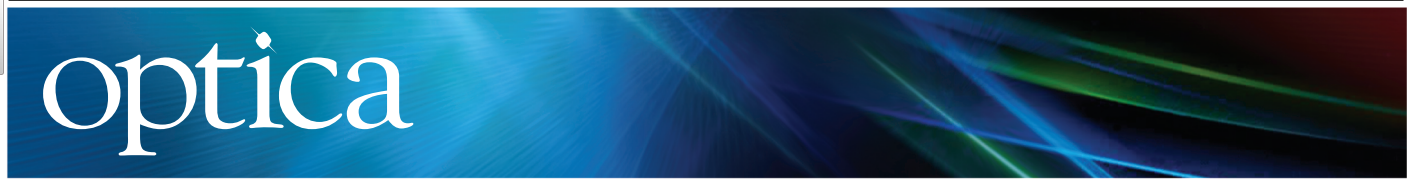
Research Article Vol. 7, No. 5 / May 2020 / Optica 551



Accelerating recurrent Ising machines in photonic integrated circuits

Mihika Prabhu,[1,†,*] Charles Roques-Carmes,[1,8,†] Yichen Shen,[2,4,9,†] Nicholas Harris,[1,3] Li Jing,[4] Jacques Carolan,[1] Ryan Hamerly,[1,5] Tom Baehr-Jones,[6] Michael Hochberg,[6] Vladimir Čeperić,[4] John D. Joannopoulos,[4,7] Dirk R. Englund,[1] and Marin Soljačić[1,4]

[1]MIT Research Lab of Electronics, 50 Vassar St, Cambridge, Massachusetts 02139, USA
[2]Lightelligence, 268 Summer Street, Boston, Massachusetts 02210, USA
[3]Lightmatter, 61 Chatham St 5th floor, Boston, Massachusetts 02109, USA
[4]MIT Department of Physics, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA
[5]NTT Research, 1950 University Ave #600, East Palo Alto, California 94303, USA
[6]Nokia Corporation, 171 Madison Ave 1100, New York, New York 10016, USA
[7]Institute of Soldier Nanotechnologies, 500 Technology Square, Cambridge, Massachusetts 02139, USA
[8]e-mail: chrc@mit.edu
[9]e-mail: ycshen@mit.edu
*Corresponding author: mihika@mit.edu

Received 23 December 2019; revised 4 April 2020; accepted 19 April 2020 (Doc. ID 386613); published 18 May 2020

Conventional computing architectures have no known efficient algorithms for combinatorial optimization tasks such as the Ising problem, which requires finding the ground state spin configuration of an arbitrary Ising graph. Physical Ising machines have recently been developed as an alternative to conventional exact and heuristic solvers; however, these machines typically suffer from decreased ground state convergence probability or universality for high edge-density graphs or arbitrary graph weights, respectively. We experimentally demonstrate a proof-of-principle integrated nanophotonic recurrent Ising sampler (INPRIS), using a hybrid scheme combining electronics and silicon-on-insulator photonics, that is capable of converging to the ground state of various four-spin graphs with high probability. The INPRIS results indicate that noise may be used as a resource to speed up the ground state search and to explore larger regions of the phase space, thus allowing one to probe noise-dependent physical observables. Since the recurrent photonic transformation that our machine imparts is a fixed function of the graph problem and therefore compatible with optoelectronic architectures that support GHz clock rates (such as passive or non-volatile photonic circuits that do not require reprogramming at each iteration), this work suggests the potential for future systems that could achieve orders-of-magnitude speedups in exploring the solution space of combinatorially hard problems.

https://doi.org/10.1364/OPTICA.386613

1. INTRODUCTION

Combinatorial optimization is critical for a broad array of tasks, including artificial intelligence, bioinformatics, cryptography, scheduling, trajectory planning, and circuit design [1–3]. However, combinatorial problems typically fall into the nondeterministic-polynomial hard (NP-hard) problem class, becoming computationally intractable at large scale for traditional algorithms. This challenge motivates the search for alternatives to conventional (von Neumann) computing architectures that can efficiently solve such problems. The Ising problem, which consists of finding the ground state spin configuration of a quadratic Hamiltonian defined by a symmetric matrix $\mathbf{K}$ and spins of unit amplitude $\sigma_{1\le i\le N}\in\{-1,1\}$,

$$H^{(K)}(\vec{\sigma})=-\frac{1}{2}\sum_{1\le i,j\le N}\sigma_i K_{ij}\sigma_j, \tag{1}$$

has garnered significant attention, as many other combinatorial problems can be polynomially reduced to an Ising problem [4,5]. Therefore, any technique for finding the ground state of arbitrary Ising problems, which is an NP-hard computational task, may extend to a wide range of other computationally intensive optimization problems.

