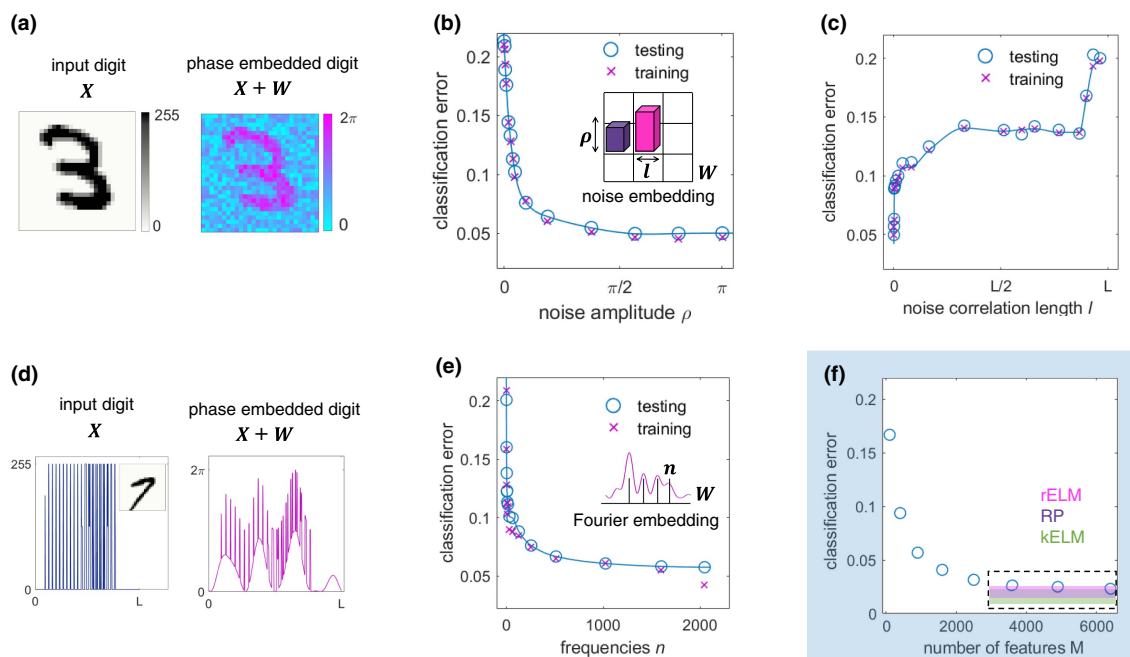

Fig. 2. Learning ability of the PELM architecture. The optical computing scheme is evaluated on the MNIST data set by varying the encoding properties and feature space. (a) Input digit and 2D phase mask showing its encoding by noise embedding: the input signal overlaps with a disordered matrix. PELM training and testing error for noise embedding when varying the (b) noise amplitude and (c) its correlation length, for $M = 1600$. (d) Input vector encoded over a carrier signal (Fourier embedding). (e) Classification error versus the number of frequencies of the embedding signal. (f) Minimum testing error with the increasing number of features M. The indicated accuracies are the best ones reported with random ELM (rELM) [47], random projections (RP) [41], and kernel ELM (kELM) [47] on the same task.

intrinsic in any optical setup, e.g., a distortion of the optical wavefront.

Another key hyperparameter of the scheme is the number of features (channels) M, which sets the dimensionality of the feature (measurement) space. Figure 2(f) reports the testing error with varying M. The performance rapidly increases with the number of features, whereas only hundreds of channels are necessary for a proficient PELM. For $M \gtrsim N/20$, the error approaches 2%, a value very close to the optimal accuracy achievable with learning machines based on the same paradigm [47]. In these digital machines, the main computational cost both for training and performing the classification is imputable to large matrix multiplications and their processing with nonlinear functions. In the optical device we carry out, these operations occur simply by free-space light propagation and detection. The only calculation we kept on digital hardware is the offline training via simple regression.

3. EXPERIMENTAL DEVICE

The experimental setup for the free-space photonic extreme learning machine is illustrated in Fig. 3(a) and detailed in Appendix B. A phase-only SLM encodes both the input data set and the embedding matrix on the phase spatial profile of the laser beam. In practice, distinct attributes of a data set are arranged in separate blocks, a data sample corresponds to a set of phase values on these blocks, and a preselected phase mask is superimposed [inset in Fig. 3(a)]. Phase information self-mixes during light propagation, a linear process that corresponds to the matrix multiplication in Eq. (2). The far field is detected

on a camera that performs the nonlinear activation function. As shown in Fig. 3(b), the resulting acquired image is a saturated function of the optical amplitude in the lens focal plane. From this data matrix, we extract the M channels to construct the PELM feature space. An example of an input data measured on a feature space is reported in Fig. 3(c). In analogy with the numerical implementation, the device is trained by finding the weights β_i that, when applied the optical output, allow performing the classification.

