Ultrafast neural sampling with spiking nanolasers

Ivan K. Boikov[1,*], Alfredo de Rossi[1] and Mihai A. Petrovici[2]

[1] Thales Research & Technology, Palaiseau Cedex, 91767, France

[2] Department of Physiology, University of Bern, Bern, 3012, Switzerland

[*] mail@ikboikov.net

January 27, 2025

Abstract

Owing to their significant advantages in terms of bandwidth, power efficiency and especially speed, optical neuromorphic systems have arisen as interesting alternatives to conventional semiconductor devices. Recently, photonic crystal nanolasers with excitable behaviour were first demonstrated. Depending on the pumping strength, they emit short optical pulses – spikes – at various intervals on a nanosecond timescale. In this theoretical work, we show how networks of such photonic spiking neurons can be used for Bayesian inference through sampling from learned probability distributions. We provide a detailed derivation of translation rules from conventional sampling networks such as Boltzmann machines to photonic spiking networks and demonstrate their functionality across a range of generative tasks. Finally, we provide estimates of processing speed and power consumption, for which we expect improvements of several orders of magnitude over current state-of-the-art neuromorphic systems.

Introduction

In the overwhelming majority of artificial neural networks (ANNs) in use today, neuronal outputs are determined by smooth and continuous activation functions, and their values are updated synchronously for multiple neurons within a network. In contrast, biological neurons, especially in the mammalian cortex, communicate asynchronously with short, stereotypical pulses called "spikes". While the non-differentiability of such spike-based codes indeed makes them more difficult to train using error backpropagation and therefore less attractive for conventional deep learning (however, see [1–3]), they possess several properties that are extremely relevant in physical neuronal networks, biological and artificial alike. First, they can be much more efficient than time-continuous or rate codes: as information becomes implicitly encoded in the timing of the spikes, rather than the value of neuronal outputs, a spike code can save both energy and bandwidth. Thus, spiking neurons have become the de-facto standard model for neuromorphic systems, both purely digital [4–6] and mixed-signal [7–10]. Second, due to the all-or-nothing nature of spikes, the transmission of information is significantly less prone to disruption by noise. This makes spiking neural networks particularly attractive for mixed-signal devices, whose analog components (neurons and synapses) invariably introduce spatio-temporal noise into the network dynamics.

By associating neuronal spikes with binary states, one can link spiking neural networks to Boltzmann machines (BMs) [11–13]. This machine learning model forms the basis for powerful variants such as deep belief networks [14] and autoencoders [15], which have been used for applications ranging from acoustic modeling [16] to medical diagnostics [17] and quantum tomography [18]. By moving to the spiking domain, one can harness the speed and efficiency of neuromorphic substrates for such applications. Indeed, first demonstrations of such spike-based sampling (SBS) networks on highly accelerated neuromorphic hardware have already shown promising results [19–21].

However, despite the impressive results of electric and electronic networks, electrical interconnects between computing elements limit bandwidth, latency and energy efficiency [22]. In contrast, photonic transmission of information offers significant advantages in terms of loss and speed. Not only is the issue of Joule heating alleviated, but the improved fan-in/-out by wavelength multiplexing also allows large-bandwidth light-speed transmission of signals. A photonic equivalent of a neuron is a laser with an excitable response [23–25]; interaction between such neurons with highly accelerated intrinsic dynamics is being developed [26]. Among a variety of possible implementations, a semiconductor laser with a saturable absorber (SA) is of particular interest here [27] because of the strong analogy with the leaky integrate-and-fire (LIF) neuron model [28].

Semiconductor lasers can be miniaturized and integrated in a photonic circuit and need tiny amounts of power (about 100 µW) for operation [29]. These are essential requirements for advancing the neuromorphic state of the art. Very recently, this technology was shown to produce an excitable response by introducing a section behaving as an SA [30], making it promising for the implementation of efficient integrated photonic spiking neurons (PSNs). In this work, we demonstrate the capability of such PSN networks to implement BMs through SBS.

Results

Preliminaries

Whenever a neuron fires a spike, it enters a refractory period of some duration τ during which it cannot spike again. Thus, a spike can be interpreted as representing the onset of an "active" state

$$z_k(t) = 1 \Leftrightarrow \text{neuron has fired in } (t - \tau; t] , \quad (1)$$

with $z_k(t) = 0$ otherwise. In BMs, the probability of neuron ensemble being in a particular state $\boldsymbol{z}$ follows the Boltzmann distribution $p(\boldsymbol{z}) \propto \exp\left[-E(\boldsymbol{z})/k_\mathrm{B}T\right]$, where the energy E is defined as

$$E = -\sum_{k<j} W_{kj} z_k z_j - \sum_k b_k z_k ,$$

with neuronal biases b_k and synaptic weights $W_{kj} = W_{jk}$, k_B is the Boltzmann constant, and T is the ensemble (Boltzmann) temperature. Here, we disregard the units, assume $k_\mathrm{B}T = 1$ and omit this term for brevity.

In SBS, a spiking neural network approximates $p(\boldsymbol{z})$ through a time-continuous analogon of Gibbs sampling [11]. A particularly interesting variant builds on LIF sampling (LIFS) neurons [12, 13], as they represent a de-facto standard model across the vast majority of neuromorphic platforms. In LIFS, the required stochasticity is generated by adding (or exploiting pre-existing) noise on the neuronal membranes, which consequently follow the Ornstein-Uhlenbeck process

$$\tau_\mathrm{m} \dot{u}_k = (b - u) + \sum_j \sum_{t_j^\mathrm{spike}} W_{kj} \kappa(t - t_j^\mathrm{spike}) + \sigma_W \, \mathrm{d}W \quad (2)$$

with an autocorrelation determined by the neuronal membrane time constant τ_m ($\mathrm{d}W$ represents a Wiener process scaled by a noise amplitude σ_W). The interaction between neurons is mediated by an additive postsynaptic potential (PSP) kernel κ that is triggered by an incoming spike at time t^spike.

Spiking nanolasers

The PSNs discussed in this work are semiconductor lasers composed of two sections: a gain section and an SA. The two sections form a single resonator and are therefore spanned by a single laser mode (Fig. 1a). The gain section is pumped to reach local electron population inversion and stimulated emission of photons. The SA is not pumped, such that here absorption always dominates, yet saturating as the flux of absorbed photons increases. As the pump rate increases beyond a certain threshold, the absorption is overtaken by the stimulated emission; akin to negative differential resistance in electronic oscillators, this introduces a positive feedback which initiates the emission of a pulse. In turn, this depletes the population of excited electron-hole pairs, hence the gain, and the laser is shut off. Therefore, the emission of a new pulse immediately following a previous one is strongly suppressed.
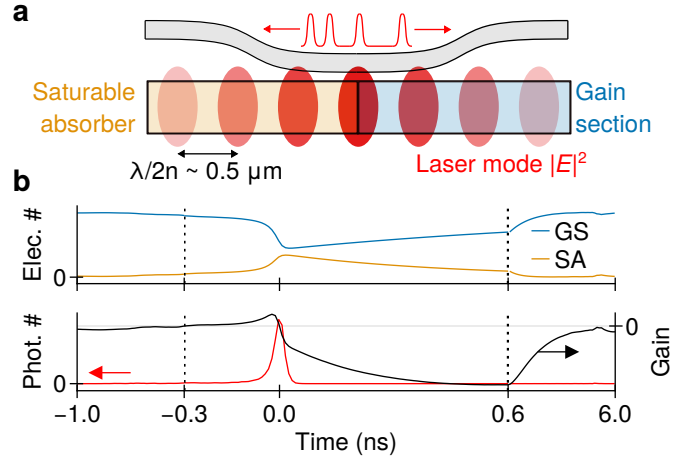

Figure 1: Spiking nanolasers. **(a)** Schematic of a PSN coupled to a waveguide (gray). Here, the laser is a photonic crystal; its modes are standing-wave and optical spikes are coupled out to the waveguide in both directions, E is an electric field, λ is the wavelength in vacuum, and n is the refractive index of the absorber. **(b)** Optical spike emission. Dashed lines separate sections with different scaling of the time axis.

After some time, the pumping re-establishes the initial gain, and the nanolaser can spike again. This process (Fig. 1b) bears a strong resemblance to the generation of action potentials and the subsequent refractoriness found in biological neurons, as described by the Hodgkin-Huxley model [31].

The spiking dynamics in a semiconductor laser with an SA is described by the Yamada model [27, 32], which ignores spontaneous emission and therefore assumes a perfectly deterministic response. Here, we consider a laser with a single optical mode; its field is distributed over a volume comparable to λ^3, where λ is the wavelength. Moreover, the active region, where electron and hole pairs are created, is even smaller. Under these conditions, spontaneous emission is not negligible, as the fraction of it going into the mode – the spontaneous emission factor β – is between 0.1 and 1.0, whereas in macroscopic semiconductor lasers it is 10^{-4} or less [33]. Therefore, in nanolasers, the average number of photons is much smaller, and the relative noise due to the granularity is much larger, which disrupts the otherwise deterministic spiking at regular intervals controlled by pumping strength.

