nature computational science

Perspective

https://doi.org/10.1038/s43588-025-00800-1

Probabilistic photonic computing for AI

Received: 6 November 2024

Accepted: 3 April 2025

Published online: 23 May 2025

Check for updates

Frank Brückerhoff-Plückelmann[1,2,3], Anna P. Ovvyan[1,2], Akhil Varri[1], Hendrik Borras[4], Bernhard Klein[4], Lennart Meyer[2], C. David Wright[5], Harish Bhaskaran[6], Ghazi Sarwat Syed[3], Abu Sebastian[3], Holger Fröning[4] & Wolfram Pernice[1,2]

Probabilistic computing excels in approximating combinatorial problems and modeling uncertainty. However, using conventional deterministic hardware for probabilistic models is challenging: (pseudo) random number generation introduces computational overhead and additional data shuffling. Therefore, there is a pressing need for different probabilistic computing architectures that achieve low latencies with reasonable energy consumption. Physical computing offers a promising solution, as these systems do not rely on an abstract deterministic representation of data but directly encode the information in physical quantities, enabling inherent probabilistic architectures utilizing entropy sources. Photonic computing is a prominent variant of physical computing due to the large available bandwidth, several orthogonal degrees of freedom for data encoding and optimal properties for in-memory computing and parallel data transfer. Here, we highlight key developments in physical photonic computing and photonic random number generation. We further provide insights into the realization of probabilistic photonic processors and their impact on artificial intelligence systems and future challenges.

The use of artificial neural networks (ANNs) has been steadily increasing in various application domains, including autonomous driving[1], medical diagnosis[2] and natural language processing[3]. As utilization of these applications gains popularity, so does the demand for energy-efficient, low-latency hardware capable of handling the intensive computations required by ANNs. Particularly in safety-critical applications[4], there is an immediate need for reliable uncertainty estimation, especially for scenarios outside the training distribution. Traditional hardware solutions, such as central processing units, graphics processing units and tensor processing units (TPUs), are inherently deterministic, making them ill suited for evaluating probabilistic models without incurring substantial computational overhead and increased latency.

Physical neuromorphic computing approaches emulate the working principles of biological brains by directly encoding the data in physical quantities instead of abstract representations of data. In particular, neuromorphic architectures leverage massively parallel in-memory computation, which offers a promising route to reducing the latency and power consumption of AI accelerators[5]. Moreover, neuromorphic systems are designed to incorporate tunable stochasticity, aligning with the neuroscience principle of free energy minimization[6,7]. This built-in stochasticity is essential for effective uncertainty estimation and solution approximation. As computation is implemented via manipulating physical quantities in an analog way, physical entropy sources for true random number generation can be directly linked to the computation without inducing further overhead. For example, thermal noise in a magnetic tunnel junction can be exploited to realize a probabilistic matrix weight based on a tunable probabilistic conductance distribution[8].

Photonic physical computation systems are especially intriguing due to the large available bandwidth—for example, more than 4 THz considering only the conventional communication band ranging from 1,530 nm to 1,565 nm—and multiple orthogonal degrees of freedom to encode and manipulate data in parallel. This has fueled the development of photonic hardware accelerators leveraging amplitude[9] and

[1]Physical Institute, University of Münster, Münster, Germany. [2]Kirchhoff-Institute for Physics, University of Heidelberg, Heidelberg, Germany. [3]IBM Research Europe, Rüschlikon, Switzerland. [4]Institute of Computer Engineering, University of Heidelberg, Heidelberg, Germany. [5]Department of Engineering, University of Exeter, Exeter, UK. [6]Department of Materials, University of Oxford, Oxford, UK. e-mail: wolfram.pernice@kip.uni-heidelberg.de

phase encoding[10] and orthogonal encodings such as time–frequency[11] or amplitude–frequency and amplitude–mode[12,13]. In addition to the variety of physical quantities for data encoding and computation, photonics also offers optimal properties for data transmission and storage, with no charging of wires and higher long-time memory stability—for example, in the case of phase-change memory[9]. Similarly, photonics offers multiple methods for true random number generation, each linked to different physical quantities such as phase and intensity[14,15]. This enables a high compatibility with the various encoding and computation schemes: for instance, intensity-based photonic crossbar arrays can directly leverage intensity noise[16,17].

Here we offer our viewpoint on recent advancements in physical probabilistic computing, with emphasis on photonic implementations. We examine current architectures from both hardware and software perspectives and highlight their fundamental working principles. Our focus is on the seamless integration of entropy sources with high-speed photonic processors and their utilization on a system level. Additionally, we offer an outlook on the impact of photonic probabilistic processors on AI and discuss the challenges lying ahead.

Beyond conventional deterministic computing

Specialized digital deterministic processors such as TPUs power state-of-the-art neural networks and keep pushing electronic hardware to its physical limits to deal with the massive computational workload at reasonable energy levels and timescales[18]. Emerging analog computing architectures promise to complement conventional systems for dedicated tasks such as massively parallel, low-latency matrix multiplication[5,11,19]. To harness the advantages of probabilistic computing and design networks that incorporate stochasticity as one of the cornerstones of neuromorphic design, novel processors are required[20]. Since physical entropy sources lie at the heart of probabilistic sampling, their direct integration with physical computing architectures is especially intriguing.

Physical computing

Physical processors offer tremendous potential to augment conventional hardware due to their fundamentally different architecture, as illustrated in Fig. 1. In contrast to the abstract data encoding in digital processors, physical processors encode the information directly in a physical quantity, for example the phase, amplitude or length of a pulse[5,9,10,21]. Similarly, computation is directly implemented by manipulating the corresponding quantity—for example, changing the amplitude of a pulse by a fixed factor to realize a multiplication. In contrast, digital processors use complex concatenations of logic gates to realize the same operation on the abstract digital data encoding. This has a severe impact on the overall properties of the system. Physical systems are inherently prone to noise, as shot noise arising from the discrete nature of photons and electrons imposes a lower limit. Further, excess noise is present in practical systems—for example, due to spontaneous emission of photons or Johnson–Nyquist noise in electronic circuits. Thus, physical processors, which directly encode the information in a physical quantity, will always exhibit noise and limited precision. Notably, this does not preclude them from achieving digital-level performance for specific deterministic applications such as ANNs[22–24]. In contrast, the abstract (binary) encoding in digital systems effectively removes noise from the computation. Even though the state of a single bit is still linked to a physical quantity (for example, the voltage), it is a binary system. Therefore, it is not affected as long as the physical noise cannot cause a bit flip that is not detected by parity checks, thus enabling high-precision computing. This fundamental difference explains why conventional digital processors are ill suited for probabilistic computing and sampling, even though there are efforts to realize probabilistic bits[25], whereas it is an inherent property of physical processors. The challenge and at the same time the unique opportunity lies within making the physical noise accessible and tunable such that it can be used for probabilistic modeling.

In addition to the fundamental difference from digital computing regarding data representation, physical computing also inherently paves the way to massively parallel low-latency computation. Neither data movement nor computation is linked to a clock rate; instead, the information carrier—for example, the optical pulse—propagates through the physical system that implements the required mathematical operations—that is, manipulations of the physical quantity encoded on the information carrier. As several information carriers must propagate through the physical system at the same time and the propagation time is in general short in comparison with the clock rate of the interface, low-latency parallel computation is an inherent feature. In contrast, a digital system would perform the same operations on subsequent time bins given by the clock rate, thus resulting in a substantially larger latency.

Photonic architectures

The bosonic nature of photons makes them ideal information carriers in combination with low-loss and broadband optical interconnects. In addition, light pulses propagate in vacuum at around 300,000 km s^{-1} and are only slowed down by a factor of two in typical waveguide materials such as silicon nitride, reducing the inherent latency of a physical processor. As mentioned earlier, photonics offers a variety of physical quantities that can be used to encode data, including frequency[11,26], power[9,27], phase[10], polarization modes[13] and spatial modes[22]. Similarly, nonlinear material responses may be used to manipulate the frequency of a pulse[28], a tunable absorber and phase shifter can change the power and phase of a pulse[29,30] and tunable converters and multiplexers can change the mode states[13,31]. Combination of the two leads to a variety of integrated photonic processors including waveguide crossbar arrays[9,32], Mach–Zehnder interferometer meshes[10,33], microring weight banks[27,34] and diffractive networks[35,36].

