Editors' Suggestion

Neuromorphic Computing in Ginzburg-Landau Polariton-Lattice Systems

Andrzej Opala,[1,*] Sanjib Ghosh,[2] Timothy C.H. Liew,[2] and Michał Matuszewski[1]

[1] *Institute of Physics, Polish Academy of Sciences, Al. Lotników 32/46, PL-02-668 Warsaw, Poland*

[2] *Division of Physics and Applied Physics, Nanyang Technological University, Singapore 637371, Singapore*

(Received 15 August 2018; revised manuscript received 24 March 2019; published 13 June 2019)

The availability of large amounts of data and the necessity of processing it efficiently have led to the rapid development of machine-learning techniques. To name a few examples, artificial-neural-network architectures are commonly used for financial forecasting, speech and image recognition, robotics, medicine, and even research. Direct hardware for neural networks is highly sought for overcoming the von Neumann bottleneck of software implementations. Reservoir computing (RC) is a recent and increasingly popular bioinspired computing scheme that holds promise for efficient temporal information processing. We demonstrate the applicability and performance of RC in a general complex Ginzburg-Landau lattice model, which adequately describes the dynamics of a wide class of systems, including coherent photonic devices. In particular, we propose that the concept can be readily applied in exciton-polariton lattices, which are characterized by unprecedented photonic nonlinearity, opening the way to signal processing at rates of the order of 1 Tbit s^{-1}.

DOI: 10.1103/PhysRevApplied.11.064029

I. INTRODUCTION

In contrast to single- or multilayer (deep) feedforward networks, recurrent neural networks (RNNs) may exhibit complex internal-state dynamics. This makes them particularly efficient in the analysis of time-dependent signals that require memory, such as speech or text recognition and interpretation [1]. On the other hand, the training of RNNs is difficult and not always convergent. The idea of reservoir computing (RC), also known as echo-state networks or liquid-state machines, may be viewed as a generalization of the RNN concept, inspired by the internal structure of the brain [2–5].

The core of the system is formed by a recurrent network of nodes (neurons) that are connected to each other (see Fig. 1). This network, called the reservoir, is static, as its connections are unchanged during training. This is important for both the feasibility of a physical implementation and the convergence of the training procedure. Typically, the input signal represented by $u_i(t)$ is multiplied by random weights and injected into reservoir nodes. The subsequent dynamics of the amplitudes in the nodes (called neuron activations) are delivered to output neurons y_i that are used, for example, for classification or prediction of time-dependent signals. In the classification task, the output neuron with the highest activation is identified with the predicted class. A supervised training procedure consists of presenting many different inputs to the system and adjusting the output weights to minimize the average error. The output signals are linear superpositions of the amplitudes of the reservoir, which results in a training procedure that is both efficient and convergent [6].

The simple scheme described above has proved to be surprisingly efficient in various machine-learning tasks. The performance of RC in time-series prediction and speech recognition is particularly well described [2,7–12]. Notably, software implementation of RC has won a financial time-series prediction competition [13]. To date, RC has been successfully realized not only as a software but also as hardware implementations in systems including semiconductor chips [14], memristor arrays [15,16], optoelectronic [7,17,18] and all-optical systems [19], mechanical oscillators [20], and even in a bucket of water [21]. For a comprehensive review, see Ref. [22]. Photonic systems are particularly interesting as they hold promise for huge signal-processing rates [7,8,11,23,24].

In many machine-learning approaches, input data are transformed in a multidimensional space according to a nonlinear map. This nonlinear mapping in RC is typically provided by the complex dynamics of the reservoir, although nonlinear read-out has also been implemented [11]. To be useful for computing, the network has to posses several characteristics. First, the size of the reservoir must be sufficiently large so that it is able to perform the desired computation within its fixed internal structure. The so-called echo-state property [5] is related to the stability of the system. The state of the reservoir must be determined solely by the history of the signal u_i, but for sufficiently

*opala@ifpan.edu.pl

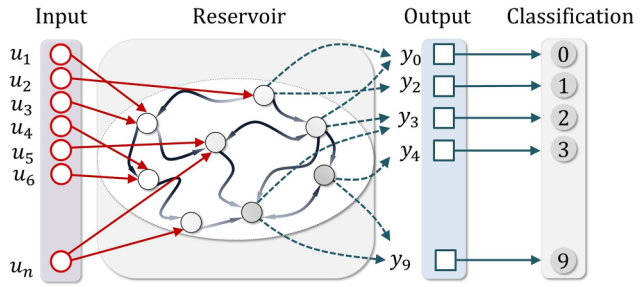

FIG. 1. The concept of RC. A single input layer $u_i(t)$ is used to excite signals in the nonlinear reservoir consisting of hidden nodes. Signals propagate between the nodes, which are connected to each other with random weights. This performs a nonlinear transformation of input in a high-dimensional space. The evolution of node amplitudes is collected in the output layer y_i and used for classification or prediction. In contrast to standard recurrent neural networks, the input weights and connections within the reservoir are static. Only the output weights are trained using a convergent regression procedure.

long evolution it should not depend on the initial conditions or signals from the distant past. On the other hand, the dynamics must be sufficiently nonlinear so that the state of the reservoir allows for the separation of signals that differ by a small amount. The optimal working point appears to be placed close to a certain stability threshold [6,11]. Importantly, it has been demonstrated that powerful computation can be achieved in a variety of different designs and systems. For instance, according to the original idea, reservoir nodes are connected to each other with random weights [2]. However, various other designs have proved to be efficient as well [7,11,12].

In this work, we consider the implementation of RC in systems described by the complex Ginzburg-Landau equation (CGLE), which is one of the fundamental models of wave phenomena [25]. It provides a universal description of weakly nonlinear spatiotemporal systems that are invariant under a global gauge change $\psi \to \psi e^{i\phi}$, where typically ψ is a slowly varying envelope of an oscillatory wave packet. Its range of applications spans from the description of hydrodynamic systems and chemical reactions to superconductors and superfluids, to ultracold quantum gases, and to lasers [25–27]. We propose to implement reservoir dynamics in a lattice of weakly coupled traps with nearest-neighbor couplings. According to our estimates, semiconductor exciton-polariton microcavity systems appear to be a promising platform for RC with very high signal-processing data rates, which can be achieved due to the extremely strong optical nonlinearity on a picosecond time scale [28].