We test the optical device on various computing tasks and data sets. The aim is to point out that, when the PELM is fully effective on a given task, it can be also easily and rapidly retrained for a different application. The main results are summarized in Fig. 4 and demonstrate the learning capability of the optical device. We first use the MNIST data set to prove classification on a large-scale multiclass problem. When trained using $M = 4096$ output channels, the PELM reaches a mean classification accuracy of 92% (± 0.005) over $N_t = 10,000$ test images [Figs. 4(a) and 4(b)]. We obtain accuracy that exceeds recent optical convolutional processors [15] and is comparable with optical deep neural networks [8], but without the fabrication of specific optical layers and the heavy training they require. The best classification performance is not sensitive to the embedding matrix employed on the encoder, which experimentally confirms our numerical findings (Fig. 2). Figure 4(b) shows a confusion matrix measured using the random embedding method, which gives 92.18% accuracy, compared with the 92.06% [Fig. 4(a)] when using Fourier embedding. Our free-space PELM hence surpasses diffractive neural networks [8]
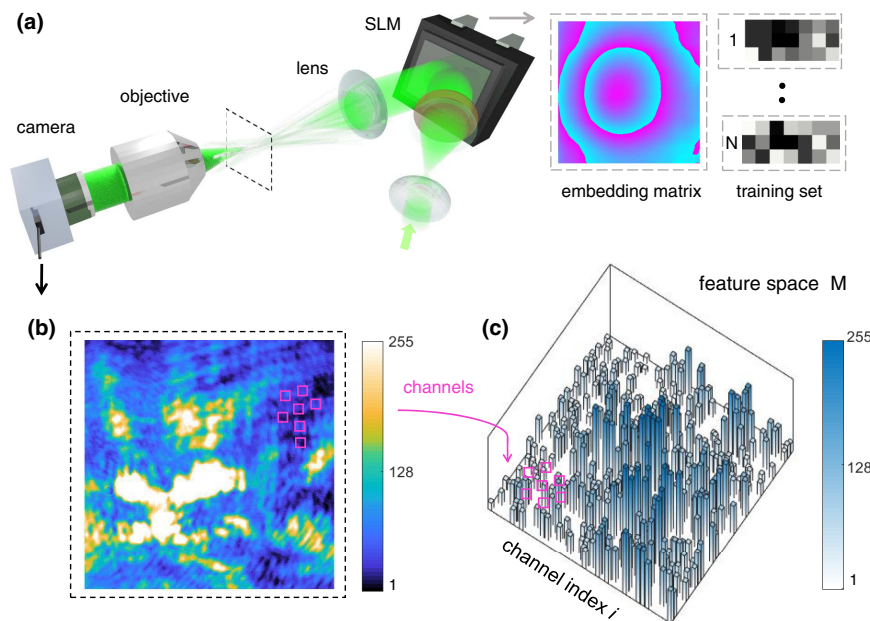

Fig. 3. Experimental implementation. (a) Sketch of the optical setup. A phase-only spatial light modulator (SLM) encodes data on the wavefront of a 532 nm continuous-wave laser. The far field in the lens focal plane is imaged on a camera. Insets show a false-color embedding matrix and training data encoded as phase blocks, respectively. (b) Detected spatial intensity distribution for a given input sample. White-colored areas reveal camera saturation in high-intensity regions, which provides the network nonlinear function. Pink boxes show some of the M spatial modes (blocks of pixels) that are used as readout channels. (c) Example of an input data in a feature space of dimension $M = 256$, as projected by the optical device. Each bar represents an output channel, and training consists in finding the vector that properly tunes all the bar heights.
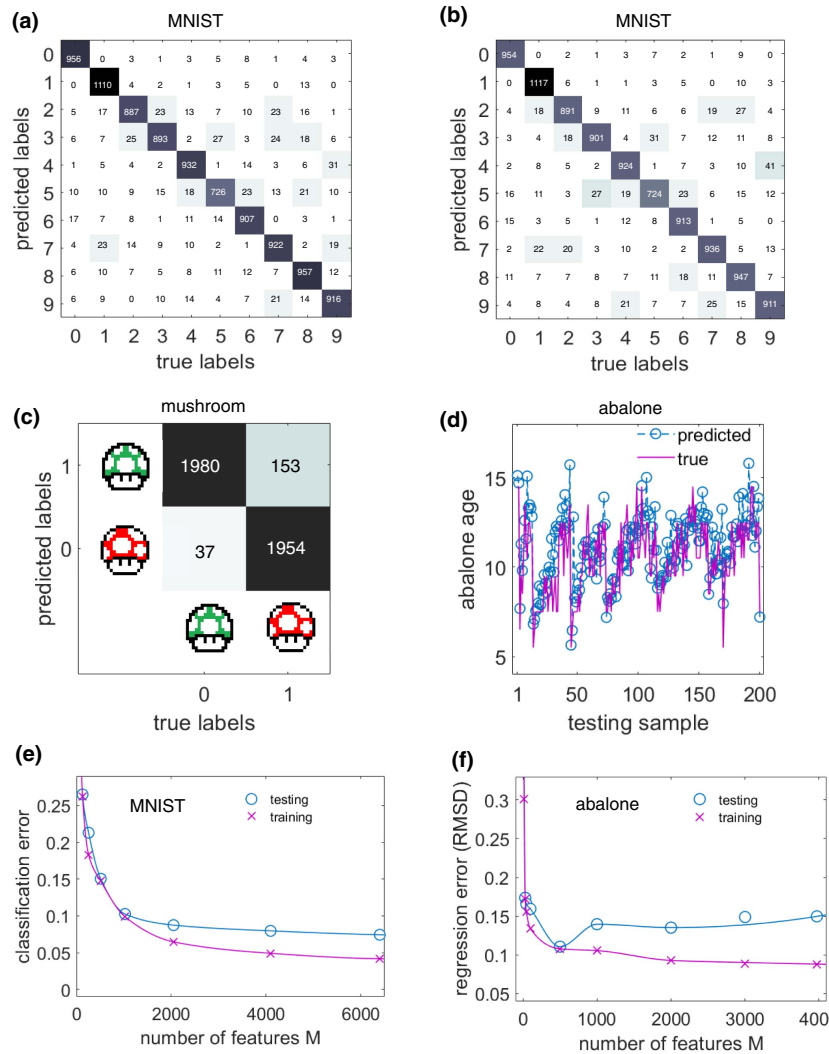
Fig. 4. Experimental performance of the PELM on classification and regression tasks. Confusion matrices on the MNIST data set for a free-space PELM, which makes use of (a) Fourier and (b) random embedding (92.18% and 92.06% accuracy, $M = 4096$). (c) Performance on the mushroom binary classification problem (95.4%). (d) Optical predictions and true values for the abalone data set. (e) Classification and (f) regression error as a function of the number of features. Rapid convergence to optimal performance is found. Experimental results are compared with numerical simulations and training errors.