While there exist many reviews explaining the different computation schemes[19,37], it is important to consider the fundamental constraints and advantages of photonic processors to lend a perspective on their utilization for probabilistic computing. The concepts of analog processors favor an application-specific circuit design. The input data are encoded in physical quantities and the computation units apply transforms to the input. While some parameters, such as matrix weights in an analog matrix–vector multiplication (MVM) accelerator, can be tuned, the overall functionality remains fixed. This makes mixed-signal processors particularly valuable, combining efficient analog computation for core tasks with flexible digital units for additional processing[5]. Photonic implementations, while offering low transmission losses and high bandwidth, are relatively large in size. The width of dielectric waveguides is on the order of the effective wavelength inside the medium, that is, on a 0.4-μm to 1-μm scale considering the default C band[38]. As guiding is based on total internal reflection, smooth bend radii are required for routing, where the minimal bend radius depends on the refractive index contrast and might range between 5 μm for the silicon-on-insulator platform and 30 μm for the silicon nitride on insulator platform. Active components such as laser sources, modulators and photodetectors also tend to be on a 100-μm scale. While plasmonic waveguiding can strongly reduce the feature size, ohmic losses markedly increase the propagation loss and cancel out one of the main advantages of photonic waveguides[39]. Guiding light in an integrated dielectric waveguide is possible with propagation losses as low as 1.77 dB m^{-1} (ref. 40), making large physical circuit designs feasible, which pairs well with the large feature sizes of photonic components that are compatible with larger fabrication nodes—for example, 180-nm processes. A major challenge, however, is the large size of photonic components. Physical photonic computing would need to overcome great challenges to closely pair with memory densities of physical electronic processors, that is, having a matrix cell well below 1 μm^2. On the other hand, photonic systems are ideal for large circuit designs due to optimal propagation properties and

https://doi.org/10.1038/s43588-025-00800-1

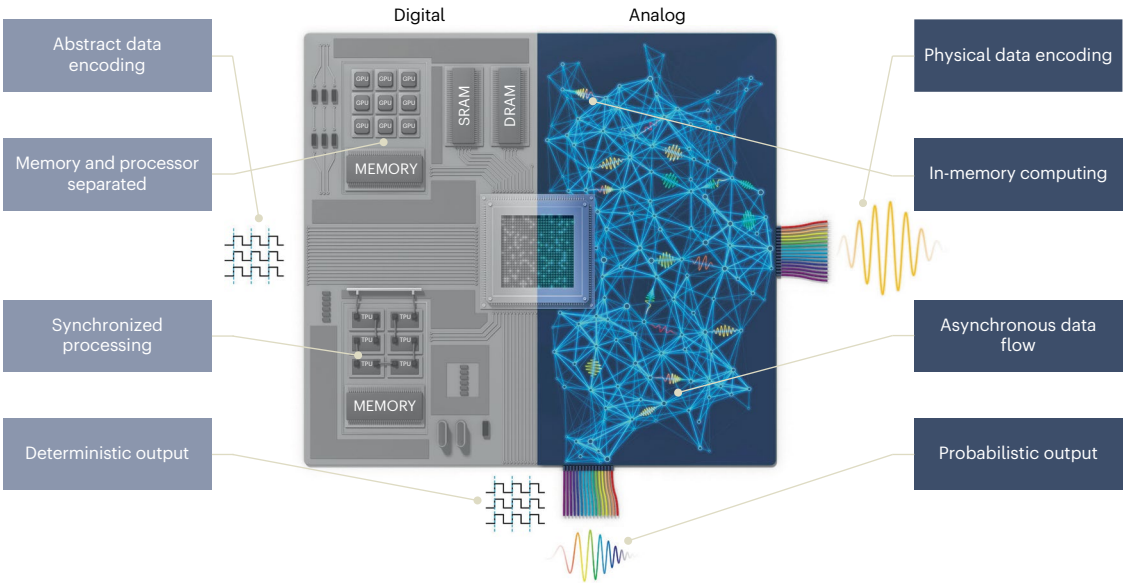

Fig. 1 | Route towards probabilistic computing. Analog computing is a promising avenue to overcome the bottleneck arising from probabilistic sampling in digital hardware. The sketch illustrates the main differences between the architectures. First, the abstract (binary) data encoding is intrinsically designed for deterministic high-precision computing. In contrast, physical computing is inherently probabilistic as the information is directly encoded in physical quantities—for example, the amplitude and frequency of a pulse. Second, in-memory computing, for example by attenuation or frequency filtering, favors parallel low-latency computation. Instead, conventional architectures exhibit a separation between memory and processing elements, which requires additional data shuffling. Similarly, analog processors evolve continuously in time in contrast to discrete clock-based digital operation schemes, enabling an asynchronous data flow within the system. By design, the digital system provides a deterministic output again and must rely on pseudorandom generation to simulate randomness. Instead, the analog system only applies a physical transformation—for example, a frequency chirp—to the physical input state.

coarser feature sizes. Thus, complex and compute-heavy tasks that do not require a large amount of memory are an ideal fit for photonic processors. Overall, photonics is a compelling platform for physical computing, driven by multiple degrees of freedom for information encoding and manipulation in combination with the optimal waveguiding properties. These attributes collectively contribute to a reduced latency and minimal power consumption, making photonics an interesting approach for advanced computational hardware.

Photonic entropy sources

Entropy sources lie at the heart of probabilistic computing. Independent of the type of physical computing, there are two approaches for direct integration: first, integration within the computation units[8,41,42]; second, as an innate feature of the physical information carrier[16,17,43]. While the first approach offers more flexibility for modeling the stochasticity, it strongly increases the complexity and potentially the energy dissipation of the circuit, since there are typically more computation units than carriers. In addition, random fluctuations in material properties might be bandwidth limited. As the power spectral density of a random process is directly linked to its autocorrelation function via the Wiener–Khinchin theorem, a small bandwidth $\Delta\nu$ leads to a large correlation time $\tau_c \approx 1/\Delta\nu$. Since subsequent samples must be uncorrelated for many applications, the correlation time induces an upper limit of the achievable sampling rate. Photonics offers various possibilities for innate random physical information carriers, including phase and intensity noise of classical light sources and quantum fluctuations on the single-photon level.

Entropy sources are essential for probabilistic computing architectures, enabling sampling from probability density functions. While digital architectures rely on costly pseudorandom number generation, photonic processors inherently provide this capability. This section explores three optical true random number generation mechanisms, sketched in Fig. 2, which form the foundation of probabilistic computing architectures.

Phase noise

Pure phase noise manifests in a finite coherence time of an optical source while the second order of coherence, given by the power fluctuations, remains unity for all time delays. Practically, phase noise is present in laser systems and is primarily caused by spontaneous emission within the cavity in two ways. First, spontaneous emission can directly change the phase of the outgoing laser beam; second, spontaneous emission can deplete the excited state of the gain medium and thus randomly change the refractive index along the laser cavity[44]. The second effect is mostly present in semiconductor lasers[45]. As phase noise at optical frequencies around 200 THz is not directly measurable via an intensity measurement, interferometric measurement schemes are deployed, as illustrated in Fig. 2a. By superimposing a single laser with a delayed copy of itself (homodyne detection) or utilizing the difference between the phase of two separate lasers (heterodyne detection), computing tasks, such as physical random number generation, can be realized[15,46–50]. While true random number generators based on photonic phase noise have been successfully realized, their deployment within computation schemes has been elusive up until now, as lasers are often optimized to exhibit a low phase noise and linewidth—for example, for sensing—instead of featuring a large noise bandwidth for probabilistic computing. Naturally, physical computation schemes deploying the phase as an information carrier, for example Mach–Zehnder interferometer meshes[10], would be a suitable fit.

