Photonic probabilistic machine learning using quantum vacuum noise

Seou Choi [1] ✉, Yannick Salamin [1,2], Charles Roques-Carmes [1,3] ✉, Rumen Dangovski [1,4], Di Luo [4,5,6], Zhuo Chen [2,4], Michael Horodynski [2], Jamison Sloan[1], Shiekh Zia Uddin [1,2] & Marin Soljačić [1,2]

Probabilistic machine learning utilizes controllable sources of randomness to encode uncertainty and enable statistical modeling. Harnessing the pure randomness of quantum vacuum noise, which stems from fluctuating electromagnetic fields, has shown promise for high speed and energy-efficient stochastic photonic elements. Nevertheless, photonic computing hardware which can control these stochastic elements to program probabilistic machine learning algorithms has been limited. Here, we implement a photonic probabilistic computer consisting of a controllable stochastic photonic element – a photonic probabilistic neuron (PPN). Our PPN is implemented in a bistable optical parametric oscillator (OPO) with vacuum-level injected bias fields. We then program a measurement-and-feedback loop for time-multiplexed PPNs with electronic processors (FPGA or GPU) to solve certain probabilistic machine learning tasks. We showcase probabilistic inference and image generation of MNIST-handwritten digits, which are representative examples of discriminative and generative models. In both implementations, quantum vacuum noise is used as a random seed to encode classification uncertainty or probabilistic generation of samples. In addition, we propose a path towards an all-optical probabilistic computing platform, with an estimated sampling rate of ~1 Gbps and energy consumption of ~5 fJ/MAC. Our work paves the way for scalable, ultrafast, and energy-efficient probabilistic machine learning hardware.

Probabilistic machine learning can accelerate image generation[1,2], heuristic optimization[3,4], and probabilistic inference[5,6] by leveraging stochasticity to encode uncertainty and enable statistical modeling[7,8]. These approaches are well suited for real-life applications which must account for uncertainty and variability, including autonomous driving[9], medical diagnosis[10], and drug discovery[11]. However, digital complementary metal-oxide-semiconductor (CMOS) technology requires extensive resource overhead to simulate randomness and control probabilities, which leads to significantly increased power consumption and decreased operational speed[12]. These challenges have sparked recent proposals for beyond-CMOS hardware such as low-barrier magnetic tunnel junctions[13] and diffusive memristors[14]—both of which leverage intrinsic noise as a source of randomness.

Concurrently, optical neural networks (ONNs)[15,16] have shown remarkable progress in energy efficiency[17,18], speed[19], and bandwidth[20] for solving deterministic tasks such as image classification[21] and

[1]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]E. L. Ginzton Laboratories, Stanford University, Stanford, CA, USA. [4]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Cambridge, MA, USA. [5]Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Department of Physics, Harvard University, Cambridge, MA, USA. ✉e-mail: seouc130@mit.edu; chrc@stanford.edu
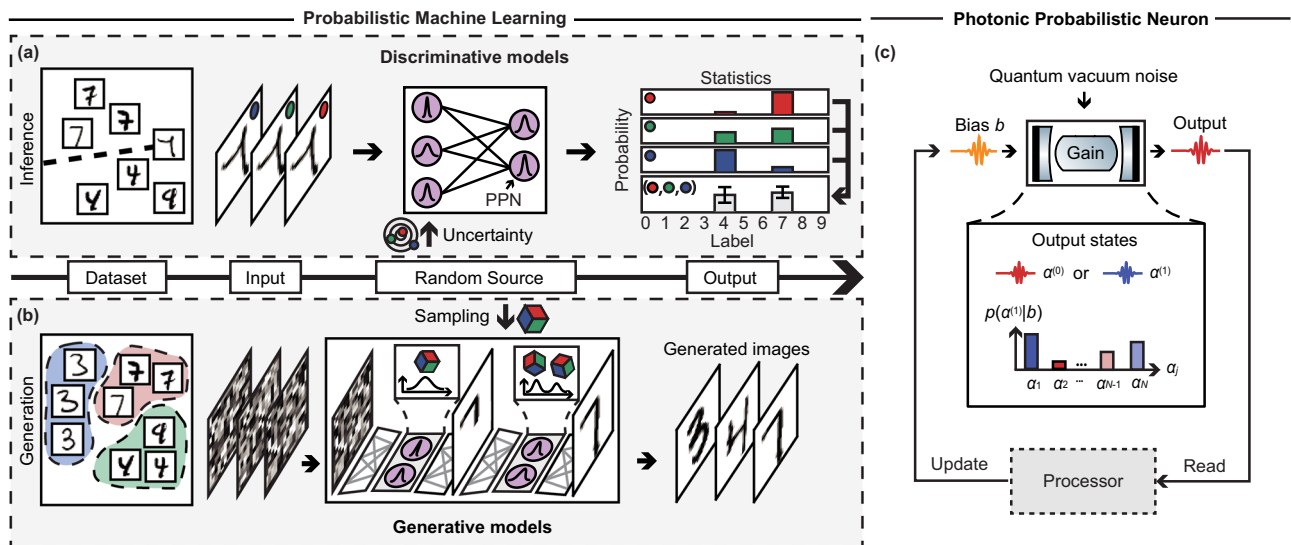
Fig. 1 | Probabilistic machine learning with stochastic photonic elements.
Probabilistic machine learning enabled by physical random sources, solving **a** inference and **b** generation tasks. Neural networks learn a decision line for inference tasks and overall distribution for generation tasks. **a** Random sources encode uncertainty in neural network parameters, allowing statistical interpretation on inference results. **b** Stochastic image generation seeded by random sources samples new images from certain probability distributions stored in neural networks. Both computational tasks require controllable stochastic photonic elements which can learn probability distribution and perform statistically independent sampling, which we refer to as photonic probabilistic neurons (PPNs). **c** Schematic of PPNs. One of the output states ($\alpha^{(0)}$ or $\alpha^{(1)}$) of a multistable optical system is randomly selected from a certain state probability distribution $p(\alpha^{(1)}|b)$ controlled by a bias field level b. Subsequently, a processing unit reads the output state and updates the bias value b for the next sampling. N independent outcomes can be sampled from different probabilities by time-multiplexing the bias signal. Optical elements in (**c**), originally published by GWoptics; released under a Creative Commons Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0). MNIST images in (**a** and **b**), originally published by LeCun, et al.[37]; released under a Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0).