There is currently no known efficient classical algorithm to find the exact ground state of an arbitrary Ising graph, so heuristic and meta-heuristic algorithms are often implemented as a means of quickly obtaining approximate solutions [6]. Various physical systems have been proposed as Ising machines, as the evolution of many natural systems (ferromagnets [7], lipid membranes [8], random photonic networks [9], etc.) can be described by Hamiltonians similar to the one in Eq. (1). Parallel machines provide additional advantages by reducing the correlation between consecutive samples and preventing premature trapping in local

minima by applying many local modifications simultaneously [10] and can also efficiently explore the phase space with many independent searches running in parallel [1,11,12].

This observation motivates the development of (re)configurable parallel analog machines, such as programmable nanophotonic processors [13–17]. The computational speedup with these machines is still polynomial (since analog machines also suffer from the $P = NP$ paradigm [18]). However, when operating at a fast clock rate (GHz), afforded by a photonic implementation, these optical architectures could enable orders-of-magnitude speedups against conventional solvers [17].

A number of photonic Ising machines have been developed in the past few years, based on various physical systems including degenerate optical parametric oscillators (OPOs) [19,20], fiber networks [21,22], injection-locked laser networks [23,24], and spatial multiplexing [25]. Some of these (classical) Ising machines even demonstrated performance superior to that of certain quantum annealers on dense graphs [26]. Experimental OPO demonstrations [20,27–29] and recent theoretical proposals [30] have shown great potential for finding optimal cuts of MAX-CUT problems with large numbers of spins. However, all-optical OPO machines face scaling limitations due to dispersion and decoherence of the time-division-multiplexed pulses in increasingly large resonators [28]. Additionally, some hybrid systems implement couplings via matrix multiplication in a field-programmable gate array (FPGA), and thus face scaling issues similar to those found in traditional computing for large-scale multiplication [20]. While massive spatial free-space multiplexing approaches [25] achieve a very large number of spins ($N \sim 10^4$), the spin couplings are limited to N degrees of freedom [instead of $O(N^2)$ for arbitrary Ising problems], and their clock rate is inherently slow (~kHz). These Ising machines also differ in the way they compute: while the vast majority of them belongs to the class of heuristic solvers, the OPO-based approach has been shown to operate as a negative-temperature simulated annealing machine [31], and the approach presented in Refs. [17,25] and in the present work belong to the class of (positive temperature) Markov chain Monte Carlo algorithms [1].

Here, we experimentally demonstrate a photonic recurrent Ising sampler for probabilistically finding the ground state of an arbitrary Ising problem. Using a programmable silicon-on-insulator nanophotonic processor [14,15,17,32–34], we leverage schemes for decomposing any unitary matrix into a mesh of linear optical components [35,36] to enable sampling of arbitrary Ising graphs. Furthermore, recent work on parallel photonic circuits for optical neural networks has demonstrated the capability for $O(N)$ time-scaling for vector-matrix multiplication with an array of tunable Mach–Zehnder interferometers (MZIs) [14]. While digital electronics such as application-specific integrated circuits (ASICs) or FPGAs can also exploit parallelization to achieve the same time scaling, the architecture we demonstrate is additionally compatible with quasi-passive photonic processing. As a result, the time per iteration can be reduced from 5–100 ns in parallelized digital hardware to 0.1–1 ns in passive integrated photonics [17].

2. METHODS

A. Theory

The proof-of-principle integrated nanophotonic recurrent Ising sampler (INPRIS) is based on a recently proposed parallel photonic recurrent network designed to find the ground state energies of Ising problems [17]. The conceptual structure of a single algorithm step of the N-spin photonic recurrent network is shown in Fig. 1(b). At time step t, the spin state vector $\vec{S}^{(t)} \in \{0, 1\}^N$ (equivalent to $\sigma_i^{(t)} = 2S_i^{(t)} - 1 \in \{-1, 1\}$) is encoded in the amplitudes of N coherent optical signals at the input. During each algorithm step, the optically encoded spin state vector propagates through an array of optical components that encodes an arbitrarily reconfigurable $N \times N$ optical matrix [14,15,35], $\mathbf{C}$, that is fixed and dependent on the problem-specific Ising coupling matrix, $\mathbf{K}$. The output of the matrix multiplication is noisy, being perturbed by a Gaussian noise source with standard deviation ϕ. This Gaussian perturbation to the signal could be realized with electrical or optical modulation of the refractive index, or via quantum detection noise [27]. The output of the noisy matrix multiplication is then fed to an optoelectronic threshold operation, $f_{\vec{\theta}_{\text{th}}}$, where $\vec{\theta}_{\text{th}}$ is a vector of threshold values, that converts the continuous output vector back into a binary spin state, $\vec{S}^{(t+1)} \in \{0, 1\}^N$. In combination, the sampler calculates the input spins for the subsequent algorithm step using the following expression:

$$\vec{S}^{(t+1)} = f_{\vec{\theta}_{\text{th}}}\left(\mathcal{N}\left(\mathbf{C}\vec{S}^{(t)}, \phi\right)\right), \tag{2}$$

where $\mathcal{N}(\mu, \phi)$ is the normal distribution with mean, μ, and standard deviation, ϕ.

