

Uncertainty Reasoning with Photonic Bayesian Machines

F. Brückerhoff-Plückelmann^{1†}, H. Borras^{2†}, S. U. Hulyal^{3†}, L. Meyer¹, X. Ji³
J. Hu³, J. Sun³, B. Klein², F. Ebert¹, J. Dijkstra¹, L. McRae¹, P. Schmidt¹
T. J. Kippenberg^{3*}, H. Fröning², W. Pernice^{1*}

¹Kirchhoff-Institute for Physics, University of Heidelberg; Heidelberg, 69120, Germany.

²Institute of Computer Engineering, University of Heidelberg; Heidelberg, 69120, Germany.

³Institute of Physics, Swiss Federal Institute of Technology, Lausanne, CH-1015 Lausanne, Switzerland.

*Corresponding author. Email: wolfram.pernice@kip.uni-heidelberg.de, tobias.kippenberg@epfl.ch

†These authors contributed equally to this work.

Artificial intelligence (AI) systems increasingly influence safety-critical aspects of society, from medical diagnosis to autonomous mobility, making uncertainty awareness a central requirement for trustworthy AI. We present a photonic Bayesian machine that leverages the inherent randomness of chaotic light sources to enable uncertainty reasoning within the framework of Bayesian Neural Networks. The analog processor features a 1.28 Tbit/s digital interface compatible with PyTorch, enabling probabilistic convolutions processing within 37.5 ps per convolution. We use the system for simultaneous classification and out-of-domain detection of blood cell microscope images and demonstrate reasoning between aleatoric and epistemic uncertainties. The photonic Bayesian machine removes the bottleneck of pseudo random number generation in digital systems, minimizes the cost of sampling for probabilistic models, and thus enables high-speed

trustworthy AI systems.

Artificial intelligence (AI) is reshaping science, technology, and society, yet its growing impact raises pressing ethical and practical concerns, when important decisions rely on opaque models. Particularly safety-critical AI systems need to be able to quantify and communicate uncertainty when exposed to noisy inputs (aleatoric uncertainty) and unknown scenarios (epistemic uncertainty) as sketched in Fig. 1(a). For example, an autonomously driving car might be confronted with bad weather conditions and unfamiliar objects on the street in real-word scenarios. In these cases conventional deterministic networks struggle with indicating uncertainty and thus would likely report a wrong prediction without warning (1, 2). Transparent estimation of uncertainty is therefore central to building trustworthy AI systems. One promising approach lies in treating the network parameters as probabilistic distributions rather than deterministic values, thereby natively encoding uncertainties within the model parameters (3, 4). Instead of finding parameters that maximize the accuracy of the network, the distributions are inferred via Bayesian Inference (5), see Fig. 1(b). While different approaches to implement Bayesian Neural Networks (BNNs) exist, their practical use has been limited by the severe computational cost of sampling-based inference. Each prediction requires multiple stochastic forward passes, leading to large latency overheads that render BNNs too slow in practice for complex real-time decision-making. Addressing these limitations would not only accelerate AI, but also pave the way for a new generation of intelligent systems capable of reasoning under uncertainty instead of ignoring it.

The central challenge in practical BNN implementations lies in how randomness is generated and processed. On digital hardware, pseudo-random number generation (PRNG) and repeated sampling introduce substantial computational and latency bottlenecks. Leveraging intrinsic analog noise within physical hardware accelerators presents an attractive alternative to directly represent probabilistic weight distributions (6), thereby eliminating the overhead of digital PRNG during prediction. However, practical deployment requires tunable and controllable noise characteristics (7). Electronic implementations based on memristive elements can enable noise shaping by using multiple cells to independently encode the mean and variance of a conductance distribution (8, 9). Besides challenges in increased circuit complexity, device variability and drift, the limited electronic bandwidth constrains these analog systems typically to 100 ns time scales (9, 10) and thus sampling rates in the tens of MHz. In contrast, photonic systems inherently provide high-bandwidth entropy

sources for GHz sampling rates where the absence of capacitive charging enables processing speeds on a sub nanosecond timescale (11). As illustrated in Fig. 1(c), these systems are further capable of generating multiple uncorrelated chaotic carriers from a single source (12, 13). Despite this potential, a scalable, low-latency probabilistic hardware accelerator has remained out of reach.

In this work, we introduce a low-latency photonic Bayesian machine that accelerates hybrid BNNs composed of deterministic and probabilistic layers by performing convolution operations with probabilistic weights at a rate of 37.5 ps per convolution. We deploy the photonic system for blood cell classification, evaluating its capability to capture epistemic uncertainty using microscope images of erythroblasts that were excluded from the training dataset. Furthermore, we demonstrate comprehensive uncertainty reasoning through evaluation on a community-standard benchmark (14). Specifically, we train a network exclusively on the MNIST dataset, while we assess uncertainty disentanglement by introducing aleatoric uncertainty via the Ambiguous-MNIST dataset and epistemic uncertainties via the Fashion-MNIST dataset during prediction with the photonic Bayesian machine.

Results

We implement a BNN for uncertainty aware image classification based on DenseNet (16) and MobileNetV1 (17). The BNN features a single probabilistic layer which is offloaded to the photonic Bayesian machine, removing the bottleneck of digital pseudo random number generation. In the photonic Bayesian machine we combine spectral encoding of the stochastic weights sketched in Fig. 1(c) with dispersion-based high-speed photonic computing. For a single probabilistic layer, we encode all weights simultaneously in the emission of a broadband chaotic light source. The power fluctuations are uncorrelated (12) and tunable, thus directly compatible with the stochastic variational inference (SVI) framework. We compute convolutions with the encoded probabilistic weights by encoding the input vector simultaneously on all spectral channels with a broadband electro-optic modulator (EOM) and by coupling the frequency encoded weight domain with the time encoded input data domain via a chip-integrated chirped grating. Due to the compact design of the grating and the absence of capacitive charging in the photonic compute system, each probabilistic convolution is calculated optically within 37.5 ps. Similarly, the integrated chirped grating enables