II. MODEL

We consider the discrete version of the CGLE, which describes a system enclosed in a simple two-dimensional array of weakly coupled traps:

$$\frac{d\psi_n}{dt} = W^{\text{in}}_{nm} u_m - i \sum_{m=nn} W_{nm}\psi_m + \left(\gamma - \Gamma|\psi_n|^2 - ig|\psi_n|^2\right)\psi_n, \qquad (1)$$

where the first term on the right-hand side corresponds to coherent signal injection, with W^{in}_{nm} being the mask applied to the signal u_m (see Fig. 2). The coupling coefficients between nearest-neighbor lattice sites are denoted by W_{nm}, $\gamma = P - \kappa$ is the gain coefficient, in general equal to the difference between the pumping rate (P) and the linear decay rate (κ), Γ is the nonlinear decay

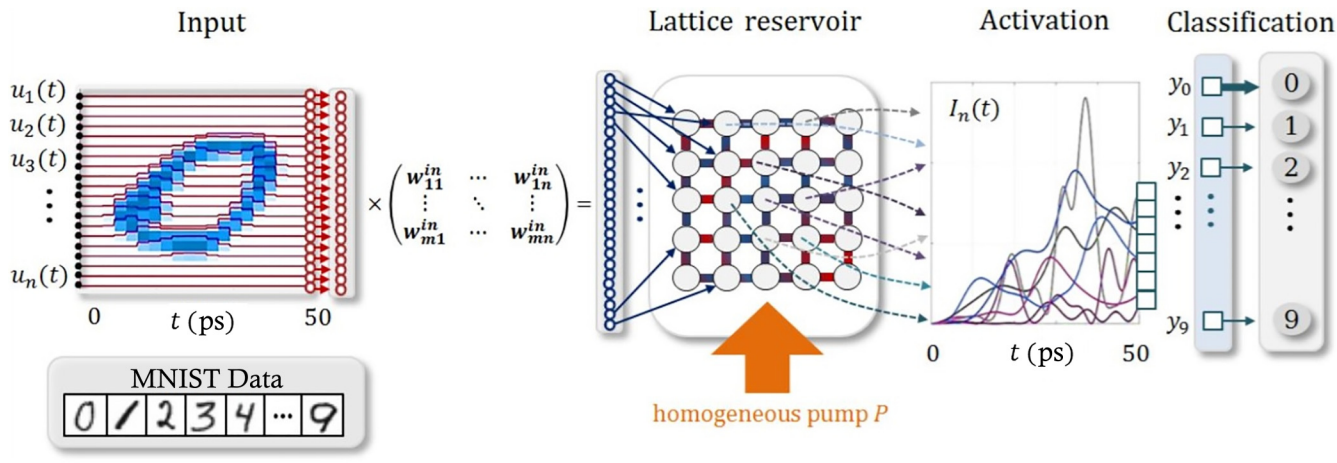

FIG. 2. A scheme for a handwritten digit classification task. The data are convoluted with random weights and imprinted on the lattice by driving each of the lattice sites. At the same time, the system is pumped to maintain a dynamic state close to the stability (or lasing) threshold. The resulting density $I_n(t) = |\psi_n(t)|^2$ in each node (activations) is recorded at the end of the sequence and used for classification of the input.

rate, and g is the conservative nonlinear coefficient. The geometry of the lattice and the scheme of the experiment designed for a classification task are shown in Fig. 2. In our numerical simulations, we consider a simple rectangular $N \times N$ lattice reservoir with random positive symmetric nearest-neighbor couplings, which is a natural choice for photonic systems such as microcavities, photonic crystals, or waveguide arrays. Note that simpler geometries, such as a one-dimensional lattice with identical couplings and even single-site systems, have been demonstrated to perform well in experimental tests [7,9,12,17]. The two-dimensional multisite system has, however, an advantage in terms of increased efficiency.

The parameters of the model (1) used in numerical simulations are chosen to correspond to a lattice of coupled semiconductor exciton-polariton microcavities [29–32]. Exciton-polaritons are composite quantum quasiparticles of semiconductor excitations and photons in the strong-coupling regime [33–35]. They exhibit an interesting combination of properties of matter and light. The extremely low effective mass of polaritons, of the order of 10^{-4} times the electron mass, results from the photonic component and allows for the effective transport across the lattice on a picosecond time scale. On the other hand, the interaction resulting from the exciton component provides unprecedented instantaneous nonlinearity g, orders of magnitude stronger than in other photonic systems [28]. These properties have been used recently to demonstrate remarkable phenomena including nonequilibrium condensation and lasing and superfluidity of polaritons [36–39], as well as realization of ultrafast all-optical switches [40–42]. Recently, a single-layer neural-network design has been proposed [43]. Note that reservoir computing should not be confused with the exciton reservoir, which consists of incoherent particles in the lasing regime. Here, we neglect the influence of the exciton reservoir on the dynamics of the system, which can be achieved by an appropriate pumping scheme [44]. Our results apply to a range of other systems due to the universality of the CGLE and its scaling properties, which allows them to be conveyed to other systems with different values of physical parameters in the CGLE (1) (details of the rescaling are given in the Appendix A).

III. RESULTS

We present results for handwritten digit recognition using the modified National Institute of Standards and Technology (MNIST) data set, which is one of the standard tests of pattern recognition in machine learning. Additional simulations for the Mackey-Glass nonlinear system prediction task and speech recognition are presented in Appendices D and E. The goal of the MNIST recognition task is to classify the digits of various writers using the recorded gray-scale images. The data set contains 70,000 digits and in our simulations for training (4000 digits) and testing (1000 digits) we use randomly picked subsets of the whole set. Each digit consists of 20×20 gray-scale pixels.

We convert each image of a digit into temporal signals according to the scheme illustrated in Fig. 2, with pixels in each row converted into step-wise signals $u_i(t)$, where $i = 1, \ldots, 20$ is the row number. The temporal length τ corresponding to a single pixel is adjusted to achieve the optimal recognition rate. This parameter must be adjusted so that the time scales of the reservoir and of the input signal are compatible, but the overall performance is not very sensitive to its value. The signal vector $u_i(t)$ is multiplied by a random constant matrix $\mathbf{W}^{\text{in}}$ of dimension $20^2 \times N^2$, which has the purpose of both distributing the information across the lattice and adjusting the length of the input vector to the number of pillars in the lattice N^2. The incoming signal initiates dynamics of complex reservoir amplitudes $\psi_n(t)$, which correspond to neuron activations. The squared modulus of each lattice site is recorded at the end of the sequence. A linear transformation is used to translate the read-out into output neuron activations, $y_j = \sum_n W^{\text{out}}_{jn} |\psi_n(t_E)|^2$, where $j = 0, \ldots, 9$ and $t_E = 20\tau$ is the length of the sequence. In the training phase, logistic regression is used to obtain optimal output weights $\mathbf{W}^{\text{out}}$ (see Appendix C). During testing, the obtained $\mathbf{W}^{\text{out}}$ are used to classify the digit. Ideally, the result is equal to $d_j = 1$ for the correct digit and $y_j = 0$ for all other digits. In practice, all d_i have mixed values and we choose the one that is the highest as the predicted digit.

An example of the resulting error rates is presented in Fig. 3. The average accuracy for this particular 9×9 lattice is 89.2%, higher than that obtained by a linear classifier [45]. Additionally, we note that due to the reduced dimensionality of the output (N^2) with respect to the input

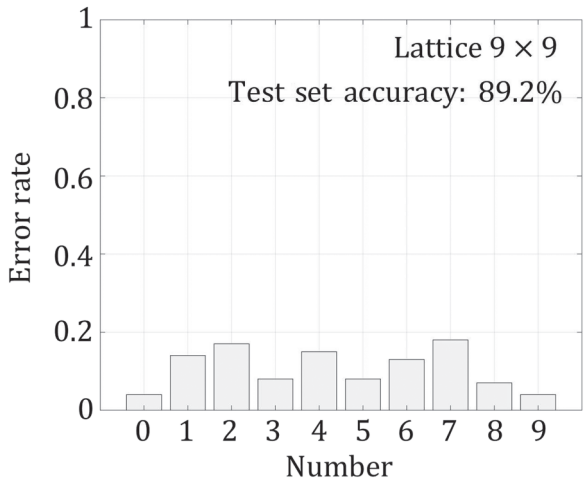

FIG. 3. The classification error rate for the MNIST data set for each handwritten digit, in the case of a 9×9 lattice with 81 nodes.