A rigorous description of noise requires a quantum mechanical formalism, which is exceedingly complicated for semiconductor systems. Therefore, the semiconductor laser is often approximated as a homogeneously broadened two-level system, as intraband scattering is fast enough such that electrons and holes are in thermal equilibrium [34, 35]. The light-matter interaction within a semiconductor can therefore be described as a collection of dipoles interacting with the same optical mode. With some approximations, such as fast dephasing of the polarization with respect to the damping rate of carriers and

photons, the quantum mechanical descriptions lead to rate equations [36, 37].

The rate equations describe the evolution in time of the populations of excited dipoles n_e and photons S in a cavity mode due to multiple processes. Stimulated emission and absorption are described by a term GS, where $G = \gamma_r(2n_e - n_0)$ represents the gain, n_0 is the total number of dipoles, and γ_r is the radiative transition rate. Spontaneous emission and photon damping are described by $\gamma_r n_e$ and γS. Since $n_e \leq n_0$, the maximum gain is limited to $\gamma_r n_0$; this condition is ensured by an optical pumping term $\gamma_p(n_0 - n_e)$, where γ_p is the pumping rate and $n_0 - n_e$ represents pumping saturation.

The stochastic equations are formed by including the Langevin forces $F_i(t)$ which are random variables with zero mean and auto-/cross-correlation strengths $\langle F_i(t)F_j(t)\rangle = 2D_{ij}$. The Langevin forces are calculated consistently with the rate equations [37] based on the McCumber noise model [38] (see Nanolaser noise model in Methods). Here, we extend the model by including two separate sections, one providing gain with an excited population n_e, and another one representing the SA, denoted with the suffix a, with an excited population n_a, both interacting with the same mode:

$$\dot{S} = GS + \gamma_r n_e + \gamma_{r,a} n_a + F_S(t) \,,$$
$$\dot{n}_e = -\gamma_r(2n_e - n_0)S - \gamma_t n_e + \gamma_p(n_0 - n_e) + F_e(t) \,, \quad (3)$$
$$\dot{n}_a = -\gamma_{r,a}(2n_a - n_{0,a})S - \gamma_{t,a} n_a + F_a(t) \,,$$

where $G = \gamma_r(2n_e - n_0) + \gamma_{r,a}(2n_a - n_{0,a}) - \gamma$ is the net gain including photon damping, and other parameters are defined in Table 1 (see Nanolaser with quantum wells in Methods). For clarity, we normalize γ_p by the threshold pumping strength in the absence of noise γ_p^{thr} (see Nanolaser with quantum wells in Methods) which we later refer to as the "threshold". It is important to note that with noise, spike emission can also occur with $\gamma_p < \gamma_p^{thr}$.

Fig. 2a shows a time trace of the membrane potential of a single LIFS neuron with parameters from [13] and Gaussian noise. This serves as a reference for the behavior of PSNs, which we discuss below. In Fig. 2b we show time traces of PSN gain with pumping far from and close to the spiking threshold, respectively. In this system, the gain is a stochastic variable as it depends on stochastic electron densities. As a result, when a PSN is not refractory, the gain undergoes a random walk as shown in Fig. 2c, similarly to the LIFS membrane potential. The ISI distribution, gain autocorrelation and probability density function of a PSN are shown in Fig. 2d, e and f, respectively. These are close to the corresponding properties of LIFS neurons, with small deviations explained by the additional nonlinear terms of the PSN state equations.

It has been pointed out that the granularity of light challenges numerical integration based on the Langevin forces [39–41]. Therefore, we further cross-checked the results by implementing a rigorous discrete PSN model, as proposed in [40] and concluded that both methods lead to the same statistical properties of the PSN (see Discrete nanolaser model in Methods).
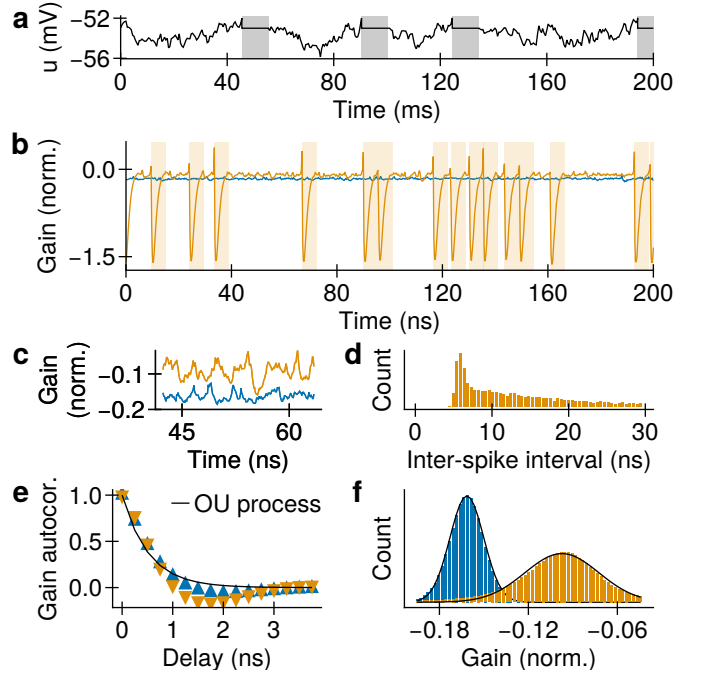

Figure 2: PSNs emulate LIFS neurons. (a) Membrane potential of a LIF neuron. Shading represents the "active" state after a spike emission. Below: properties of a PSN far from (blue, $\gamma_p = 0.955\gamma_p^{thr}$) and close to (orange, $\gamma_p = 0.995\gamma_p^{thr}$) the spiking threshold. (b) Gain time traces. After a spike is emitted, the gain is reduced drastically, and the PSN is considered "active" for a time shown with shading. (c) Gain time traces between spike emissions. (d) ISI histogram. (e) Autocorrelation of the gain. For each spike at t_s, the interval $(t_s - 0.5\tau; t_s + \tau)$ was omitted from the analysis to exclude the highly nonlinear regime dynamics of the spiking process. The case far from the threshold is fitted with an Ornstein-Uhlenbeck process, as also obeyed by the free membrane potential of LIFS neurons. (f) Histograms of gain values between spike emissions and fit with normal distributions (black lines).

Networks of spiking nanolasers

Interference-based interaction between PSNs is challenging: their bistability, combined with a potentially large number, will complicate the necessary resonance alignment. For this reason, in this work we assume that PSNs are incoherent, and their interaction is mediated by photodiodes. Spikes are extracted from a PSN by coupling to a waveguide. Connection weights can be implemented optically using a waveguide crossbar array [42]. There, each output waveguide incoherently combines optical signals from input waveguides, each weighted by a set of couplers. Due to a lack of interference, these couplers can only implement non-negative synaptic weights, but in BMs, weights can be negative as well. For this reason, we follow [43] and use balanced photodetectors, which require an $N \times 2N$ crossbar array for N PSNs (see Fig. 3a). As a result, optical spikes are converted into electrical current
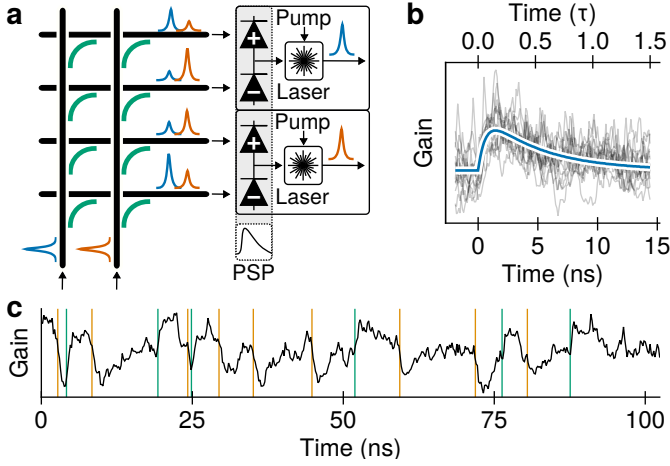
Figure 3: Setup of a PSN network. **(a)** Spiking nanolaser interaction using a waveguide crossbar array. Couplers (green) define the ratio of spike power sent to a balanced photodiode: one photodiode increases the electric current through the PSN, and the other one reduces it, corresponding to an excitatory or an inhibitory synapse, respectively. Limited electrical circuit bandwidth leads to filtering of the spike, resulting in an almost alpha-shaped PSP. **(b)** Change of the nanolaser gain upon a reception of a spike (i.e. a PSP). The blue line shows an alpha-shaped fit. **(c)** Gain time trace of a PSN affected by incoming spikes (vertical lines) from two other PSNs: one is connected with a positive weight (excitatory, green) and the other with a negative one (inhibitory, orange).

that changes the pumping strength by

$$\Delta\gamma_{p,k}(t) = \sum_j \kappa_{kj} \int \mathrm{LPF}(t - t^*) S_j(t^*)\mathrm{d}t^* , \quad (4)$$

where κ_{kj} represents a coupling strength incorporating relevant factors such as the coupling of a PSN to a waveguide and the photodiode responsivity, and $\mathrm{LPF}(\Delta t)$ is the impulse response function of the photodiodes that we assume to be a low-pass filter with a timescale τ_U due to their limited bandwidth. The change of current induces a change of gain shown in Fig. 3b; the change can be positive (excitatory) or negative (inhibitory), as shown in Fig. 3c.