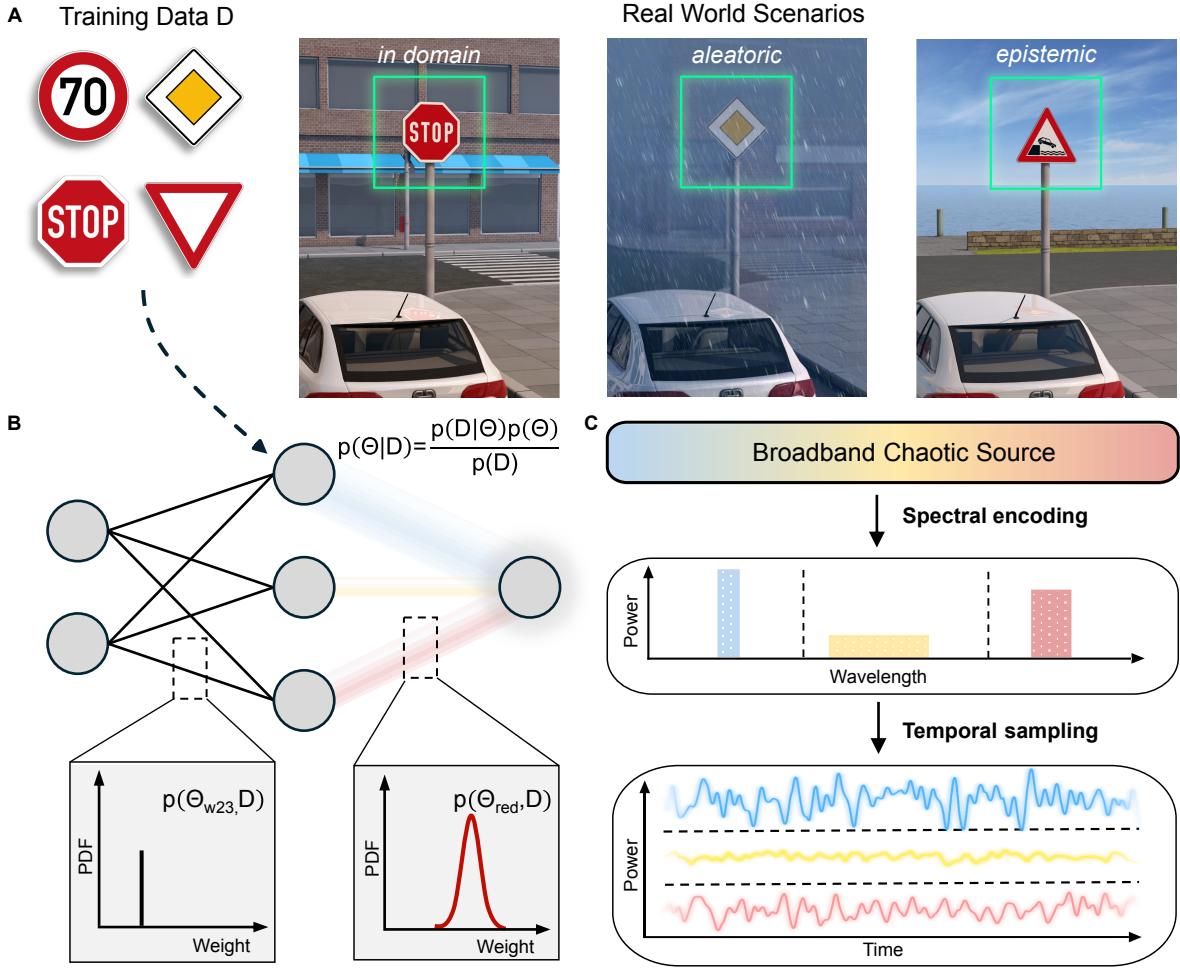


Figure 1: Towards trustworthy artificial intelligence. (a) In real world applications, AI systems are exposed to aleatoric uncertainties, for example due to limited vision, and to epistemic uncertainties arising from completely unknown inputs. Quantifying these uncertainties is crucial to properly deploy the prediction of AI systems. (b) Bayesian Neural Networks infer probability distributions over model parameters from the training data, rather than single point estimates as in conventional networks, providing a principled framework for uncertainty reasoning. Only a subset of network weights needs to be stochastic (15). Inference requires sampling these distributions and repeating the forward pass multiple times, incurring substantial computational overhead. (c) Photonic systems offer a direct, high-speed way to encode such distributions in the physical domain. A single broadband chaotic light source can be spectrally divided into many independent entropy sources, each realizing one stochastic weight. Random numbers are sampled by measuring the optical power, whose mean and variance are programmed by adjusting each channel's bandwidth and optical power.

a sub-100 ns latency, crucial for safety critical real time applications.

System Architecture

Our photonic system combines an erbium-based amplified spontaneous emission (ASE) source with the concept of frequency time interleaving (11) to compute convolutions as sketched in Fig. 2(a). We shape the emission spectrum into nine different frequency channels, centered around 194 THz with a spacing of 403 GHz. Each frequency channel encodes one single probabilistic weight, with its mean and standard deviation determined by the total optical power and bandwidth of that channel, respectively (also see Supplementary Fig. S2 (18)). The bandwidth of each individual frequency channel is chosen within a 25 GHz to 150 GHz range to enable a change in standard variation by about 68 percent. We temporally modulate the input vector on all frequency channels simultaneously with a high-speed EOM. An 80 GSPS digital to analog converter (DAC) with 8 bit resolution drives the EOM, using three samples per encoded vector component. A waveguide-integrated chirped grating made from thin-film silicon nitride, Fig. 2(b), induces a frequency dependent group delay of -93.1 ps /THz implementing a temporal shift of 1 symbol (3 samples) between adjacent frequency channels, Fig. 2(e). The broadband, time-shifted signal is then recorded with a high-speed photodetector. The total output power represents the convolution of the input vector with the nine probabilistic weights. We sample the signal with an 80 GSPS, 8 bit analog to digital converter (ADC), achieving an effective throughput of about 26.7 billion probabilistic convolutions per second. We iteratively program the bandwidth and optical power of each frequency channel simultaneously by computing test convolutions and calculating the difference between the target weight distributions and the programmed distributions. The detailed feedback-based update rule is shown in the Supplementary Information (18). To quantify the computation error of the analog photonic system, we compare the mean and standard deviation of the measured output distribution with the target ones for different programmed weights, also see Eq. S8. We determine the computation accuracy of the photonic Bayesian machine by computing convolutions with 25 different random kernels and track the statistics of the measured and target output distribution, shown in Fig. 2(c) and Fig. 2(d). For the mean the measured computation error is 0.158 and for the standard deviation 0.266. The larger error in the standard deviation is mostly attributed to the smaller output range rather than the precision of the weight programming itself. By increasing the