Intensity noise

In contrast to phase noise, intensity noise leads to a non-constant second order of coherence, which describes the autocorrelation of the intensity for classical fields, in addition to a finite coherence time, and is directly measurable with a photodetector. Intensity noise can be easily generated over a large bandwidth by exploiting photon bunching in chaotic light sources—for example, amplified spontaneous emission (ASE) in erbium-doped fibers, laser diodes with feedback[51], superluminescent light-emitting diodes or thermal light sources. The intensity noise arises

https://doi.org/10.1038/s43588-025-00800-1

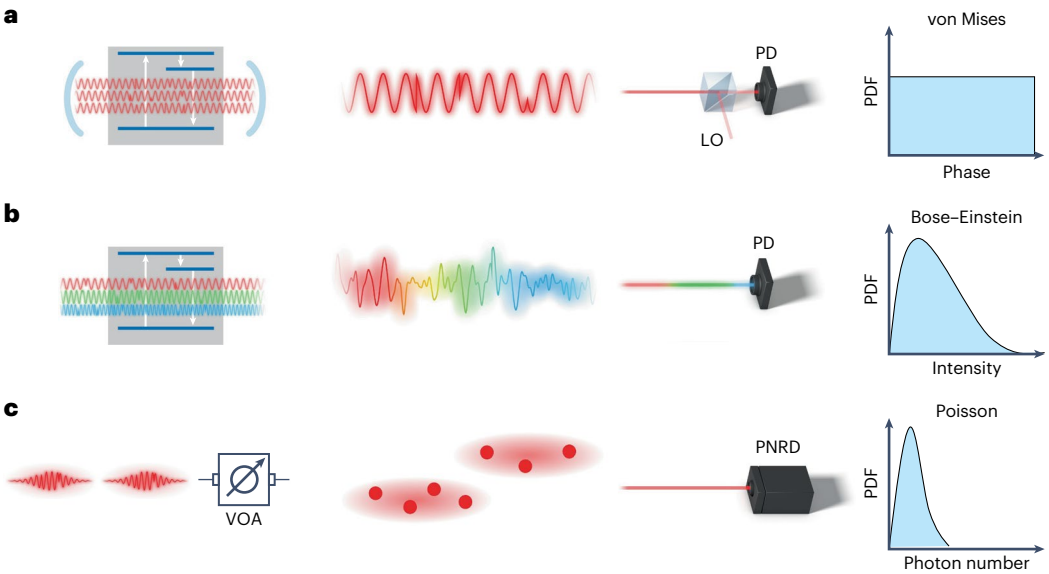

Fig. 2 | Optical random number generation. a, Spontaneous emission within laser cavities induces phase noise in the outgoing coherent laser beam, defining its linewidth. The phase noise can be extracted via homodyne or heterodyne detection schemes deploying default photodetectors (PDs) and follows a von Mises distribution. The probability density function (PDF) is uniform for mixing with an uncorrelated local oscillator (LO). **b**, Chaotic light states emitted by thermal light sources or generated by ASE can span large frequency ranges—for example, covering the full visible spectrum. They inherently feature power fluctuations due to the random beating between the various frequency components of the light state. The fluctuations can be directly measured with a photodetector and follow an M-fold Bose–Einstein distribution. **c**, Quantum fluctuations can serve as a perfect source of randomness on a single-photon level—for example, if the light state is not an eigenstate of the measurement operator. Thus, an optical random number generator might be directly implemented by generating a (weakly) coherent state with a (pulsed) laser and a variable optical attenuator (VOA) and measuring the state with a photon-number resolving detector (PNRD). The mean photon number per pulse is defined by the laser power and the VOA, but the measured number of photons in each pulse will follow Poisson statistics.

from the chaotic beating between the various frequency components of the chaotic light state resulting in an M-fold Bose–Einstein distribution, as sketched in Fig. 2b. The degeneracy factor M depends on the number of independent coherence cells within the measurement time and can be experimentally tuned by choosing the optical bandwidth of the chaotic carrier and the bandwidth of the electronic detection circuit[52]. An important property is that the s.d. of the power fluctuations is proportional to the mean power, $\sigma \approx \mu/\sqrt{M}$, thus the s.d. of the probability distribution can be modified at optical data transmission frequencies easily exceeding 10 GHz (ref. 16). As the correlation time of the fluctuation is inversely proportional to the optical bandwidth, ultrahigh-speed physical random number generation can be realized[14,53–55]. There exist two approaches for direct integration into photonic processors. One is to build a mixed system using both coherent light states and chaotic states. In this way, the low-intensity noise of the coherent state may represent the mean of the distribution whereas the chaotic light state leads to noise around some mean, together enabling the modeling of distributions[43]. The drawback is the complexity of the physical system, as different states of light with different bandwidths need to be properly tuned with respect to each other and propagated through the system, practically limiting the scalability. Another approach would be to solely rely on CHAOTIC light states, markedly simplifying optical management in the circuit because chaotic light sources tend to be easier to realize as no cavity is required, and the same light source can be deployed for each information carrier instead of sufficiently detuned coherent lasers[16].

Quantum fluctuations

Apart from exploiting the randomness of macroscopic systems manifested in the coherence properties of coherent and chaotic light at high photon numbers, the intrinsic randomness of quantum mechanics can be directly harnessed. For example, weak coherent states measured in the photon-number basis exhibit a Poisson distribution, which can be exploited as a quantum random number generator as shown in Fig. 2c. Eaton et al., for instance, used transition-edge sensors as photon-number-resolving detectors to measure weak coherent states with photon numbers ranging from 0 to 100, thereby generating unbiased random numbers[56]. Similar experiments have been conducted using silicon photomultipliers, and these systems may eventually be fully integrated with photon-number-resolving superconducting nanowire single-photon detectors[57,58]. There are two approaches to leveraging these quantum fluctuations—that is, the measurement of light states that are not eigenstates of the measurement operator—for computation. First, a processor might be designed to compute directly on quantum states, which leads to the field of photonic quantum computing, as demonstrated, for example, by Madsen et al.[59]. Second, classical light states could be coupled to quantum states, allowing the measured fluctuations to be tuned. In this case, the fluctuations would then be coupled to a macroscopic quantity, resulting in the measured probability distribution[60]. The work of Choi et al. characterizes the basic required components and proposes a potential design for a future all-optical system, although a functional prototype is not realized yet[61].

All three entropy sources eventually enable probabilistic computing, but require different computing architectures: for example, phase noise will not be directly usable in incoherent, purely intensity-based computation schemes. Comparing the three different approaches, leveraging quantum fluctuations requires more sensitive readout electronics to ensure that the tunable photonic noise is not covered by static electronic ground noise. Regarding the two classical approaches, generating chaotic light states with high-bandwidth intensity noise only requires a pumped gain medium, strongly simplifying circuit design and packaging.