and competes with cutting-edge photonic hardware in terms of accuracy. In particular, convolutional opto-electronic setups also achieve a coarse accuracy of 92% [19]. Ultrafast photonic processors reach 95% precision with the help of electronic layers [15], whereas similar photonic accelerators operate with 88% accuracy [16].

By only updating the input training set and the output weights, the PELM can be quickly reconfigured for classifying different objects. We consider a binary class problem, i.e., the mushroom task, where the goal is to separate poisonous from edible mushrooms from a series of features (see Appendix C for details). Figure 4(c) shows that, on this data set, we achieve 95.4% accuracy, using $N_t = 4124$ test samples and $N = 4000$ training points, which is close to the precision of digital ELMs [36]. The excellent properties of the experimental device as a classifier generalizes to regression problems. We test the abalone data set (see Appendix C), a task where the input data ciphers a sea snail, and the output is expected to furnish its age. It is a solid

benchmark since it has low-dimensional inputs ($L = 8$) but contains extreme events that are difficult to predict. Figure 4(d) shows experimental abalone predictions obtained via a feature space of dimension $M = 2000$. The testing error measured as normalized root mean square displacement (NRMSD) is ≈0.12, i.e., regression is performed with remarkable accuracy by using only linear optical propagation [44].

A key point of the PELM architecture is a testing error that rapidly converges to its optimal value as the feature space dimension increases. Experimental results demonstrating this property are shown in Figs. 4(e) and 4(f), respectively, for the MNIST and abalone data set. Good classification/regression performance is maintained even for a very small number of trained channels, if compared with the data set size. For instance, useful abalone predictions can be obtained with only $M = 128$ channels, i.e., with a training process that consists only in inverting a modest matrix (128×128 elements). Heuristically, we find that the accuracy reaches a limiting value

as $M \sim N/20$, in agreement with numerical simulation for the same machine hyperparameters. The same behavior is found for training errors, which indicates that input data are already completely separated by the optical setup.

It is worth noting that the device does not need training at each use. Once we learn a given data set, we store the corresponding output weights and program a different problem; then, we are able to accomplish any of the posed tasks on-demand. The diverse classification and regression instances in Fig. 4 can be tested without any sequential order, simply specifying the type of task they belong to.

4. DISCUSSION

A. Improving the Photonic Hardware

The limit provided by training errors in Figs. 4(e) and 4(f) allows us to identify the various factors toward improving the computational resolution of the optical machine. We ascribe residual errors to practical nonidealities of the device. The finite precision of both encoding and readout, including noise in optical intensities, introduces errors that are absent in the digital implementation [48]. Considering a camera precision of 8 bit in the numerical model, we find a relevant decrease of the maximum accuracy. This indicates that improved performance is attainable by fine-tuning the optical components. Flickering effects of the phase modulator, which give a tiny but variable inaccuracy each time the machine is interrogated, are identified as another relevant source of error. Moreover, optical and mechanical fluctuations of the tabletop optical line during training can result in further inconsistencies. These fluctuations can be compensated in future experiments using incremental learning [36], a refined technique that allows us to adaptively adjust the weights while training is ongoing. Training by use of sparse regression methods may be also useful when dealing with larger data sets.