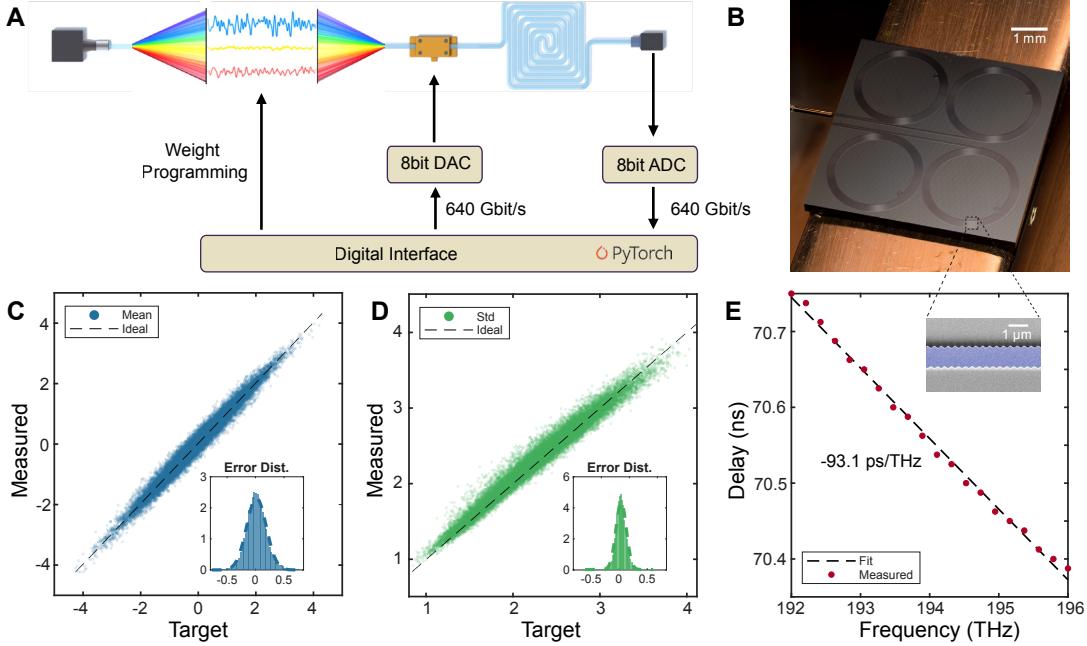


Figure 2: Photonic Bayesian Machine. (a) We encode the probabilistic weights by amplitude-bandwidth shaping the broadband chaotic emission of an ASE source. Afterwards, we encode the input vector with an EOM, apply a frequency dependent group delay and read out the signal with a photodetector. A high-speed DAC and ADC interface the analog compute unit with the digital system running PyTorch. (b) We use an integrated chirped grating to realize the frequency dependent group delay with overall low latency. The chip is fabricated on a low-loss silicon nitride platform. (c), (d) For 25 different kernels, we compare the measured output distributions with the target output distributions. The photonic Bayesian machine exhibits a computation error of 0.158 regarding the mean and 0.266 regarding the standard deviation of the output distribution. (e) We measure the delay between the DAC output signal and the ADC input signal for different center frequencies. Due to the chirped gratings, the photonic Bayesian machine exhibits a frequency dependent group delay of -93.1 ps/THz . The inset shows a scanning electron microscope image of the chirped grating; the periodicity is varied along the length of the spiral to sweep the center of reflection band.

maximal channel bandwidth, the error in the standard deviation could be reduced at the expense of the overall number of weight channels.

Bayesian Neural Network Architecture

To properly express uncertainty with the photonic Bayesian machine we hand-craft a Bayesian Neural Network. Its architecture features a single probabilistic layer and is based on DenseNet (16) for its skip connections and MobileNetV1 (17) for its Depthwise Separable (DWS) convolutions, of which we implement a variation shown in Fig. 3. We implement the whole network in a structure comparable to a Dense Block of DenseNet, while simultaneously feature pooling. The single probabilistic layer is selected to maximize its effectiveness within the Bayesian Neural Network, thereby ensuring the most efficient utilization of the photonic Bayesian machine’s computational capabilities. We further improve hardware compatibility, by favoring highly grouped convolutions over standard ones, which reduces the number of unique weights. The full architecture is shown in Fig. 3. We develop a PyTorch (19, 20) functional for the photonic Bayesian machine, such that the digital layers can seamlessly communicate with the photonic hardware layer. Then, we execute the full hardware and software pipeline of the BNN on both a CPU and the photonic accelerator, offloading all probabilistic operation to the photonic domain. Given the physical nature of the entropy source, we use Stochastic Variational Inference to infer the weight distributions from the training data (21). To ensure compatibility with gradient descent-based optimization, we develop a differentiable surrogate model for the system, approximating the physical weights as Gaussian distributions, also see the Supplementary Information (18). Here we make extensive use of straight-through estimators to simulate the limited hardware accuracy during the forward pass, while gradients remain unaffected. Simultaneously, we deploy automatic differentiation frameworks such as AutoGrad, allowing for a high degree of flexibility in the design process. During prediction, we replace the surrogate model with the photonic Bayesian machine. We sample ten times from the output distribution of the photonic system for each computation and evaluate the rest of the pipeline accordingly. Consequently, for each input image, we get N=10 different prediction scores for each of the output classes. From these statistics, we compute the total uncertainty of the BNN

