# Photonic Bayesian Neural Network Using Programmed Optical Noises

Changming Wu, Xiaoxuan Yang, Yiran Chen, and Mo Li

*(Invited Paper)*

***Abstract*—The Bayesian neural network (BNN) combines the strengths of neural networks and statistical modeling in that it simultaneously performs posterior predictions and quantifies the uncertainty of the predictions. Integrated photonics has emerged as a promising hardware platform of neural network accelerators capable of energy-efficient, low latency, and parallel computing. However, photonic neural networks demonstrated to date are mostly deterministic network models. Here, we extend the photonic neural network to a statistical model and proposed a photonic Bayesian neural network (P-BNN) architecture based on the integrated photonic platform and harnessing the inherent optical noises. The Bayesian neuron is realized by controlling the probability distribution of the signal-amplified spontaneous emission (signal-ASE) beat noise. We show the P-BNN's advantages in making predictions using the posterior distribution by simulating a p-BNN to perform handwritten number classification tasks. The simulation results show that the proposed P-BNN not only makes successful predictions on the expected images from the test dataset but also detects and rejects the unexpected images outside the training datasets. The P-BNN architecture is compatible with on-chip optical amplifiers and can be scaled up using current and emerging integrated photonics technologies, thus is promising for practical neural network applications.**

***Index Terms*—Integrated optics, neural network, optical computing.**

THE classical neural networks could universally approximate a continuous function deterministically while the statistical models make predictions and generate data by specifying a set of statistical assumptions between parameters. The Bayesian neural network (BNN) combines the strengths of neural networks and statistical modeling by adopting the statistical methodology and extending classical neural networks (such as deep neural networks (DNNs)) with posterior inference. Unlike classical DNNs, in the training phase, BNNs learn from the observations and update the distribution of their parameters. During the inference phase, BNNs generate a complete posterior distribution and produce probabilistic guarantees on the predictions [1]. It means that, in the parameter space, BNNs can provide statistical insights into the nature and distribution of the neural network's learned parameters. These features make the BNN highly attractive in applications such as medical diagnosis [2], [3], junk mail filtering [4], environment modeling [5], traffic flow forecasting [6], [7], and more [8], [9].

Integrated photonics provides a scalable analog computing hardware platform that has significant advantages in power efficiency, communication latency, and parallelism as compared to its electronic counterpart. These advantages specifically match the requirements of the neural network hardware accelerators, hence integrated photonic neural networks recently have become a focus of intense research and commercial interests [10], [11], [12], [13]. However, photonic neural networks reported to date are predominantly deterministic neural networks, such as DNNs and convolutional neural networks (CNNs) [14], [15], [16], [17], [43], [44], [45], [46], [47]. Stochastic networks, such as the Bayesian network, have not been explored in the photonic domain. Here, we propose an integrated photonic BNN (P-BNN) architecture, which consists of two independent integrated photonic accelerators and utilizes the signal-ASE beat noise in the optical amplifiers to construct a probabilistic Bayesian connection. The advantage of the P-BNN is demonstrated in simulation by performing number classification inferences using inputs of both the expected images from the training dataset and unexpected images out of the training datasets. The proposed P-BNN architecture extends current photonic neural networks models to a statistical BNN model and is compatible with the current integrated photonics technologies of large-scale photonic circuits with wavelength division multiplexing (WDM) and integrated semiconductor optical amplifiers [18], [19] (SOAs).

We build photonic Bayesian neurons with probabilistic connections using signal-ASE beat noise. As shown in Fig. 1(a), a conventional neuron in DNNs maps the input data $x$ to the output $y$ following the equation $y = f(w \cdot x + b)$, where weight $w$ and the bias $b$ are trainable but definite parameters, and $f$ is a pre-defined

Manuscript received 5 July 2022; revised 24 September 2022; accepted 24 October 2022. Date of publication 28 October 2022; date of current version 23 November 2022. The work of Changming Wu and Mo Li was supported in part by ONR MURI under Grant N00014-17-1-2661 and in part by National Science Foundation under Grant CCF-2105972. The work of Xiaoxuan Yang and Yiran Chen was supported in part by NSF under Grant 1955246 and Grant 1910299 and in part by ARO under Grant W911NF-19-2-0107. *(Corresponding authors: Changming Wu; Mo Li.)*

Changming Wu is with the Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195 USA (e-mail: cmwu@uw.edu).

Xiaoxuan Yang and Yiran Chen are with the Department of Electrical and Computer Engineering, Duke University, Durham, NC 27708 USA (e-mail: xiaoxuan.yang1@duke.edu; yiran.chen@duke.edu).

Mo Li is with the Department of Electrical and Computer Engineering, University of Washington, Seattle, WA 98195 USA, and also with the Department of Physics, University of Washington, Seattle, WA 98195 USA (e-mail: moli96@uw.edu).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/JSTQE.2022.3217819.