Probabilistic processors

Deterministic architectures augmented with (pseudo) random number generators are a powerful framework for various application fields

https://doi.org/10.1038/s43588-025-00800-1

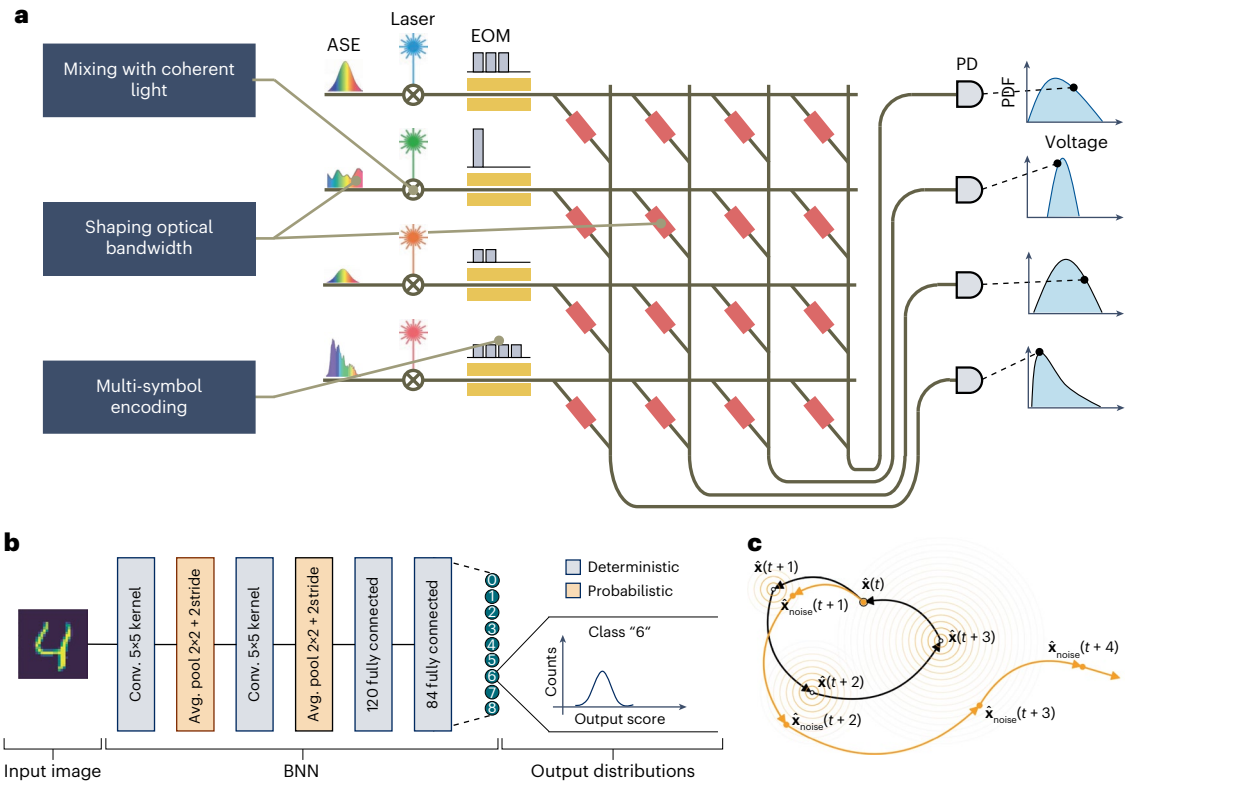

Fig. 3 | Photonic probabilistic computing. a, There exist several possibilities to enhance power-encoding-based photonic crossbar arrays with tunable stochasticity. First, we can mix coherent carriers (lasers) with chaotic carriers (ASE sources) and use the power ratio to encode the mean and variance of the distributions. Second, we can rely on chaotic carriers alone and modify the degeneracy factor of the underlying Bose–Einstein distributions. Since the optical bandwidth determines the properties of the distribution, shaping the optical bandwidth, that is, applying a wavelength-dependent filter, enables noise tunability. Since the noise bandwidth is larger than the modulation speed, subsequent input symbols are uncorrelated. Thus, using several symbols to encode the input enables tuning of the mean, given by the sum of the symbols, and tuning of the s.d., given by the number of symbols used. The full probabilistic MVM is implemented by a default optical transmission measurement—for example, with a photonic crossbar array. Here, each input pulse, corresponding to one input vector component, is equally distributed to the four output columns and the power individually attenuated by tunable attenuators marked in red. After weighting, all pulses in one column are summed by superposition and measured by a photodetector to obtain the result of one scalar product. Reading out all four photodetectors gives the result of the full MVM. **b**, The probabilistic processor can be deployed to implement a probabilistic layer in a feedforward architecture such as a BNN. **c**, Injecting noise in recurrent systems such as Hopfield networks or those based on vector symbolic architectures greatly improves their performance. Deterministically updating the state $\hat{\mathbf{x}}$ in each time state (black line) can result in convergence to a local minimum. In contrast, injecting noise at each step increases the search space, helping the system escape those minima (orange line). Due to the sequential nature (the second timestep can only be computed after the first), these architectures also strongly benefit from the low latency and high parallelism of photonic processors. Panels reproduced from: **b**, ref. 16 under a Creative Commons licence CC BY 4.0; **c**, ref. 68, Springer Nature Limited.

including Bayesian neural networks (BNNs)[62,63], generative models[64,65] and heuristic methods[66]. Due to the deterministic nature of digital processing, pseudorandom number generation imposes a substantial bottleneck that is detrimental to real-time tasks such as confidence estimation during autonomous driving and might induce unreasonable convergence times for heuristic optimization. Inspired by the progress in analog, electric probabilistic computing for BNNs[41,67] and in-memory optimization[68,69], we lend our perspective on the utilization of integrated photonic probabilistic processors deploying chaotic light as the entropy source. Such processors concurrently allow for MVM and massively parallel random number generation in a single clock cycle, providing a marked speedup in comparison with pseudorandom number generation and material-based entropy sources[8,16]. Figure 3a sketches multiple methods to obtain tunable stochasticity in a photonic crossbar array solely based on power encoding. First, we can mix chaotic light with coherent light and thus control the mean and variance of the distributions by adapting the power ratio between them[43]. To reduce the complexity of the circuit, we can also utilize solely chaotic carriers[32]. However, this directly links the mean to the s.d. of the distributions that are encoded, strongly limiting the flexibility. As the degeneracy factor of the underlying Bose–Einstein distribution depends on the number of independent coherence cells within the measurement interval, we need to access this quantity. Multisymbol waveform encodings can effectively change the measurement interval at GS s^{-1} rates by adapting the pulse length and height[16]. In contrast, changing the optical bandwidth will change the size of the coherence cells without requiring multisymbol encodings.

Table 1 provides an overview of various probabilistic photonic computing prototypes at different development stages. Scaling benefits from using a single carrier (bandwidth), as this simplifies carrier generation, circuit design and system stabilization. Probabilistic expressiveness also affects scalability—for example, generating static output noise for Ising models is simpler than implementing tunable stochastic matrix weights for BNNs. All approaches achieve/envision sampling rates in the gigasamples per second range, eliminating the bottleneck found in digital architectures. While a digital central processing unit can, for example, also perform matrix multiplications at gigasamples per second rates, the probabilistic sampling operations slow overall performance to megasamples per second rates[16]. The key advantage of probabilistic photonic hardware is its ability to overcome this limitation. The following sections explore models that could benefit from this advancement.

Table 1 | Overview of different probabilistic photonic architectures

	Entropy source	Application	Sampling rate	Scalability
Choi et al.[61]	Quantum vacuum noise	Stochastic binary neural networks	10 kS s^{-1}, 1 GS s^{-1a}	Free-space prototype
Wu et al.[17]	ASE–ASE beating	Generative adversarial networks	4 kS s^{-1}, 40 GS s^{-1a}	No. of carriers$\propto$O(N)
Brückerhoff-Plückelmann et al.[16]	ASE–ASE beating	Stochastic BNN inputs	70.4 GS s^{-1}	No. of carriers$\propto$O(1)
Wu et al.[43]	ASE–signal beating	Stochastic BNN weights	40 GS s^{-1a}	No. of carriers$\propto$O(N)
Lightelligence[93]	Electro-optic interface noise	Ising problem	1 GS s^{-1}	No. of carriers: $\propto$O(1)

In this table, we compare different architectures in terms of entropy source, application, sampling rate and scalability. The number of different optical carriers within the system is a limiting factor for the scalability, as more carriers must be generated in different frequency bands and the spectral response of the circuit must be stabilized. [a]Estimated/extrapolated values.

Probabilistic deep neural networks

Combining deep neural networks (DNNs) with random number generators for probabilistic sampling can greatly improve the training process (for example, by augmenting the input data or enforcing a smoother latent space in encoder–decoder architectures), and can enhance the performance of generative models. Probabilistic photonic computing, with its ability to simultaneously perform ultrafast random number generation and computation, offers great potential to accelerate these networks. Here we will discuss the benefits of probabilistic photonic computing on generative tasks, focusing on diffusion models and variational autoencoders (VAEs).