speech recognition[22]. An important feature of ONNs is the inherent presence of noise in their operation. Therefore, photonic computing hardware typically implements computational tasks that are robust to optical noise[16]. ONNs have also been explored in regimes where deterministic tasks are performed with high accuracy, despite the presence of high levels of inherent noise[18]. Conversely, ONNs in which optoelectronic noise is intentionally added have also been proposed for optimization[23] and generative networks[24]. Interestingly, quantum optics offers a natural source of randomness in the ground state of electromagnetic field, known as quantum vacuum noise[25–27]. This intrinsic noise source is ubiquitous in optics and has been used to achieve high-data rate random number generation[28,29]. In addition, optical systems influenced by quantum vacuum noise have shown natural abilities to generate probability distributions[30–32], which are of strong interest for computing applications[13,14]. However, the experimental demonstration of a photonic probabilistic machine learning system has remained elusive so far, mostly due to the lack of *programmable* stochastic photonic elements.

Here, we experimentally demonstrate a probabilistic computing platform utilizing photonic probabilistic neurons (PPNs). Our PPN is implemented as a biased degenerate optical parametric oscillator (OPO), which leverages quantum vacuum noise to generate a probability distribution encoded by a bias field. We realized a hybrid optoelectronic probabilistic machine learning system which combines time-multiplexed PPNs and electronic processors with algorithm-specific measurement-and-feedback strategies. We demonstrate probabilistic inference of MNIST-handwritten digits with a stochastic binary neural network (SBNN), highlighting how quantum vacuum noise can encode classification uncertainty in discriminative models. Additionally, we showcase the generation of MNIST-handwritten digits with a pixel convolutional neural network (pixelCNN), demonstrating how statistical sampling in generative models can be facilitated by quantum vacuum noise. Furthermore, we provide a thorough discussion of the potential of an all-optical probabilistic machine learning

system, offering a possible performance enhancement by a factor of 100 in both speed and energy over traditional CMOS implementations, thereby opening new avenues in high-speed, energy-efficient computing applications.

Results

Probabilistic computing with time-multiplexed PPNs
We first provide a brief overview of two probabilistic machine learning models and their optical implementation with PPNs (Fig. 1).

Discriminative models learn decision lines that encode classification boundaries between different images (Fig. 1a, left)[33]. Probabilistic neural networks (Fig. 1a, middle) then impart statistical properties onto network parameters (e.g., weight uncertainty[5] or layer nodes[34]). Therefore, the network can provide a statistical ensemble of classification results, which are shown as different probabilities of the image classified to certain labels (Fig. 1a, right). Probabilistic inference can quantify classification uncertainty, which becomes critical for ambiguous images located near the decision boundary[35,36].

On the other hand, generative models learn the underlying probability distribution of the training dataset (e.g., images) in order to create new ones (Fig. 1b, left)[33]. When generating new images, generative models use random sources to seed stochastic image sampling based on the probability distribution learned by the network (Fig. 1b, middle). As a result, images with different labels can be generated (Fig. 1b, right).

In both of these computational tasks, probabilistic machine learning requires stochastic photonic elements whose probability distribution can be tuned, and that can perform statistically independent sampling. We refer to the optical implementation of this capability as PPNs (purple circles in Fig. 1a, b).

The proposed PPN is depicted in Fig. 1c. The building block consists of a synchronously pumped degenerate OPO[30]. An OPO consists of a nonlinear medium (e.g., second-order nonlinear crystal, down converting photon frequency) and an optical cavity surrounding it.
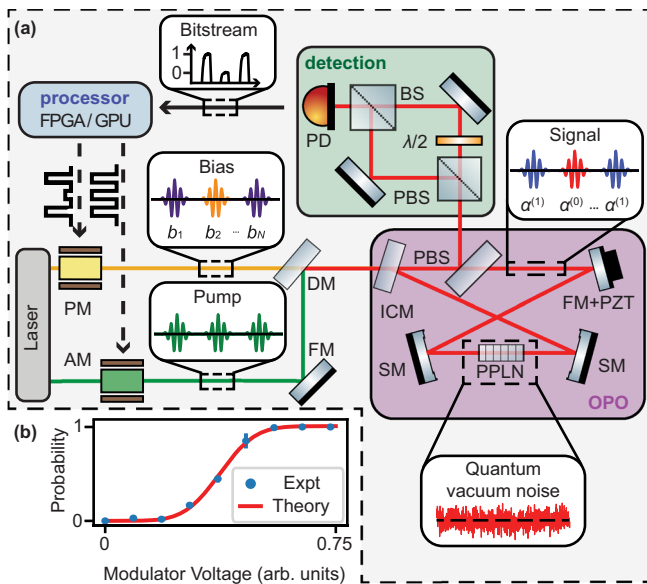
Fig. 2 | Experimental demonstration of a photonic probabilistic computer.
a Experimental setup, consisting of an ultrafast laser pumping a nonlinear cavity, and homodyne detection to measure the phase of the OPO signal. Electronic processing units (FPGA/GPU) generate electrical signals to tune the probability. AM amplitude modulator, PM phase modulator, FM flat mirror, DM dichroic mirror, ICM in-coupling mirror, (P)BS (polarization) beam-splitter, SM spherical mirror, PPLN periodically poled lithium niobate nonlinear crystal. PZT piezoelectric actuator, $\lambda/2$ half waveplate, PD photodiode. **b** Modulator voltage–probability relationship. Error bar represents the standard deviation. Optical or electronic elements in (**a**), originally published by GWoptics; released under a Creative Commons Attribution-NonCommercial 3.0 Unported (CC BY-NC 3.0).