This implementation of connections induces an almost alpha-shaped interaction kernel, again similar to typical interaction kernels in both biological and abstract (LIF) neuronal networks. For optimal performance of a network, the timescale of $\Delta\gamma_{p,k}(t)$ can require optimization. Here, it depends on τ_U, which we assume to be tunable, as discussed further below.

Membrane potential of spiking nanolasers

Consider a network of nanolasers and its k-th PSN is not currently spiking, i.e. its gain does a random walk (see Fig. 2c). Then, Eq. (3) is approximated by the following stochastic equation for the gain (see Simplified gain

equation in Methods):

$$\mathrm{d}G_k \approx [-(G_k - G_{p,k})/\tau_G + 2\gamma_r n_0 \Delta\gamma_{p,k}]\,\mathrm{d}t + \sigma_G\,\mathrm{d}W_k ,$$
$$(5)$$

where G_p is a drift term depending on the pumping strength, and the second term corresponds to the PSN interaction (see Eq. (4)). This equation is identical to that of the membrane potential of a LIFS neuron (see Eq. (2)) with two assumptions. First, $n_a \ll n_e$, which holds during non-spiking with moderate pumping strength. Second, is that changes of n_e due to incoming spikes need to be small, otherwise the interaction between PSNs deviates from linearity (see Simplified gain equation in Methods). Based on the formal equivalence between the two equations, we propose to use the gain G as the membrane potential of a PSN:

$$u_k(t) = (G_k(t) - G_0)/\partial_u G ,$$

where G_0 is the gain for which $p(z = 1) = 0.5$ (which requires the yet unknown refractory period τ), and $\partial_u G$ is a proportionality coefficient. Therefore, we must to find these three parameters to relate membrane potentials of LIFS neurons and PSNs.

Consider a single LIFS neuron with a logistic activation function $p(z = 1|\bar{u}) = \sigma(\bar{u})$. We expect a similar behaviour from a single PSN, where $\bar{G} = G_p$ (see Eq. (5)). In Fig. 4a we sweep the pumping strength G_p; for each G_p, we solve Eq. (3) for a sufficiently long time and estimate the activation function using Eq. (1) and the law of large numbers:

$$p(z = 1|G_p) \approx \bar{z}(G_p) ,$$

where $\bar{z} = \langle z(m \cdot \tau)\rangle_m$ is an average PSN activation discretized with a time step τ. The goal is then to ensure that

$$\bar{z}(G_p) \approx \sigma\left[(G_p - G_0)/\partial_u G\right] ,$$

which is an optimization problem for the three variables τ, G_0 and $\partial_u G$. With it solved, we find that the PSN activation function matches the logistic activation function of LIFS neurons well. Here, $\tau = 11$ ns, which corresponds to a sampling rate of approximately 0.1 GHz. Based on the similarities between the expressions for membrane potentials of LIFS neurons (Eq. (2)) and PSNs (Eq. (5)) we find the following translation rules between the bias and the pumping strength:

$$b_k = (G_{p,k} - G_0)/\partial_u G . \quad (6)$$

Connections between LIFS neurons induce a change of their membrane potentials according to Eq. (2). Consider two LIFS neurons connected unidirectionally, i.e. only $W_{12} \neq 0$, with $b_2 = 2$. This way, the second neuron acts as the source of spikes, while the first is the recipient. Sweeping b_1 and W_{12} we obtain a set of curves $p(z_1 = 1|b_1, W_{12})$ shown with lines in Fig. 4b. We now aim for a similar behaviour in PSNs.

We convert b_1 and b_2 to PSN pumping strength according to Eq. (6). We then sweep κ_{12} to obtain a set of curves
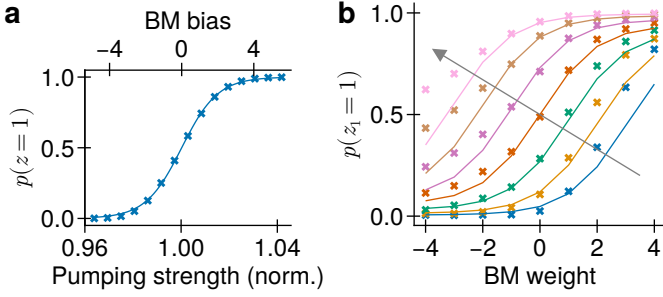
4

Figure 4: Translation of Boltzmann parameters to PSN parameters. (a) Spiking nanolaser activation function (crosses) with a logistic function fit (line). (b) Impact of connection weight on activation of a receiving neuron. Lines and crosses show activation of a BM neuron and a nanolaser, respectively. Each color corresponds to a receiving neuron bias $b_1 = -3, -2, \ldots 3$ increasing along the arrow.

$p(z_1 = 1|G_{p,1}(b_1), \kappa_{12})$. The goal is to find a proportionality coefficient ξ such that

$$p(z_1 = 1|G_{p,1}(b_1), \xi\kappa_{12}) \approx p(z_1 = 1|b_1, W_{12}) . \quad (7)$$

The result is given in Fig. 4b. We find that for limited biases and weights, PSNs replicate the behaviour of LIFS neurons well. Based on Eq. (7), we find the translation rule for weights:

$$W_{kj} = \xi\kappa_{kj} . \quad (8)$$

Optical sampling from Boltzmann distributions

To investigate the accuracy of sampling with PSNs, we first consider sampling from predefined Boltzmann distributions over a small set of binary random variables. Following [13], biases and weights were drawn from Beta distributions: $b_k \sim 1.2(\mathcal{B}(0.5, 0.5) - 0.5)$ and $W_{kj} \sim \beta(\mathcal{B}(0.5, 0.5) - 0.5)$, where β controls the range of weights and is either 0.6, 1.2 or 2.4. The generated biases and weights are translated to PSN parameters using Eqs. (6) and (8). The sampled distributions p are compared to the target distributions p^* by means of the Kullback-Leibler divergence $D_{KL}(p \parallel p^*)$.

First, we optimize the PSP timescale controlled by τ_U. On one hand, τ_U must be small enough such that a PSP does not last longer than the refractory period τ. On the other hand, too short τ_U will make the PSP short, but strong, which can break the operating regime assumptions (see Simplified gain equation in Methods). In LIFS neurons, the PSP timescale is ideally slightly shorter than the refractory period [13, Fig. 7]; we use this as a starting point. In Fig. 5a we sweep τ_U and track the sampling accuracy for all considered β. We find that the optimal τ_U is approximately 0.37τ. Figure 3b shows a PSP with such a timescale, and indeed, the PSP becomes negligible after $t = \tau$.

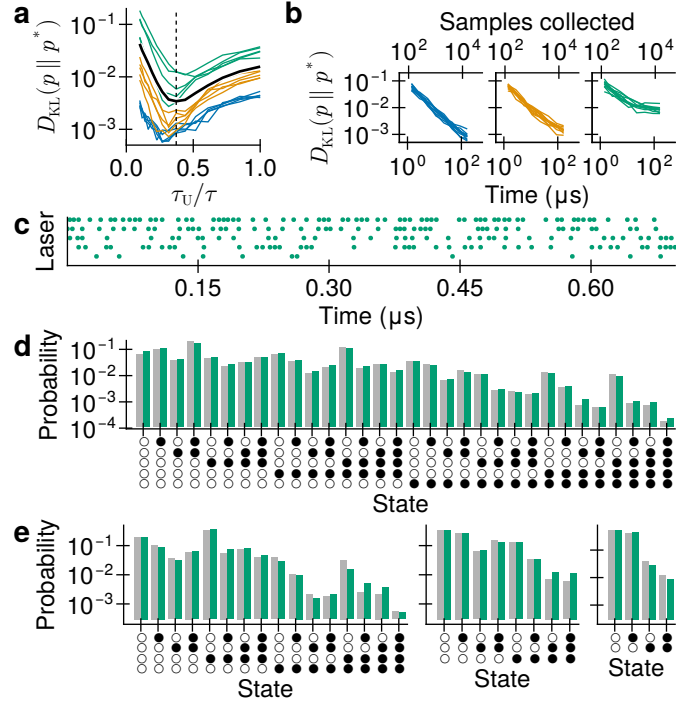We proceed to sample from sets of 10 random Boltzmann distributions for different β. Fig. 5b shows the con-

Figure 5: Sampling from random Boltzmann distributions with PSN networks. (a) Optimization of the photodiode timescale: τ_U is swept and sampling performance computed for PSN networks with maximum weights of 0.6 (blue), 1.2 (orange) and 2.4 (green). The solid black line is the mean D_{KL} for each τ_U. The dashed line shows $\tau_U = 0.37\tau$. (b) Convergence of sampling from 10 random Boltzmann distributions with weights up to 0.6, 1.2 and 2.4 from left to right. (c) Spike raster during sampling from a Boltzmann distribution with weights up to 2.4. (d) Sampling result for a Boltzmann distribution in (c). Gray: analytical distribution, green: sampling result. (e) Sampling from conditional distributions. From left to right: $p(z_{1345}|z_2 = [1])$, $p(z_{245}|z_{13} = [1, 0])$ and $p(z_{12}|z_{345} = [1, 1, 1])$. Colors match (d).

vergence of the sampling procedure towards the target distribution. We find that within each set the convergence is almost identical. In Fig. 5d we compare a distribution of samples to the exact distribution for $\beta = 2.4$, and we note a very close match. Figure 5c shows a raster of spikes during this sampling.