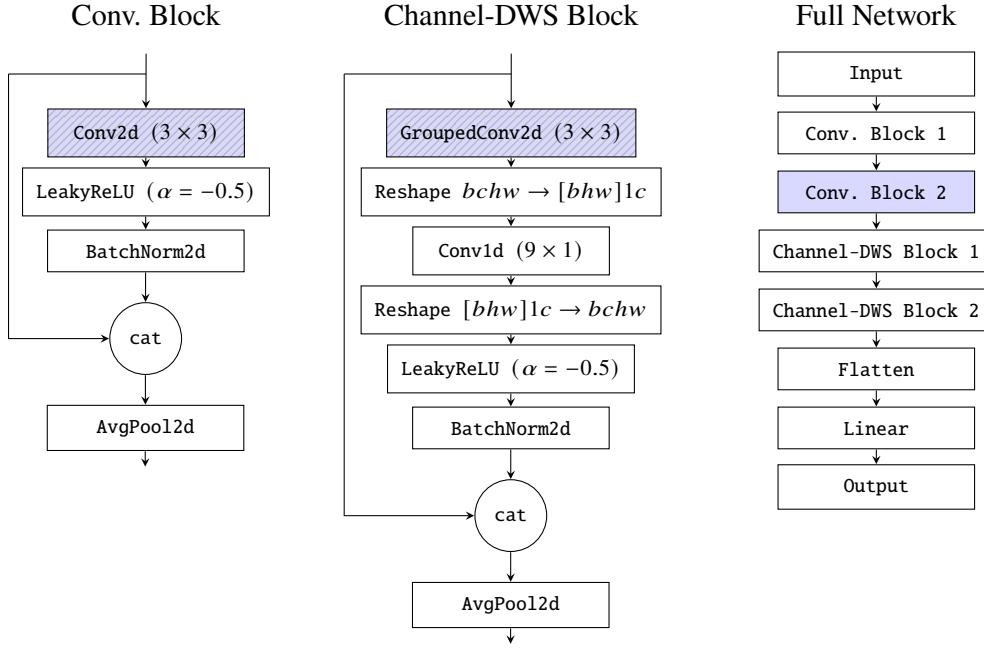


Figure 3: Network Architecture. We use a convolutional neural network to classify images. Two different types of convolutional blocks with skip connections are used to stack in total six convolutional layers, followed by a final linear layer. The size of the final linear layer depends on the number of different classes within the training dataset. All skip connections use concatenation over the channel dimension, as described by (16). The DWS convolution operates as two convolutions, where the first is fully grouped, while the second one is a 1D Convolution over the channel dimension. To achieve this the conventional tensor shape of *Batch* (b), *Channel* (c), *Height* (h), *Width* (w), is reordered to move the channels into the last dimension, while simultaneously moving all other parts into the first dimension and inserting a place holder dimension, denoted by l . The architecture flexibly supports probabilistic layers in each block, denoted by the hatched, blue area. In the full network we execute the probabilistic convolutional block marked in blue on the photonic Bayesian machine.

via the Shannon Entropy H:

$$H = - \sum_{c=1}^C \left[\frac{1}{N} \sum_{n=1}^N p(y_n = c \mid x, \theta_n) \cdot \log \left(\frac{1}{N} \sum_{n=1}^N p(y_n = c \mid x, \theta_n) \right) \right] \quad (1)$$

and the Softmax Entropy SE describing the aleatoric uncertainties:

$$SE = - \frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C p(y_n = c \mid x, \theta_n) \cdot \log(p(y_n = c \mid x, \theta_n)) \quad (2)$$

The difference between Shannon Entropy and Softmax Entropy is the Mutual Information (MI), which captures epistemic uncertainties.

Blood Cell Classification

To demonstrate the efficacy of the photonic Bayesian machine under safety critical conditions, we utilize it to detect uncertainties during AI assisted medical diagnosis. We deploy the hybrid BNN architecture shown in Fig. 3 to classify microscope images of Basophils (class label 0), Eosinophils (1), immature Granulocytes (2), Lymphocytes (3), Monocytes (4), Neutrophils (5) and Platelets (6) sketched in Fig. 4(a) (22). We infer the distribution of the network parameters only from the training data using the SVI framework (21), also see the Supplementary Information (18) for a detailed explanation of the implementation with PyTorch and Pyro (23). Figure 4(b) shows the evolution of the standard deviation of three exemplary weight distributions of the BNN. During prediction, we also show the network images of Erythroblasts, a red blood cell precursor state, not included in the train dataset and thus exhibiting high epistemic uncertainty. By evaluating the output distributions of the BNN during prediction, we distinguish between in-domain (ID) and the out-of-domain (OOD) blood cell images. For an ID test sample as the Eosinophil in Fig. 4(e), the network confidently predicts the same correct class for all ten samples of the output distribution. In contrast, for an OOD Erythroblasts image as in Fig. 4(f), the network confidently predicts a different class for some samples, leading to a high Mutual Information indicating epistemic uncertainty. We use the MI as a metric to realize an OOD classifier. The network rejects a test picture if its output distribution exhibits a MI above a certain threshold. In this way, the BNN communicates the uncertainty in its prediction to a practitioner, informing them to seek further assessments. Figure 4(c) shows the true positive rate plotted vs the false positive rate for different MI thresholds,