(20^2), the regression performed during the learning phase requires much less computation time. This is important if the minimization of error, performed offline, turns out to be the most time-consuming part of teaching. At the same time, no offline computations are required during testing and the recognition rate is limited only by the reservoir dynamics time scale.

We use realistic parameters that correspond to experiments performed in gallium arsenide polariton lattices [31,32], with couplings W_{nm} distributed at random between zero and 0.165 meV, $g = 0.25$ μeV, $\Gamma = 0.1$ μeV, and $\tau = 2.5$ ps. Signal processing on a picosecond time scale can be achieved due to the use of a photonic system with a very strong nonlinearity in the regime of quantum coupling of light and matter. According to our simulation, the typical rate at which data can be fed into each lattice site and processed efficiently is one byte (understood as a unit of information) every few tens of picoseconds. This estimation is in agreement with numerous time-resolved experiments in polariton systems performed in the nonlinear regime [46–50]. In contrast to single-site RC systems, here signal processing is performed at N^2 nodes in a truly parallel manner. A lattice of a 100 nodes should enable a realistic data rate, of the order of 1 Tbit s^{-1}, in a micrometer-sized system. This compares favorably even with state-of-the-art optoelectronic [8] and passive-photonic-microcircuit [11] RC implementations, which have recently achieved 10 Gbit s^{-1} data rates.

Figure 4 shows the dependence of the error rate on the effective gain parameter γ. RC systems display optimal performance when, in the absence of an input signal, the system is stable but close to an instability threshold. In the case of the CGLE, at the zero gain point $\gamma = 0$, the trivial solution ($\psi_n = 0$) loses stability and a new stationary state appears. This threshold is interpreted in the exciton-polariton context as the onset of polariton lasing. It is clear from Fig. 4 that an optimal working point is found close to zero gain or at the lasing threshold.

We present the dependence of the error rate on the size of the lattice in Fig. 5. In this figure, the error rates are calculated as a result of averaging over ten different simulations corresponding to different random weights in the reservoir $\mathbf{W}$ and in the input matrix $\mathbf{W}^{\text{in}}$. Since the input data u_i is convoluted with random weights before being used for excitation, adjusting the number of rows in the rectangular matrix $\mathbf{W}^{\text{in}}$ allows for imprinting the data on an arbitrary-sized lattice. While in the case $N > 20$ redundancy in the input is unavoidable, the same signal may be processed in various ways in different parts of the reservoir, which leads to improvement of the overall performance. Indeed, as shown in Fig. 5, the error rate consistently decreases with N and the error rate for a 50×50 lattice is as low as 5.0%, similar to a feedforward neural network with a single hidden layer [45]. The recognition rate in the case of a 9×9 lattice is equal to 89.2%, similar to that of a memristor-array reservoir computer of comparable size [15], but the read-out vector size is much smaller in our case (100 variables versus a 176×10 network).

As the error rate decreases with the system size, one can expect that the larger Hilbert space of quantum systems might offer increased performance [10] compared to their classical counterparts. Indeed, in Appendix F, we find a reduced error rate if one has access to additional nonclassical observables. However, given the additional complexity of measuring these quantities, the quantum advantage would not necessarily be more practical.

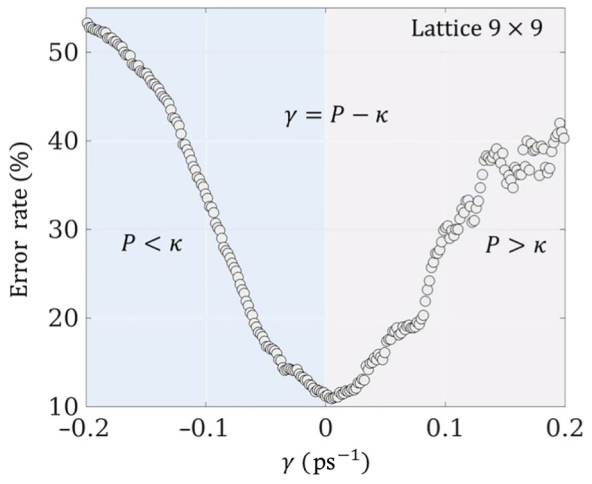

FIG. 4. The dependence of the error rate on the pumping bias parameter γ. The optimal working conditions appear to be close to the stability threshold, $\gamma \approx 0$.

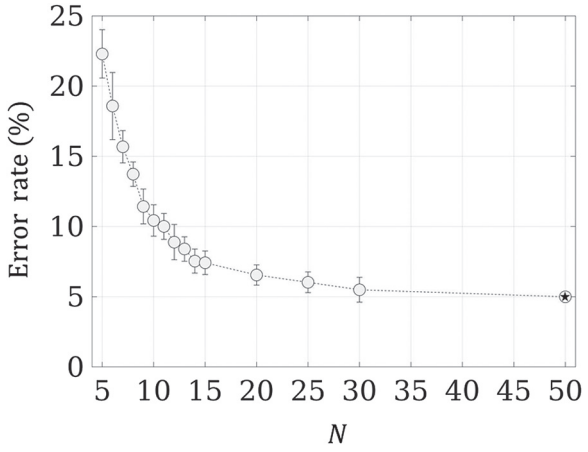

FIG. 5. The dependence of the error rate on the linear size of the lattice N, with the total number of lattice sites equal to N^2. The data points are obtained by averaging over ten random realizations of the reservoir-weights matrix $\mathbf{W}$ and the input matrix $\mathbf{W}^{\text{in}}$. The error bars indicate the standard deviation of the error rate. Note that the point marked with a star corresponds to a single realization.

IV. DISCUSSION

In this work, we demonstrate the applicability of the RC framework in the wide class of systems described by the CGLE. We would like to point out that the majority of the previously considered systems, including optoelectronic ones, were based on amplitude-dependent nonlinearity. On the other hand, the investigated CGLE model describes evolution of a complex wave function, where the couplings between nodes are represented by imaginary values, as in the second term in Eq. (1), and the nonlinearity relies on phase modulation (the g interaction constant), which is very different from the amplitude nonlinearity. Our results extend the applicability of RC to an important class of weakly nonlinear wave systems with gauge invariance, which include coherent photonic systems. Importantly, our scheme is robust to both disorder and dissipation, which are usually hindrances in information-processing schemes. Here, dissipation is useful for ensuring the echo-state property and a disordered network of connections is part of the design. We also demonstrate the robustness of the model against spatiotemporal noise (see Appendix B).