The phase of the initial optical field is random due to electromagnetic field fluctuations inside the cavity (quantum vacuum noise). When the power of the pump laser exceeds a certain threshold power, phase-sensitive gain of the OPO allows the initial state to fall into one of the bistable output states with either phase (0 rad, or π rad)[28]. In other words, quantum vacuum noise acts as a perfect random source that manifests itself in the output phase. In fact, this random source is an intrinsic noise source ubiquitous in quantum optics[25–27]. When a vacuum-level external bias field b is introduced in the OPO cavity, the probability distribution of the output steady states can be coherently controlled[30]. Specifically, our OPO-based PPN encodes a Bernoulli trial $B(p)$ with binary outcomes having probability p and $1 - p$. Independent random sampling and processing can be realized by time-multiplexing the bias signal, resulting in N independent outcomes with encoded probabilities as depicted by different heights in Fig. 1c.

The experimental system realizing the PPN, and its implementation into a probabilistic computing system, is shown in Fig. 2. The system consists of three modules: biased OPO (purple area), detection (green area), and processing unit (blue area). We time-multiplex OPO signals with an amplitude modulator along the pump path to sample multiple binary outputs from a single optical cavity at a rate of 10 kHz. This bit rate is chosen to ensure the statistical independence of each PPN[28,30]. We use a homodyne detector to measure the optical phase of the steady state and map it to the corresponding bit value (i.e., 0 rad → 0 and π rad → 1).

During each cycle, a bit is measured by the homodyne detector (value 0 or 1), conditioned on the bias value b. This bit, or a collection of bit values ("bitstream"), is then fed into an electronic processing unit to update the bias field value and sample the PPN in the next cycle. In our experiment, the processing unit is taken as either a field-programmable gated array (FPGA) or a graphics processing unit (GPU). The FPGA is more adapted for real-time bitstream processing

and control of the optical system, while the GPU can accelerate complex machine learning algorithms such as image generation at the cost of a slower system control.

Individual p_i values are encoded in the phase of the bias field b_i by applying a calibrated square-wave voltage to a phase modulator in the bias line path. The voltage–probability relation provided by the phase modulator is shown in Fig. 2b. This relation is used in the following computing experiments to control the bias voltage. A detailed description of the experimental setup is discussed in Supplementary Note 1.

Photonic probabilistic computer for image classification
We now perform probabilistic image classification of MNIST-handwritten digits[37] using a pre-trained SBNN model on our optical probabilistic computing platform (Fig. 3a). SBNN encodes inference uncertainty by substituting deterministic layer nodes (as found in conventional fully connected neural networks) with stochastic binary nodes[38]. In a conventional, fully connected neural network, the jth node value in the $(n + 1)$th layer $X_{j,n+1}$ can be calculated in two steps: (1) matrix–vector multiplication (MVM) between weight matrix W and the nth layer X_n ($z_{j,n} \equiv \sum_i W_{j,i} X_{i,n}$); followed by (2) a nonlinear activation function $\sigma(\cdot): X_{j,n+1} = \sigma(z_{j,n})$.

Within our SBNN model, each layer node is represented by a PPN, and a single layer (yellow areas in Fig. 3a) is described as a bitstream of time-multiplexed PPNs. Because of the nonlinear nature of the bias-probability relationship (Fig. 2b), sampling a binary output $X_{j,n}$ with our PPN from the given bias $b_{j,n}$ (or equivalently bias modulator voltage $V_{j,n}$ in our experiment), naturally corresponds to passing a nonlinear activation function: $X_{j,n} = B(p_{j,n}) = B[\sigma(V_{j,n})]$. Modulator voltage $V_{j,n}$ is calculated via MVM between the weight W_{n-1} and the $(n - 1)$th layer X_{n-1} (gray areas in Fig. 3a, which is performed by the FPGA in our experiment). In other words, each PPN node binarizes the input, which consists of a weighted sum of previous layer nodes, with probability $p_{j,n}$. Because of the stochastic nature of the nodes, their probabilities change for every inference, leading to a probabilistic interpretation of classification results for an identical input image (Fig. 3a, right).

To perform image classification of MNIST-handwritten digits with our optical SBNN, we first binarize original MNIST-handwritten digits (Fig. 3a, left). Original MNIST-handwritten digits (grayscale, pixel values ranging from 0 to 255) are normalized between 0 and 1. The resulting pixel values represent the probability value for each PPN. The grayscale images are binarized by sampling the PPNs. The binary images are propagated through the network (784 → 128 → 64 → 10), with real-time communication between PPNs and the FPGA. The output layer $O_{0,1,...,9}$ is used to interpret the classification result, higher O_j corresponding to the higher probability of image representing digit "j". The network is pre-trained in silico and the weights are implemented on the FPGA. A detailed description of the training process and how the FPGA communicates with the optical setup is in Supplementary Note 2.