After many algorithm steps, the output spin state probability distribution converges to an Ising Gibbs distribution (independent of the initial spin state) [17]:

$$\lim_{t\to\infty} P(\vec{\sigma}^{(t)}) \propto \exp[-\beta H_L(\vec{\sigma}^{(t)})], \tag{3}$$

$$H_L(\vec{\sigma}) = -\frac{1}{\beta}\sum_i \log\cosh\left(\beta\sum_j J_{ij}\sigma_j\right) \tag{4}$$

$$\approx \beta H^{(\mathbf{J}^T\mathbf{J})}(\vec{\sigma}), \tag{5}$$

where $\mathbf{C} = 2\mathbf{J}$ is the matrix implemented by the optical circuit, and $\beta = 1/(k\phi)$ (k being a constant that depends only on the noise distribution present in the system). The effective Hamiltonian of the system H_L can be approximated by an Ising Hamiltonian with weights $\mathbf{J}^2$ for small β (corresponding to large noise, ϕ) and symmetric $\mathbf{J}$. Therefore, to probe a symmetric Ising coupling matrix, $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}^\dagger$, with $\mathbf{U}\mathbf{D}\mathbf{U}^\dagger$ being the eigenvalue decomposition of $\mathbf{K}$, we program the optical matrix to implement the square root of $\mathbf{K}$. The square root motivates us to express the optically applied diagonal matrix in the form $\mathbf{D}_\alpha = \text{Re}(\sqrt{\mathbf{D} + \alpha\mathbf{\Delta}})$. The term $\alpha\mathbf{\Delta}$ is a diagonal offset matrix with $\mathbf{\Delta} > 0$, and α is a scalar dropout parameter that allows us to tune the dimensionality of the ground state search by selectively dropping out lower negative eigenvalues (the choice of parameters is described in Ref. [17]). In the following, we will refer to $\alpha = 1$ as *no dropout* (none of the eigenvalues is dropped out, as $\mathbf{D} + \alpha\mathbf{\Delta}$ is positive semidefinite) and $\alpha = 0$ as *dropout* (the negative eigenvalues of $\mathbf{D} + \alpha\mathbf{\Delta}$ are set to zero by taking the real part). In our algorithm, the eigenvalue dropout parameter is a fixed transformation on the coupling matrix, unlike dropout encountered in machine learning, which is a dynamic technique to prevent overfitting when training neural networks [37].

To illustrate an example ground state search of the parallel photonic network architecture, we simulate a two-dimensional