Next, we demonstrate Bayesian inference by sampling from conditional probability distributions. We split the five neurons in two arbitrary groups Y and X. The first group is clamped to an arbitrarily chosen state y, and the neurons in the second group are free; their probability distribution is the conditional probability distribution $p(X|Y = y)$. In PSNs, we clamp the state by significantly reducing or increasing the pumping strength. Figure 5e shows the close match between the correct conditionals and those sampled with our PSNs. We therefore conclude that a network of PSNs can sample accurately from a wide range of Boltzmann distributions over small state spaces, as well as accurately perform Bayesian inference therein.
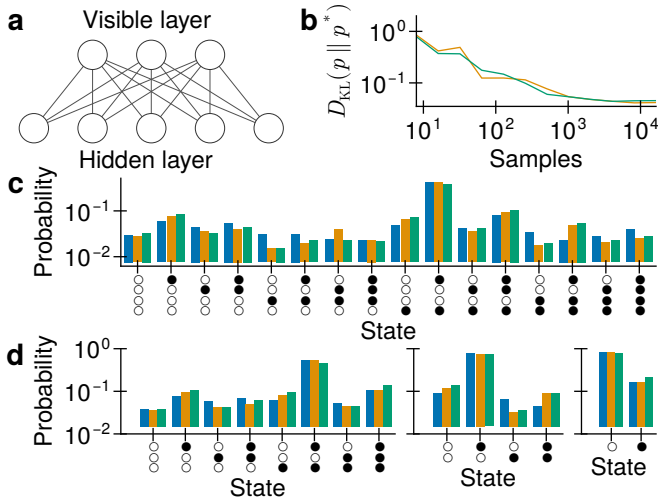
5

Figure 6: Sampling from an arbitrary distribution with a network of PSNs. **(a)** Architecture of the implemented PSN network. Each line represents a pair of symmetric synaptic connections. **(b)** Convergence of sampling with an RBM (orange) and a PSN network (green). **(c)** Comparison of the target distribution (blue) to the distributions sampled by the RBM and the PSN network. **(d)** Sampling from conditional distributions. From left to right: $p(z_{134}|z_2 = [0])$, $p(z_{24}|z_{13} = [1,0])$ and $p(z_3|z_{124} = [1,0,1])$. Colors match (c).

Optical sampling from arbitrary distributions

Boltzmann distributions are only a subset of all possible probability distributions over binary variables. However, the "fully visible" sampling networks described above can be extended by adding "hidden" neurons that are not observed during sampling. Consequently, the probability distribution of the visible layer becomes a marginal distribution over the full state space, which can, in principle, take any shape, given a large enough hidden space. To simplify training and improve convergence, a hierarchical network structure is preferable, with no horizontal connections within individual layers [14]. Here, we emulate such a two-layer restricted Boltzmann machine (RBM) with an equivalently structured PSN network (see Fig. 6a). Given enough hidden neurons, an RBM can sample from any distribution with arbitrary precision [44]. Consequently, by implementing such an RBM with PSNs, optical sampling from arbitrary distributions can be achieved.

For the target distribution $p^*(z)$ we choose four binary variables with the probability of each state sampled from the inverse continuous uniform distribution and normalized such that their sum is unity. The probability distribution is shown in Fig. 6c.

Using the contrastive divergence algorithm [45], we train an RBM with 4 visible and 10 hidden neurons. Its sampled distribution over visible neurons $p_{\mathrm{RBM}}(z)$ is shown in Fig. 6c.

The parameters of the trained RBM were then trans-

ferred to a network of 14 PSNs. Its sampled distribution over visible neurons $p_{\mathrm{PSN}}(z)$ is also shown in Fig. 6c. We find that the accuracy of the PSN network is very close to that of the RBM, and sampling from the target distribution $p^*(z)$ and several conditionals shown in Fig. 6d is correspondingly accurate.

Optical probabilistic inference

In this section, we demonstrate Bayesian inference from incomplete information with PSN networks. For better visualization, we chose three images of digits "0", "3" and "4" from the MNIST dataset [46], rescaled to 12×12 and with brightness rounded to zero or unity. These images were mapped to a fully connected network of 144 PSNs, with one pixel assigned to each neuron. The PSN network was then trained as an associative memory to store the prepared images (see Probabilistic inference training in Methods). Similar to the simulations described in previous sections, we first trained an equivalent BM with wake-sleep and then mapped the resulting parameters to the PSN network.

First, we assess the mixing capability of the network by observing its "dreaming" phase (corresponding to the sleep phase during training). Without external input, the PSN network correctly samples from the prior and switches randomly between states forming the three images with approximately equal probability. Figure 7a shows a two-dimensional projection of PSN samples onto vectors corresponding to the images (see Probabilistic inference visualization in Methods). We found that the result is close to that of the BM, and in most cases the network switches between the numbers every few samples (Fig. 7b), indicating good mixing between the states.

Next, we assess the inference capability of the network in a pattern completion scenario. By applying additional bias to a few neurons that are only active for two out of the three patterns, we provide informative, but limited input to the network. We chose five pixels that are black for "0" and "3", but not "4" (Fig. 7c), and apply an additional positive bias to the corresponding neurons. The biases of other neurons are unchanged, i.e. no information is given. Such an input is ambiguous w.r.t. "0" and "3", but incompatible with "4". As a result, the PSN network only generates complete images of "0" and "3", and randomly switches between them.

Learning from data

Sampling from arbitrary distributions implies the ability to sample from distributions dictated by real-world data. Here, we learn a generative model of handwritten digits based on the MNIST dataset [46]. Following [19], we round the brightness of each pixel to minimum or maximum.

To work with this dataset, we use a hierarchical sampling network: an RBM with three layers: visible, hidden and label (Fig. 8a). The brightness of pixels is mapped to an activity of a neuron in the visible layer. The activity
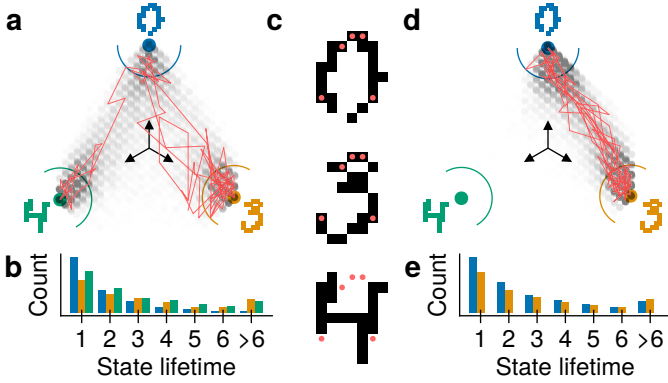
Figure 7: Bayesian inference with a network of PSNs. **(a)** Two-dimensional projection of network states after sampling from the prior trained to store images of digits "0", "3" and "4". Colored dots show projections of the images. More samples are shown with darker dots; those inside the half-circles are considered to be close to a corresponding digit. The red line shows a trajectory of 100 samples. **(b)** Histogram of time in samples spent inside the half-circles in (a) (colors match). **(c)** Input for Bayesian inference. The five red markers show the pixels where positive bias is applied. Such an input is ambiguous w.r.t. "0" and "3", but incompatible with "4". **(d,e)** Same as (a,b) when provided the input shown in (c).

in the label layer shows which digit is represented in the visible layer at the current point in time. For example, if visible neurons form an image of "0", only the label neuron corresponding to a "0" will spike.