Digital Object Identifier 10.1109/JSTQE.2022.3217819

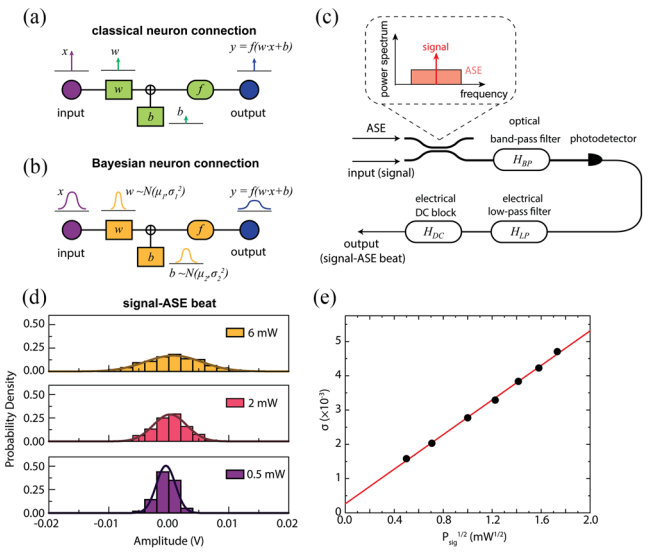



Fig. 1. (a) and (b) Comparison between a classical neuron connection in a DNN (a) and a Bayesian neuron connection in a BNN (b) Unlike the classical neuron maps the input value to the output value deterministically, the Bayesian neuron trims the input data distribution into the output distribution. The model distribution parameters $\mu$ and $\sigma$ are updated in a Bayesian neuron during training rather than the single weight value $w$. (c) Schematic of utilizing the signal-ASE beat noise to build an optical Bayesian neuron. The optically filtered signal and the ASE noise are mixed and then detected by photodetectors. After a DC block, the random electrical signals are sampled by an oscilloscope. (d) Histograms of signal-ASE beat noise distribution with a fixed time-averaging ASE power when the signal power is 0.5 mW, 2 mW and 6 mW respectively. The standard deviation $\sigma$ is increased with the increasing signal power. (e) The linear dependence between the $\sigma$ and the field intensity of the input signal mimics the Bayesian neuron connection.

nonlinear function. This mapping is deterministic in that once the network parameters are fixed, a neuron will produce identical output for the same input during every inference. In contrast, the statistical nature of a BNN requires each trainable parameter in a Bayesian neuron to have a probability distribution rather than a constant, as illustrated in Fig. 1. More description of the differences between the classical neural networks and BNNs is included in the Supplementary Information. A Bayesian neuron connection can also be described as $y = f(w \cdot x + b)$. However, the parameters $w$ and $b$ are not constants. Instead, without loss of generality, the weight $w \sim N(\mu_1, {\sigma_1}^2)$ and the bias $b \sim N(\mu_2, {\sigma_2}^2)$ are assumed to be probabilistic following specific statistical models, which we assume Gaussian distributions in this work. Therefore, the output $y$ of the Bayesian neuron is stochastic and varies each time. During training, both the mean $(\mu_1, \mu_2)$ and the standard deviation $(\sigma_1, \sigma_2)$ of the Gaussian distributions are optimized. The Bayesian neuron thus modifies the input data distribution to the output distribution [20].

The linear component of the Bayesian neuron connection, $w \cdot x + b$, will have a Gaussian probability distribution $N(\mu_1 \cdot x + \mu_2,\ {\sigma_1}^2 \cdot |x|^2 + {\sigma_2}^2)$. The mean value $N(\mu_1 \cdot x + \mu_2, 0)$ is programmed using classical photonic neurons [12], [21], [22]. We demonstrate programming the standard deviation in the distribution $N(0, {\sigma_1}^2 \cdot |x|^2 + {\sigma_2}^2)$ in a Bayesian neuron utilizing the "signal-ASE beat noise", which is a ubiquitous noise source in optical communication systems where optical amplifiers are needed [23]. In experiments, we generate the ASE noise from erbium-doped fiber amplifiers (EDFA) and combine it with the input signal (Fig. 1(c)). The broadband ASE noise and the monochromatic optical signal are optically filtered and beat at photodetectors, generating baseband random electrical photocurrents within the photodetector bandwidth. The DC photocurrent is filtered by a DC block with a cut-off frequency of 0.1 MHz, passing only the stochastic photocurrent variances and sampled by an oscilloscope. Fig. 1(d) plots the histograms of the measured signal-ASE beat probability distribution when the time-averaged ASE power is fixed (at 0.04 mW) and the signal power is 0.5 mW, 2mW, and 6 mW, respectively. The probability density function approximates a zero-mean Gaussian distribution well, with the standard deviation $\sigma$ proportionally increasing with the increasing signal power. Therefore, we show that this signal-ASE beat noise resembles the random distribution input needed in a Bayesian neuron connection. The standard deviation of the signal-ASE beat noise $\sigma_{\text{sig-ASE}}$ is theoretically given by [24]:

$$\sigma^2_{\text{sig-ASE}} = \frac{2R^2 \text{P}_{\text{sig}} \text{P}_{\text{ASE}}}{M},$$

where $R$ is the responsivity of the photodetector, $P_{\text{sig}}$ and $P_{\text{ASE}}$ are the optical signal power and the time-averaging ASE noise power, respectively, and the $M$ is a configuration constant $M = \frac{2B_O}{B_{LP}}$ determined by the optical filter bandwidth $B_O$ and the photodetector bandwidth $B_{LP}$. The resemblance between the Bayesian neuron connection and the signal-ASE beat noise allows us to encode the standard deviation $\sigma_1$ of the weight parameter $w$ by tuning the time-averaging power of the ASE noise power $P_{\text{ASE}}$ and encoding the input $x$ to the $P_{sig}{}^{1/2}$. The standard deviation $\sigma_2$ of bias parameter $b$ is added to the neuron by combining the ASE-signal random noise with an additional optical random number generator [25]. The linear dependence between the $\sigma_{\text{sig-ASE}}$ and the $P_{\text{sig}}{}^{1/2}$ agrees with the theory (see Fig. 1(e)). The small intercept at zero $P_{\text{sig}}$ shown in Fig. 1(e) comes from other noise origins such as the ASE-ASE beat noise and the shot noise [24], [26], [27]. It is neglected in later discussion since it can be easily compensated in optical neural network implementation by offsetting the input (see Supplementary Information for detailed description).

The proposed P-BNN architecture consists of an input layer, an output layer, and multiple hidden layers in between, as in Fig. 2(a). The input received by the $m_{\text{th}}$ Bayesian neuron in the $j_{\text{th}}$ layer is calculated by applying the nonlinear function on a weighted sum of the input signals from neurons in the previous layer:

$$y^j_m = f\left(\sum_n w^{j-1}_{mn} x^{j-1}_n + b^{j-1}_m\right),$$

where $x^{j-1}_n$ is the input to the $n_{\text{th}}$ neuron in the previous layer, $w^{j-1}_{mn}$ is the weight connecting the $m_{\text{th}}$ neuron in the $j_{\text{th}}$ layer and the $n_{\text{th}}$ neuron in the previous layer, $b^{j-1}_m$ is the bias. They both have the Gaussian distributions:

$$w^{j-1}_{mn} \sim N\left(\mu^{j-1}_{mn},\ {\sigma^{j-1}_{mn}}^2\right),$$

$$b^{j-1}_n \sim N\left(\mu^{j-1}_{b,n},\ {\sigma^{j-1}_{b,n}}^2\right).$$

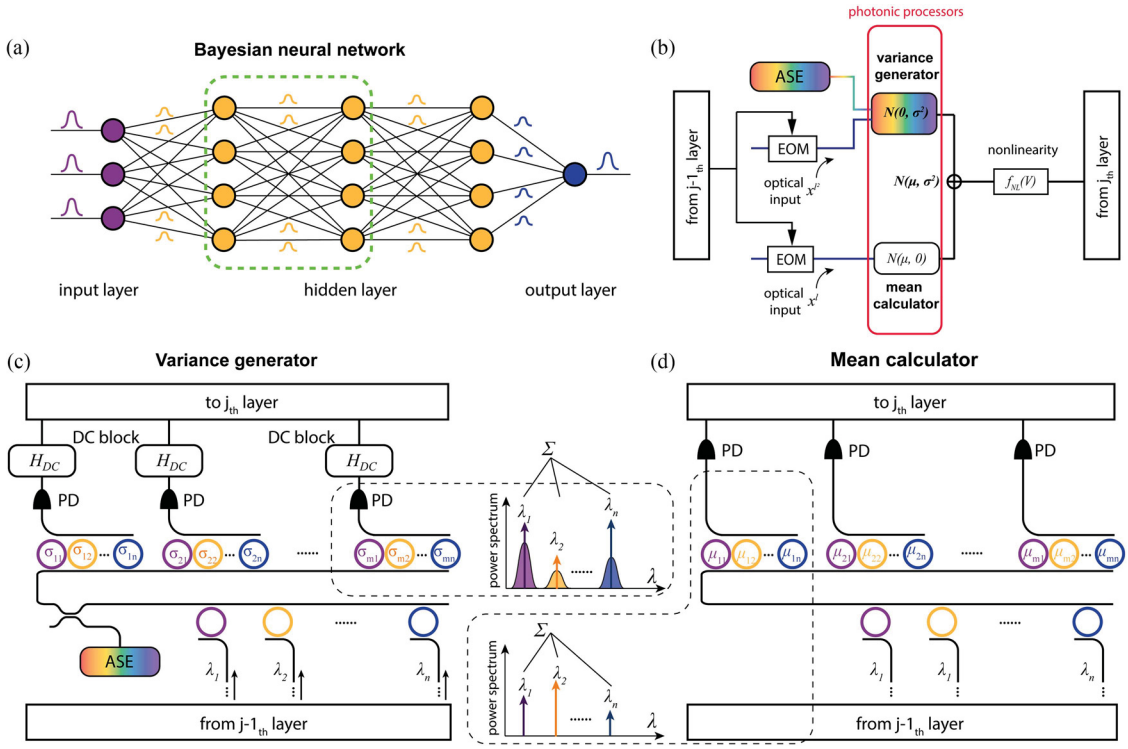



Fig. 2. (a) General BNN architecture diagram. Each trainable parameter in the BNN has a probability distribution attached to it. The final prediction of BNN is made by averaging over all individual sampled predictions. (b) Decomposition of one layer of the BNN. The input data is sent into two individual photonic processors to program the mean value and the standard deviation individually. The output optical signals are then converted into the electrical domain using photodetectors and combined. After post-processing with nonlinearity, the data is converted back into the optical domain and then transfer to the next layer. (c) The schematic of the variance generator which aims to program neuron connection weight component following the distribution $N(0,\ {\sigma_m^j}^2)$. A collector made from the rings collects the input signals from the previous layer multiplexed to a bus waveguide. The optical input signals in the bus waveguide then are combined with a broadband ASE noise source. A distributor demultiplexes signal and ASE of multiple wavelengths and distributed them into individual output neurons in this layer. The corresponding joint signal-ASE beat random noise is detected by a single photodetector (PD). The coupling efficiency of each ring is used to encode the standard deviation in each Bayesian neuron connection weight distribution. (d) The schematic of the mean calculator aims to program the mean of each Bayesian neuron connection weight. The structure of this processor is similar to the processor shown in (c) The coupling efficiency of each ring in the distributor encodes the mean value in each Bayesian neuron.

The weighted sum generates the data following the Gaussian distribution

$$N\left(\mu_m^j, {\sigma_m^j}^2\right) = N\left(\sum_n \mu_{mn}^{j-1} x_n^{j-1} + \mu_{b,n}^{j-1}, \sum_n {\sigma_{mn}^{j-1}}^2 |x_n^{j-1}|^2 + {\sigma_{b,m}^{j-1}}^2\right)$$

for a given input $x_n^{j-1}$, while the input value $x_n^{j-1}$ varies at each access following its probability distribution.

We propose a P-BNN architecture using two individual photonic processors, as shown in Fig. 2(b). The mean calculator computes the mean value $N(\mu_m^j,\ 0)$, and the variation generator generates random output with distribution $N(0,\ {\sigma_m^j}^2)$. The output optical signals are separately measured using photodetectors and combined to achieve the desired distribution of $N(\mu_m^j, {\sigma_m^j}^2)$. After post-processing with nonlinearity, the data is converted back into the optical domain and then transferred to the next layer. Fig. 2(c) plots the schematic of the variation generator which programs the neuron connection weight component following the zero-mean Gaussian distribution $N(0,\ {\sigma_m^j}^2)$. The outputs from a previous layer in various wavelength channels are coupled onto a bus waveguide using a WDM collector based on a bank of micro-ring filters. The optical input signals in the bus waveguide are combined with a broadband ASE noise source, then demultiplexed and distributed to each neuron in the same layer using WDM distributors based on micro-ring filters. The center wavelength of each micro-ring in the distributor matches the wavelength of the corresponding wavelength channel in the collector [28]. The coupling efficiency between the feeding waveguide and micro-ring controls the time-averaged ASE power $P_{\mathrm{ASE}}$ and signal power $P_{\mathrm{sig}}$ in each wavelength channel, consequently controlling the standard deviation of the signal-ASE beat random output. The standard deviation ${\sigma^j}_m$ for each Bayesian neuron connection's weight distribution thus can be programmed by the coupling efficiency.

In practice, the coupling efficiency could be programmed by controlling the gap between the ring and the bus waveguide in a passive photonic processor or by using tunable micro-rings in a state-of-the-art active photonic processor [21], [24], [29], [30]. The P-BNN supports both offline and online training modes.