To illustrate the theoretical performance of the proposed system, we estimate the data-processing rate and compare it with hardware implementations of reservoir neural networks realized experimentally. As the MNIST digit-recognition task has not been implemented in many works on RC, we use the TI 46 speech-recognition task as a benchmark. The details of the implementation in the case of an exciton-polariton network are given in Appendix E. Our numerical simulations demonstrate an estimated processing rate of 1.6×10^{10} words/s, compared to the 2500 words/s reported in a complementary metal-oxide semiconductor (CMOS) field-programmable gate array (FPGA) implementation of liquid-state machines [51] and the record high processing rate of 7.7×10^5 words/s achieved in an optoelectronic-delay-line system [7]. The very high estimated processing rate of exciton-polariton systems results from the strong optical nonlinearity on a picosecond time scale and the parallel processing in each node of the lattice.

In terms of scalability of the system size, we note that polariton lattices with several thousands of pillar nodes have already been fabricated and investigated experimentally [52]. One can also estimate the physical size scalability with respect to other (non-RC) neuromorphic implementations. For example, the IBM TrueNorth chip [53] contains 1 million neurons on an approximate surface of 2 cm^2, which gives an average of 5000 neurons/mm^2. A typical size of a polariton pillar node is 10 μm^2, which gives an estimate of 100,000 nodes/mm^2. While nodes in the RC framework are not equivalent to neurons in other architectures, which usually have some tunability, this estimate indicates that—at least in some applications—polariton RC could be competitive with respect to state-of-the-art neuromorphic systems.

From the point of view of energy efficiency, an important advantage of exciton-polariton systems is that they belong to the class of photonic (neutral-particle) systems, which do not suffer from radiative heating, an important issue that is limiting further development of CMOS and other electronic technologies. Since excitons are neutral particles, they also do not contribute to radiative heating. Energy loss in polariton lattices results mostly from the escape of photons through imperfect microcavity mirrors. However, this does not impose a fundamental limit on the energy efficiency (contrary to radiative heating in electronic systems) and several solutions can be proposed to suppress this loss channel. For example, transverse-photon modes trapped by total internal reflection can be used to reduce polariton decay through the mirrors [28].

ACKNOWLEDGMENTS

A.O. acknowledges support from the National Science Center, Poland via Grant No. 2016/22/E/ST3/00045. M.M. acknowledges support from the National Science Center, Poland via Grant No. 2017/25/Z/ST3/03032, under the QuantERA program. S.G. and T.L. acknowledge support from the Singapore Ministry of Education (via the AcRF Tier 2 Grant No. MOE2017-T2-1-001).

APPENDIX A: SCALING PROPERTIES OF THE COMPLEX GINZBURG-LANDAU EQUATION

The discrete CGLE with noise reads as follows:

$$\frac{d\psi_n}{dt} = W^{\text{in}}_{nm} u_m - i \sum_{m=nn} W_{nm}\psi_m + \left(\gamma - \Gamma|\psi_n|^2 - ig|\psi_n|^2\right)\psi_n + D\xi_n(t). \quad \text{(A1)}$$

The simplicity of Eq. (A1) allows for the rescaling of physical coefficients using two arbitrary scaling parameters, τ and α, according to $t = \tau\tilde{t}$, $\psi_n = \alpha^{1/2}\tilde{\psi}_n$, $u_n = \alpha^{1/2}\tilde{u}_n$, $\gamma = \tilde{\gamma}/\tau$, $W^{\text{in}} = \tilde{W}^{\text{in}}/\tau$, $W_{nm} = \tilde{W}_{nm}/\tau$, $g = \tilde{g}/(\tau\alpha)$, and $\Gamma = \tilde{\Gamma}/(\tau\alpha)$. The dynamics of the system with tildes will be identical to those of the original one except for the difference in the time scale and amplitude of the wave function. It follows that the only relevant parameters that govern the qualitative behavior of the system are the ratio g/Γ and the relative values of coefficients γ, W^{in}_{nm}, W_{nm}, and $\xi_n(t)$. In numerical simulations, the random values in the nearest-neighbor weight matrix $\mathbf{W}_{nm}$ and in $\mathbf{W}^{\text{in}}$ matrix are created by the uniform pseudorandom-number generator. The function $\xi_n(t)$ represents the Langevin noise, with D being the noise strength.

APPENDIX B: THE EFFECT OF NOISE

The additional Langevin noise term $\xi_n(t)$ represents noise in the system that is random in space and time. The effect of noise on the prediction error for the MNIST and speech-recognition tasks is presented in Figs. 6 and 8, respectively.

APPENDIX C: LOGISTIC REGRESSION

We use a logistic-regression algorithm to train the readout function for the digit-recognition task. Values between 0 and 1 are assigned by a linear-regression classifier to the output values for each vector $\mathbf{y}$ containing N^2 elements.

We introduce the hypothesis function $h_\Theta(\mathbf{y})$, given by

$$h_\Theta(\mathbf{y}) = (\boldsymbol{\Theta}^{\mathrm{T}}\mathbf{y}), \tag{C1}$$

where $\boldsymbol{\Theta}$ is the weights vector. For classification of the hypothesis representation, we introduce the function $g(z)$:

$$h_\Theta(\mathbf{y}) = g(\Theta^{\mathrm{T}}\mathbf{y}), \tag{C2}$$

where $g(z)$ is given by the logistic function

$$g(z) = \frac{1}{1+e^{-z}}. \tag{C3}$$

Combining the above equations, the hypothesis function reads as follows:

$$h_\Theta(\mathbf{y}) = \frac{1}{1+e^{-\boldsymbol{\Theta}^{\mathbf{T}}\mathbf{y}}}. \tag{C4}$$

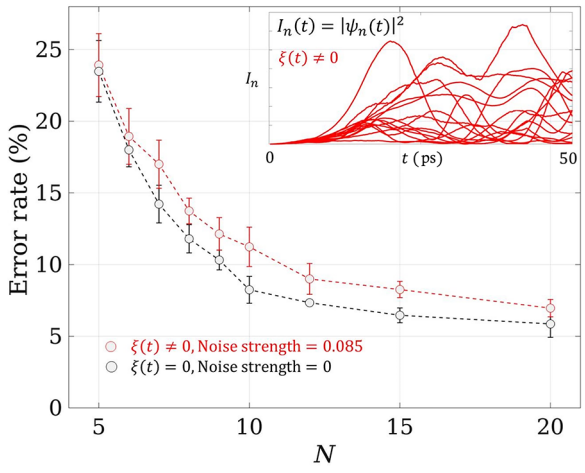

FIG. 6. The dependence of the error rate on the linear size of the lattice N for the MNIST recognition task, in the cases with and without noise. The data are averaged over five random realizations of the reservoir-weights matrix $\mathbf{W}$, the input matrix $\mathbf{W}^{\mathrm{in}}$, and the noise $\xi(t)$. The parameters are the same as in Fig. 5. The inset plot shows the dynamics of the density $I_n(t)$ in $n = 15$ nodes with nonzero noise.