In this section, we consider three tasks: completion, guided dreaming and classification. For completion, visible neurons are clamped to brightness of corresponding pixels except for a bottom-right quadrant, which is assumed "obscured" and remains free alongside other neurons. The task is for the free visible neurons to complete the obscured quadrant. For guided dreaming, one label neuron is clamped, and the visible neurons are expected to form an image of a corresponding digit. For classification, visible neurons are clamped to brightness of corresponding pixels, and the most active label neuron on average is taken as the answer ("winner-takes-all").

For these tasks, we use the BM parameters from [48] (Fig. 8b,c), as they were already optimized for LIFS networks and are therefore expected to be favorable for an implementation with PSNs. This network is composed of 1194 neurons: 784 in the visible, 400 in the hidden and 10 in the label layers, respectively. Its implementation with PSNs thus represents a scaling test for our general approach.

The simulation of PSNs during the tasks starts with a burn-in phase, where no input is provided. Then, the inputs are provided sequentially, with each subsequent input following immediately after the previous one.

For pattern completion with PSNs, 20 samples were drawn for each image. This is a difficult inference task, as the network should not simply produce an average image,

but instead needs to adapt to the style of each individual input sample. Figure 8d shows the results for a few images from the MNIST testing dataset. We found that in most cases, the obscured parts were completed to a large degree of accuracy.

For guided dreaming, label neurons were clamped for 200τ. To enforce top-down control (from labels to pixels), we strengthened the weights between the hidden and label layers by a factor of 2, while marginally reducing the others by 10% (see Discussion). In Fig. 8e, we show how the PSN network can thereby be used for generating images from all learned classes.

While BMs are designed primarily as generative networks, especially when trained purely with contrastive Hebbian methods and without dedicated backpropagation-based fine-tuning, input classification can still be viewed as a form of Bayesian inference; thus, it is instructive to compare classification performance between the original BM and its PSN implementation.

Simulating the processing of 10,000 images of the testing MNIST dataset with a PSN network of this size is computationally intensive; we therefore drew samples until the saturation of the classification convergence curve (Fig. 8f). In this case, we drew 50 samples for each image, yielding an average classification accuracy of 87.0%. The RBM is much less demanding; we could thus draw 500 samples, obtaining 87.8% accuracy. Their confusion matrices are compared in Fig. 8g; we note their similarity, as well as the only marginal performance loss caused by the exchange of substrates.

Interestingly, during the completion task, the networks perform classification as well. However, the accuracy of the PSN network is reduced to 75.0%, compared to 85.6% for the RBM (Fig. 8f); we expect this to be mitigated by further fine-tuning or direct in-situ training of the PSN network (see Discussion). However, in both cases, we observe that the PSN network only needs a few samples to converge to a solution, which is faster than the RBM and can prove beneficial in time-constrained scenarios.

Discussion

In this work, we have demonstrated the feasibility of spike-based sampling using networks of photonic spiking neurons. We have rigorously derived an analogy between the gain dynamics of a two-section semiconductor laser and the membrane dynamics of biological neurons, both above and below the spiking threshold. Using the resulting translation rules, we have mapped the learned parameters of Boltzmann machines to the corresponding quantities in nanolaser networks and have demonstrated accurate sampling across varied tasks of different scale and complexity.

While the analogy between PSNs, ideal LIFS neurons and ideal BM neurons represents a good approximation, we take note of two explicit differences. First, the refractoriness in PSNs is a result of a steep drop in the gain. This is more akin to the strong, but relative refractoriness of biological neurons, which is not as absolute as the
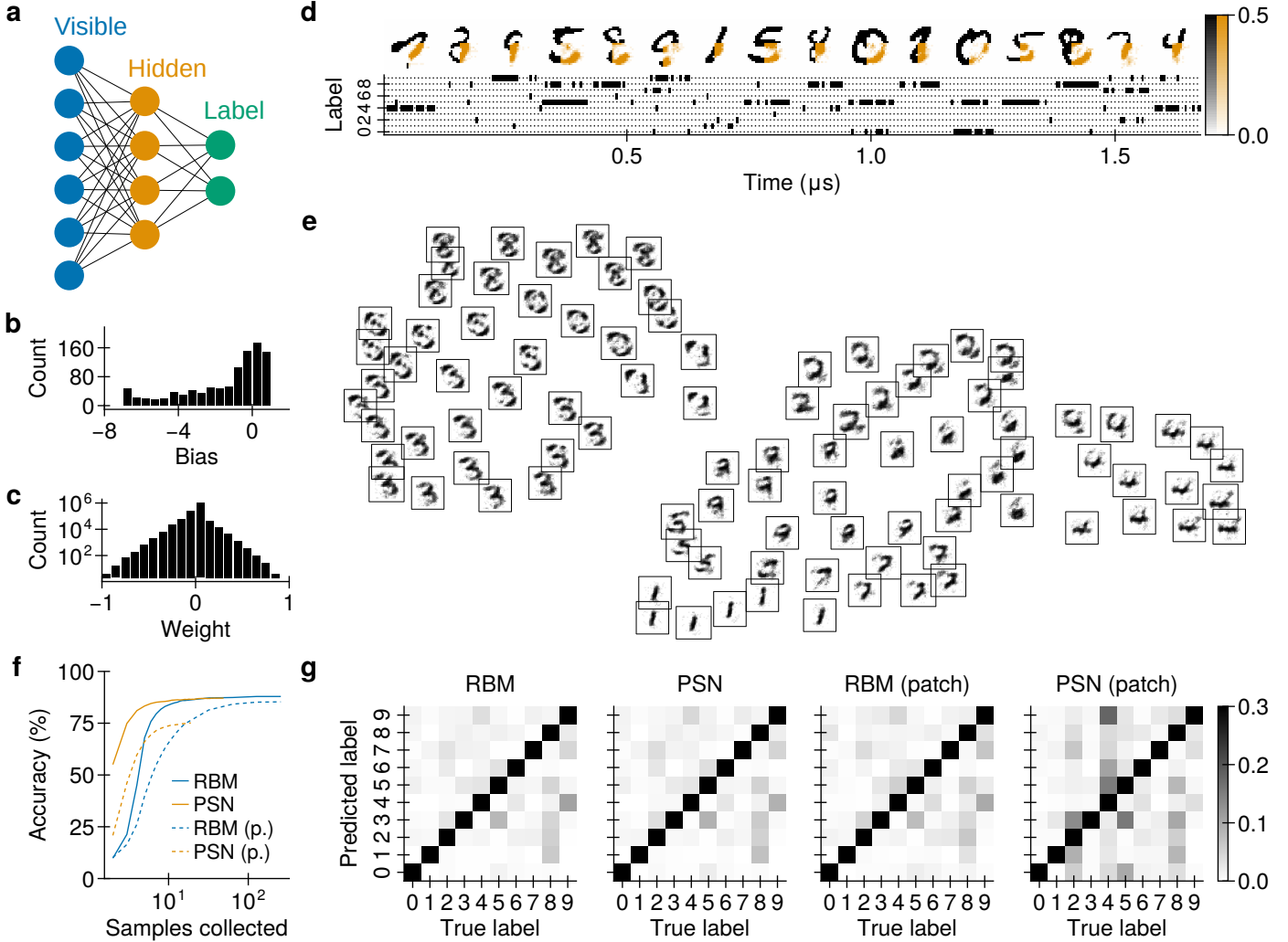
Figure 8: Network of spiking nanolasers applied on the MNIST dataset. **(a)** Scheme of an hierarchical sampling network. Each line represents a symmetric connection. **(b)(c)** Histograms of Boltzmann machine parameters. Strongly negative biases (-30) were omitted from the figures. **(d)** Time trace of spiking nanolaser-based MNIST completion with a bottom-right quarter patch occlusion. Restored pixels are orange. **(e)** Guided dreaming. Each image corresponds to a separate dream. For each dream, the activity of neurons in the visible layer was averaged over the last 20τ of each dream. The image positions correspond to projection in two dimensions using T-SNE [47]. **(f)** Convergence of classification with an RBM and a corresponding PSN network with complete images (solid lines) and with the bottom-right quarter patch occluded (dashed lines). **(g)** Confusion matrices for classification (left pair) (right pair).

refractoriness assumed by the LIFS model. Second, the PSP shape in PSN networks is close to an alpha function. This represents a deviation from the interaction kernels in BMs, which are rectangular. Nevertheless, neither of these properties is significantly detrimental to the ultimate network sampling accuracy. This is in line with observations from [11] and [13], which also explicitly address the issues of relative refractoriness and PSP shape. We expect the accuracy to further improve when parameter translation is replaced by in-situ training of PSN hardware.

The use of photonic timescales fosters a large improvement of sampling speeds compared to biological or electronic timescales. Even for highly accelerated neuromorphic systems such as BrainScaleS-1 [19] and BrainScaleS-2 [10], convergence speeds for sampling from small Boltzmann distributions over 5 random variables amount to ca. 10 seconds. With PSNs, these convergence times drop by over 4 orders of magnitude to ca. 10^2 microseconds. This acceleration factor would directly translate to a corresponding decrease in times-to-solution for any neuromorphic applications of SBS. Beyond the examples of Bayesian inference discussed here, these include tasks as diverse as stochastic constrained optimization [49] or quantum tomography [20, 21].