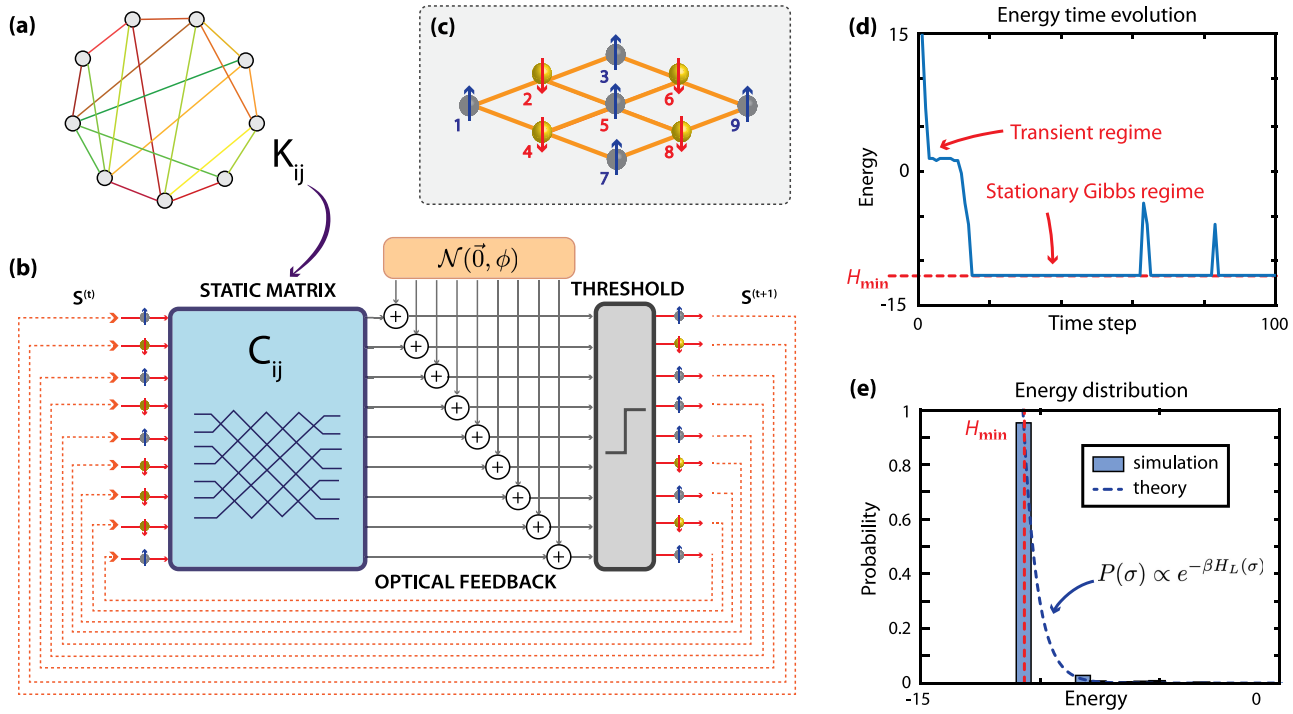


Fig. 1. Photonic recurrent Ising sampler. (a) An arbitrary Ising graph described by a coupling matrix **K** is processed to produce the fixed transformation for the recurrent sampler, shown in (b). (b) Single algorithm iteration of the photonic recurrent Ising sampler. An in-phase optical signal encoding a spin state is fed to an optical matrix multiplication unit encoding $\mathbf{C} = 2\sqrt{\mathbf{K}}$, where **K** is the coupling matrix of the desired Ising graph. The output signal is noisy, with a distribution that is Gaussian with standard deviation ϕ, and goes through an analog nonlinear unit before being fed back to the chip input. Considering as an example a nine-spin 2D antiferromagnetic graph, with coupling and ground state shown in (c), the simulated energy evolution as a function of time is shown in (d). (e) Simulated energy distribution of the optical output, which converges in probability to the Gibbs distribution of the associated Ising problem [Eq. (3)], for which the ground state (c) is exponentially more likely than higher-energy states at low temperatures.

antiferromagnet with periodic boundary conditions and identical nearest-neighbor couplings, $K_{ij} = -1$, shown in Fig. 1(c). After several algorithm steps, the optical output converges to a probability distribution that matches the Gibbs distribution with energy described in Eq. (3). Thus, the ground state with energy $H_{\min}$ [every spin having opposite direction to its nearest neighbors, shown in Fig. 1(c)] is most likely, which can be readily observed in the simulated optical output [Fig. 1(d)]. The expected Gibbs-like energy distribution is also readily observable on the simulated energy output distribution [Fig. 1(e)].

B. Experimental Demonstration

The experimental INPRIS demonstration presented in this work finds the ground state of various four-spin Ising problems with arbitrary graph couplings, using noise as a resource to enable convergence to the ground state. The optical matrix multiplication update step, $\mathbf{C}\vec{S}^{(t)}$ [from Eq. (2)], is broken down into three stages, motivated by the eigenvalue decomposition of the Ising weight matrix **K**. The two unitary matrix multiplication steps, $\mathbf{U}^{\dagger}$ and **U**, are implemented in a programmable nanophotonic processor (PNP) [15,16] comprising a unitary mesh of tunable MZIs [Fig. 2(b)]. The diagonal matrix $\mathbf{D}_{\alpha}$ is currently performed on an external CPU but could also be implemented optically with an electro-optic attenuator, a modulator, or a single MZI [14,38–41]. In future implementations, the optical diagonal matrix should also be tunable, as $\mathbf{D}_{\alpha}$ is a function of the eigenvalue dropout hyperparameter, α. For this demonstration, we consider only two values $\alpha = \{0, 1\}$, but a more in-depth discussion of a proposed future implementation is presented in Supplement 1, Section 3.

Our experimental setup (Fig. 2) consists of a 1550 nm laser feeding one port of the PNP, then routed [red MZIs in Fig. 2(b)] to a 5×5 unitary processor comprising 13 individually thermally tunable MZIs [42] [green MZIs in Fig. 2(b)], controlled by a microcontroller. The output intensities of each spatial waveguide in the PNP are measured by an InGaAs photodiode array and subjected to the set of nonlinear transformations outlined in Eq. (2) and Fig. 2(c) (phase-intensity reconstruction, addition of extrinsic Gaussian noise, and nonlinear threshold unit), currently implemented in an external CPU. The output of these electronic transformations then determines the PNP phase setting at the next algorithm step, and so forth.