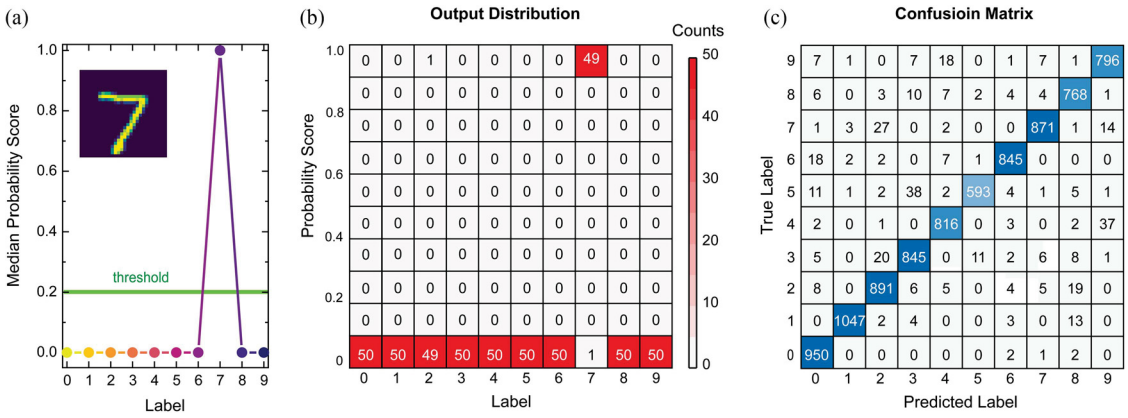



Fig. 3. Number classification using the photonic BNN. (a) The median of SoftMax probability scores of each class when using BNN to identify an image of "7". A total of 50 inferences were made using the same input image. The prediction is "undecided" if the maximum median score is smaller than the threshold (0.2). In this demonstration, the class labeled as number "7" has the maximum median score of over 0.9. Inset: The input image of "7" for classification. (b) The output probability score distribution of each class. Class 7 has high SoftMax probability scores (over 0.9) in 49 predictions over the total 50 inferences by the BNN, while the other classes have near-zero output scores. (c) The confusion matrix for the MNIST test dataset using the photonic BNN. The overall accuracy is 95.91%.

In the offline training mode, all the parameters of P-BNN are mapped to the device after being trained on a digital computer, the coupling efficiency of each ring only needs to be set once. While in the online training mode, all the parameters of P-BNN are updated in real-time thus the tunable rings are necessary. Since the center frequency differences between the micro-rings are much larger than photodetector bandwidth, the signal-ASE beat noise from each wavelength channel in the distributor is uncorrelated. Therefore, the summation of the noise from multiple channels can be achieved using one single photodetector (see Supplementary Information for detailed description). Fig. 2(d) plots the schematic of the mean calculator. The structure of the mean calculator is similar to the variance generator as shown in Fig. 2(c). The coupling efficiency of each ring in the distributor encodes the mean value $\mu_m^j$ in each Bayesian neuron (Fig. 2(d)).

For a demonstration, we simulate the performance of a two-layer P-BNN and train it to classify hand-written numbers in the MNIST dataset. The $28 \times 28$ pixels, 8-bit grayscale images are fed into the input layer, connected to a hidden layer with 1024 nodes. The final output is a 10-element vector representing the output SoftMax probability score for 10 classes (each class labels a number from 0 to 9), respectively. Unlike a DNN which only selects the class with the largest output score as the final prediction from a single inference [14], [31], [32], [33], [34], the BNN makes predictions by sampling the distribution of output scores over multiple inferences and selecting the class with the highest median SoftMax probability score. In our simulation, we sample the network with the same input 50 times with the network parameters varying each time following their probability distribution. A threshold of 0.2 is set here for the maximum medium score. The prediction is "undecided" if the maximum median score is smaller than the threshold. Fig. 3(a) plots the median of the SoftMax probability scores over 50 inferences using the P-BNN to identify an image of "7". The P-BNN gives a correct prediction of class 7, as the outcome of class 7 has the maximum median score of over 0.9 while the median score of other classes is close to 0.0. The prediction can also be confirmed through the probability score distribution of each class (see Fig. 3(b)). Class 7 has high SoftMax probability scores (over 0.9) in 49 predictions over the total 50 inferences by the BNN, while the other classes have near-zero output scores in all 50 inferences. More MNIST number classification examples could be found in Supplementary Information. The confusion matrix for MNIST number classification is shown in Fig. 3(c), the overall accuracy for our photonic BNN is 95.9%.

Conventional neural networks such as DNNs make deterministic predictions based on the prior knowledge learned from the training dataset. They cannot make judgments about the origin of the inputs and will give wrong predictions when the input is out of the expected dataset. These wrong predictions from unexpected input cannot be distinguished by post-processing. The advantage of the BNN compared to the DNN is that it produces probabilistic guarantees on predictions based on posteriors, which allows the BNN to filter out unexpected input instead of giving a wrong prediction. We demonstrate this advantage by using the P-BNN which is trained for number classification tasks to recognize the capital letter "A". As shown in Fig. 4(a), the median SoftMax probability scores of all classes are below the threshold, the P-BNN thus gives the final prediction as "undecided". Moreover, more insights can be revealed through the output probability score distribution of each class, as shown in Fig. 4(c). Class 1, 5 and 7 have a higher probability to obtain high output scores compared to other classes, indicating similar sub-features extracted by the BNN between the letter "A" and the numbers "1", "5", and "7". These similar sub-features are intuitively highlighted in Fig. 4(b). More classification results of unexpected inputs in our simulation could be found in Supplementary Information.