To test the performance of our optical SBNN, a batch of grayscale MNIST-handwritten digits (100 images) from the test set is selected. By binarizing each grayscale MNIST-handwritten digit 10 times to encode statistical uncertainty, we prepared 1000 binarized MNIST-handwritten digits in total to be classified by our optical SBNN. While propagating to the output layer, PPNs in the input and hidden layers encode the uncertainty by stochastically sampling the binary values from given probabilities. Once the output layer is reached, we can collect the statistics from 10 different inference results for each input image. Confusion matrices in Fig. 3b show that the overall experimental classification accuracy (96.5%) is in close agreement with the accuracy obtained from the numerical simulations for the single batch (97.0%) and total test images (98.3%) (see Supplementary Note 2). The classification accuracy of our photonic probabilistic computing hardware is also comparable with that of other optical computing platforms reaching more than 95%[21,39,40].
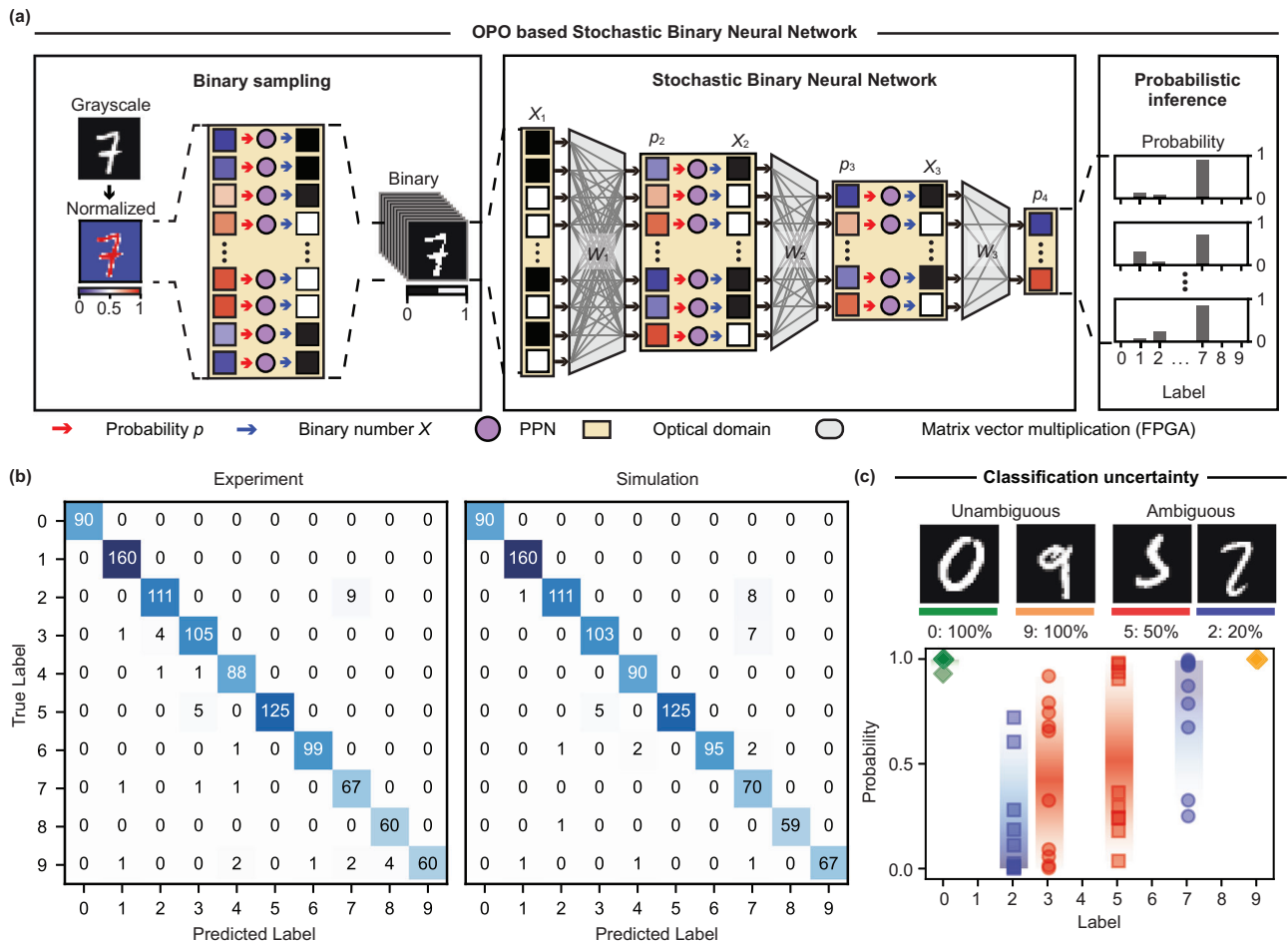
Fig. 3 | Probabilistic inference uncertainty encoded by quantum vacuum noise.
a Hybrid photonic-electronic architecture for stochastic binary neural networks (SBNNs). Original MNIST grayscale handwritten digit is binarized 10 times with PPNs and each binarized image propagates through SBNN (left panel). Binary nodes are sampled by PPNs and their corresponding p values are evaluated by FPGA (middle panel). Because the nodes are stochastic, inference results vary (right panel). **b** Confusion matrices of image classification results. A total of 1000 binary images (100 grayscale testing images × 10 times of binarization = 1000 input images) are tested. **c** Diagnosing inference results with the aid of quantum vacuum noise. Breadth in probability and low classification accuracy reflect the ambiguity of the input image. MNIST images in (**a** and **c**), originally published by LeCun, et al.[37]; released under a Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0).