To maximize the number of spins that can be handled by a single PNP, we broke down the algorithm step shown in Fig. 1(b) into multiple runs on the PNP, such that each run required only one matrix to be encoded on the chip [shown in Fig. 2(c)]. During each run, the PNP implements a unitary matrix of the form $\mathbf{h}_i\mathbf{V}\mathbf{R}[S^{(\tau)}]$ [$i \in 1, 2$, Fig. 2(c)], where $\mathbf{R}[S^{(\tau)}]$ rotates the PNP input $(1, 0, 0, 0, 0)^T$ into vector $S^{(\tau)}$, τ is the integer (resp. half-integer) number of algorithm steps, and $S^{(\tau)}$ is the current spin state (resp. an intermediate result after multiplication by $\mathbf{D}\mathbf{U}^{\dagger}$). **V** corresponds to one of the two unitary matrices involved in the eigenvalue decomposition of **K** ($\mathbf{V} \in \{\mathbf{U}, \mathbf{U}^{\dagger}\}$), and $\mathbf{h}_1$ (resp. $\mathbf{h}_2$) is a homodyne matrix interfering PNP outputs 1 with 2 and 3 with 4 (resp. 2 with 3 and 4 with 5). We use the PNP output 1 as an idler signal of known amplitude and phase to reconstruct the

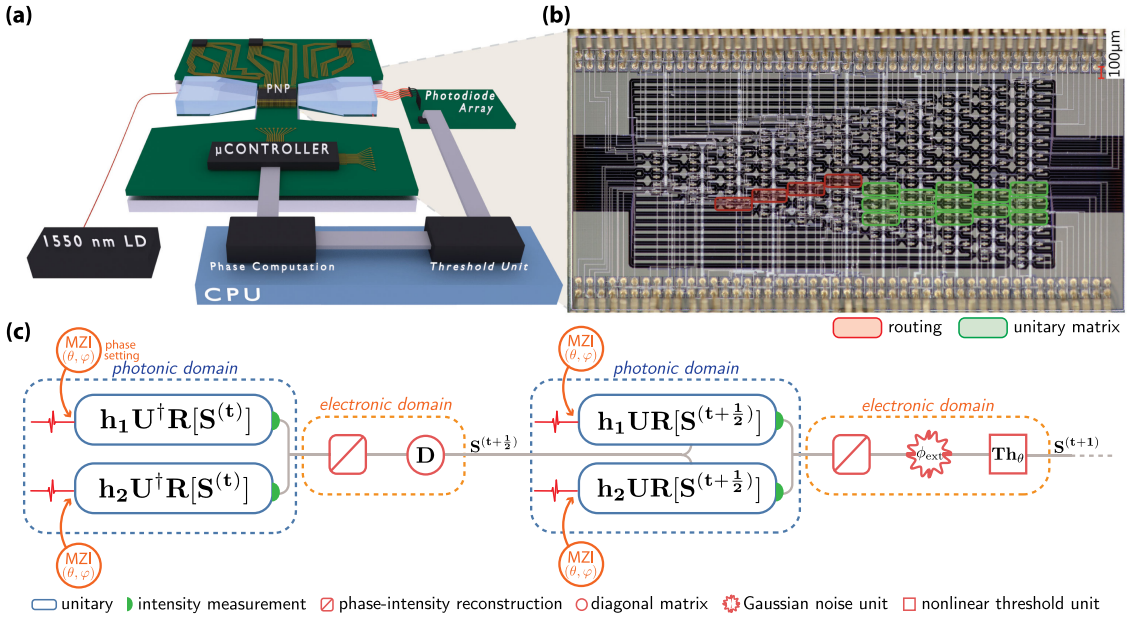


Fig. 2. Experimental realization of integrated nanophotonic coherent Ising sampler. (a) A 1550 nm laser diode is coupled to a single input port of a silicon-on-insulator programmable nanophotonic processor (PNP). Output mode intensities are measured with an InGaAs photodiode array, then processed in the electronic domain to determine PNP phase settings for the next iteration. (b) The PNP comprises 88 Mach–Zehnder interferometers with 176 individually controlled thermal phase shifters [15,16] and encodes a circuit consisting of input routing (red) and a U(5) unitary matrix (green). (c) Each algorithm step, shown conceptually in Fig. 1(b), requires four passes through the PNP chip. Each use of the chip performs a unitary matrix product of a state-preparation rotation matrix $\mathbf{R}$, the desired unitary ($\mathbf{U}$ or $\mathbf{U}^\dagger$), and one of two homodyne detection matrices ($\mathbf{h_1}$ or $\mathbf{h_2}$). Phase-intensity reconstruction, a diagonal matrix multiplication, Gaussian noise addition, and a nonlinear threshold unit are applied in the electronic domain.

phase and amplitude of every other output in a cascaded manner (more details on the experimental setup is given in Supplement 1, Section 1).