Diffusion models for image generation

Diffusion models[64] are widely used in generative modeling for producing high-fidelity images, such as those in DALL-E[70]. These models work by iteratively adding Gaussian noise to data and then learning to reverse the noise process to generate new samples. During inference, diffusion models require extensive sampling from noise distributions at each timestep and passing of these samples through a trained convolutional neural network, often a U-Net[71]. This sequential process makes image generation computationally intensive.

Probabilistic photonic computing can accelerate diffusion models by enabling ultrafast, high-quality random number generation, as shown in Fig. 2. Additionally, depending on matrix sizes and data flow, convolutional and pooling operations following sampling can be implemented optically at speeds of tens of gigasamples per second. Recent demonstrations of optical nonlinearity in silicon photonics[72] also enable sequential neural network computation directly within the optical domain, reducing unnecessary conversions between optics and electronics. This acceleration could enable real-time image generation and expand the use of diffusion models in time-sensitive applications such as video synthesis and interactive media.

Variational autoencoders for data imputation

VAEs[65] are a class of generative models that use a neural network (encoder) to transform high-dimensional input data into a lower-dimensional latent space, typically represented by probability distributions, such as Gaussian distributions. Samples are drawn from these distributions and passed through a decoder network to reconstruct the input data, enabling the generation of new samples. For example, datasets often have missing or incomplete data due to limitations in measurement technologies, sample quality or cost constraints. Here, VAEs can be used for data imputation by learning the underlying distributions in the latent space[73,74].

Photonic processors are intriguing for VAEs, especially if the input state is optical. Instead of measuring the full input state, an optical encoder block compresses the state first, and only the latent space is captured and further processed digitally[75]. Similarly, optical encoder and decoder blocks can enhance the performance of fiber transmission systems[76]. While these systems exhibit noise due to their analog nature, probabilistic photonic processors also enable tunable noise in the latent space and could be generalized to arbitrary scenarios. As for diffusion networks, implementing the neural network in this context poses challenges in synchronizing the hybrid optoelectronic approach to maximize the use of computing resources.

Bayesian neural networks

BNNs fundamentally differ from other (probabilistic) DNNs in the sense that the network parameters are given by the distribution inferred from the known training data and not by a minimal cost point estimate[63]. Introducing probabilistic mappings within the architecture allows for additional degrees of freedom, for modeling uncertainty, as illustrated in Fig. 3b. This probabilistic approach helps BNNs quantify both aleatoric (data-related) and epistemic (model-related) uncertainty, making them particularly useful in safety-critical and decision-making applications[63]. However, the sequential operations and large state associated with pseudorandom number generation pose a substantial bottleneck. For example, sampling rates for Mersenne Twister pseudorandom number generation are only on the order of 10 MS s^{-1}. Physical computing can remove this bottleneck, as it is inherently probabilistic. In the electronic domain, Bonnet et al. exploited the programming noise of memristors and phase-change materials within a crossbar array (Fig. 3a) to implement a BNN[67]. To avoid the latency and potentially limited endurance associated with continuous reprogramming, Liu et al. utilized the thermal noise in magnetic tunnel junctions to enable probabilistic in-memory computing with a sampling rate of 500 MS s^{-1} (ref. 8) In contrast, photonics offers the potential for sampling rates exceeding 10 GS s^{-1} (Table 1), as the correlation time of the entropy source is linked not to material properties but to the coherence properties and bandwidth of the optical carrier, which easily exceeds 1 THz (ref. 16). The drawback is that the memory density of integrated photonic processors is substantially smaller than that of their electronic counterparts, limiting the size of probabilistic mappings. Similarly, a large portion of the overall architecture likely needs to be deterministic to provide the required memory capacity, with only a few probabilistic mappings. This trend aligns well with ongoing research in BNNs, which focuses on the number and placement of probabilistic layers within the architecture to maintain sufficient functionality[77]. In the following we present two real-world examples where both reasoning about uncertainty is crucial and the resources available for predictions are limited. Here the combination of BNNs, which enable reliable reasoning about uncertainty, and fast and efficient photonic computing is extremely promising.

Computer vision models for safe autonomous driving

Autonomous driving is a core application of modern AI systems, where computer vision models have become prevalent due to their practical advantages in terms of hardware and sensor requirements. Some manufacturers, such as Tesla, have even developed entire autonomous driving systems that rely solely on camera input. For safety-critical

applications, uncertainty estimation in predictions based on visual inputs is essential, as it enables the system to gauge confidence in its predictions. As a specific example, Kendall and Gal demonstrate uncertainty estimation in monocular depth regression and semantic segmentation using BNNs[4]. Here incorporating uncertainty helps to assess prediction reliability. This capacity helps to identify when additional caution or alternative actions are necessary, enhancing overall safety.

Bayesian models for medical imaging support

High-resolution medical imaging, such as magnetic resonance imaging and computed tomography scans, plays a crucial role in guiding medical professionals' attention to potential anomalies or critical artifacts within the body. In this safety-critical domain, providing real-time diagnostic support is highly valuable for enabling interactive and precise diagnoses. However, delivering this level of responsiveness is challenging due to the immense data volume inherent to high-resolution images. Lambert et al. provide an overview of how BNNs can allow physicians to understand the reliability of the model's predictions and aid in decision-making[78]. Photonic probabilistic computing has the potential to make real-time feedback feasible.

Energy-based neural networks and resonator networks

Energy-based neural networks and resonator networks reuse output states as input states, enabling a temporal evolution of the system. These architectures are particularly intriguing for photonic computing when the dominant computational workload involves (probabilistic) MVMs with constant matrix elements[79]. Hopfield networks and Boltzmann machines, for which the Nobel Prize in Physics was awarded in 2024[80], highlight their compatibility with photonic computing[81,82]. Both are fully connected recurrent architectures that can be used, for example, as an associative memory or for optimization tasks. Thus, the latency for a single MVM is crucial, a good match for photonic processors operating at subnanosecond MVM latencies. Depending on the application, network parameters can be trained using, for instance, the Hebbian learning rule for associative memory[83], or they may be directly derived from the energy function that describes the optimization problem[69]. Since the energy function of Hopfield networks and Boltzmann machines resembles that of the Ising model, a variety of problems, such as the traveling salesman or max-cut problem, can be approximately solved using these recurrent architectures[84–87]. Hopfield networks evolve deterministically, posing a high risk of converging to local minima in the energy landscape. In contrast, the probabilistic Boltzmann machine can escape these minima[87,88]. Figure 3c illustrates the temporal evolution in a deterministic and probabilistic architecture. Due to the noise injection in each timestep, the probabilistic architecture can eventually diverge from the local minima. In addition to classical Hopfield networks and Boltzmann machines, emerging vector symbolic architectures and resonator networks are an intriguing application for physical probabilistic computing as MVMs are the dominant computational workload[68]. These architectures use quasi-orthogonal vectors, also known as code vectors, to encode different symbols. The quasi-orthogonality enables parallel computation on the superposition of different symbols, which greatly improves the performance for high-dimensional factorization[89,90]. Moreover, neurosymbolic AI, which combines standard ANNs and vector symbolic architectures, allow for abstract reasoning and thus set new state-of-the-art performance in fluid intelligence tests such as Raven's progressive matrices[91].

Probabilistic recurrent architectures impose two major computational challenges. First, their operation is sequential, meaning that latency must be minimized, in contrast to maximizing throughput in parallel architectures. Second, probabilistic sampling is required at each timestep. Physical in-memory computing is ideal for reducing latency since it is limited only by clock speed at the input/output interface and during postprocessing, and not at the computation stage. Khaddam-Aljameh et al. demonstrated MVMs with a 256 × 256 matrix within 127 ns on an electronic crossbar array[92], while photonics holds promise for subnanosecond latency due to higher bandwidth and ideal memory properties that enable pulse amplitude modulation instead of pulse width modulation. Prototypes, such as the Photonic Arithmetic Computing Engine from Lightelligence, achieve 150 ps optical latency for a 64 × 64 MVM and a latency of 3 ns for a single probabilistic iteration of the corresponding Ising problem[93]. Integrating a photonic entropy source directly into processors, as depicted in Fig. 3a, eliminates the need for electronic probabilistic sampling and even enables stochasticity that is tunable during the temporal evolution of the system.