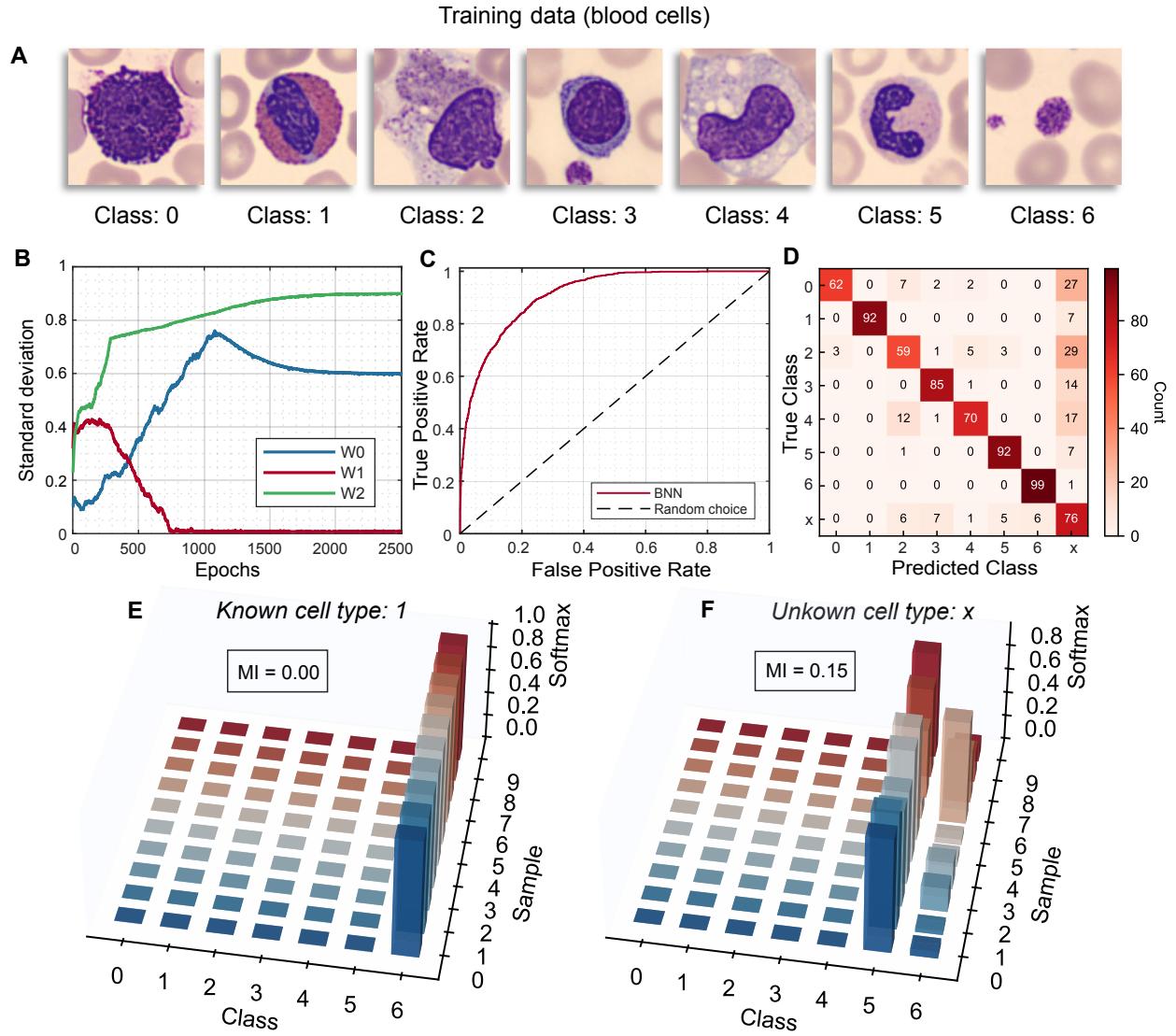


Figure 4: Blood Cell Classification. (a) We infer the BNN weight distributions from labeled microscope images of various red and white blood cells. (b) During Stochastic Variational Inference, both the mean and standard deviation of each weight distribution are learned from the training data to represent the model’s knowledge and uncertainty. (c) We use Mutual Information to detect and reject unknown cell types during prediction; as the MI threshold decreases, both false-positive and true-positive rates rise, shown as a ROC curve. (d) With an MI threshold of 0.0185, we reject most unknown cell-type images and increase the ID classification accuracy from 90.26% to 94.62%. (e),(f) For each image, we draw ten samples from the BNN output. Known cell types yield consistent predictions, while an unknown Erythroblast image produces varied predictions across samples, giving a high MI.

resulting in an area under receiver operating characteristic (AUROC) of 91.16 %, highlighting the OOD detection performance of the BNN. At the same time the treatment of uncertainty also helps to improve the accuracy of ID classification. Here the BNN rejects unclear cases which are flagged with high uncertainty, improving the prediction quality from 90.26 % to 94.62 % for an optimal MI threshold of 0.0185. Figure 4(d) shows the full confusion matrix including the Erythroblasts images labelled with “x”.

Uncertainty Reasoning

Apart from detecting uncertainty in safety critical tasks, the photonic Bayesian machine further enables uncertainty reasoning, differentiating between epistemic uncertainties arising from unknown input classes and aleatoric uncertainties from factually unclear inputs. Based on the identified type of uncertainty, the full AI system can adapt its behavior to effectively address the situation. We test the photonic Bayesian machine on a community benchmark for this problem (14), inferring the model distribution from the MNIST dataset, and using Ambiguous-MNIST and Fashion-MNIST during prediction, sketched in Fig. 5(a). Notably we train on MNIST only, while in literature the training is often done on Ambigious-MNIST as well (14, 24, 25). We thus implement a training which fully excludes uncertainty samples from the training procedure, faithfully mimicking real-world settings. In contrast, related work has shown that using uncertainty samples can further improve uncertainty scores, albeit open questions remain regarding representativeness of these uncertainty samples. During prediction, we again sample ten times from the output distribution for each test input image. The network confidently predicts the correct class with each sample for an ID image, Fig. 5(b), and predicts a different class for different samples for an OOD image, Fig. 5(c), leading to a high MI. In contrast, for an input image from the Ambiguous-MNIST dataset, each sample of the output distribution exhibits a large Softmax Entropy, Fig. 5(d), indicating a high aleatoric uncertainty. By computing both the MI and the Softmax Entropy for each test image, we obtain three clusters shown in Fig. 5(e), each corresponding to one of three datasets. In this way, the photonic Bayesian machine not only detects uncertainty but also reasons which type of uncertainty it is. While we note, that the MI and SE are loosely correlated to each other in Fig. 5(e), the network is still able to effectively differentiate between aleatoric and epistemic uncertainty. Overall, the classification accuracy for the MNIST dataset is 96.01 % and 99.7 % with OOD rejection. The aleatoric uncertainty detector