The cost function is described by the following equation:

$$J(\Theta) = \frac{1}{n}\sum_{i=1}^{n}\left\{-\mathbf{x}^{(i)}\log\left[h_\Theta(\mathbf{y}^{(i)})\right] - \left(1-\mathbf{x}^{(i)}\right)\log\left[1-h_\Theta\left(\mathbf{y}^{(i)}\right)\right]\right\}, \tag{C5}$$

where n is the number of samples and $\mathbf{x}^{(i)}$ is the correct output for given input states $\mathbf{y}^{(i)}$. We minimize the cost function by the gradient method:

$$\frac{\partial J(\boldsymbol{\Theta})}{\partial \boldsymbol{\Theta}_j} = \frac{1}{m}\sum_{i=1}^{n}\left[h_\Theta\left(\mathbf{y}^{(i)}\right) - \mathbf{x}^{(i)}\right]\mathbf{y}_j^{(i)}. \tag{C6}$$

Weights are calculated using the MATLAB 2016 software with the function "FMINCG()," written by Carl Edward Rasmussen.

APPENDIX D: THE MACKEY-GLASS PREDICTION TASK

Using the reservoir, we want to predict the solution of the Mackey-Glass equation:

$$\frac{\partial z}{\partial t} = \frac{\alpha\, z(t-\tau_{\mathrm{MG}})}{1+z^{\beta}(t-\tau_{\mathrm{MG}})} - \gamma_{\mathrm{MG}}\, z(t), \tag{D1}$$

which is a nonlinear differential equation with time-delay feedback. We use the parameters $\alpha = 0.2$, $\gamma_{\mathrm{MG}} = 0.1$, $\tau_{\mathrm{MG}} = 17$, and $\beta = 10$. Using a set of training data, we would like to find output weights W^{out} such that

$$y(t) = z(t) - 1 = \sum_n W_n^{\mathrm{out}}|x_n(t)|^2, \tag{D2}$$

where the feedback $u = \sum_n W_n^{\mathrm{out}}|x_n(t-\Delta t)|^2$, in which Δt is a small time step. As a performance measure, we use

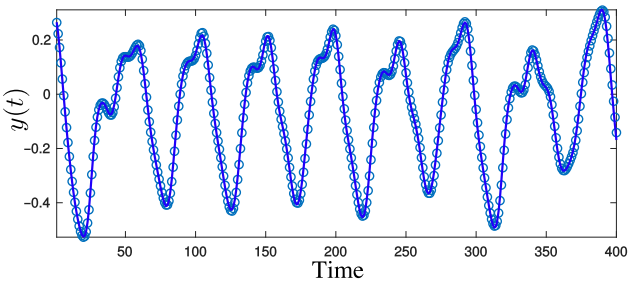

FIG. 7. The Mackey-Glass prediction task performed by a reservoir computer formed with a Ginzburg-Landau lattice. We show $y(t)$ evaluated from Eq. (D1) (points) and predicted by the reservoir computer (solid line) as functions of the time t. The overall prediction error $\sigma = 3\times10^{-4}$ (NRSE value). Here, we consider $N = 16$, $\Gamma = 5$, $\gamma = 10^{-4}$, $g = 2$, and a training data set of 1000 time steps.

the normalized mean-square error (NRSE):

$$\sigma = \frac{\sum_i [y(t_i) - y_p(t_i)]^2}{\sum_i [y(t_i) + y_p(t_i)]^2}, \tag{D3}$$

where $y(t_i)$ and $y_p(t_i)$ are the true solution of the Mackey-Glass equation and the corresponding prediction from the reservoir computer at time t_i, respectively.

In Fig. 7, we show the performance of our reservoir computer (a Ginzburg-Landau lattice with $N = 16$) for the Mackey-Glass prediction task. For a linear lattice size $N = 16$, we find an NRSE of $\sigma = 3 \times 10^{-4}$, which is similar to the values reported in Ref. [54].

APPENDIX E: SPEECH RECOGNITION

Here, we consider the task of isolated spoken-digit recognition, which has been a commonly considered benchmarking task for RC systems [11,17,18]. We use a standard data set, which was collected at Texas Instruments (TI) in 1980 (the NIST TI 46 corpus, which is available from the Linguistic Data Consortium). The data are a set of ten isolated spoken digits (0–9) by eight different female individuals. In the training set, each individual utters a digit ten times, resulting 800 spoken digits in total. After training the network, we evaluate the success rate of our system by using an additional 560 spoken digits (ten digits spoken seven times by the eight individuals).

Each recorded piece of speech is sampled at 12.5 kHz and is then converted into a cochleagram using the Lyon cochlear-ear model [55], previously identified as a good form of preprocessing for speech recognition [56]. These cochleagrams are used as input to the reservoir computer using the same scheme as illustrated in Fig. 2. At a given time t, the input temporal signals $u_n(t)$ represent a column of the cochleagram data with $n = 1, \ldots, 78$ (row number). The full set of cochleagram data are sent through 24 (column number) time steps. The elements of the constant matrix $\mathbf{W}^{\text{in}}$ of size $78 \times N^2$ are chosen at random in the range ± 0.5 (78×24 is the dimension of the cochleagram data and N is the system size). We read out the computed $|\psi_n|^2$ after sending a full cochleagram signal. The final output is then obtained with the linear transform $y_j = \sum_n W^{\text{out}}_{jn} |\psi_n|^2$, where $j = 0, 1, \ldots, 9$. Using the training data set, we obtain the optimal output weights $\mathbf{W}^{\text{out}}$. In the test phase, we use the 560 test data. The output is recognized as the digit j if y_j is the maximum among $j = 0, 1, \ldots, 9$. The error rate in recognizing the spoken digits is presented in Fig. 8. As an example for comparison to other works, we obtain smaller error rates than the reported values for the same task in Ref. [56].

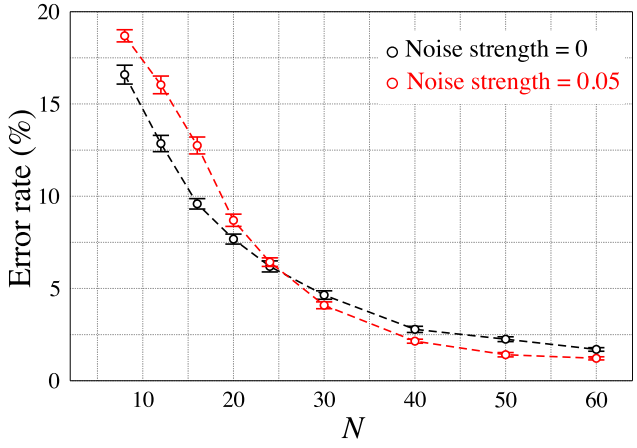

FIG. 8. The linear lattice size N versus the error rate for isolated spoken-digit recognitions with (red) and without (black) noise. We find that a noisy system has a lower error rate than that of a system without noise when N is large. This could be due to the overfitting of the speech data with a large number of degrees of freedom (for large N), which typically diminishes in the presence of random noise. The data are averaged over ten random realizations of $\mathbf{W}$ and $\mathbf{W}^{\text{in}}$. The parameters are $\Gamma = 5$, $\gamma = 10^{-4}$, $g = 2$, and noise strengths $D = 0$ (black) and $D = 0.05$ (red).