Consequently, a single algorithm step of the INPRIS, shown in Fig. 1(b), is performed with four runs on the PNP in our proof-of-concept experiment (two unitary matrices, each requiring two homodyne measurements to reconstruct the amplitude and phase of the output vector). The required reconfiguration of the PNP phases for each of these runs, due to the state-dependent $\mathbf{R}[S^{(\tau)}]$ and homodyne matrices $\mathbf{h}_i$, introduces an intrinsic noise on the measured outputs of the photonic system with standard deviation ϕ_{int}, which depends on the single-shot fidelity of the encoded PNP transformations (the measured single-shot fidelity of the PNP was 91.6%—Supplement 1, Section 2). We emphasize that, unlike our experimental demonstration of optical neural networks [14], a deterministic PNP crosstalk correction step is performed just once before each PNP run, and that additional phase shifter optimization using a detection feedback loop is not required to optimize unitary fidelity. This high single-shot fidelity allows us to achieve low enough noise levels to probe Gibbs distribution in their ordered phase, while circumventing the increased complexity required to perform fidelity optimization at each iteration. This single-shot scheme mimics a passive or low-bandwidth photonic system, which motivates our treatment of non-idealities in the detected PNP outputs due to MZI phase settings as a separate intrinsic noise at every algorithm step. We also perturb the magnitude of each detected PNP output at the end of the algorithm step with an extrinsic zero-mean Gaussian noise with standard deviation ϕ_{ext} via the CPU [see Gaussian noise unit in Fig. 2(c)]. Assuming the intrinsic contribution to the noise is also Gaussian, these two independent sources of noise add to yield a total noise level $\phi = \sqrt{\phi_{\text{int}}^2 + \phi_{\text{ext}}^2}$.

3. RESULTS

A. Ground State Search

Figures 3(b)–3(f) show our main experimental results: the noise-dependent ground state population for various four-spin Ising models. Each data point is averaged over 10 solver instances with randomized initial spin states and 100 algorithm steps per instance. By comparing simulation results (green curves correspond to ideal phase and amplitude reconstruction, while blue incorporate our experimental homodyne scheme) with the larger number of iterations in Figs. 3(b)–3(f), we observe that 100 algorithm steps are enough to converge to the ground state with high probability. In addition, we observe that our samples within the first 100 iterations show increased ground state probability over a random search [righthand plots in Figs. 3(b)–3(f)], with probabilities that approach a Gibbs distribution. While we do observe that the ground state is the most probable state for many of these graphs after 100 algorithm steps, there is a graph-dependent burn-in time required for the sampler to converge to steady-state dynamics [17]. The samples generated during the burn-in time are included in the histograms in Figs. 3(b)–3(f) and are responsible for some decrease in ground state probability. Comparisons between the measured energy state probabilities and a Gibbs distribution for all sampled graphs and dropout levels are presented in Supplement 1, Section 5.

Since we can adjust the total noise in the system by tuning the extrinsic noise, ϕ_{ext}, we can characterize the intrinsic noise