B. Comparison with Digital Kernel Methods

The learning principle of the PELM lies on the basis of various kernel methods [49]. In kernel classification, mapping of the input data to the high-dimensional feature space is not explicit, and correlations between samples are evaluated through a kernel function over the input space. Such a kernel $\mathbf{K}$ contains the inner product of all pairs of data points in the feature space, i.e., all the information sufficient to perform data separation. When trained via ridge regression, the output $\mathbf{Y}$ of a kernel classifier can be generally expressed as

$$\mathbf{Y} = \mathbf{K}(\mathbf{K} + c\mathbf{I})^{-1}\mathbf{T}, \qquad (3)$$

where $\mathbf{K} = k(\mathbf{X}_i, \mathbf{X}_j)$ is an $N \times N$ kernel matrix, with k being a given kernel function. Comparing Eqs. (3) and (1) indicates that, in our PELM scheme, the free-space optical setup acts as an effective photonic kernel [41]. To provide a comparison with standard digital kernels, we have evaluated their performance on the MNIST data set. We exploit the modularity of the approach, i.e., we use the same training algorithm [Eq. (3)] with different kernel functions. We focus on the relevant case of the element-wise Gaussian kernel $k(\mathbf{X}_i, \mathbf{X}_j) = \exp(-\gamma\|\mathbf{X}_i - \mathbf{X}_j\|^2)$, with tunable parameter γ. Using this kernel, we obtain an average classification accuracy of 96%, to be compared with the 85% given by a linear kernel ($\mathbf{K} = \mathbf{X}\mathbf{X}^T$).

Smaller classification errors can be obtained with more elaborate kernel functions [41]. Although the evaluation of the kernel matrix elements can be time-efficient, training requires storing and inverting this large matrix ($N \times N$), which becomes unfeasible on large-scale data sets. This enormous memory consumption represents a major drawback, which makes digital kernels energetically inefficient with respect to the photonic implementation. In fact, in the PELM scheme, we make use of explicit feature mapping, and the output matrix used for training has a much smaller size ($M \times M$). This implies that the computational cost for training the device does not depend on the data set size N but only on the selected number of channels M. The advantage of the proposed approach extends also to kernel methods based on different algorithms. For instance, a support-vector machine on MNIST takes about 10^7 multiply-adds (MACs) per recognition [50], whereas in our PELM the optical processors enables adequate classification of the single handwritten digit using only 10^4 digital MAC operations.

C. Application Potentials

Besides computing effectiveness comparable with its digital counterpart, the PELM hardware offers several advantages that are promising for fast processing of big data, especially for inputs that naturally present themselves as optical signals. In fact, unlike deep optical neural networks [51,52], training is simple and scales favorably with the data set size. It can also be performed online with affordable computational resources. Once trained, forward information propagates optically in a fully passive way and in free space. This can provide a key advantage with respect to ELM algorithms and kernel methods both in terms of scalability and latency. In fact, the matrix multiplication in Eq. (2), which maps input data to the feature space, on a digital computer requires a time and memory consumption that grows quadratically with the data dimension L. The optical hardware performs this operation totally in parallel, independently of the input data size, and without power dissipation. Since only M scalar multiplications are necessary from the optical detection to the final classification, the PELM has a low latency, which is independent on the data-set dimension. This is a crucial property in applications that require fast responses, such as in real-time computer vision. The main speed limitation of our free-space PELM is related to the frame rate of the input optical modulator. Liquid-crystal SLMs currently have typical frame rates of the order of 100 Hz, but phase modulators based on micro-electro-mechanical, electro-optical, or acousto-optical technologies are approaching GHz frequencies [53,54]. With similar components available, PELM hardware could perform $R = L \times M \times 10^9$ operations per second (OPS), a value that, for large L, can overcome the current limits of electronic computing (peta OPS).

The absence of any optical element or medium along the network path, which differentiates our scheme from all other optical neuromorphic devices previously realized, is a valuable aspect for various reasons. The tiny optical absorption and scattering of light in air imply that our scheme needs low-power optical sources (µW or lower), i.e., it requires minimal energy consumption. More importantly, our device is extremely stable to mechanical and optical perturbations, because its operation does not depend on components that must be kept completely

static after training. This suggests that the proposed optical processor is promising for use in dynamic environments. In fact, tolerance to external disturbances requires only the relative alignment between the SLM plane and the camera plane, which can be carried out with mechanical or optical stabilization. We thus expect the free-space PELM can operate even when being subject to shocks. For small perturbations, tolerance is already guaranteed by the finite spatial extent of the input and output modes. Specifically, the output channels are made by averaging over several camera pixels (see Appendix B), which reduces the effect of vibrations. Therefore, we expect that re-training the device would be necessary only when the perturbation is large enough that a change of the channel positions cannot be retrieved via some feedback mechanism.

In view of edge-computing applications, the PELM can be further adapted toward ease of use and energy efficiency. In fact, using optical signals from the environment as input data, the encoding operation performed by the SLM is reduced to the sole embedding. This operation can be integrated directly into optical propagation, removing the need for a phase modulator. For example, we can divide the coherent input field and recombine it after a digital micromirror device (DMD) has performed the embedding. The intensity extracted from a selected plane along the optical path gives the output matrix, which can be enriched with phase and spectral information. Similar schemes promise a miniaturization of our free-space device and its operation as a fast and stable optical processor.