In conclusion, we harness the signal-ASE beat noise in the photonic system to construct a photonic BNN. In demonstrations based on simulation, the P-BNN not only makes correct predictions on the images from the test dataset but filters out the unexpected images as "undecided" as well. Assuming a moderate data rate of 10 Gbits/sec and 4 WDM channels, the computing density of P-BNN can reach 10 TOPS/mm$^2$ (Tera-Operations

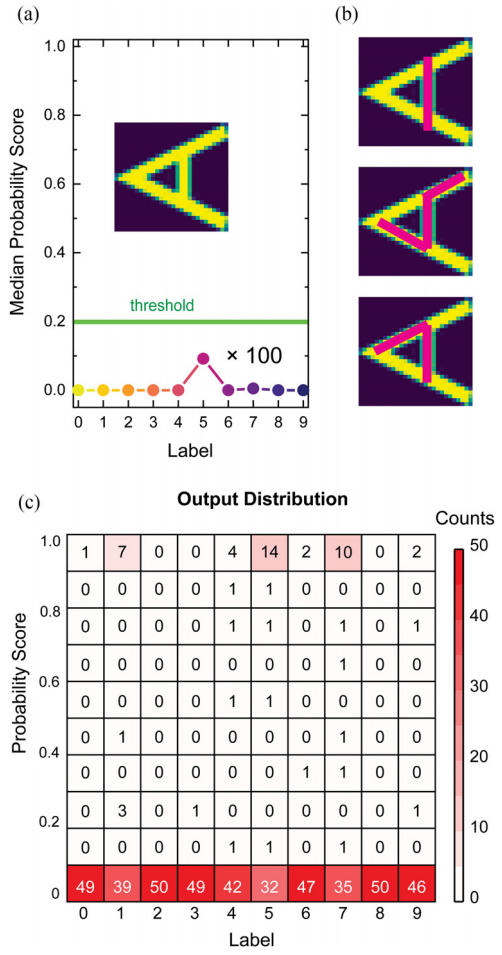



Fig. 4. Image classification using the same P-BNN but with unexpected input. (a) The median of SoftMax probability scores of each class when using the BNN to identify an image of "A". The BNN is the same as the network used in Fig. 3 and is only designed for number classification. A total of 50 inferences were made using the same input image of "A". In this demonstration, the median scores of all classes are below the threshold, thus the final prediction is "undecided". The scores are scaled up by a factor of 100 for visual clarity. Inset: The input image of "A" for classification. (b) Schematic of the features of numbers 1, 5 and 7 that could be extracted from the image of the "A". (c) The output probability score distribution of each class. Classes 1, 5 and 7 have a higher chance to obtain a high probability score compared to other classes, indicating some similar features between the image of "A" and the numbers 1, 5, and 7.

per second per mm$^2$), significantly higher than that of state-of-the-art digital processors [35]. The efficiency and speed of P-BNN will further benefit from the hybrid photonic-electronic system, which can optimally balance the energy advantages of photonic systems while realizing flexible non-linearity, connectivity, and training precision using integrated circuits. Increasing the operation data rate will increase the signal-ASE beat noise to the ASE-ASE beat noise ratio which helps to improve the programming accuracy of the P-BNN. Meanwhile, the feasible size of a cascaded ring resonator architecture of the P-BNN is limited by the insertion loss of the ring resonators. Scaling up to a large network thus faces the challenge of optical power attenuation unless using on-chip optical amplification, which is not yet available [36]. On-chip optical amplification that is compatible with silicon photonics requires co-integration of active optical gain material on silicon photonic structures. Current approaches toward scalable on-chip optical amplifiers involve either III–V to silicon wafer bonding [37] (heterogeneous integration), or co-packaging with precise assembly [38], [39] (hybrid integration). Quantum dots are potentially another promising approach as they can be largely grown directly onto silicon [40], [41], [42]. Our proposed P-BNN architecture leverages intrinsic ASE noise in on-chip optical amplifiers rationally and thus will show great significance in practical photonic neural network applications after overcoming the scaling challenge.