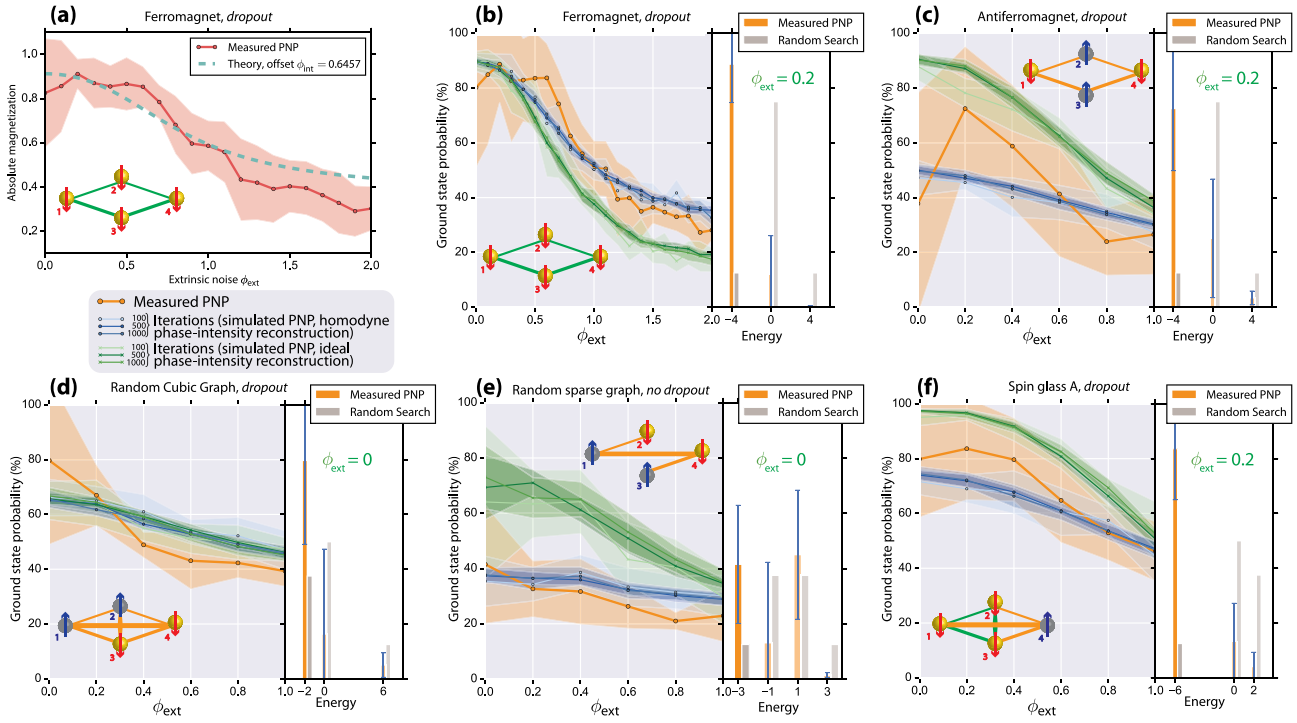


Fig. 3. Evaluating physical observables and finding ground states with the integrated nanophotonic coherent Ising sampler. (a) Experimental evaluation of the magnetization of a two-dimensional ferromagnet (red dots). The fit with theory allows us to evaluate intrinsic noise on the PNP $\phi_{int} = 0.6457$. (b)–(f) Left: ground state probability as a function of extrinsic noise ϕ_{ext} (orange), compared to simulated PNP with homodyne (blue) and ideal (green) phase-intensity reconstruction. Right: energy histograms for value of ϕ_{ext} showing best performance and comparison with random search. Inset: schematic representation of the Ising model being modeled by the PNP. Green (resp. orange) lines between nodes represent +1 (resp. −1) couplings.

level, ϕ_{int}, by measuring the noise-dependent evolution of physical observables. Conventional physical observables describing spin glasses (energy, magnetization, heat capacity, susceptibility, etc.) can be calculated by using the thresholded PNP outputs as samples of the Gibbs distribution [Eq. (3)]. Typically, these samples would be simulated using a heuristic algorithm [1,3,43], for instance, with conventional simulations run on a computer. The measured absolute magnetization of a four-spin ferromagnet with dropout, measured by our PNP proof of concept, is shown in Fig. 3(a), which fits its theoretical value with a resulting intrinsic noise $\phi_{int} = 0.6457$. The simulated values (green and blue plots) in Figs. 3(b)–3(f) are calculated using the total noise level with the experimentally fitted ϕ_{int}. This noise level naturally present in the photonic system can be leveraged to drive the ground state search, as can be seen in Figs. 3(b)–3(f), which show the ground state population for a variety of graphs (couplings shown in inset) as a function of the extrinsic noise standard deviation ϕ_{ext} applied in the CPU. We notice that large ground state probabilities are attained for $\phi_{ext} = 0$, meaning that the regime of optimal total noise level is attained with little to no further addition of extrinsic noise in the system. This observation suggests that the intrinsic noise present in the hardware is already very close to the optimal total noise level. Decreasing the intrinsic dynamic noise within the system would enable sampling in regimes with lower total noise, where an optimum for the added extrinsic noise can be found. This decrease in precision can be addressed by using current sources to modulate photonic active components instead of voltage drivers. In addition, dedicating separate photonic circuits for passive and active tasks will also reduce overall system infidelity. While noise is usually considered a nuisance in most physical (computing) systems, our experimental demonstration relies on physical noise, coming from finite single-shot fidelity of the PNP. This sheds new light on noise as a computational resource in physical computing systems. When injecting extrinsic (CPU-applied) noise to the outputs at each algorithm step, we observe that the ground state probability remains large and agrees with simulations of the PNP that account for homodyne detection and phase-intensity reconstruction. The ideal performance of the PNP is shown in green for various numbers of algorithm steps per sampler instance, thus demonstrating the effect of our cascaded homodyne detection scheme on the ground state convergence probability of the INPRIS. Overall, the ground state probability is ≥80% for most graphs at some dropout level ($\alpha = 0$ or $\alpha = 1$) and low extrinsic noise levels, largely outperforming random search algorithms. We also observe that experiments *with dropout* usually outperform those *without dropout*, as expected from theory [17]. Extended experimental results on ground state probabilities for all generated graphs and dropout levels are available in Supplement 1, Section 5.