Future trajectories of probabilistic photonic processors

Since noise is inherent to all physical processes that underlie computation, an exciting direction is to explore options that do not suppress noise but rather leverage it as a source of stochasticity. For instance, noise can be interpreted as random variability that facilitates stochastic sampling, potentially reducing the need for complex and computationally expensive sampling techniques typically required in purely digital implementations. By reconceptualizing noise as a resource rather than an error, photonic analog computing may open new pathways for applications that rely on probabilistic reasoning, particularly in energy-constrained environments where digital sampling methods may be impractical.

However, developing photonic entropy sources and processors from research demonstrators to proof-of-concept prototypes that can have an impact on present-day computational challenges requires a sophisticated design flow, as sketched in Fig. 4. Defining application requirements is the foundational element that shapes the entire design flow. Key factors include the type of problem in terms of network architecture, availability of training data, enforced reliability standards, and inference performance targets, such as throughput, latency and efficiency—all of which will influence the overall design.

From a system design perspective, hardware non-idealities, such as fabrication imperfections, can negatively impact the performance of analog photonic circuits. Therefore, the software must account for these uncertainties during training. Hardware-aware training, which incorporates simulation or experimental data from devices, is gaining traction as a robust solution[94,95]. If the system exhibits noise sources and device variabilities that cannot be captured during hardware-aware training, in situ training[22,24,63,96–98] is another promising approach. Here backpropagated gradients are experimentally calculated on chip, inherently accounting for hardware noise during training.

Moreover, control algorithms must be implemented in software to manage temperature fluctuations in sensitive optical components such as ring resonators. For instance, algorithms such as proportional–integral–derivative control can be implemented more compactly on chip, using photodetectors as integrators, requiring careful consideration before chip tapeout. Peripheral hardware choices, including the laser source, receiver and data converters, will also affect the system's effective number of bits[99]. Similarly, hardware-aware training based on techniques such as quantization-aware training or post-training quantization is required to maintain inference accuracy at lower effective number of bits.

Finally, the heterogeneous integration of optical sampling, photonic tensor cores, electronic tensor cores and shared memory resources necessitates complex scheduling, partitioning and pipelining during inference. These operations are often handled by off-chip peripherals, such as field-programmable gate arrays, limiting demonstrations to smaller scales. Large-scale in-memory computing architectures for DNN workloads can be realized by connecting multiple analog crossbar arrays, digital compute cores and memory cores[100,101]. This architecture features a dense 2D mesh of on-chip interconnects

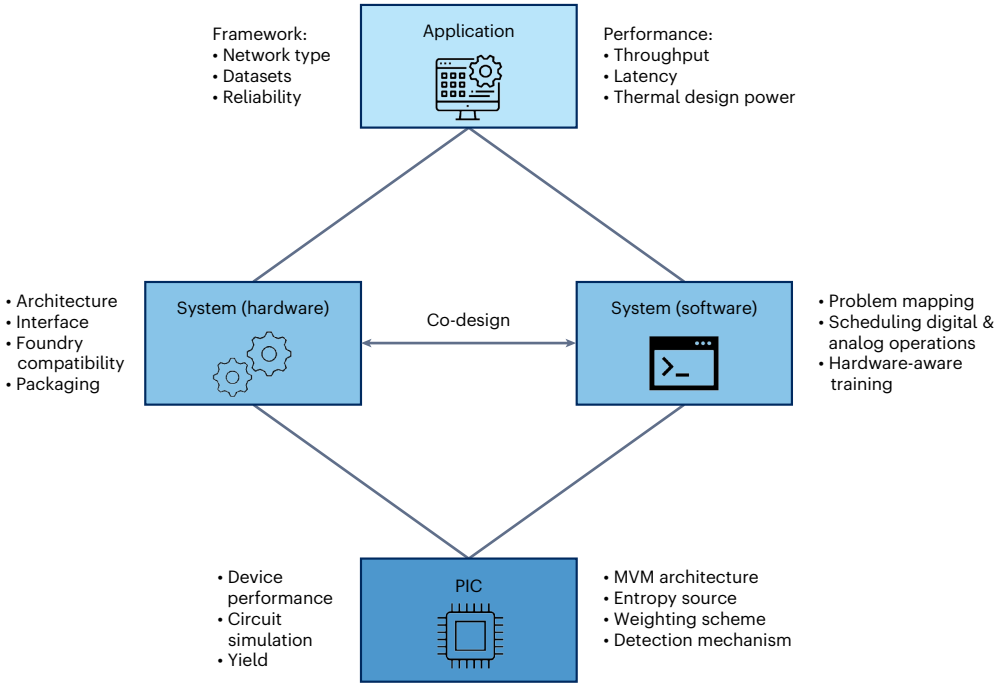

Fig. 4 | Hardware–software co-design. Developing probabilistic photonic processors from research experiments to proof-of-concept prototypes requires a sophisticated design strategy. First, the top application layer defines system requirements in terms of performance metrics and the general framework. The middle layer involves the interdependent development of hardware and software. Co-design ensures that the hardware aligns with software requirements, while the software accounts for hardware limitations and imperfections. Finally, the system architecture determines the design parameters of the photonic integrated circuit (PIC). On the basis of computing needs, such as memory density and noise encoding flexibility, the appropriate MVM architecture and photonic implementation are selected.

for ultrafast communication between cores, along with power management solutions and resource allocation for various DNNs across multiple chips. Similar architectures could be envisioned for photonic probabilistic computing, adding further complexity to the software–hardware co-design with the inclusion of sampling operations.

Overall, the combination of probabilistic and photonic computing presents promising avenues for enabling more secure and accurate machine learning algorithms, as well as realizing their deployment on fast and efficient hardware. While current examples are often at the laboratory-demonstration stage, the groundwork has been laid for future large-scale implementations. Interdisciplinary collaboration between photonic system architects and computer scientists in conjunction with industry partners to define the application requirements and provide the required devices on a scalable foundry level will be crucial to enable the next generation of proof-of-concept prototypes.

References

1. Rao, Q. & Frtunikj, J. Deep learning for self-driving cars: chances and challenges. In *2018 IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS)* 35–38 (IEEE, 2018).
2. Ker, J. & Wang, L. Deep learning applications in medical image analysis. *IEEE Access* **6**, 9375–9389 (2018).
3. Arkhangelskaya, E. O. & Nikolenko, S. I. Deep learning for natural language processing: a survey. *J. Math. Sci.* **273**, 533–582 (2023).
4. Kendall, A. & Gal, Y. What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems 30*, 5575–5585 (NIPS, 2017).
5. Le Gallo, M. et al. A 64-core mixed-signal in-memory compute chip based on phase-change memory for deep neural network inference. *Nat. Electron.* **6**, 680–693 (2023).
6. Friston, K. et al. The free energy principle made simpler but not too simple. *Phys. Rep.* **1024**, 1–29 (2023).
7. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11**, 127–138 (2010).
8. Liu, S. et al. Bayesian neural networks using magnetic tunnel junction-based probabilistic in-memory computing. *Front. Nanotechnol.* **4**, 1021943 (2022).
9. Feldmann, J. et al. Parallel convolutional processing using an integrated photonic tensor core. *Nature* **589**, 52–58 (2021).
10. Shen, Y. et al. Deep learning with coherent nanophotonic circuits. *Nat. Photon.* **11**, 441–446 (2017).
11. Xu, X. et al. 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* **589**, 44–51 (2021).
12. Brückerhoff-Plückelmann, F. et al. Broadband photonic tensor core with integrated ultra-low crosstalk wavelength multiplexers. *Nanophotonics* **11**, 4063–4072 (2022).
13. Wu, C. et al. Programmable phase-change metasurfaces on waveguides for multimode photonic convolutional neural network. *Nat. Commun.* **12**, 96 (2021).
14. Cao, G., Zhang, L., Huang, X., Hu, W. & Yang, X. 16.8 Tb/s true random number generator based on amplified spontaneous emission. *IEEE Photon. Technol. Lett.* **33**, 699–702 (2021).
15. Huang, M., Chen, Z., Zhang, Y. & Guo, H. A phase fluctuation based practical quantum random number generator scheme with delay-free structure. *Appl. Sci.* **10**, 7 (2020).
16. Brückerhoff-Plückelmann, F. et al. Probabilistic photonic computing with chaotic light. *Nat. Commun.* **15**, 10445 (2024).
17. Wu, C. et al. Harnessing optoelectronic noises in a photonic generative network. *Sci. Adv.* **8**, eabm2956 (2022).
18. Mehonic, A. & Kenyon, A. J. Brain-inspired computing needs a master plan. *Nature* **604**, 255–260 (2022).
19. Zhou, H. et al. Photonic matrix multiplication lights up photonic accelerator and beyond. *Light Sci. Appl.* **11**, 30 (2022).
20. Schuman, C. D. et al. Opportunities for neuromorphic computing algorithms and applications. *Nat. Comput Sci.* **2**, 10–19 (2022).