In this work, we have considered photonic neuronal samplers ranging in size from a few PSNs up to more than a thousand. While the implementation of individual components has seen recent experimental validation, the large-scale implementation of PSN networks in integrated photonics faces several challenges, which we discuss below.

We have assumed photonic crystal nanolasers as PSNs. Their small footprint promises high-density integration, but power dissipation then becomes an important issue. It is therefore necessary to reduce the amount of power required for PSN operation. Electrically driven nanolasers with the pump threshold of 10 µA have been recently demonstrated [50]. The photonic crystals nanolasers similar to those used in this work can also be pumped electrically, requiring about 100 µA each [29]. We estimate that for the MNIST tasks, 1194 nanolasers would require $3{\times}3$ mm^2 of chip space and, excluding optical and electrical losses and controller power consumption, approximately 700 mW of power: 100 mW for pumping and 600 mW for the amplification of spikes. With a sampling rate of 0.1 GHz, such a network of PSNs is equivalent to an RBM running at 64 TFLOP/s and 95 TFLOP/J.

The photonic crystal nanolasers are complex optical structures that require a dedicated fabrication process. A recently demonstrated technique of micro transfer printing, where each cavity is transferred from a wafer on the chip, has been shown to be scalable [51].

The implementation of dense programmable optical interconnects is one of the most challenging and actively pursued goals. In this work, the largest interconnect matrix considered was 784×400. As the PSN network is incoherent, the implementation of positive and negative weights would require balanced photodetectors which doubles the number of required matrix outputs (see Fig. 3a).

Table 1: Parameters of the PSN model.

Parameter	Value	Definition
γ	0.2 THz	photon damping rate
γ_r	$2.79 \times 10^{-6}\gamma$	transition rate
$\gamma_\mathrm{r,a}$	$5.31 \times 10^{-6}\gamma$	same, for the SA
γ_t	$1.28 \times 10^{-3}\gamma$	carrier damping rate
$\gamma_\mathrm{t,a}$	$1.01 \times 10^{-3}\gamma$	same, for the SA
χ_g	3	differential gain ratio
n_0	1.02×10^6	gain section dipole count
$n_\mathrm{0,a}$	8.20×10^5	same, in the SA

A naïve implementation with integrated programmable switch matrices [52] would be limited by chip space. The largest implementations currently demonstrated, such as a 240×240 port switch [53], would be still several times below the requirement. However, the number of necessary input ports can be significantly reduced by frequency multiplexing. Three-dimensional interconnects [54] are also promising for more extreme cases and multi-chip systems.

Methods

Nanolaser with quantum wells

The rate Eq. (3) describe the interaction of light with quantum dots [36, 37]. These semiconductor nanostructures, akin to artificial atoms, localize carriers within a few nanometers, and therefore are well modelled with a two-level electronic system. However, this work builds on the experimental results reported in [30], where the gain material consists of quantum wells. There, unlike in quantum dots, the gain depends nonlinearly on the density of carriers in the conduction band (i.e. the population of excited dipoles in our model divided by the volume of the section). This nonlinear dependence is approximated by a piecewise linear function. The slope is larger at lower carrier density, with a ratio χ_g. To ensure that $n_\mathrm{e} \leq n_0$, we corrected the pumping term in Eq. (3) by replacing $n_\mathrm{e} \to 2n_\mathrm{e}/(\chi_\mathrm{g} + 1)$.

The pump rate at threshold γ_p is computed from Eq. (3) by assuming a steady state without noise, $G = 0$ and $S \approx 0$, which leads to $\gamma_\mathrm{p}^\mathrm{thr} = \gamma_\mathrm{t} n_\mathrm{e}^\mathrm{thr}/[n_0 - 2n_\mathrm{e}^\mathrm{thr}/(\chi_\mathrm{g} + 1)]$, where $n_\mathrm{e}^\mathrm{thr} = (\gamma + \gamma_\mathrm{ra}n_\mathrm{0,a} + \gamma_\mathrm{r}n_0)/2\gamma_\mathrm{r}$.

Nanolaser noise model

The stochastic Eq. (3) are composed of deterministic ("drift") and stochastic ("diffusion") parts that can be represented in matrix form:

$$\mathrm{d}u = \mu(u,t)\,\mathrm{d}t + \sigma(u,t)\,\mathrm{d}W\ ,$$

where $u = [S, n_\mathrm{e}, n_\mathrm{a}]^\mathrm{T}$. In this work, the diffusion term follows the approach in [38, A.13.1.2] based on the McCumber noise model [55] and is comprised of five Langevin forces corresponding to the following groups of processes:

1. electron-photon interaction in the gain section,

2. same, in the SA,

3. electronic processes inside the gain section,

4. intrinsic optical loss,

5. electronic processes inside the SA.

The Langevin forces are stochastic processes represented with Wiener processes with zero average and cross-correlation $2D_{ij}$; for $i = j$, the latter represents the autocorrelation or noise spectral density. For the groups of processes described above, we find

$$2D_{SS}^{e} = \gamma_r n_0 S + \gamma_r n_e ,$$
$$2D_{SS}^{a} = \gamma_{r,a} n_{0,a} S + \gamma_{r,a} n_a ,$$
$$2D_{ee}^{o} = \gamma_p \left(n_0 - \frac{2n_e}{1 + \chi_g} \right) + (\gamma_t - \gamma_r) n_e ,$$
$$2D_{SS}^{\gamma} = \gamma S ,$$
$$2D_{aa}^{o} = (\gamma_{t,a} - \gamma_{r,a}) n_a ,$$

where n_e, n_a and S are the averaged (not stochastic) values. This way, the diffusion matrix becomes:

$$\sigma(u, t) =$$
$$\begin{bmatrix} \sqrt{2D_{SS}^{e}} & \sqrt{2D_{SS}^{a}} & & \sqrt{2D_{SS}^{\gamma}} \\ -\sqrt{2D_{SS}^{e}} & & \sqrt{2D_{ee}^{o}} & \\ & -\sqrt{2D_{SS}^{a}} & & \sqrt{2D_{aa}^{o}} \end{bmatrix} .$$

The equations are integrated using the *SKSROCK* solver from the *DifferentialEquations.jl* library [56] in the Julia programming language [57].

Simplified gain equation

Here, we derive the Eq. (5). Consider a network of PSNs. A derivative of the gain of a PSN is given in Eq. (3):

$$\dot{G} = 2\gamma_r \dot{n}_e + 2\gamma_{r,a} \dot{n}_a .$$

Assume the PSN is not currently emitting a spike, i.e. its gain does a random walk shown in Fig. 2c, in which case, $S \approx 0$. Then, substitute the derivatives of electron populations from Eq. (3):

$$\dot{G} \approx -2\gamma_r \gamma_t n_e + 2\gamma_r (\gamma_p + \Delta\gamma_p(t))(n_0 - n_e) -$$
$$- 2\gamma_{r,a} \gamma_{t,a} n_a + 2\gamma_r F_e(t) + 2\gamma_{r,a} F_a(t) =$$
$$= -2\gamma_r n_e (\gamma_t + \gamma_p) + 2\gamma_r \gamma_p n_0 - 2\gamma_{r,a} \gamma_{t,a} n_a +$$
$$+ 2\gamma_r \Delta\gamma_p(t)(n_0 - n_e) + 2\gamma_r F_e(t) + 2\gamma_{r,a} F_a(t) .$$

where $\Delta\gamma_p(t)$ is given in Eq. (4). Then, replace $2\gamma_r n_e = G + \gamma - \gamma_{r,a}(2n_a - n_{0,a}) + \gamma_r n_0$:

$$\dot{G} \approx -(\gamma_t + \gamma_p)(G + \gamma - \gamma_{r,a}(2n_a - n_{0,a}) + \gamma_r n_0) -$$
$$+ 2\gamma_r \Delta\gamma_p(t)(n_0 - n_e) + 2\gamma_r \gamma_p n_0 - 2\gamma_{r,a} \gamma_{t,a} n_a +$$
$$+ 2\gamma_r F_e(t) + 2\gamma_{r,a} F_a(t) .$$

Rearranging the terms we find

$$\dot{G} \approx -(\gamma_t + \gamma_p)G + 2\gamma_{r,a}(\gamma_t + \gamma_p - \gamma_{t,a})n_a -$$
$$- (\gamma_t + \gamma_p)(\gamma + \gamma_{r,a} n_{0,a} + \gamma_r n_0) + 2\gamma_r \gamma_p n_0 -$$
$$+ 2\gamma_r(n_0 - n_e)\Delta\gamma_p(t) + 2\gamma_r F_e(t) + 2\gamma_{r,a} F_a(t) .$$