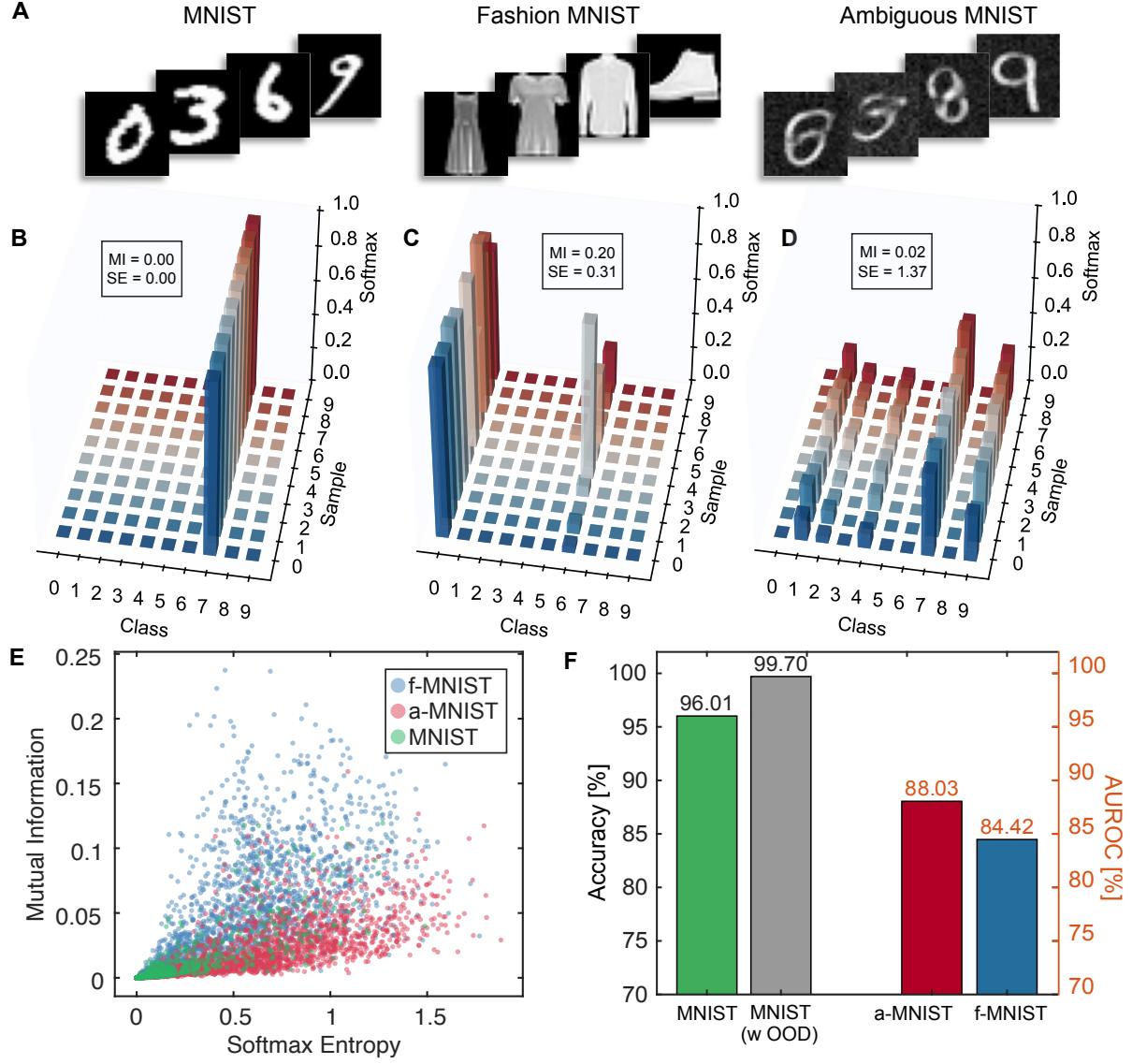


Figure 5: Uncertainty disentanglement. (a) We infer the weight distributions of the BNN from the MNIST dataset containing handwritten numbers and introduce epistemic uncertainty during prediction with images of different fashion items and aleatoric uncertainties by ambiguous number like images. (b)-(d) For the ID data the BNN predicts the correct result with high confidence for each sample, and for the OOD data a different result for each sample. In contrast, each sample shows high Softmax Entropy for ambiguous input data. (e) Computing the MI and Softmax Entropy for each sample separates the three different kinds of input data during prediction, allowing to both detect and analyze uncertainties. (f) The accuracy for the MNIST dataset increases from 96.01 % to 99.7 % when enabling OOD rejection at a MI threshold of 0.00308, the epistemic and aleatoric uncertainty detector achieve an AUROC of 84.42 % and 88.03 % respectively.

achieves an AUROC of 88.03 % on the Ambiguous-MNIST dataset and an AUROC of 84.42 % on the Fashion-MNIST dataset.

Discussion

While many approaches to implement Bayesian Neural Networks exist, each comes with its trade-offs. Markov chain Monte Carlo methods are mathematically most rigorous for analytically intractable tasks by sampling from the true posterior of the model parameters. However, their computational demands make them unpractical for larger network architectures (5). Ensemble based approaches like Deep Ensembles (3) are a practical tradeoff, less mathematically grounded, but more handleable, since existing tool chains can be reused and a model simply needs to be trained multiple times. While these approaches try to estimate the true posterior, their memory demands for storing many sets of network parameters are significant. In contrast, Stochastic Variational Inference approximates the posterior distribution with parametrized distributions. The distribution parameters, for example mean and variance, are trained via stochastic gradient descent, enabling scaling to deep neural networks (6), while approximating the true posterior.

Leveraging the tunable stochasticity and large bandwidth of chaotic light sources as variational distributions, the photonic Bayesian machine realizes ultra-fast probabilistic convolution processing. Amplified spontaneous emission in an erbium doped fiber serves as true random number generator, passing the state-of-the-art National Institute of Standards and Technology (NIST Special Publication 800-22) tests for entropy sources (26). Modelling the physical weights via Gaussian variational distributions enables inferring the weight distributions from the training data via the principled framework of Stochastic Variational Inference (21). Similarly, the low computation error and linear weight response allows for surrogate-based digital training, decoupling the training process from the actual physical hardware and thus avoiding the per circuit retraining of hardware-in-the-loop routines (27). In combination with PyTorch compatibility, the photonic Bayesian machine serves an ultrafast probabilistic hardware accelerator. By co-designing a Bayesian Neural Network for the photonic Bayesian machine, we demonstrate the efficacy of the machine in multiple contexts. Most notably, the combination of the chaotic light-based entropy source and SVI is sufficient for uncertainty reasoning, distinguishing between aleatoric and epistemic uncertainties without being

exposed to corresponding uncertainty samples at any point of the training routine.