Figure 3c shows how our probabilistic neural network can diagnose the reliability of inference results by harnessing quantum vacuum noise. Unlike deterministic neural networks, the variability of layer nodes in SBNNs results in different probability for each inference. One of the factors that can potentially degrade the classification performance is the ambiguity of the image (i.e., how close the image is to the decision boundary, as shown in Fig. 1a). By encoding uncertainty during inference, our photonic probabilistic computing hardware suggests all possible labels that ambiguous images can be classified. We choose two ambiguous images and two unambiguous images from the test dataset and plot the probability of each binarized grayscale MNIST-handwritten digit being classified under a certain label. Because we binarized 10 images each, 10 different probability values are shown for each label.

Three different scenarios are described in Fig. 3c. Unambiguous images such as "0" and "9" (achieving 100% of classification accuracy) show relatively consistent classification results with probabilities of correct classification close to 1. In this scenario, probabilistic neural networks show similar behavior to deterministic neural networks, which always give the same classification result with a fixed probability. When the input image becomes ambiguous (image "5" underlined in red, achieving 50% of classification accuracy), our SBNN model indicates that the image can be either "3" or "5". Accordingly, the

distribution of probabilities on each label broadens with its average value close to 50%. The worst case scenario is depicted by image "2" (underlined in blue), showing low overall accuracy (20%) and strong inconsistency in classification results. Such scenario clearly showcases how probabilistic sampling can provide additional information to the end-user. Classification results for labels that are not included in Fig. 3c can be found in Supplementary Note 2.

Offering both overall accuracy and statistics of classification results, probabilistic neural networks can diagnose inference results by providing a confidence level of the decision. The total classification result for each input image can be found in Supplementary Note 2.

Generating images from quantum vacuum noise with photonic generative models

We now turn to the demonstration of generative models with our photonic probabilistic computing platform (Fig. 4), demonstrating the use of quantum optical randomness as a source for generative machine learning models. We use a type of autoregressive model (pixelCNN), which models a conditional probability of a current pixel value from previous pixels[41].

Our implementation protocol for pixelCNN with PPNs is described in Fig. 4a. A binary image with the first $N-1$ pixels $X_{i \leq N-1}$
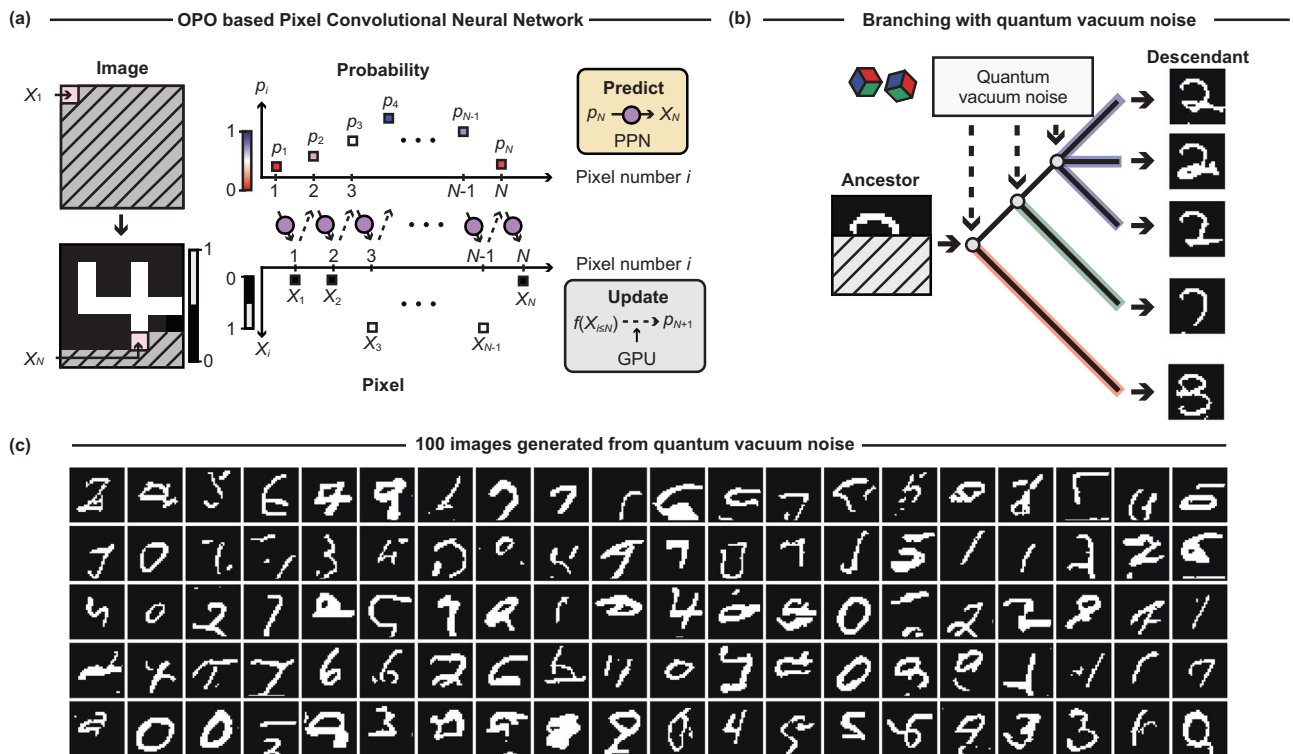
Fig. 4 | Images sampled from quantum vacuum noise. a PixelCNN generating binary MNIST-handwritten digits. PPNs sample the pixel value X_N from the given p_N value and GPU calculates the p_{N+1} value for the next pixel from the previous pixels $X_{i \leq N}$. **b** Branching off to different MNIST-handwritten digits, guided by quantum vacuum noise. Stochastic sampling allows generation of images with different digits and features. **c** Hundred generated images from pixelCNN, starting from a complete empty image. MNIST images in the left side of (**b**), originally published by LeCun, et al.[37]; released under a Creative Commons Attribution-ShareAlike 3.0 Unported (CC BY-SA 3.0).