The second term on the first line is negligible compared to the first, as during the random walk $n_a \ll n_e$. The terms on the second line can be considered a drift term:

$$G_p = \gamma + \gamma_{r,a} n_{0,a} + \gamma_r n_0 + 2\tau_G \gamma_r \gamma_p n_0 .$$

where $\tau_G = 1/(\gamma_t + \gamma_p)$. This way, the dynamical equation becomes

$$dG \approx [-(G - G_p)/\tau_G + 2\gamma_r(n_0 - n_e)\Delta\gamma_p(t)] dt + \sigma_G dW .$$

Here, the interaction term can be nonlinear as n_e changes due to incoming spikes; we assume that such interaction is weak. Moreover, during the walk, the gain section is close to transparency, i.e. $n_e \approx n_0/2$. Finally, we find

$$dG \approx [-(G - G_p)/\tau_G + 2\gamma_r n_0 \Delta\gamma_p(t)] dt + \sigma_G dW .$$

Discrete nanolaser model

In this work, we assumed a continuous model, i.e. S, n_e and n_a are continuous variables and stochastic processes are approximated by Langevin forces. However, the particles are discrete, and so are the processes. In a typical laser, the number of photons and electron-hole pairs in lasers is large, and such a model is a good approximation. However, for nanolasers this can be disputed, given their small volume and the operation regime close to the threshold. Therefore, we consider a discrete model based on [37, 58].

We consider all processes separately – 10 total – as stochastic with rates: $\gamma_{r/r,a} S n_{e/a}$ for stimulated emission in the gain section / SA, $\gamma_{r/r,a} S(n_{0/0,a} - n_{e/a})$ for photon absorption in the gain section / SA, γS for optical loss, $\gamma_{r/r,a} n_{e/a}$ for spontaneous emission in the gain section / SA, $\gamma_{t/t,a} n_{e/a}$ for nonradiative recombination and out-of-mode spontaneous emission in the gain section / SA, $\gamma_p(n_0 - 2n_e/(\chi_g + 1))$ for pumping. When an event happens, a single particle is added or removed from appropriate variables. For example, for stimulated emission in the SA, $S \to S + 1$ and $n_a \to n_a - 1$. Such a simulation is considerably more computationally demanding, but is rigorous and more accurate for this system.

The simulation was carried out using the *SSAStepper* solver from the *DifferentialEquations.jl* library [56]. Fig. 9 shows a comparison between the models. We find that they give very similar results in the operating regime of interest.

Probabilistic inference training

The approach used here follows [13]. A fully visible BM was trained to store three digits – "0", "3" and "4" – taken from the MNIST dataset and scaled down to 12×12. The intensity of each pixel, ranged from zero to unity, was rounded to 0.05 or 0.95. This way, we define the target statistics $\langle z_k \rangle_{tgt}$ and $\langle z_k z_j \rangle_{tgt}$ for the BM. We start with arbitrary weights w_{kj} and biases b_j and collect a sufficient number of samples to estimate $\langle z_k \rangle$ and $\langle z_k z_j \rangle$. Then, we refine the BM parameters using the following update rules:
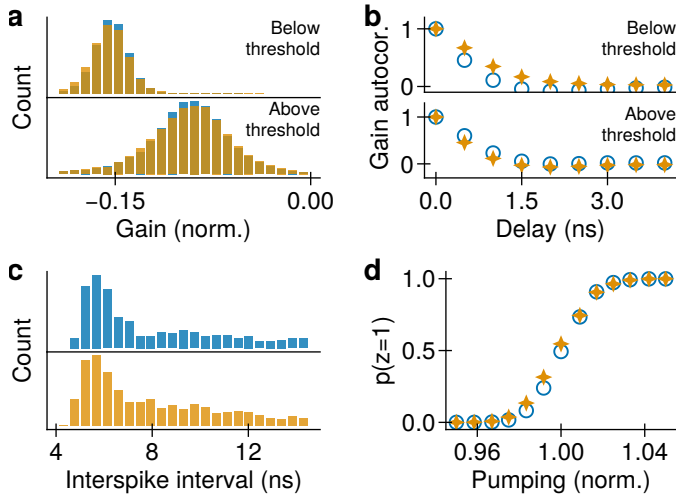
Figure 9: Comparison of the simplified continuous model used throughout the article (blue) and the rigorous discrete model (orange). **(a)** Distribution of gain outside the refractory period. Below and above the threshold correspond to normalized pumping of 0.96 and 1.0. **(b)** Same, autocorrelation of gain. **(c)** ISI distribution for normalized pumping of 1.0. **(d)** Activation curves assuming identical τ used in Fig. 4.

$\Delta b_k \propto \langle z_k \rangle_{\text{tgt}} - \langle z_k \rangle$ and $\Delta W_{kj} \propto \langle z_k z_j \rangle_{\text{tgt}} - \langle z_k z_j \rangle$. Then, sampling and refinement is repeated until a satisfactory result is achieved.

Probabilistic inference visualization

The projection procedure is identical to [13], but also takes into account the non-orthogonality of the basis vectors $\boldsymbol{B^0}$, $\boldsymbol{B^3}$ and $\boldsymbol{B^4}$:

$$\hat{M}_{\text{proj}} = \begin{pmatrix} \sin(\varphi_{\text{B}}^0) & \sin(\varphi_{\text{B}}^3) & \sin(\varphi_{\text{B}}^4) \\ \cos(\varphi_{\text{B}}^0) & \cos(\varphi_{\text{B}}^3) & \cos(\varphi_{\text{B}}^4) \end{pmatrix} \left(\hat{B}^\top \hat{B} \right)^{-1} ,$$

where $\hat{B} = (\boldsymbol{B^0}, \boldsymbol{B^3}, \boldsymbol{B^4})$.

Data availability

The parameters of the RBM used for sampling from an arbitrary distribution, the distribution itself and the BM used for probabilistic inference are available upon a reasonable request. The RBM used for the MNIST tasks is in the supporting dataset of [48]. The MNIST dataset is publicly available [46].

Code availability

Simulation code is available upon a reasonable request.

Acknowledgment

References

[1] J. Göltz et al. "Fast and energy-efficient neuromorphic deep learning with first-spike times". In: *Nature machine intelligence* 3.9 (2021), pp. 823–835.

[2] T. C. Wunderlich and C. Pehle. "Event-based backpropagation can compute exact gradients for spiking neural networks". In: *Scientific Reports* 11.1 (2021), p. 12829.

[3] A. Stanojevic et al. "An exact mapping from ReLU networks to spiking neural networks". In: *arXiv preprint arXiv:2212.12522* (2022).

[4] P. A. Merolla et al. "A million spiking-neuron integrated circuit with a scalable communication network and interface". In: *Science* 345.6197 (2014), pp. 668–673. DOI: 10.1126/science.1254642.

[5] S. B. Furber et al. "The spinnaker project". In: *Proceedings of the IEEE* 102.5 (2014), pp. 652–665.

[6] M. Davies et al. "Loihi: A neuromorphic manycore processor with on-chip learning". In: *Ieee Micro* 38.1 (2018), pp. 82–99.

[7] T. Pfeil et al. "Six networks on a universal neuromorphic computing substrate". In: *Frontiers in neuroscience* 7 (2013), p. 11.

[8] S. Moradi et al. "A scalable multicore architecture with heterogeneous memory structures for dynamic neuromorphic asynchronous processors (DYNAPs)". In: *IEEE transactions on biomedical circuits and systems* 12.1 (2017), pp. 106–122.

[9] A. Neckar et al. "Braindrop: A mixed-signal neuromorphic architecture with a dynamical systems-based programming model". In: *Proceedings of the IEEE* 107.1 (2018), pp. 144–164.

[10] S. Billaudelle et al. "Versatile emulation of spiking neural networks on an accelerated neuromorphic substrate". In: *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE. 2020, pp. 1–5.

[11] L. Buesing et al. "Neural dynamics as sampling: a model for stochastic computation in recurrent networks of spiking neurons". In: *PLoS computational biology* 7.11 (2011), e1002211. DOI: 10.1371/journal.pcbi.1002211.

[12] M. A. Petrovici et al. "Stochastic inference with deterministic spiking neurons". In: *arXiv preprint arXiv:1311.3211* (2013).

[13] M. A. Petrovici et al. "Stochastic inference with spiking neurons in the high-conductance state". In: *Physical Review E* 94.4 (2016), p. 042312. DOI: 10.1103/PhysRevE.94.042312.

[14] G. E. Hinton, S. Osindero, and Y.-W. Teh. "A fast learning algorithm for deep belief nets". In: *Neural computation* 18.7 (2006), pp. 1527–1554. DOI: 10.1162/neco.2006.18.7.1527.

[15] G. E. Hinton and R. R. Salakhutdinov. "Reducing the dimensionality of data with neural networks". In: *science* 313.5786 (2006), pp. 504–507.