The photonic Bayesian machine features a direct path towards both hardware and architectural integration. Apart from enabling scalability via foundry processes, integration reduces the latency of the photonic system. By realizing the frequency delay of -93.1 ps/THz on-chip via a 5.68 cm long chirped grating we reduce the pulse propagation time by more than three orders of magnitude in comparison to fiber-based approaches (11). Similarly, the ASE based entropy source can be integrated on chip via erbium doped silicon nitride waveguides (28) or active electrically pumped gain media on indium phosphide (29). Independent of the number of probabilistic weights, the photonic Bayesian machine only requires a single high-speed digital to analog converter and analog to digital converter. In combination with the broadband modulation above 100 GHz bandwidth based on integrated electro optic modulators (30) and photodetection on similar frequency scales (31), the photonic Bayesian machine can be seamlessly integrated in todays predominantly digital computing architectures.

By leveraging noise inherent to photonic systems instead of compensating it using costly methods, the presented photonic Bayesian machine successfully overcomes the bottleneck of random number generation in SVI-based probabilistic modeling. In combination with ultrafast sampling in photonic systems, fast computation of large probabilistic neural networks is enabled, paving the way for a new generation of trustworthy AI systems.

References and Notes

1. D. Amodei, *et al.*, Concrete Problems in AI Safety (2016), doi:10.48550/arXiv.1606.06565, <https://arxiv.org/abs/1606.06565>.
2. B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and Scalable Predictive Uncertainty Estimation Using Deep Ensembles, in *Advances in Neural Information Processing Systems*, vol. 30 (2017), pp. 6402–6413, <https://proceedings.neurips.cc/paper/2017/file/9ef2ed4b7fd2c810847ffa5fa85bce38-Paper.pdf>.
3. A. Kendall, Y. Gal, What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, in *Advances in Neural Information Processing Systems*, vol. 30 (2017), pp. 5574–5584, <https://papers.nips.cc/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html>.
4. B. Lambert, F. Forbes, S. Doyle, H. Dehaene, M. Dojat, Trustworthy Clinical AI Solutions: A Unified Review of Uncertainty Quantification in Deep Learning Models for Medical Image Analysis. *Artificial Intelligence in Medicine* **150**, 102830 (2024), doi:10.1016/j.artmed.2024.102830, <https://www.sciencedirect.com/science/article/pii/S0933365724000721>.
5. L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, M. Bennamoun, Hands-On Bayesian Neural Networks—A Tutorial for Deep Learning Users. *IEEE Computational Intelligence Magazine* **17** (2), 29–48 (2022), doi:10.1109/MCI.2022.3155327, <https://doi.org/10.1109/MCI.2022.3155327>.
6. C. Blundell, J. Cornebise, K. Kavukcuoglu, D. Wierstra, Weight Uncertainty in Neural Networks, in *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, Lille, France), vol. 37 of *Proceedings of Machine Learning Research* (2015), pp. 1613–1622, <https://proceedings.mlr.press/v37/blundell15.html>.
7. F. Brückerhoff-Plückelmann, *et al.*, Probabilistic Photonic Computing for AI. *Nature Computational Science* **5** (5), 377–387 (2025), doi:10.1038/s43588-025-00800-1, <https://doi.org/10.1038/s43588-025-00800-1>.

8. S. Liu, *et al.*, Bayesian Neural Networks Using Magnetic Tunnel Junction-Based Probabilistic In-Memory Computing. *Frontiers in Nanotechnology* **4**, 1021943 (2022), doi:10.3389/fnano.2022.1021943, <https://www.frontiersin.org/articles/10.3389/fnano.2022.1021943>.
9. Y. Lin, *et al.*, Uncertainty Quantification via a Memristor Bayesian Deep Neural Network for Risk-Sensitive Reinforcement Learning. *Nature Machine Intelligence* **5**, 714–723 (2023), doi:10.1038/s42256-023-00680-y, <https://doi.org/10.1038/s42256-023-00680-y>.
10. M. Le Gallo, *et al.*, A 64-Core Mixed-Signal In-Memory Compute Chip Based on Phase-Change Memory for Deep Neural Network Inference. *Nature Electronics* **6**, 680–693 (2023), doi:10.1038/s41928-023-01010-1, <https://doi.org/10.1038/s41928-023-01010-1>.
11. X. Xu, *et al.*, 11 TOPS Photonic Convolutional Accelerator for Optical Neural Networks. *Nature* **589** (7840), 44–51 (2021), doi:10.1038/s41586-020-03063-0, <https://doi.org/10.1038/s41586-020-03063-0>.
12. F. Brückerhoff-Plückelmann, *et al.*, Probabilistic Photonic Computing with Chaotic Light. *Nature Communications* **15** (1), 10445 (2024), doi:10.1038/s41467-024-54931-6, <https://doi.org/10.1038/s41467-024-54931-6>.
13. C. Wu, *et al.*, Harnessing Optoelectronic Noises in a Photonic Generative Network. *Science Advances* **8** (3), eabm2956 (2022), doi:10.1126/sciadv.abm2956, <https://doi.org/10.1126/sciadv.abm2956>.
14. J. Mukhoti, A. Kirsch, J. van Amersfoort, P. H. S. Torr, Y. Gal, Deep Deterministic Uncertainty: A New Simple Baseline, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), pp. 24384–24394, doi:10.1109/CVPR52729.2023.02336.
15. M. Sharma, S. Farquhar, E. Nalisnick, T. Rainforth, Do Bayesian Neural Networks Need To Be Fully Stochastic?, in *Proceedings of the 26th International Conference on Artificial Intelligence and Statistics*, vol. 206 of *Proceedings of Machine Learning Research* (2023), pp. 7694–7722, <https://proceedings.mlr.press/v206/sharma23a.html>.