21. Marković, D., Mizrahi, A., Querlioz, D. & Grollier, J. Physics for neuromorphic computing. *Nat. Rev. Phys.* **2**, 499–510 (2020).
22. Wang, T. et al. An optical neural network using less than 1 photon per multiplication. *Nat. Commun.* **13**, 123 (2022).
23. Sludds, A. et al. Delocalized photonic deep learning on the internet's edge. *Science* **378**, 270–276 (2022).
24. 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. *Nat. Commun.* **16**, 359 (2025).
25. 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).
26. Brückerhoff-Plückelmann, F. et al. Event-driven adaptive optical neural network. *Sci. Adv.* **9**, eadi9127 (2023).
27. Tait, A. N. et al. Neuromorphic photonic networks using silicon photonic weight banks. *Sci. Rep.* **7**, 7430 (2017).
28. Li, G. H. Y. et al. All-optical, ultrafast energy-efficient ReLU function for nanophotonic neural networks. *Nanophotonics* **12**, 847–855 (2022).
29. Grottke, T., Hartmann, W., Schuck, C. & Pernice, W. H. P. Optoelectromechanical phase shifter with low insertion loss and a 13π tuning range. *Opt. Express* **29**, 5525–5537 (2021).
30. Ríos, C. et al. In-memory computing on a photonic platform. *Sci. Adv.* **5**, eaau5759 (2019).
31. Xu, R. et al. Mode conversion trimming in asymmetric directional couplers enabled by silicon ion implantation. *Nano Lett.* **24**, 10813–10819 (2024).
32. Dong, B. et al. Partial coherence enhances parallelized photonic computing. *Nature* **632**, 55–62 (2024).
33. Hamerly, R., Bandyopadhyay, S. & Englund, D. Asymptotically fault-tolerant programmable photonics. *Nat. Commun.* **13**, 6831 (2022).
34. Tait, A. N. et al. Microring weight banks. *IEEE J. Sel. Top. Quantum Electron.* **22**, 312–325 (2016).
35. Hu, J. et al. Diffractive optical computing in free space. *Nat. Commun.* **15**, 1525 (2024).
36. Farhat, N. H., Psaltis, D., Prata, A. & Paek, E. Optical implementations of the Hopfield model. *Appl. Opt.* **24**, WB3 (1985).
37. Shastri, B. J. et al. Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* **15**, 102–114 (2021).
38. Bogaerts, W. et al. Silicon microring resonators. *Laser Photon Rev.* **6**, 47–73 (2012).
39. Messner, A. et al. Plasmonic, photonic, or hybrid? Reviewing waveguide geometries for electro-optic modulators. *APL Photon.* **8**, 10 (2023).
40. Bose, D. et al. Anneal-free ultra-low loss silicon nitride integrated photonics. *Light Sci. Appl.* **13**, 156 (2024).
41. Sebastian, A. et al. Two-dimensional materials-based probabilistic synapses and reconfigurable neurons for measuring inference uncertainty using Bayesian neural networks. *Nat. Commun.* **13**, 6139 (2022).
42. Dutta, S. et al. Neural sampling machine with stochastic synapse allows brain-like learning and inference. *Nat. Commun.* **13**, 2571 (2022).
43. Wu, C., Yang, X., Chen, Y. & Li, M. Photonic Bayesian neural network using programmed optical noises. *IEEE J. Sel. Top. Quantum Electron.* https://doi.org/10.1109/JSTQE.2022.3217819 (2023).
44. Tran, M. A., Huang, D. & Bowers, J. E. Tutorial on narrow linewidth tunable semiconductor lasers using Si/III–V heterogeneous integration. *APL Photon.* **4**, 11 (2019).
45. Henry, C. H. Theory of the linewidth of semiconductor lasers. *IEEE J. Quantum Electron* **18**, 259–264 (1982).
46. Lovic, V., Marangon, D. G., Lucamarini, M., Yuan, Z. & Shields, A. J. Characterizing phase noise in a gain-switched laser diode for quantum random-number generation. *Phys. Rev. Appl.* **16**, 054012 (2021).
47. Álvarez, J.-R., Sarmiento, S., Lázaro, J. A., Gené, J. M. & Torres, J. P. Random number generation by coherent detection of quantum phase noise. *Opt. Express* **28**, 5538 (2020).
48. Qi, B., Chi, Y.-M., Lo, H.-K. & Qian, L. High-speed quantum random number generation by measuring phase noise of a single-mode laser. *Opt. Lett.* **35**, 312–314 (2010).
49. Nie, Y. Q. et al. The generation of 68 Gbps quantum random number by measuring laser phase fluctuations. *Rev. Sci. Instrum.* **86**, 063105 (2015).
50. Guo, H., Tang, W., Liu, Y. & Wei, W. Truly random number generation based on measurement of phase noise of a laser. *Phys. Rev. E* **81**, 051137 (2010).
51. Sciamanna, M. & Shore, K. A. Physics and applications of laser diode chaos. *Nat. Photon.* **9**, 151–162 (2015).
52. Goodman, J. *Statistical Optics* (John Wiley & Sons, 2000).
53. Guo, Y. et al. 40 Gb/s quantum random number generation based on optically sampled amplified spontaneous emission. *APL Photon.* **6**, 6 (2021).
54. Zhang, L. et al. 640-Gbit/s fast physical random number generation using a broadband chaotic semiconductor laser. *Sci. Rep.* **8**, 4–11 (2017).
55. Shen, B. et al. Harnessing microcomb-based parallel chaos for random number generation and optical decision making. *Nat. Commun.* **14**, 4590 (2023).
56. Eaton, M. et al. Resolution of 100 photons and quantum generation of unbiased random numbers. *Nat. Photon.* **17**, 106–111 (2023).
57. Mattioli, F. et al. Photon-number-resolving superconducting nanowire detectors. *Supercond. Sci. Technol.* **28**, 10 (2015).
58. Aungskunsiri, K. et al. Quantum random number generation based on multi-photon detection. *ACS Omega* **8**, 35085–35092 (2023).
59. Madsen, L. S. et al. Quantum computational advantage with a programmable photonic processor. *Nature* **606**, 75–81 (2022).
60. Roques-Carmes, C. et al. Biasing the quantum vacuum to control macroscopic probability distributions. *Science* **381**, 205–209 (2023).
61. Choi, S. et al. Photonic probabilistic machine learning using quantum vacuum noise. *Nat. Commun.* **15**, 7760 (2024).
62. Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann, 1988).
63. Jospin, L. V. et al. Hands-on Bayesian neural networks — A tutorial for deep learning users. *IEEE Comput. Intell. Mag.* **17**, 29–48 (2022).
64. Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. *Adv. Neural Inf. Process. Syst.* **33**, 6840–6851 (2020).
65. Kingma, D. P. & Welling, M. Auto-Encoding Variational Bayes. *Camb. Explor. Arts Sci.* https://doi.org/10.61603/ceas.v2i1.33 (2014).
66. Fahlman, S. E., Hinton, G. E. & Sejnowski, T. J. Massively parallel architectures for AI: NETL, Thistle, and Boltzmann machines. In *Proc. AAAI-83 Conference* (AAAI-Press) 109–113 (1983).
67. Bonnet, D. et al. Bringing uncertainty quantification to the extreme-edge with memristor-based Bayesian neural networks. *Nat. Commun.* **14**, 7530 (2023).
68. Langenegger, J. et al. In-memory factorization of holographic perceptual representations. *Nat. Nanotechnol.* **18**, 479–485 (2023).
69. Sarwat, S. G., Kersting, B., Moraitis, T., Jonnalagadda, V. P. & Sebastian, A. Phase-change memtransistive synapses for mixed-plasticity neural computations. *Nat. Nanotechnol.* **17**, 507–513 (2022).