5. CONCLUSION

Nowadays, artificial intelligence is permeating the field of photonics [55,56] and viceversa [57]. Optical platforms are enabling computing functionalities with unique features, ranging from photonic Ising machines for complex optimization problems [58–65] to optical devices for cryptography [66]. We have realized a novel photonic neuromorphic computing device that is able to perform classification and feature extraction only by exploiting optical propagation in free space. Our results demonstrate that fabricated optical networks, or complex physical processes and materials, are not mandatory ingredients for performing effective machine learning on a optical setup. All the essential factors for learning can be included in optical propagation via encoding and decoding methods. On this principle, we demonstrated a photonic extreme-learning machine that, given its unique adaptability and intrinsic stability, is attractive for future processing of optical data in real-time and dynamic conditions. More generally, our approach envisions the exceptional possibility of harnessing any wave system as an intelligent device that learns from data, in photonics, and far beyond.

APPENDIX A: NETWORK DETAILS

1. ELM Framework

The basic ELM structure is a single-layer feedforward neural network in which hidden nodes are not tuned [35]. For real-valued random internal weights, the scheme matches single-layer random neural networks [37,38]. Considering an $N \times L$ input data set $\mathbf{X}$ and one output mode, the output function is

$$\mathbf{Y} = \mathbf{H}\boldsymbol{\beta} = \sum_{j=1}^{M} H_j(\mathbf{X})\beta_j, \qquad \textbf{(A1)}$$

where $\boldsymbol{\beta}$ is the weight vector determined by training, and $\mathbf{H}(\mathbf{X}) = \mathbf{H}$ is the $N \times M$ hidden-layer matrix outcome. $\mathbf{H}(\mathbf{X}) = [\mathbf{g}(\mathbf{X}_1), ..., \mathbf{g}(\mathbf{X}_N)]$, where $\mathbf{g}(\mathbf{x})$ maps the input sample $\mathbf{x}$ from the L-dimensional input space to the M-dimensional feature space. Under appropriate conditions, the machine is able to interpolate any continuous function (universal interpolator) and acts as a universal classifier [36]. Given the target labels $\mathbf{T}$, training corresponds to solving the ridge regression problem: $\mathrm{argmin}_{\boldsymbol{\beta}}(\|\mathbf{H}\boldsymbol{\beta} - \mathbf{T}\|^2 + c^{-1}\|\boldsymbol{\beta}\|^2)$, where c is a parameter controlling the trade-off between the training error and the regularization. This constrained optimization can be recast as a dual optimization problem [36]. A solution that is computationally affordable for large data sets is given by Eq. (1): $\boldsymbol{\beta} = (\mathbf{H}^{\mathrm{T}}\mathbf{H} + c\mathbf{I})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{T}$. In this case, matrix inversion involves the $M \times M$ matrix $\mathbf{H}^{\mathrm{T}}\mathbf{H}$, which makes the method scalable and effective for large-scale applications. The output function of the classifier is thus

$$\mathbf{Y} = \mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{H} + c\mathbf{I})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{T}. \qquad \textbf{(A2)}$$

In the case of a single-output node performing binary classification [as for the problem in Fig. 4(c)], the decision function is $f(\mathbf{X}) = \mathrm{sign}\{\mathbf{H}(\mathbf{X})[\mathbf{H}^{\mathrm{T}}(\mathbf{X})\mathbf{H}(\mathbf{X}) + c\mathbf{I}]^{-1}\mathbf{H}^{\mathrm{T}}(\mathbf{X})\mathbf{T}\}$. It generalizes for a multiclass classifier with multiple output nodes as $f(\mathbf{X}) = \mathrm{arg\,max}_k \mathbf{Y}_k(\mathbf{X})$, where $\mathbf{Y}_k(\mathbf{X})$ denotes the output $\mathbf{Y}$ of the kth node.

2. Encoding Methods

We consider phase encoding of the input data in all the presented results (phase-only light modulation). The embedding matrix $\mathbf{W}$ is a fixed signal, independent of the specific input data, which characterizes the phase encoder and modifies the PELM feature space. In the noise embedding method, $\mathbf{W}$ is chosen as a uniformly distributed random matrix with maximum amplitude ρ (noise level) made by blocks of size l (noise correlation length). In the Fourier embedding method, we construct the embedding matrix $\mathbf{W} = [W_1, ..., W_n]$ from n frequencies, $W_k = \sum_{\omega=1}^{n}(a_\omega/n)\exp(i\omega k/n)$, with coefficients a_ω of equal amplitude and arbitrary phase. Both the embedding signal and input data are encoded within the phase interval $[0, \pi]$. In experiments, the preselected embedding matrix is discretized in gray levels and superimposed to each input data.