## REFERENCES

[1] D. J. C. MacKay, "Bayesian neural networks and density networks," *Nucl. Inst. Methods Phys. Res., A*, vol. 354, pp. 73–80, 1995.
[2] M. E. Billah and F. Javed, "Bayesian convolutional neural Network-based models for diagnosis of blood cancer," *Appl. Artif. Intell.*, vol. 36, 2022, Art. no. 2011688, doi: 10.1080/08839514.2021.2011688.
[3] M. Nour, Z. Cömert, and K. Polat, "A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization," *Appl. Soft Comput.*, vol. 97, 2020, Art. no. 106580.
[4] K. Manjusha and R. Kumar, "Spam mail classification using combined approach of Bayesian and neural network," in *Proc. Int. Conf. Comput. Intell. Commun. Netw.*, 2010, pp. 145–149, doi: 10.1109/CICN.2010.
[5] P. A. Aguilera, A. Fernández, R. Fernández, R. Rumí, and A. Salmerón, "Bayesian networks in environmental modelling," *Environ. Modelling Softw.*, vol. 26, pp. 1376–1388, 2011.
[6] Y. Gu et al., "An improved bayesian combination model for short-term traffic prediction with deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 3, pp. 1332–1342, Mar. 2020.
[7] Y. Xie, D. Lord, and Y. Zhang, "Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis," *Accident Anal. Prevention*, vol. 39, pp. 922–933, 2007.
[8] C. Capdevila, F. G. Caballero, and C. García De Andrés, "Determination of MS temperature in steels: A Bayesian neural network model," *ISIJ Int.*, vol. 42, pp. 894–902, 2002.
[9] S. Arangio and J. L. Beck, "Bayesian neural networks for bridge integrity assessment," *Struct. Control Health Monit.*, vol. 19, pp. 3–21, 2012.
[10] R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," *Phys. Rev. X*, vol. 9, 2019, Art. no. 021032.
[11] T. Wang et al., "An optical neural network using less than 1 photon per multiplication," *Nature Commun.*, vol. 13, 2022, Art. no. 123.
[12] B. J. Shastri et al., "Photonics for artificial intelligence and neuromorphic computing," *Nature Photon.*, vol. 15, pp. 102–114, 2021.
[13] G. Wetzstein et al., "Inference in artificial intelligence with deep optics and photonics," *Nature*, vol. 588, pp. 39–47, 2020.
[14] Y. Shen et al., "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, pp. 441–446, 2017.
[15] C. Wu et al., "Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network," *Nature Commun.*, vol. 12, 2021, Art. no. 96.
[16] J. Feldmann et al., "Parallel convolutional processing using an integrated photonic tensor core," *Nature*, vol. 589, pp. 52–58, 2021.
[17] X. Xu et al., "11 TOPS photonic convolutional accelerator for optical neural networks," *Nature*, vol. 589, pp. 44–51, 2021.
[18] M. P. Chang, C. L. Lee, B. Wu, and P. R. Prucnal, "Adaptive optical self-interference cancellation using a semiconductor optical amplifier," *IEEE Photon. Technol. Lett.*, vol. 27, no. 9, pp. 1018–1021, May 2015.
[19] M. P. Chang, N. Wang, B. Wu, and P. R. Prucnal, "A simultaneous variable optical weight and delay in a semiconductor optical amplifier for microwave photonics," *J. Lightw. Technol.*, vol. 33, no. 10, pp. 2120–2126, May 2015.
[20] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 1613–1622.
[21] C. Huang et al., "Demonstration of scalable microring weight bank control for large-scale photonic integrated circuits," *APL Photon.*, vol. 5, 2020, Art. no. 040803.
[22] Z. Ying et al., "Electronic-photonic arithmetic logic unit for high-speed computing," *Nature Commun.*, vol. 11, 2020, Art. no. 2154.

[23] T. Umeki, H. Takara, Y. Miyamoto, and M. Asobe, "3-dB signal-ASE beat noise reduction of coherent multi-carrier signal utilizing phase sensitive amplification," *Opt. Exp.*, vol. 20, pp. 24727–24734, 2012.

[24] D. M. Baney, P. Gallion, and R. S. Tucker, "Theory and measurement techniques for the noise figure of optical amplifiers," *Opt. Fiber Technol.*, vol. 6, pp. 122–154, 2000.

[25] C. Wu et al., "Harnessing optoelectronic noises in a photonic generative network," *Sci. Adv.*, vol. 8, 2022, Art. no. eabm2956.

[26] W. Chen, R. S. Tucker, X. Yi, W. Shieh, and J. S. Evans, "Optical signal-to-noise ratio monitoring using uncorrelated beat noise," *IEEE Photon. Technol. Lett.*, vol. 17, no. 11, pp. 2484–2486, Nov. 2005.

[27] R. Hui and M. O'Sullivan, "Characterization of optical devices," in *Fiber Optic Measurement Techniques*, 1st ed. Cambridge, MA, USA: Academic, 2009, ch. 3, pp. 259–363.

[28] J. Feldmann, N. Youngblood, C. D. Wright, H. Bhaskaran, and W. H. P. Pernice, "All-optical spiking neurosynaptic networks with self-learning capabilities," *Nature*, vol. 569, pp. 208–214, 2019.