APPENDIX F: AN EXAMPLE OF COMPUTING WITH QUANTUM RESERVOIRS

Let us introduce a quantum version of the reservoir network considered in the main paper, described by the master equation for the system density matrix $\boldsymbol{\rho}$:

$$i\hbar \frac{\partial \boldsymbol{\rho}}{\partial t} = i\hbar(\gamma_Q/2) \sum_n \left(2\hat{a}_n \boldsymbol{\rho} \hat{a}_n^\dagger - \hat{a}_n^\dagger \hat{a}_n \boldsymbol{\rho} - \boldsymbol{\rho} \hat{a}_n^\dagger \hat{a}_n\right) + \left[\hat{H}, \boldsymbol{\rho}\right], \tag{F1}$$

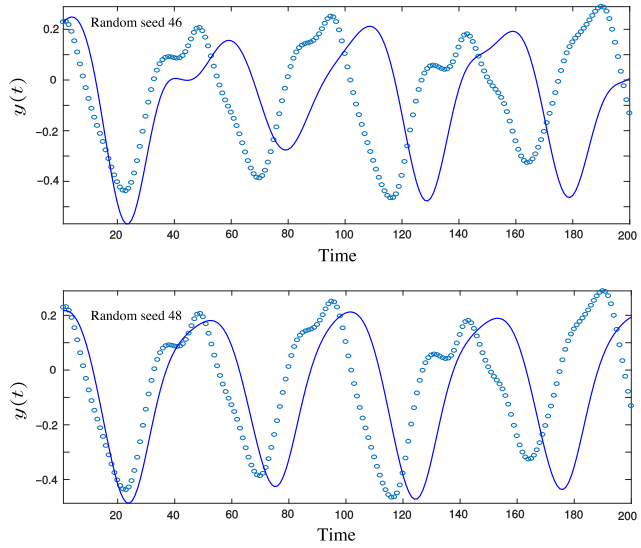

FIG. 9. A typical prediction (solid lines) from a classical reservoir of size 2×3 compared with the Mackey-Glass data (circles). The error $\sigma \sim 0.5$ (NRSE value). We use $\Gamma = 5$, $\gamma = 10^{-4}$, $g = 2$ and a training data set of 200 time steps.

where the Hamiltonian is given as follows:

$$\hat{H} = \sum_{\langle n,m\rangle} J_{nm}\left(\hat{a}_n^\dagger \hat{a}_m + \hat{a}_m^\dagger \hat{a}_n\right) + U\sum_n \hat{a}_n^\dagger a_n^\dagger \hat{a}_n \hat{a}_n + u\sum_n \left(F_n \hat{a}_n^\dagger + F_n^* \hat{a}_n\right) \tag{F2}$$

in which the hopping amplitudes J_{nm} and pump strengths F_n are chosen at random. In a straightforward analogy with the classical version, we may define

$$y^{\text{out}}(t) = \sum_n W_n^{\text{out}} \text{Tr}[\boldsymbol{\rho}(t)\hat{a}_n^\dagger \hat{a}_n] \equiv \sum_n W_n^{\text{out}} \langle \hat{a}_n^\dagger(t)\hat{a}_n(t)\rangle \tag{F3}$$

and the feedback $u = y^{\text{out}}(t - \Delta t)$. However, such a setting has no quantum advantage, as the number of outputs $y^{\text{out}}(t)$ remains the same and $\langle \hat{a}_n^\dagger(t)\hat{a}_n(t)\rangle$ is an effectively classical quantity, which does not itself represent any quantum correlations.

To make use of the larger Hilbert space of the quantum system and access its potentially nonclassical correlations, we consider the quantum-entanglement measures S_{mn} for the continuous variables between two sites m and n [57,58] as additional measurable quantities:

$$S_{mn} = V(\hat{p}_m - \hat{p}_n) + V(\hat{q}_m + \hat{q}_n), \tag{F4}$$

where the amplitude operator $\hat{p}_n = (\hat{a}_n + \hat{a}_n^\dagger)/2$, the phase operator $\hat{q}_n = (\hat{a}_n - \hat{a}_n^\dagger)/(2i)$, and the variance of an operator $V(\hat{\mathcal{O}}) = \langle \hat{\mathcal{O}}^2\rangle - \langle \hat{\mathcal{O}}\rangle^2$. Assuming that these quantities are experimentally accessible, we define

$$y^{\text{out}}(t) = \sum_n W_n^{\text{out}} \langle \hat{a}_n^\dagger(t)\hat{a}_n(t)\rangle + \sum_{mn} D_{nm}^{\text{out}} S_{mn}(t), \tag{F5}$$

with the feedback $u = y^{\text{out}}(t - \Delta t)$. We can add the same entanglement measures to the output of the classical reservoir. However, S_{mn} remains 1 for any two sites n and m due to the absence of entanglement in a classical system. We study these systems considering a 2×3 lattice. Given that the Hilbert space of H is large, only this small lattice can be simulated within our available computational resources.

A small classical system (2×3) has limited accuracy with a NRSE $\sigma \sim 0.5$ (see Fig. 9). The same real-sized quantum system obtains better results ($\sigma \sim 0.01$), as shown in Fig. 10, due to the larger size of its Hilbert space and the larger number of available output quantities. It should be noted that here we consider a system in the strongly interacting regime, to obtain nonclassical correlations. We also neglect the feedback caused by the process of measuring the quantum system, which implies the use of many copies of the system, as considered in Ref. [10].

As a consistency check, let us now take our quantum reservoir in the regime $\max[J_{nm}] \ll \gamma_Q$, where the hopping

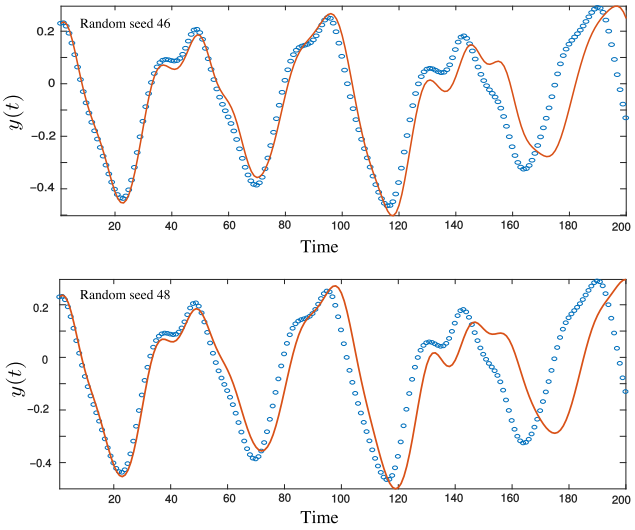

FIG. 10. A typical prediction (solid lines) from a quantum reservoir of size 2×3 compared with the Mackey-Glass data (circles). The error $\sigma \sim 0.01$ (NRSE value). We use $U = 5\gamma_Q$, random J uniformly distributed in $[\pm\gamma_Q/2]$, and a training data set of 200 time steps. We see that a quantum reservoir has a greater prediction capability than that of a classical one of the same size.