APPENDIX B: EXPERIMENTAL SETUP AND DEVICE TRAINING

A continuous-wave laser beam with wavelength $\lambda = 532$ nm is expanded and polarized, and illuminates a reflective phase-only SLM (Hamamatsu X13138, 1280 × 1024 pixels, 12.5 μm pixel pitch, 60 Hz maximum frame rate), which encodes input data within an embedding phase mask by pure phase modulation. By grouping several SLM pixels, the modulator active area is divided into L input nodes, with each node having 210 phase levels equally distributed in the 0–2π interval. Phase-modulated light propagates in free space through a focusing plano-convex lens ($f = 150$ mm). The optical field in the

lens focal plane is collected by an imaging objective (NA = 0.4) and detected by a CCD camera with 8 bit (256 gray levels) intensity sensitivity. To obtain a nonlinear saturated function of the transmitted field, the camera exposure is manually increased to obtain overexposed images. We note that similar performance is found using the camera in the unsaturated regime, which corresponds to a square nonlinear response. M output channels are preselected within the camera region of interest, where the signal in each channel is obtained by binning over few camera pixels (10×10 for MNIST classification) to reduce detection noise.

Training is performed by loading one by one the N input samples on the SLM and keeping fixed the embedding matrix. At each training step, values from the M output channels are acquired [Fig. 3(c)] and stored on a conventional computer controlling the setup. The readout weights are obtained by applying Eq. (1) on the entire set of measurements. They are readily used in the testing phase, where each of N_t testing samples is sent through the photonic machine and passively classified by weighing the detected output. Different tasks can also be performed on-demand without retraining the device. As expected, practical effects limit the time for which the device maintains its ability to perform well on a given task without retraining. We found that good performance is maintained for more than 1 h. On longer times, laser fluctuations and local variations of the SLM response due to thermal effects and liquid crystal relaxation become relevant.

APPENDIX C: DATA SETS FOR CLASSIFICATION AND REGRESSION

Recognition of handwritten digits is tested on the MNIST data set, a standard benchmark for multiclass classification. The data set, which includes 10 classes, consists of 60,000 training samples (N) and $N_t = 10{,}000$ digits for testing, each with size $L = 28 \times 28$. Although state-of-the-art convolutional neural networks reach accuracies exceeding 99.8% on MNIST (https://github.com/Matuzas77/MNIST-0.17), the task remains the basic test for any novel machine learning device. In fact, superior algorithms are application-specific and require massive data processing.

The mushroom (https://archive.ics.uci.edu/ml/datasets/Mushroom) is a binary class data set with relatively large size and low dimension. It includes 8124 samples with $L = 22$ features in random order. The goal is to separate edible from poisonous mushrooms. A typical ELM accuracy is 88.9 % for a split ratio $N/N_t \approx 0.23$ [36].

The abalone data set (https://archive.ics.uci.edu/ml/datasets/Abalone) is one of the mostly used benchmarks for machine learning and concerns the identification of sea snails in terms of age and physical parameters. Each training point has $L = 8$ attributes, and the entire data set has 8177 samples. Digital ELMs report testing errors around 0.07 for $N/N_t \approx 2$. Errors are evaluated using the root mean square displacement

$$\mathrm{RMSD} = \sqrt{\textstyle\sum_i^{N_t} (Y_i - T_i)^2 / N_t}.$$

Disclosures. The authors declare no conflicts of interest.

Data Availability. The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

1. S. Haykin, *Neural Networks and Learning Machines* (Pearson Prentice Hall, 2008).
2. E. Strubell, A. Ganesh, and A. McCallum, "Energy and policy considerations for deep learning in NLP," arXiv:1906.02243 (2019).
3. G. Wetzstein, A. Ozcan, S. Gigan, S. Fan, D. Englund, M. Soljačić, C. Denz, D. A. B. Miller, and D. Psaltis, "Inference in artificial intelligence with deep optics and photonics," Nature **588**, 39–47 (2020).
4. N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, "Optical implementation of the Hopfield model," Appl. Opt. **24**, 1469–1475 (1985).
5. D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals," Appl. Opt. **27**, 1752–1759 (1988).
6. C. Denz, *Optical Neural Networks* (Springer, 1998).
7. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljacic, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**, 441–446 (2017).
8. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004–1008 (2018).
9. N. M. Estakhri, B. Edwards, and N. Engheta, "Inverse-designed metastructures that solve equations," Science **363**, 1333–1338 (2019).
10. J. Bueno, S. Maktoobi, L. Froehly, I. Fischer, M. Jacquot, L. Larger, and D. Brunner, "Reinforcement learning in a large-scale photonic recurrent neural network," Optica **5**, 756–760 (2018).
11. Y. Zuo, B. Li, Y. Zhao, Y. Jiang, Y. C. Chen, P. Chen, G. B. Jo, J. Liu, and S. Du, "All-optical neural network with nonlinear activation functions," Optica **6**, 1132–1137 (2019).
12. A. N. Tait, T. F. de Lima, E. Zhou, A. X. Wu, M.-A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Neuromorphic photonic networks using silicon photonic weight banks," Sci. Rep. **7**, 7430 (2017).
13. J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," Nature **569**, 208–214 (2019).
14. J. Feldmann, N. Youngblood, M. Karpov, H. Gehring, X. Li, M. Stappers, M. Le Gallo, X. Fu, A. Lukashchuk, A. S. Raja, J. Liu, C. D. Wright, A. Sebastian, T. J. Kippenberg, W. H. P. Pernice, and H. Bhaskaran, "Parallel convolutional processing using an integrated photonic tensor core," Nature **589**, 52–58 (2021).
15. X. Xu, M. Tan, B. Corcoran, J. Wu, A. Boes, T. G. Nguyen, S. T. Chu, B. E. Little, D. G. Hicks, R. Morandotti, A. Mitchell, and D. J. Moss, "11 TOPS photonic convolutional accelerator for optical neural networks," Nature **589**, 44–51 (2021).
16. X. Xu, M. Tan, B. Corcoran, J. Wu, T. G. Nguyen, A. Boes, S. T. Chu, B. E. Little, R. Morandotti, A. Mitchell, D. G. Hicks, and D. J. Moss, "Photonic perceptron based on a Kerr microcomb for high-speed, scalable, optical neural networks," Laser Photon. Rev. **14**, 2000070 (2020).
17. J. Spall, X. Guo, T. D. Barrett, and A. I. Lvovsky, "Fully reconfigurable coherent optical vector-matrix multiplication," Opt. Lett. **45**, 5752–5755 (2020).
18. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," Sci. Rep. **8**, 12324 (2018).
19. M. Miscuglio, Z. Hu, S. Li, J. George, R. Capanna, P. M. Bardet, P. Gupta, and V. J. Sorger, "Massively parallel amplitude-only Fourier neural network," Optica **7**, 1812–1819 (2020).