[29] S. Moazeni et al., "A 40-Gb/s PAM-4 transmitter based on a ring-resonator optical DAC in 45-nm SOI CMOS," *IEEE J. Solid-State Circuits*, vol. 52, no. 12, pp. 3503–3516, Dec. 2017.

[30] A. H. Atabaki et al., "Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip," *Nature*, vol. 556, pp. 349–354, 2018.

[31] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[32] X. Xu et al., "Scaling for edge inference of deep neural networks," *Nature Electron.*, vol. 1, pp. 216–222, 2018.

[33] X. Lin et al., "All-optical machine learning using diffractive deep neural networks," *Sci. (1979)*, vol. 361, pp. 1004–1008, 2018.

[34] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[35] D. Silver et al., "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play," *Science*, vol. 362, pp. 1140–1144, 2018.

[36] M. A. Al-Qadasi, L. Chrostowski, B. J. Shastri, and S. Shekhar, "Scaling up silicon photonic-based accelerators: Challenges and opportunities," *APL Photon.*, vol. 7, 2022, Art. no. 020902.

[37] W. Xie et al., "Silicon-integrated nonlinear III-V photonics," *Photon. Res.*, vol. 10, pp. 535–541, 2022.

[38] D. Liang and J. E. Bowers, "Recent progress in lasers on silicon," *Nature Photon.*, vol. 4, pp. 511–517, 2010.

[39] C. Xiang, W. Jin, and J. E. Bowers, "Silicon nitride passive and active photonic integrated circuits: Trends and prospects," *Photon. Res.*, vol. 10, pp. A82–A96, 2022.

[40] C. Shang et al., "Perspectives on advances in quantum dot lasers and integration with Si photonic integrated circuits," *ACS Photon.*, vol. 8, pp. 2555–2566, 2021.

[41] Z. Zhang et al., "Monolithic passive−Active integration of epitaxially grown quantum dot lasers on silicon," *Physica Status Solidi A*, vol. 219, 2022, Art. no. 2100522.

[42] S. Chen et al., "Electrically pumped continuous-wave III–V quantum dot lasers on silicon," *Nature Photon.*, vol. 10, pp. 307–311, 2016.

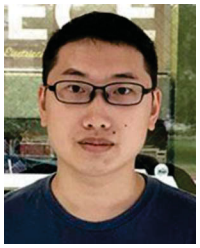
[43] W. Zhang et al., "Silicon microring synapses enable photonic deep learning beyond 9-bit precision," *Optica*, vol. 9, pp. 579–584, 2022.

[44] C. Huang et al., "A silicon photonic–electronic neural network for fiber nonlinearity compensation," *Nature Electron.*, vol. 4, pp. 847–844, 2021.

[45] C. Liu et al., "A programmable diffractive deep neural network based on a digital-coding metasurface array," *Nature Electron.*, vol. 5, pp. 113–122, 2022.

[46] T. Zhou et al., "Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit," *Nature Photon.*, vol. 15, pp. 367–373, 2021.

[47] L. G. Wright et al., "Deep neural networks trained with backpropagation," *Nature*, vol. 601, pp. 549–555, 2022.

**Changming Wu** received the B.S. degree in physics from the University of Science and Technology of China, Hefei, China, in 2014, and the M.S. degree from the Hong Kong University of Science and Technology, Hong Kong, in 2017. He is currently working toward the Ph.D. degree with the University of Washington, Seattle, WA, USA. His main research interests include integrated photonics for machine learning and optical computing using nonvolatile phase-change memory.

**Xiaoxuan Yang** received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2016, and the M.S. degree in electrical engineering from the University of California, Los Angeles, Los Angeles, CA, USA, in 2018. She is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. Her research interests include emerging nonvolatile memory technologies and hardware accelerators for deep learning applications.

**Yiran Chen** received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA, in 2005. He is currently a Professor with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA, and the Director of the NSF AI Institute for Edge Computing Leveraging the Next-generation Networks (Athena) and the NSF Industry–University Cooperative Research Center for Alternative Sustainable and Intelligent Computing, and the Co-Director of the Duke Center for Computational Evolutionary Intelligence. He has authored or coauthored one book and about 500 technical publications and has been granted 96 U.S. patents.

**Mo Li** received the B.S. degree in physics from the University of Science and Technology of China, Hefei, China, in 2001, and the Ph.D. degree in applied physics from Caltech, Pasadena, CA, USA, in 2007. He is currently a Professor with the UW Department of Electrical and Computer Engineering and UW Physics Department. He is the Principal Investigator of the Laboratory of Photonic Devices at UW. He has authored or coauthored more than 80 peer-reviewed journal papers and has been granted six US patents. His main research interests include integrated photonics, optoelectronics, optomechanics, NEMS/MEMS, spintronics and quantum photonics.