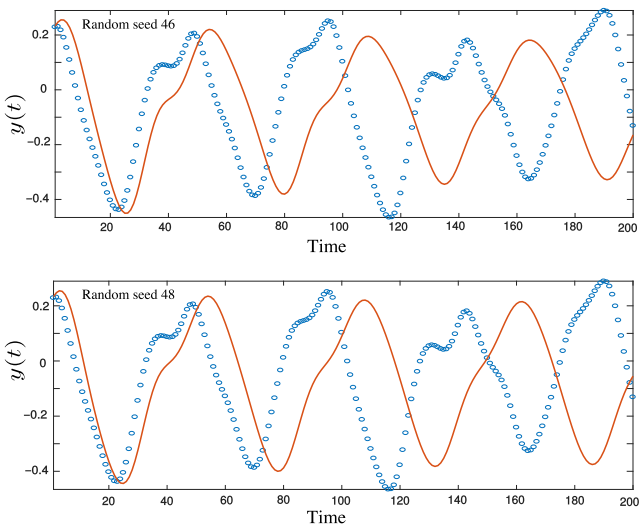

FIG. 11. Predictions (solid lines) from a highly dissipative quantum reservoir of size 2×3 compared with the Mackey-Glass data (circles). The error $\sigma \sim 0.51$ (NRSE value), similar to that of a classical reservoir. We use $U = 5\gamma_Q$, random J distributed in $[\pm\gamma_Q/8]$ and a training data set of 200 time steps. A high decay rate $\gamma_Q \gg J_{mn}$ suppresses the quantum entanglement in the reservoir and thus the quantum advantage is lost.

between the sites is weaker than the decay rate γ_Q. In this regime, the quantum entanglement is suppressed and thus we lose the advantage of using the quantum reservoir (see Fig. 11).

[1] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature **521**, 436 (2015).

[2] H. Jaeger and H. Haas, Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication, Science **304**, 78 (2004).

[3] W. Maass, T. Natschlager, and H. Markram, Real-time computing without stable states: a new framework for neural computation based on perturbations, Neural Comput. **14**, 2531 (2002).

[4] P. Enel, E. Procyk, R. Quilodran, and P. F. Dominey, Reservoir computing properties of neural dynamics in prefrontal cortex, PLoS Comput. Biol. **12**, e1004967 (2016).

[5] M. Lukoševičius and H. Jaeger, Reservoir computing approaches to recurrent neural network training, Comput. Sci. Rev. **3**, 127 (2009).

[6] M. Lukoševičius, in *Neural Networks: Tricks of the Trade: Second Edition*, edited by G. Montavon, G. B. Orr, and K.-R. Müller (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012), p. 659.

[7] 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).

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

[9] L. Appeltant, M. C. Soriano, G. van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, Information processing using a single dynamical node as complex system, Nat. Commun. **2**, 468 (2011).

[10] K. Fujii and K. Nakajima, Harnessing Disordered-Ensemble Quantum Dynamics for Machine Learning, Phys. Rev. Appl. 8, 024030 (2017), arXiv:1602.08159 [quant-ph].

[11] K. Vandoorne, P. Mechet, T. van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, Experimental demonstration of reservoir computing on a silicon photonics chip, Nat. Commun. **5**, 3541 (2014).

[12] J. Torrejon, M. Riou, F. A. Araujo, S. Tsunegi, G. Khalsa, D. Querlioz, P. Bortolotti, V. Cros, K. Yakushiji, A. Fukushima, H. Kubota, S. Yuasa, M. D. Stiles, and J. Grollier, Neuromorphic computing with nanoscale spintronic oscillators, Nature **547**, 428 (2017).

[13] http://www.neural-forecasting-competition.com/NN3/index.htm.

[14] F. Schürmann, K. Meier, and J. Schemmel, in *Advances in Neural Information Processing Systems 17*, edited by L. K. Saul, Y. Weiss, and L. Bottou (MIT Press, Cambridge, MA, 2005), p. 1201.

[15] C. Du, F. Cai, M. A. Zidan, W. Ma, S. H. Lee, and W. D. Lu, Reservoir computing using dynamic memristors for temporal information processing, Nat. Commun. **8**, 2204 (2017).

[16] D. Kudithipudi, Q. Saleh, C. Merkel, J. Thesing, and B. Wysocki, Design and analysis of a neuromemristive reservoir computing architecture for biosignal processing, Front. Neurosci. **9**, 502 (2016).

[17] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, Optoelectronic reservoir computing, Sci. Rep. 2, 287 (2012), arXiv:1111.7219.

[18] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, Photonic information processing beyond turing: an optoelectronic implementation of reservoir computing, Opt. Express **20**, 3241 (2012).

[19] F. Duport, B. Schneider, A. Smerieri, M. Haelterman, and S. Massar, All-optical reservoir computing, Opt. Express 20, 22783 (2012), arXiv:1207.1619 [physics.optics].

[20] J. C. Coulombe, M. C. A. York, and J. Sylvestre, Computing with networks of nonlinear mechanical oscillators, PLoS ONE 12, e0178663 (2017), arXiv:1704.06320.

[21] C. Fernando, and S. Sojakka, in *Advances in Artificial Life*, edited by W. Banzhaf, J. Ziegler, T. Christaller, P. Dittrich, and J. T. Kim (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003), p. 588.

[22] G. Tanaka, T. Yamane, J. Benoit Héroux, R. Nakane, N. Kanazawa, S. Takeda, H. Numata, D. Nakano, and A. Hirose, Recent advances in physical reservoir computing: a review, e-prints arXiv:1808.04962 (2018), arXiv:1808.04962 [cs.ET].

[23] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, Deep learning with coherent nanophotonic circuits, Nat. Photonics 11, 441 (2017), arXiv:1610.02365 [physics.optics].

[24] H. J. Caulfield and S. Dolev, Why future supercomputing requires optics, Nat. Photonics **4**, 261 EP (2010).

[25] I. S. Aranson and L. Kramer, The world of the complex ginzburg-landau equation, Rev. Mod. Phys. **74**, 99 (2002).

[26] J. Keeling and N. G. Berloff, Spontaneous Rotating Vortex Lattices in a Pumped Decaying Condensate, Phys. Rev. Lett. **100**, 250401 (2008).

[27] R. Labouvie, B. Santra, S. Heun, and H. Ott, Bistability in a Driven-Dissipative Super-Fluid, Phys. Rev. Lett. **116**, 235302 (2016).

[28] P. M. Walker, L. Tinkler, D. V. Skryabin, A. Yulin, B. Royall, I. Farrer, D. A. Ritchie, M. S. Skolnick, and D. N. Krizhanovskii, Ultra-low-power hybrid lightmatter solitons, Nat. Commun. **6**, 8317 EP (2015), article.

[29] N. Y. Kim, K. Kusudo, A. Löffler, S. Höfling, A. Forchel, and Y. Yamamoto, Exciton–polariton condensates near the dirac point in a triangular lattice, New J. Phys. **15**, 035032 (2013).

[30] K. Winkler, O. A. Egorov, I. G. Savenko, X. Ma, E. Estrecho, T. Gao, S. Müller, M. Kamp, T. C. H. Liew, E. A. Ostrovskaya, S. Höfling, and C. Schneider, Collective state transitions of exciton-polaritons loaded into a periodic potential, Phys. Rev. B **93**, 121303 (2016).

[31] V. G. Sala, D. D. Solnyshkov, I. Carusotto, T. Jacqmin, A. Lemaître, H. Terças, A. Nalitov, M. Abbarchi, E. Galopin, I. Sagnes, J. Bloch, G. Malpuech, and A. Amo, Spin-Orbit Coupling for Photons and Polaritons in Microstructures, Phys. Rev. X **5**, 011034 (2015).