20. C. Wu, H. Yu, S. Lee, R. Peng, I. Takeuchi, and M. Li, "Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network," Nat. Commun. **12**, 96 (2021).

21. T. W. Hughes, I. A. Williamson, M. Minkov, and S. Fan, "Wave physics as an analog recurrent neural network," Sci. Adv. **5**, eaay6946 (2019).

22. R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," Phys. Rev. X **9**, 021032 (2019).

23. T. W. Hughes, M. Minkov, Y. Shi, and S. Fan, "Training of photonic neural networks through *in situ* backpropagation and gradient measurement," Optica **5**, 864–871 (2018).

24. M. Lukosevicius and H. Jaeger, "Reservoir computing approaches to recurrent neural network training," Comput. Sci. Rev. **3**, 127–149 (2009).

25. C. Gallicchio, A. Micheli, and L. Pedrelli, "Deep reservoir computing: a critical experimental analysis," Neurocomputing **268**, 87–99 (2017).

26. G. Neofotistos, M. Mattheakis, G. D. Barmparis, J. Hizanidis, G. P. Tsironis, and E. Kaxiras, "Machine learning with observers predicts complex spatiotemporal behavior," Front. Phys. **7**, 24 (2019).

27. D. Brunner, M. C. Soriano, C. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," Nat. Commun. **4**, 1364 (2013).

28. Q. Vinckier, F. Duport, A. Smerieri, K. Vandoorne, P. Bienstman, M. Haelterman, and S. Massar, "High-performance photonic reservoir computer based on a coherently driven passive cavity," Optica **2**, 438–446 (2015).

29. L. Larger, A. Baylón-Fuentes, R. Martinenghi, V. S. Udaltsov, Y. K. Chembo, and M. Jacquot, "High-speed photonic reservoir computing using a time-delay based architecture: million words per second classification," Phys. Rev. X **7**, 011015 (2017).

30. P. Antonik, N. Marsal, D. Brunner, and D. Rontani, "Human action recognition with a large-scale brain-inspired photonic computer," Nat. Mach. Intell. **1**, 530–537 (2019).

31. A. Röhm, L. Jaurigue, and K. Lüdge, "Reservoir computing using laser networks," IEEE J. Sel. Top. Quantum Electron. **26**, 7700108 (2020).

32. U. Paudel, M. Luengo-Kovac, J. Pilawa, T. J. Shaw, and G. C. Valley, "Classification of time-domain waveforms using a speckle-based optical reservoir computer," Opt. Express **28**, 1225–1237 (2020).

33. M. Rafayelyan, J. Dong, Y. Tan, F. Krzakala, and S. Gigan, "Large-scale optical reservoir computing for spatiotemporal chaotic systems prediction," Phys. Rev. X **10**, 041037 (2020).

34. D. Pierangeli, V. Palmieri, G. Marcucci, C. Moriconi, G. Perini, M. De Spirito, M. Papi, and C. Conti, "Living optical random neural network with three dimensional tumor spheroids for cancer morphodynamics," Commun. Phys. **3**, 160 (2020).

35. G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: theory and applications," Neurocomputing **70**, 489–501 (2006).

36. G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," IEEE Trans. Syst., Man, Cybern. B, Cybern. **42**, 513–529 (2012).

37. W. F. Schmidt, M. A. Kraaijveld, and R. P. Duin, "Feed forward neural networks with random weights," in *International Conference on Pattern Recognition* (IEEE Computer Society, 1992).

38. Y. H. Pao, G. H. Park, and D. J. Sobajic, "Learning and generalization characteristics of the random vector functional-link net," Neurocomputing **6**, 163–180 (1994).