specified is given as an input to the network. In principle, N can be any natural number, $N = 1$ corresponding to the case when pixelCNN creates an image only using quantum vacuum noise as a random seed. When the input image is given, a pre-trained pixelCNN model in the GPU evaluates p_N to be encoded on the PPN from previous pixels $X_{i \leq N-1}$, generating a binary number for the Nth pixel (X_N). The probability p_{N+1} is now computed based on previous pixel values $X_{i \leq N}$. This process is repeated until the full image is generated ($28 \times 28 = 784$ pixels). Our hybrid optoelectronic computing system can generate new images using quantum vacuum noise as a random seed. Details of network structure and training method can be found in Supplementary Note 3.

Different MNIST-handwritten digits, all generated from the same incomplete input image, highlight how quantum vacuum noise enables stochastic image sampling (Fig. 4b). Although they all start from the same "ancestor" image, the multiple stochastic samples of pixel values from the PPNs branch off into different MNIST-handwritten digits with different labels ("descendant" images). It is also possible to generate different images with the same label (which is likely to be labeled as "2").

We produced 100 examples of handwritten digit images from quantum vacuum noise using our photonic probabilistic computing platform (Fig. 4c). This was done by initiating an empty image as an input to our optical pixelCNN. We also test the negative-log-likelihood (NLL) of the generated images NLL $\equiv -\sum_i \{X_i \ln(p_i) + (1 - X_i) \ln(1 - p_i)\}$, where the sum runs over $i = 1, ..., 784$ pixel indices. A lower value of NLL indicates statistical similarity to the distribution of training images, yielding 71.1 ± 18.8 for our experimental results and 64.9 ± 15.4 for numerical simulations. This shows that our system has learned an accurate representation of the image distribution. Details of the performance of image generations can be found in Supplementary Note 3.

Discussion

In our demonstration of photonic probabilistic machine learning, the speed and energy efficiency were limited by the PPN sampling rate and data transfer bandwidth between electronic processors and PPNs. In the following, we propose an all-optical probabilistic computing platform which can overcome these challenges, and evaluate the potential benefit in terms of speed and energy efficiency compared to the electronic state of the art.

To increase sampling rate and reduce energy consumption, we propose an all-optical implementation. For instance, PPNs can be implemented with injection-seeded vertical-cavity surface-emitting lasers, reaching >1 Gbps[42] and providing energy-efficient operation[43]. Fast control of the probability and state detection can be achieved with high-bandwidth modulators and detectors[44–48], suggesting that PPNs achieving 1 Gbps sampling rate are within reach (detailed explanations can be found in Supplementary Note 4).

Furthermore, our programmable stochastic element naturally implements an all-optical nonlinearity through the bias-probability relationship, which has been a historical challenge in the implementation of energy-efficient all-optical ONNs[15]. Typically, ONNs rely on optoelectronic measurement-feedback schemes to update the network layers[39,49]. Conversely, in the proposed scheme, an optical signal (*vacuum-level bias*) controls the nonlinearity of the layer. Because the bias signal can be derived directly from the accumulated PPN outputs, bypassing active components, the scheme can reduce energy consumption per multiply-accumulate (MAC) operation to as low as ~5 fJ/MAC. State-of-the-art stochastic electronic devices, such as low-barrier magnetic tunnel junctions and diffusive memristors integrated with conventional CMOS technologies are expected to achieve ~0.1 Gbps[50,51] and consume ~900 fJ/MAC[38]. Comparatively, our proposed photonic platform can be ~ ×10 faster and ~ ×100 more energy

efficient. A detailed discussion of this all-optical probabilistic computing platform is found in Supplementary Note 4.

We now compare the speed and energy performance of our photonic platform to a state-of-the-art FPGA[52,53], in an image classification task considering a binary neural network. The deterministic FPGA implementation demonstrated a classification of ~1.6 million images per second with ~23 W power consumption. Adopting the network structure of our SBNN model in Fig. 3, we can calculate the computation time and the number of MAC operations required for each inference. Our estimation gives ~4 ns and ~10^5 MAC operations per classification, which result in ~250 million image classifications per second with a power consumption of ~0.1 W. Therefore, the suggested all-optical probabilistic computing hardware could perform ×100 faster while consuming ×100 less power. Detailed discussion can be found in Supplementary Note 4.

One of the possible extensions of our work is to train the network physically[54,55]. This becomes critical when an accurate digital modeling of the physical system becomes challenging due to its complexity. Without an additional cost of simulating randomness in digital models, several training methods which resort to stochasticity, including stochastic gradient descent[56], dropout[34], and noise injection[57] could potentially be realized with PPNs. Harnessing quantum vacuum noise in optical elements for both training and testing, our PPNs will pave the way of implementing all-optical probabilistic physical neural networks, which can benefit state-of-the-art machine learning applications including large language models[58] and diffusion models[59].