[32] C. E. Whittaker, E. Cancellieri, P. M. Walker, D. R. Gulevich, H. Schomerus, D. Vaitiekus, B. Royall, D. M. Whittaker, E. Clarke, I. V. Iorsh, I. A. Shelykh, M. S. Skolnick, and D. N. Krizhanovskii, Exciton Polaritons in a Two-Dimensional Lieb Lattice with Spin-Orbit Coupling, Phys. Rev. Lett. **120**, 097401 (2018).

[33] J. J. Hopfield, Theory of the contribution of excitons to the complex dielectric constant of crystals, Phys. Rev. **112**, 1555 (1958).

[34] C. Weisbuch, M. Nishioka, A. Ishikawa, and Y. Arakawa, Observation of the Coupled Exciton-Photon Mode Splitting in a Semiconductor Quantum Microcavity, Phys. Rev. Lett. **69**, 3314 (1992).

[35] A. Kavokin, J. J. Baumberg, G. Malpuech, and F. P. Laussy, *Microcavities* (Oxford University Press, Inc., New York, NY, USA, 2008).

[36] I. Carusotto and C. Ciuti, Quantum fluids of light, Rev. Mod. Phys. **85**, 299 (2013).

[37] J. Kasprzak, M. Richard, S. Kundermann, A. Baas, P. Jeambrun, J. M. J. Keeling, F. M. Marchetti, M. H. Szymanska, R. Andre, J. L. Staehli, V. Savona, P. B. Littlewood, B. Deveaud, and L. S. Dang, Bose-einstein condensation of exciton polaritons, Nature **443**, 409 (2006).

[38] A. Amo, J. Lefrère, S. Pigeon, C. Adrados, C. Ciuti, I. Carusotto, R. Houdré, E. Giacobino, and A. Bramati, Superfluidity of polaritons in semiconductor microcavities, Nat. Phys. **5**, 805 (2009).

[39] H. Deng, H. Haug, and Y. Yamamoto, Exciton-polariton bose-einstein condensation, Rev. Mod. Phys. **82**, 1489 (2010).

[40] D. Ballarini, M. De Giorgi, E. Cancellieri, R. Houdré, E. Giacobino, R. Cingolani, A. Bramati, G. Gigli, and D. Sanvitto, All-optical polariton transistor, Nat. Commun. **4**, 1778 EP (2013), article.

[41] T. Gao, P. S. Eldridge, T. C. H. Liew, S. I. Tsintzos, G. Stavrinidis, G. Deligeorgis, Z. Hatzopoulos, and P. G. Savvidis, Polariton condensate transistor switch, Phys. Rev. B **85**, 235102 (2012).

[42] A. Amo, T. C. H. Liew, C. Adrados, R. Houdré, E. Giacobino, A. V. Kavokin, and A. Bramati, Excitonpolariton spin switches, Nat. Photonics **4**, 361 EP (2010).

[43] T. Espinosa-Ortega and T. C. H. Liew, Perceptrons with Hebbian learning based on wave ensembles in spatially patterned potentials, Phys. Rev. Lett. 114, 118101 (2015), arXiv:1408.6949 [cond-mat.dis-nn].

[44] H. Ohadi, A. Dreismann, Y. G. Rubo, F. Pinsker, Y. del Valle-Inclan Redondo, S. I. Tsintzos, Z. Hatzopoulos, P. G. Savvidis, and J. J. Baumberg, Spontaneous Spin Bifurcations and Ferromagnetic Phase Transitions in a Spinor Exciton-Polariton Condensate, Phys. Rev. X **5**, 031002 (2015).

[45] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE **86**, 2278 (1998).

[46] N. Bobrovska, M. Matuszewski, K. S. Daskalakis, S. A. Maier, and S. Kéna-Cohen, Dynamical instability of a nonequilibrium exciton-polariton condensate, ACS Photonics **5**, 111 (2018).

[47] G. Nardin, G. Grosso, Y. Léger, B. Pietka, F. Morier-Genoud, and B. Deveaud-Plédran, Hydrodynamic nucleation of quantized vortex pairs in a polariton quantum fluid, Nat. Phys. **7**, 635 EP (2011), article.

[48] M. Abbarchi, A. Amo, V. G. Sala, D. D. Solnyshkov, H. Flayac, L. Ferrier, I. Sagnes, E. Galopin, A. Lemaître, G. Malpuech, and J. Bloch, Macroscopic quantum self-trapping and Josephson oscillations of exciton polaritons, Nat. Phys. **9**, 275 EP (2013).

[49] P. Cilibrizzi, H. Sigurdsson, T. C. H. Liew, H. Ohadi, A. Askitopoulos, S. Brodbeck, C. Schneider, I. A. Shelykh, S. Höfling, J. Ruostekoski, and P. Lagoudakis, Half-skyrmion spin textures in polariton microcavities, Phys. Rev. B **94**, 045315 (2016).

[50] M. Pieczarka, M. Syperek, L. Dusanowski, A. Opala, F. Langer, C. Schneider, S. Höfling, and G. Sek, Relaxation oscillations and ultrafast emission pulses in a disordered expanding polariton condensate, Sci. Rep. **7**, 7094 (2017).

[51] Q. Wang, Y. Li, and P. Li, in *2016 IEEE International Symposium on Circuits and Systems (ISCAS)* (2016), p. 361.

[52] M. Milićević, T. Ozawa, P. Andreakou, I. Carusotto, T. Jacqmin, E. Galopin, A. Lemaître, L. L. Gratiet, I. Sagnes, J. Bloch, and A. Amo, Edge states in polariton honeycomb lattices, 2D Mater. **2**, 034012 (2015).

[53] P. A. Merolla, J. V. Arthur, R. Alvarez-Icaza, A. S. Cassidy, J. Sawada, F. Akopyan, B. L. Jackson, N. Imam, C. Guo, Y. Nakamura, B. Brezzo, I. Vo, S. K. Esser, R. Appuswamy, B. Taba, A. Amir, M. D. Flickner, W. P. Risk, R. Manohar, and D. S. Modha, A million spiking-neuron integrated circuit with a scalable communication network and interface, Science **345**, 668 (2014). http://science.sciencemag.org/content/345/6197/668.full.pdf.

[54] D. Li, M. Han, and J. Wang, Chaotic time series prediction based on a novel robust echo state network, IEEE Trans. Neural Netw. Learn. Syst. **23**, 787 (2012).

[55] R. Lyon, in *ICASSP'82. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 7 (1982), p. 1282.

[56] D. Verstraeten, B. Schrauwen, D. Stroobandt, and J. V. Campenhout, Isolated word recognition with the liquid state machine: a case study, Inf. Process. Lett. **95**, 521 (2005), Applications of Spiking Neural Networks.

[57] L.-M. Duan, G. Giedke, J. I. Cirac, and P. Zoller, Inseparability Criterion for Continuous Variable Systems, Phys. Rev. Lett. **84**, 2722 (2000).

[58] R. Simon, Peres-Horodecki Separability Criterion for Continuous Variable Systems, Phys. Rev. Lett. **84**, 2726 (2000).