39. J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," Neural Process. Lett. **9**, 293–300 (1999).

40. S. An, W. Liu, and S. Venkatesh, "Face recognition using kernel ridge regression," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2007), pp. 1–7.

41. A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Dremeau, S. Gigan, and F. Krzakala, "Random projections through multiple optical scattering: approximating kernels at the speed of light," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2016), pp. 6215–6219.

42. S. Sunada, K. Kanno, and A. Uchida, "Using multidimensional speckle dynamics for high-speed, large-scale, parallel photonic computing," Opt. Express **28**, 30349–30361 (2020).

43. G. Marcucci, D. Pierangeli, and C. Conti, "Theory of neuromorphic computing by waves: machine learning by rogue waves, dispersive shocks, and solitons," Phys. Rev. Lett. **125**, 093901 (2020).

44. U. Tegin, M. Yildirim, I. Oguz, C. Moser, and D. Psaltis, "Scalable optical learning operator," arXiv:2012.12404 (2020).

45. 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**, 457 (2021).

46. J. W. Goodman, *Introduction to Fourier Optics* (Roberts & Company, 2005).

47. L. C. C. Kasun, H. Zhou, G. B. Huang, and C. M. Vong, "Representational learning with extreme learning machine for big data," IEEE Intell. Syst. **28**, 31–34 (2013).

48. D. Pierangeli, M. Rafayelyan, C. Conti, and S. Gigan, "Scalable spin-glass optical simulator," Phys. Rev. Appl. **15**, 034087 (2021).

49. J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis* (Cambridge University, 2004).

50. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE **86**, 2278–2324 (1998).

51. T. Yan, J. Wu, T. Zhou, H. Xie, F. Xu, J. Fan, L. Fang, X. Lin, and Q. Dai, "Fourier-space diffractive deep neural network," Phys. Rev. Lett. **123**, 023901 (2019).

52. Y. Luo, D. Mengu, N. T. Yardimci, Y. Rivenson, M. Veli, M. Jarrahi, and A. Ozcan, "Design of task-specific optical systems using broadband diffractive neural networks," Light Sci. Appl. **8**, 112 (2019).

53. O. Tzang, E. Niv, S. Singh, S. Labouesse, G. Myatt, and R. Piestun, "Wavefront shaping in complex media with a 350 kHz modulator via a 1D-to-2D transform," Nat. Photonics **13**, 788–793 (2019).

54. B. Braverman, A. Skerjanc, N. Sullivan, and R. W. Boyd, "Fast generation and detection of spatial modes of light using an acousto-optic modulator," Opt. Express **28**, 29112–29121 (2020).

55. D. Piccinotti, K. F. MacDonald, S. Gregory, I. Youngs, and N. I. Zheludev, "Artificial intelligence for photonics and photonic materials," Rep. Prog. Phys. **84**, 012401 (2021).

56. G. Genty, L. Salmela, J. M. Dudley, D. Brunner, A. Kokhanovskiy, S. Kobtsev, and S. K. Turitsyn, "Machine learning and applications in ultrafast photonics," Nat. Photonics **15**, 91–101 (2020).

57. S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Data-driven discovery of partial differential equations," Sci. Adv. **3**, e1602614 (2017).

58. P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, "A fully-programmable 100-spin coherent Ising machine with all-to-all connections," Science **354**, 614–617 (2016).

59. T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K.-I. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, "A coherent Ising machine for 2000-node optimization problems," Science **354**, 603–606 (2016).

60. F. Böhm, G. Verschaffelt, and G. Van der Sande, "A poor—coherent Ising machine based on opto-electronic feedback systems for solving optimization problems," Nat. Commun. **10**, 3538 (2019).

61. D. Pierangeli, G. Marcucci, and C. Conti, "Large-scale photonic Ising machine by spatial light modulation," Phys. Rev. Lett. **122**, 213902 (2019).

62. D. Pierangeli, G. Marcucci, and C. Conti, "Adiabatic evolution on a spatial-photonic Ising machine," Optica **7**, 1535–1543 (2020).

63. M. Prabhu, C. Roques-Carmes, Y. Shen, N. Harris, L. Jing, J. Carolan, R. Hamerly, T. Baehr-Jones, M. Hochberg, V. Čeperić, J. D. Joannopoulos, D. R. Englund, and M. Soljačić, "Accelerating recurrent Ising machines in photonic integrated circuits," Optica **7**, 551–558 (2020).

64. K. P. Kalinin, A. Amo, J. Bloch, and N. G. Berloff, "Polaritonic XY-Ising machine," Nanophotonics **9**, 4127–4138 (2020).

65. Y. Okawachi, M. Yu, J. K. Jang, X. Ji, Y. Zhao, B. Y. Kim, M. Lipson, and A. L. Gaeta, "Demonstration of chip-based coupled degenerate optical parametric oscillators for realizing a nanophotonic spin-glass," Nat. Commun. **11**, 4119 (2020).

66. A. Di Falco, V. Mazzone, A. Cruz, and A. Fratalocchi, "Perfect secrecy cryptography via mixing of chaotic waves in irreversible time-varying silicon chips," Nat. Commun. **10**, 5827 (2019).