[16] A.-r. Mohamed, G. E. Dahl, and G. Hinton. "Acoustic modeling using deep belief networks". In: *IEEE transactions on audio, speech, and language processing* 20.1 (2011), pp. 14–22.

[17] A. M. Abdel-Zaher and A. M. Eldeib. "Breast cancer classification using deep belief networks". In: *Expert Systems with Applications* 46 (2016), pp. 139–144.

[18] G. Carleo and M. Troyer. "Solving the quantum many-body problem with artificial neural networks". In: *Science* 355.6325 (2017), pp. 602–606.

[19] A. F. Kungl et al. "Accelerated physical emulation of bayesian inference in spiking neural networks". In: *Frontiers in neuroscience* 13 (2019), p. 1201. DOI: 10.3389/fnins.2019.01201.

[20] S. Czischek et al. "Spiking neuromorphic chip learns entangled quantum states". In: *SciPost Physics* 12.1 (2022), p. 039.

[21] R. Klassert et al. "Variational learning of quantum ground states on spiking neuromorphic hardware". In: *Iscience* 25.8 (2022).

[22] R. Ho, K. W. Mai, and M. A. Horowitz. "The future of wires". In: *Proceedings of the IEEE* 89.4 (2001), pp. 490–504. DOI: 10.1109/5.920580.

[23] H. Wünsche et al. "Excitability of a semiconductor laser by a two-mode homoclinic bifurcation". In: *Physical review letters* 88.2 (2001), p. 023901. DOI: 10.1103/PhysRevLett.88.023901.

[24] B. Romeira et al. "Excitability and optical pulse generation in semiconductor lasers driven by resonant tunneling diode photo-detectors". In: *Opt. Express* 21.18 (Sept. 2013), pp. 20931–20940. DOI: 10.1364/OE.21.020931.

[25] P. R. Prucnal et al. "Recent progress in semiconductor excitable lasers for photonic spike processing". In: *Advances in Optics and Photonics* 8.2 (2016), pp. 228–299. DOI: 10.1364/AOP.8.000228.

[26] J. Robertson et al. "Toward Neuromorphic Photonic Networks of Ultrafast Spiking Laser Neurons". In: *IEEE Journal of Selected Topics in Quantum Electronics* 26.1 (2020), pp. 1–15. DOI: 10.1109/JSTQE.2019.2931215.

[27] S. Barbay, R. Kuszelewicz, and A. M. Yacomotti. "Excitability in a semiconductor laser with saturable absorber". In: *Optics letters* 36.23 (2011), pp. 4476–4478. DOI: 10.1364/OL.36.004476.

[28] M. A. Nahmias et al. "A leaky integrate-and-fire laser neuron for ultrafast cognitive computing". In: *IEEE journal of selected topics in quantum electronics* 19.5 (2013), pp. 1–12. DOI: 10.1109/JSTQE.2013.2257700.

[29] G. Crosnier et al. "Hybrid indium phosphide-on-silicon nanolaser diode". In: *Nature Photonics* 11.5 (2017), pp. 297–300. DOI: 10.1038/nphoton.2017.56.

[30] M Delmulle et al. "Excitability in a PhC nanolaser with an integrated saturable absorber". In: *European Quantum Electronics Conference*. Optica Publishing Group. 2023, jsiii_6_2.

[31] A. L. Hodgkin and A. F. Huxley. "A quantitative description of membrane current and its application to conduction and excitation in nerve". In: *The Journal of physiology* 117.4 (1952), p. 500.

[32] M. Yamada. "A theoretical analysis of self-sustained pulsation phenomena in narrow-stripe semiconductor lasers". In: *IEEE Journal of Quantum Electronics* 29.5 (1993), pp. 1330–1336. DOI: 10.1109/3.236146.

[33] G. Björk, A. Karlsson, and Y. Yamamoto. "On the linewidth of microcavity lasers". In: *Applied physics letters* 60.3 (1992), pp. 304–306. DOI: 10.1063/1.106693.

[34] S. Wieczorek et al. "The dynamical complexity of optically injected semiconductor lasers". In: *Physics Reports* 416.1-2 (2005), pp. 1–128. DOI: 10.1016/j.physrep.2005.06.003.

[35] G. P. Agrawal. "Population pulsations and nondegenerate four-wave mixing in semiconductor lasers and amplifiers". In: *JOSA B* 5.1 (1988), pp. 147–159. DOI: 10.1364/JOSAB.5.000147.

[36] A. Moelbjerg et al. "Dynamical Properties of Nanolasers Based on Few Discrete Emitters". In: *IEEE Journal of Quantum Electronics* 49.11 (2013), pp. 945–954. DOI: 10.1109/JQE.2013.2282464.

[37] J Mork and G. Lippi. "Rate equation description of quantum noise in nanolasers with few emitters". In: *Applied Physics Letters* 112.14 (2018), p. 141103. DOI: 10.1063/1.5022958.

[38] L. A. Coldren, S. W. Corzine, and M. L. Masanovic. *Diode Laser and Photonic Integrated Circuits*. Wiley, 2012. DOI: 10.1002/9781118148167.

[39] G. L. Lippi, J. Mørk, and G. P. Puccioni. "Numerical solutions to the Laser Rate Equations with noise: technical issues, implementation and pitfalls". In: *Nanophotonics VII*. Vol. 10672. SPIE. 2018, pp. 82–95. DOI: 10.1117/12.2305948.

[40] E. C. André, J. Mørk, and M. Wubs. "Efficient stochastic simulation of rate equations and photon statistics of nanolasers". In: *Optics Express* 28.22 (2020), pp. 32632–32646. DOI: 10.1364/OE.405979.

[41] D. Elvira et al. "Higher-order photon correlations in pulsed photonic crystal nanolasers". In: *Phys. Rev. A* 84 (6 2011), p. 061802. DOI: 10.1103/PhysRevA.84.061802.

[42] S. Ohno et al. "Si microring resonator crossbar array for on-chip inference and training of the optical neural network". In: *Acs Photonics* 9.8 (2022), pp. 2614–2622. DOI: 10.1021/acsphotonics.1c01777.

[43] M. A. Nahmias et al. "A laser spiking neuron in a photonic integrated circuit". In: *arXiv preprint arXiv:2012.08516* (2020). DOI: 10.48550/arXiv.2012.08516.

[44] N. Le Roux and Y. Bengio. "Representational power of restricted Boltzmann machines and deep belief networks". In: *Neural computation* 20.6 (2008), pp. 1631–1649. DOI: 10.1162/neco.2008.04-07-510.

[45] G. E. Hinton. "Training products of experts by minimizing contrastive divergence". In: *Neural computation* 14.8 (2002), pp. 1771–1800.

[46] Y. LeCun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. DOI: 10.1109/5.726791.

[47] L. Van der Maaten and G. Hinton. "Visualizing data using t-SNE". In: *Journal of machine learning research* 9.11 (2008).

[48] A. Korcsak-Gorzo et al. "Cortical oscillations support sampling-based computations in spiking neural networks". In: *PLoS computational biology* 18.3 (2022), e1009753. DOI: 10.1371/journal.pcbi.1009753.

[49] M. Davies et al. "Advancing neuromorphic computing with loihi: A survey of results and outlook". In: *Proceedings of the IEEE* 109.5 (2021), pp. 911–934.

[50] E. Dimopoulos et al. "Electrically-Driven Photonic Crystal Lasers with Ultra-low Threshold". In: *Laser & Photonics Reviews* 16.11 (2022), p. 2200109. DOI: https://doi.org/10.1002/lpor.202200109.

[51] A. S. Greenspon et al. "Scalable construction of hybrid quantum photonic cavities". In: *arXiv preprint arXiv:2410.03851* (2024).

[52] H. Zhou et al. "Photonic matrix multiplication lights up photonic accelerator and beyond". In: *Light: Science & Applications* 11.1 (2022), p. 30.

[53] T. J. Seok et al. "Wafer-scale silicon photonic switches beyond die size limit". In: *Optica* 6.4 (2019), pp. 490–494.

[54] J. Moughames et al. "Three-dimensional waveguide interconnects for scalable integration of photonic neural networks". In: *Optica* 7.6 (2020), pp. 640–646.

[55] D. E. McCumber. "Intensity Fluctuations in the Output of cw Laser Oscillators. I". In: *Phys. Rev.* 141 (1 Jan. 1966), pp. 306–322. DOI: 10.1103/PhysRev.141.306.

[56] C. Rackauckas and Q. Nie. "DifferentialEquations.jl – a performant and feature-rich ecosystem for solving differential equations in Julia". In: *Journal of Open Research Software* 5.1 (2017), p. 15. DOI: 10.5334/jors.151.

[57] J. Bezanson et al. "Julia: A fresh approach to numerical computing". In: *SIAM review* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671.

[58] G. Puccioni and G. Lippi. "Stochastic Simulator for modeling the transition to lasing". In: *Optics Express* 23.3 (2015), pp. 2369–2374.