16. G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Densely Connected Convolutional Networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 4700–4708, doi:10.1109/CVPR.2017.243, <https://doi.org/10.1109/CVPR.2017.243>.
17. A. G. Howard, *et al.*, MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2017), <https://arxiv.org/abs/1704.04861>.
18. Materials and methods are available as supplementary material.
19. A. Paszke, *et al.*, Automatic Differentiation in PyTorch, in *Proceedings of the NIPS 2017 Workshop on Autodiff* (2017), <https://openreview.net/forum?id=BJJsrmfCZ>.
20. J. Ansel, *et al.*, PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation, in *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)* (ACM) (2024), doi:10.1145/3620665.3640366, <https://docs.pytorch.org/assets/pytorch2-2.pdf>.
21. M. D. Hoffman, D. M. Blei, C. Wang, J. Paisley, Stochastic Variational Inference. *Journal of Machine Learning Research* **14** (1), 1303–1347 (2013), <https://www.jmlr.org/papers/v14/hoffman13a.html>.
22. J. Yang, *et al.*, MedMNIST v2 – A Large-Scale Lightweight Benchmark for 2D and 3D Biomedical Image Classification. *Scientific Data* **10** (1), 41 (2023), doi:10.1038/s41597-022-01721-8, <https://doi.org/10.1038/s41597-022-01721-8>.
23. E. Bingham, *et al.*, Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research* **20** (28), 1–6 (2019), <https://www.jmlr.org/papers/v20/18-403.html>.
24. H. Sun, B. van Breugel, J. Crabbé, N. Seedat, M. van der Schaar, What is Flagged in Uncertainty Quantification? Latent Density Models for Uncertainty Categorization, in *Thirty-seventh*

Conference on Neural Information Processing Systems (2023), <https://openreview.net/forum?id=zyhxRc9bew>.

25. S. Steger, C. Knoll, B. Klein, H. Fröning, F. Pernkopf, Function Space Diversity for Uncertainty Prediction via Repulsive Last-Layer Ensembles, in *ICML 2024 Workshop on Structured Probabilistic Inference & Generative Modeling* (2024), <https://openreview.net/forum?id=FbMN9HjgHI>.
26. Y. Guo, *et al.*, 40 Gb/s Quantum Random Number Generation Based on Optically Sampled Amplified Spontaneous Emission. *APL Photonics* **6** (6), 066105 (2021), doi:10.1063/5.0040250.
27. L. G. Wright, *et al.*, Deep Physical Neural Networks Trained with Backpropagation. *Nature* **601**, 549–555 (2022), doi:10.1038/s41586-021-04223-6.
28. Y. Liu, *et al.*, A Photonic Integrated Circuit-Based Erbium-Doped Amplifier. *Science* **376** (6599), 1309–1313 (2022), doi:10.1126/science.abo2631, <https://doi.org/10.1126/science.abo2631>.
29. B. Shi, N. Calabretta, R. Stabile, InP Photonic Integrated Multi-Layer Neural Networks: Architecture and Performance Analysis. *APL Photonics* **7** (4), 046103 (2022), doi:10.1063/5.0078873, <https://doi.org/10.1063/5.0078873>.
30. P. O. Weigel, *et al.*, Bonded Thin-Film Lithium Niobate Modulator on a Silicon Photonics Platform Exceeding 100 GHz Bandwidth. *Optics Express* **26** (18), 23728–23739 (2018), doi:10.1364/OE.26.023728, <https://doi.org/10.1364/OE.26.023728>.
31. Y. Shi, X. Li, M. Zou, Y. Yu, X. Zhang, 103 GHz Germanium-on-Silicon Photodiode Enabled by an Optimized U-Shaped Electrode. *Photonics Research* **12** (1), 1–6 (2024), doi:10.1364/PRJ.495958, <https://doi.org/10.1364/PRJ.495958>.

Acknowledgments

We thank Jochen Stuhrmann, from Illustrato, for his assistance with the illustrations. The samples were partially fabricated in the EPFL Center of MicroNanoTechnology (CMi).

Funding: This work was supported by funding from the European Research Council grant 101200429 (WP) and Contract N660012424006 (NaPSAC) from the Defense Advanced Research Projects Agency (DARPA), Defense Sciences Office (DSO) (TK, WP).

Author contributions: Conceptualization: FBP, HB, LM, BK

Methodology: FBP, JH, SH, XJ, FE, JD, LM, PS, JS

Investigation: FBP, HB, LM, JH

Visualization: FBP, HB, SH, XJ

Funding acquisition: TK, HF, WP

Project administration: FBP, TK, HF, WP

Supervision: TK, HF, WP

Writing – original draft: FBP, HB

Writing – review & editing: All authors

Competing interests: The authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials.

Supplementary materials

Materials and Methods

Supplementary Text

Figs. S1 to S8

Table S1

References (42-31)

Supplementary Materials for Uncertainty Reasoning with Photonic Bayesian Machines

F. Brückerhoff-Plückelmann^{1†}, H. Borras^{2†}, S. U. Hulyal^{3†}, L. Meyer¹, X. Ji³
J. Hu³, J. Sun³, B. Klein², F. Ebert¹, J. Dijkstra¹, L. McRae¹, P. Schmidt¹
T. J. Kippenberg^{3*}, H. Fröning², W. Pernice ^{1*}

¹Kirchhoff-Institute for Physics, University of Heidelberg; Heidelberg, 69120, Germany.

²Institute of Computer Engineering, University of Heidelberg; Heidelberg, 69120, Germany.

³Institute of Physics, Swiss Federal Institute of Technology, Lausanne, CH-1015 Lausanne, Switzerland.

*Corresponding author. Email: wolfram.pernice@kip.uni-heidelberg.de, tobias.kippenberg@epfl.ch

[†]These authors contributed equally to this work.

This PDF file includes:

Materials and Methods

Supplementary Text

Figs. S1 to S8

Table S1

References (42-31)