Our platform could also be used to implement other important computational tasks. The first one is alternative interpretable neural network models with trainable activation functions[60], which could be implemented with the PPN by taking advantage of its tunable bias-probability relationship. The second one is Ising model solvers with external magnetic fields, which can be modeled by the injection of a bias field in a network of OPOs[61].

Data availability

All data supporting this work are available within the manuscript, the Supplementary Information, the online repository: https://codeocean.com/capsule/4025993/tree. Raw data generated during the study are available once requested to the corresponding authors. Correspondence and requests should be addressed to S.C. (seouc130@mit.edu) and C.R.-C. (chrc@stanford.edu).

Code availability

The code used in this study is available at https://codeocean.com/capsule/4025993/tree.

References

1. Nichol, A. et al. Glide: towards photorealistic image generation and editing with text-guided diffusion models. Preprint at https://doi.org/10.48550/arXiv.2112.10741 (2021).
2. Goodfellow, I. et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems* **27** (NIPS, 2014).
3. Roques-Carmes, C. et al. Heuristic recurrent algorithms for photonic Ising machines. *Nat. Commun.* **11**, 249 (2020).
4. Pham, D. & Karaboga, D. *Intelligent Optimisation Techniques: Genetic Algorithms, Tabu Search, Simulated Annealing and Neural Networks* (Springer Science & Business Media, 2012).
5. Blundell, C., Cornebise, J., Kavukcuoglu, K. & Wierstra, D. Weight uncertainty in neural network. In *International Conference on Machine Learning* 1613–1622 (PMLR, 2015).
6. Neal, R. M. *Bayesian Learning for Neural Networks* Vol. 118 (Springer Science & Business Media, 2012).
7. Murphy, K. P. *Probabilistic Machine Learning: An Introduction* (MIT Press, 2022).
8. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521**, 452–459 (2015).
9. Feng, D., Harakeh, A., Waslander, S. L. & Dietmayer, K. A review and comparative study on probabilistic object detection in autonomous driving. *IEEE. Trans. Intell. Transp. Syst.* **23**, 9961–9980 (2021).
10. Richens, J. G., Lee, C. M. & Johri, S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.* **11**, 3923 (2020).
11. Jiménez-Luna, J., Grisoni, F. & Schneider, G. Drug discovery with explainable artificial intelligence. *Nat. Mach. Intell.* **2**, 573–584 (2020).
12. Qin, Y. et al. A high-speed true random number generator based on unified selector-RRAM. In *IEEE Electron Device Letters* (IEEE, 2023).
13. Chowdhury, S. et al. A full-stack view of probabilistic computing with p-bits: devices, architectures and algorithms. *IEEE J. Explor. Solid State Comput. Devices Circuits* **9**, 1–11 (2023).
14. Woo, K. S. et al. Probabilistic computing using $Cu_{0.1}Te_{0.9}/HfO_2/Pt$ diffusive memristors. *Nat. Commun.* **13**, 5762 (2022).
15. Wetzstein, G. et al. Inference in artificial intelligence with deep optics and photonics. *Nature* **588**, 39–47 (2020).
16. McMahon, P. L. The physics of optical computing. *Nat. Rev. Phys.* **5**, 717–734 (2023).
17. Wang, T. et al. An optical neural network using less than 1 photon per multiplication. *Nat. Commun.* **13**, 123 (2022).
18. Ma, S.-Y., Wang, T., Laydevant, J., Wright, L. G. & McMahon, P. L. Quantum-noise-limited optical neural networks operating at a few quanta per activation. Preprint at https://doi.org/10.48550/arXiv.2307.15712 (2023).
19. Mourgias-Alexandris, G. et al. Noise-resilient and high-speed deep learning with coherent silicon photonics. *Nat. Commun.* **13**, 5572 (2022).
20. Totovic, A., Giamougiannis, G., Tsakyridis, A., Lazovsky, D. & Pleros, N. Programmable photonic neural networks combining WDM with coherent linear optics. *Sci. Rep.* **12**, 5605 (2022).
21. Bernstein, L. et al. Single-shot optical neural network. *Sci. Adv.* **9**, eadg7904 (2023).
22. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
23. Prabhu, M. et al. Accelerating recurrent Ising machines in photonic integrated circuits. *Optica* **7**, 551–558 (2020).
24. Wu, C. et al. Harnessing optoelectronic noises in a photonic generative network. *Sci. Adv.* **8**, eabm2956 (2022).
25. Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.* **69**, 37 (1946).
26. Chan, H. B., Aksyuk, V. A., Kleiman, R. N., Bishop, D. J. & Capasso, F. Quantum mechanical actuation of microelectromechanical systems by the Casimir force. *Science* **291**, 1941–1944 (2001).
27. Sandoghdar, V., Sukenik, C., Hinds, E. & Haroche, S. Direct measurement of the van der Waals interaction between an atom and its images in a micron-sized cavity. *Phys. Rev. Lett.* **68**, 3432 (1992).
28. Marandi, A., Leindecker, N. C., Vodopyanov, K. L. & Byer, R. L. All-optical quantum random bit generation from intrinsically binary phase of parametric oscillators. *Opt. Express* **20**, 19322–19330 (2012).
29. Kim, K. et al. Massively parallel ultrafast random bit generation with a chip-scale laser. *Science* **371**, 948–952 (2021).
30. Roques-Carmes, C. et al. Biasing the quantum vacuum to control macroscopic probability distributions. *Science* **381**, 205–209 (2023).
31. Wu, C., Yang, X., Chen, Y. & Li, M. Photonic bayesian neural network using programmed optical noises. *IEEE J. Sel. Top. Quantum Electron.* **29**, 1–6 (2022).
32. Ma, B., Zhang, J., Li, X. & Zou, W. Stochastic photonic spiking neuron for Bayesian inference with unsupervised learning. *Opt. Lett.* **48**, 1411–1414 (2023).