70. Ramesh, A. et al. Zero-shot text-to-image generation. *Proc. Mach. Learn. Res.* **139**, 8821–8831 (2021).
71. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI* (eds Navab, N. et al.) 12–20 (Lecture Notes in Computer Science 9351, Springer, 2015).
72. Ashtiani, F., Geers, A. J. & Aflatouni, F. An on-chip photonic deep neural network for image classification. *Nature* **606**, 501–506 (2022).
73. Qiu, Y. L., Zheng, H. & Gevaert, O. Genomic data imputation with variational auto-encoders. *GigaScience* **9**, giaa082 (2020).
74. McCoy, J. T., Kroon, S. & Auret, L. Variational autoencoders for missing data imputation with application to a simulated milling circuit. *IFAC-PapersOnLine* **51**, 141–146 (2018).
75. Wang, T. et al. Image sensing with multilayer, nonlinear optical neural networks. *Nat Photon.* **17**, 408–415 (2023).
76. Chen, Y. et al. Photonic unsupervised learning variational autoencoder for high-throughput and low-latency image transmission. *Sci. Adv.* **9**, eadf8437 (2023).
77. Sharma, M., Farquhar, S., Nalisnick, E. & Rainforth, T. Do Bayesian neural networks need to be fully stochastic? *Proc. Mach. Learn. Res.* **206**, 7694–7722 (2023).
78. Lambert, B., Forbes, F., Doyle, S., Dehaene, H. & Dojat, M. Trustworthy clinical AI solutions: a unified review of uncertainty quantification in Deep Learning models for medical image analysis. *Artif. Intell. Med.* **150**, 102830 (2024).
79. Syed, G. S. & Sebastian, A. Solving optimization problems with photonic crossbars. US patent US20230176606A1 (2021).
80. Gibney, E. & Castelvecchi, D. Physics Nobel scooped by machine-learning pioneers. *Nature* **634**, 523–524 (2024).
81. Roques-Carmes, C. et al. Heuristic recurrent algorithms for photonic Ising machines. *Nat. Commun.* **11**, 249 (2020).
82. Fan, Z., Lin, J., Dai, J., Zhang, T. & Xu, K. Photonic Hopfield neural network for the Ising problem. *Opt. Express* **31**, 21340 (2023).
83. Attneave, F., B, M. & Hebb, D. O. The organization of behavior; a neuropsychological theory. *Am. J. Psychol.* **63**, 633–642 (1950).
84. Mohseni, N., McMahon, P. L. & Byrnes, T. Ising machines as hardware solvers of combinatorial optimization problems. *Nat. Rev. Phys.* **4**, 363–379 (2022).
85. Lucas, A. Ising formulations of many NP problems. *Front. Phys.* **2**, 5 (2014).
86. Hopfield, J. J. & Tank, D. W. 'Neural' computation of decisions in optimization problems. *Biol. Cybern.* **52**, 141–152 (1985).
87. Aarts, E. H. L. & Korst, J. H. M. Boltzmann machines for travelling salesman problems. *Eur. J. Oper. Res.* **39**, 79–95 (1989).
88. Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
89. Frady, E. P., Kent, S. J., Olshausen, B. A. & Sommer, F. T. Resonator networks, 1: an efficient solution for factoring high-dimensional, distributed representations of data structures. *Neural Comput.* **32**, 2311–2331 (2020).
90. Kent, S. J., Frady, E. P., Sommer, F. T. & Olshausen, B. A. Resonator networks, 2: factorization performance and capacity compared to optimization-based methods. *Neural Comput.* **32**, 2332–2388 (2020).
91. Hersche, M., Zeqiri, M., Benini, L., Sebastian, A. & Rahimi, A. A neuro-vector-symbolic architecture for solving Raven's progressive matrices. *Nat. Mach. Intell.* **5**, 363–375 (2023).
92. Khaddam-Aljameh, R. et al. HERMES-Core-A 1.59-TOPS/mm2PCM on 14-nm CMOS in-memory compute core using 300-ps/LSB linearized CCO-based ADCs. *IEEE J. Solid-State Circuits* **57**, 1027–1038 (2022).
93. Hua, S. et al. An integrated large-scale photonic accelerator with ultralow latency. *Nature* **640**, 361–367 (2025).
94. Tsakyridis, A. et al. Photonic neural networks and optics-informed deep learning fundamentals. *APL Photon.* **9**, 1 (2024).
95. Rasch, M. J. et al. Hardware-aware training for large-scale and diverse deep learning inference workloads using in-memory computing-based accelerators. *Nat. Commun.* **14**, 5282 (2023).
96. Pai, S. et al. Experimentally realized in situ backpropagation for deep learning in photonic neural networks. *Science* **380**, 398–404 (2023).
97. Momeni, A., Rahmani, B., Malléjac, M., del Hougne, P. & Fleury, R. Backpropagation-free training of deep physical neural networks. *Science* **382**, 1297–1304 (2023).
98. Wright, L. G. et al. Deep physical neural networks trained with backpropagation. *Nature* **601**, 549–555 (2022).
99. Varri, A. et al. Noise-resilient photonic analog neural networks. *J. Lightwave Technol.* **42**, 7969–7976 (2024).
100. Jain, S. et al. A heterogeneous and programmable compute-in-memory accelerator architecture for analog-AI using dense 2-D mesh. *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.* **31**, 114–127 (2023).
101. Dazzi, M. et al. 5 Parallel Prism: a topology for pipelined implementations of convolutional neural networks using computational memory. In *33rd Conference on Neural Information Processing Systems* (NeurIPS 2019).

Acknowledgements

We thank J. Stuhrmann, from Illustrato, for his assistance with the illustrations. This research was supported by the European Union's Horizon 2020 research and innovation program (grant no. 101017237, PHOENICS project) and the European Union's Innovation Council Pathfinder program (grant no. 101046878, HYBRAIN project). We acknowledge funding support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2181/1—390900948 (the Heidelberg STRUCTURES Excellence Cluster), the Excellence Cluster 3D Matter Made to Order (EXC-2082/1—390761711) and CRC 1459 'Intelligent Matter'.

Author contributions

Conceptualization: F.B.-P., W.P. Methodology: F.B.-P., A.P.O., A.V., H. Borras, B.K., G.S.S. Investigation: F.B.-P., A.P.O., A.V., H. Borras, B.K., G.S.S. Visualization: F.B.-P., A.P.O., A.V. Funding acquisition: W.P., H.F., C.D.W., H. Bhaskaran, A.S., H.F. Project administration: W.P., H.F. Supervision: W.P., H.F., C.D.W., H. Bhaskaran, G.S.S., A.S. - Writing—original draft: F.B.-P., A.P.O., A.V. - Writing—review & editing: all authors.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence should be addressed to Wolfram Pernice.

Peer review information *Nature Computational Science* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Jie Pan, in collaboration with the *Nature Computational Science* team.