33. Jebara, T. *Machine Learning: Discriminative and Generative.* (Springer Science & Business Media, 2012).

34. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).

35. Silipo, R. & Marchesi, C. Artificial neural networks for automatic ECG analysis. In *IEEE Transactions on Signal Processing* **46**, 1417–1425 (IEEE, 1998).

36. Leibig, C., Allken, V., Ayhan, M. S., Be rens, P. & Wahl, S. Leveraging uncertainty information from deep neural networks for disease detection. *Sci. Rep.* **7**, 17816 (2017).

37. LeCun, Y. et al. Gradient-based learning applied to document recognition. In *Proc. IEEE* **86**, 2278–2324 (IEEE, 1998).

38. Li, Y. et al. Binary-stochasticity-enabled highly efficient neuromorphic deep learning achieves better-than-software accuracy. *Adv. Intell. Syst.* **6**, 2300399 (2024).

39. Zhang, H. et al. An optical neural chip for implementing complex-valued neural network. *Nat. Commun.* **12**, 457 (2021).

40. Zuo, Y. et al. All-optical neural network with nonlinear activation functions. *Optica* **6**, 1132–1137 (2019).

41. Van Den Oord, A., Kalchbrenner, N. & Kavukcuoglu, K. Pixel recurrent neural networks. In *International Conference on Machine Learning* 1747–1756 (PMLR, 2016).

42. Zhao, J. et al. Fast all-optical random number generator. Preprint at https://doi.org/10.48550/arXiv.2201.07616 (2022).

43. Chen, Z. et al. Deep learning with coherent VCSEL neural networks. *Nat. Photonics* **17**, 723–730 (2023).

44. Valdez, F., Mere, V. & Mookherjea, S. 100 GHz bandwidth, 1 volt integrated electro-optic Mach-Zehnder modulator at near-IR wavelengths. *Optica* **10**, 578–584 (2023).

45. He, M. et al. High-performance hybrid silicon and lithium niobate Mach-Zehnder modulators for 100 Gbit s^{-1} and beyond. *Nat. Photonics* **13**, 359–364 (2019).

46. Wang, C. et al. Integrated lithium niobate electro-optic modulators operating at CMOS-compatible voltages. *Nature* **562**, 101–104 (2018).

47. Lischke, S. et al. High bandwidth, high responsivity waveguide-coupled germanium pin photodiode. *Opt. Express* **23**, 27213–27220 (2015).

48. Lischke, S. et al. Ultra-fast germanium photodiode with 3-dB bandwidth of 265 GHz. *Nat. Photonics* **15**, 925–931 (2021).

49. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).

50. Jiang, H. et al. A novel true random number generator based on a stochastic diffusive memristor. *Nat. Commun.* **8**, 882 (2017).

51. Chen, X. et al. Magnetic-tunnel-junction-based true random-number generator with enhanced generation rate. *Phys. Rev. Appl.* **18**, L021002 (2022).

52. Qin, H. et al. Binary neural networks: a survey. *Pattern Recognit.* **105**, 107281 (2020).

53. Umuroglu, Y. et al. Finn: a framework for fast, scalable binarized neural network inference. In *Proc. 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays* 65 (ACM, 2017).

54. Wright, L. G. et al. Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022).

55. Momeni, A., Rahmani, B., Mallejac, M., Del Hougne, P. & Fleury, R. Backpropagation-free training of deep physical neural networks. *Science* **382**, 1297–1303 (2023).

56. Bottou, L. Large-scale machine learning with stochastic gradient descent. In *Proc. COMPSTAT 2010* 177–186 (Springer, 2010).

57. Zur, R. M., Jiang, Y., Pesce, L. L. & Drukker, K. Noise injection for training artificial neural networks: a comparison with weight decay and early stopping. *Med. Phys.* **36**, 4810–4818 (2009).

58. Naveed, H. et al. A comprehensive overview of large language models. Preprint at https://doi.org/10.48550/arXiv.2307.06435 (2023).

59. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems* 33, 6840–6851 (2020).

60. Liu, Z. et al. Kan: Kolmogorov-arnold networks. Preprint at https://doi.org/10.48550/arXiv.2404.19756 (2024).

61. Horodynski, M. et al. Stochastic logic in biased coupled photonic probabilistic bits. Preprint at https://doi.org/10.48550/arXiv.2406.04000 (2024).

Acknowledgements

Author contributions

S.C., Y.S., C.R.-C., J.S., and M.S. conceived the original idea. R.D. and D.L. contributed to the development of machine learning algorithms. S.C. and Y.S. built the experimental setup with contributions from C.R.-C., M.H., and S.Z.U.; S.C. acquired and analyzed the data. S.C. developed the code for the electronic processing unit and trained the neural networks with contributions from R.D., D.L., and Z.C.; M.S. supervised the project. The manuscript was written by S.C., Y.S., and C.R.-C., with inputs from all authors.

Competing interests

Additional information