B. Phase Space Exploration

We further investigated the impact of the homodyne reconstruction by probing the noise-dependent phase space of the two-dimensional Ising ferromagnetic problem, shown in Figure 4. The phase space population clearly shows a skew towards one of the ground states $\sigma = (-1, -1, -1, -1)$ (results are averaged over 10 random initial states). We confirm this skew originates from the cascaded homodyne detection by comparing the phase space population to simulations with (blue) and without (green) homodyne detections for the ferromagnetic graph (see Supplement

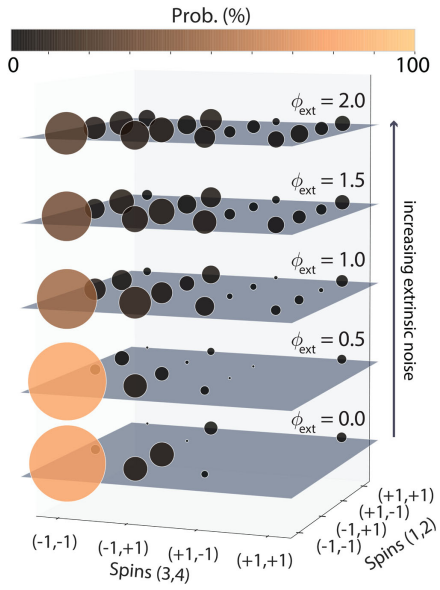


Fig. 4. Phase space probing by extrinsic noise injection on ferromagnet problem. The area of each dot is proportional to the measured mean probability of observing the system in state (y, x) at a given extrinsic noise level ϕ_{ext}, where $y = (\sigma_1, \sigma_2)$ is the state of spins 1 and 2, and $x = (\sigma_3, \sigma_4)$ is the state of spins 3 and 4.

1, Section 6). Since this skew breaks the two-fold degeneracy of the two-dimensional Ising problem, we presume that part of the phase information was lost through the reconstruction. To bypass this skew, we suggest scalable homodyne detection schemes that require more than one idler channel, in Supplement 1, Section 1. We observe that the skew does not seem to prevent the algorithm from converging to ground states but may complexify the evaluation of some physical observables. Interestingly, frozen transient states—which can be eliminated by reducing the pump amplitude—have also been observed in the OPO Ising machine [44]. As additional extrinsic noise is injected into the system, we observe that the phase space recovers its symmetry, thus providing an estimate for the amplitude of the skew in our photonic system.

Fabrication imperfections in photonic networks could be a significant bottleneck in scaling integrated nanophotonic coherent Ising samplers to large ($N > 100$) graph orders. For instance, a static skew on the phase setting or the split ratio of beam splitters will result in a static error on the effective coupling between spins, thus reshaping the Hamiltonian landscape, which could impact the algorithm efficiency [17]. Thus, the reduction of imperfections is of paramount importance in the realization of the INPRIS on large-scale static photonic networks. However, several calibration techniques have been developed to achieve precise linear optical functions with broad process tolerances [45,46]. The bandwidth and energy cost of thermal phase shifters is another bottleneck in increasing the size of photonic networks. A significant advantage of the INPRIS over learning-based methods that require reconfiguration of the PNP phases at every algorithm step [17] is that the phase reconfiguration bandwidth is irrelevant in characterizing the algorithm efficiency, since it translates only in a constant computational cost. Thus, a larger-scale INPRIS could be enabled by different architectures that do not suffer from this bottleneck and could operate at very low power levels. Examples include 3D-printed free-space optical networks [47] and silicon photonic networks combined with non-volatile phase-change materials [48].

4. DISCUSSION

To conclude, we introduce the INPRIS, a photonic circuit able to probe the Gibbs distribution of arbitrary Ising problems. We demonstrate a proof-of-concept experiment on a PNP that exhibits high success probability on a variety of four-spin graphs. We show that noise coming from the PNP-limited single-shot fidelity allows the system to operate close to its optimal noise regime, required to find ground states of Ising problems. Conversely, the influence of sources of noise of various origins (extrinsic versus intrinsic) and natures (static skew versus dynamic) is readily seen on physical observables such as magnetization, ground state population, and phase space distribution. Additionally, we observe agreement between the measured ground state probabilities and those predicted in simulation, suggesting the possibility of using simulated scaling predictions for this architecture, such as those presented in Supplement 1, Fig. 10 in [17], to motivate performance predictions of future large-scale systems with passive photonic components. We include a treatment of timing for a large-scale system in Supplement 1, Section 4.A. In fully passive networks, other sources of noise should be leveraged. For instance, recent work on large-scale photoelectric networks naturally demonstrates Gaussian noise on their outputs, arising at the quantum limit (and thus suggesting the operation of a very large-scale INPRIS that reaches the thermodynamic limit, relevant for simulations in statistical mechanics with $N \sim 10^6$, that can reach attojoule energy consumption per algorithm unit step) [27]. Furthermore, the photonic circuit program for this architecture is fixed for any given coupling configuration. This feature enables the potential for quasi-passive photonic ASICs, such as non-volatile photonic ASICs [48–50], that could deliver speed and energy savings over other physical Ising machines with active components.

Following an analysis similar to the one presented in Ref. [14], we can compare the energy consumption for a future INPRIS with that of graphics processing unit (GPU) hardware. Each iteration of the algorithm requires a matrix multiplication consisting of $2N^2$ floating-point operations (FLOPs). For a proposed passive photonic system operating at a realistic clock rate of 1 GHz, one would need to perform $2N^2 * 10^9$ FLOP/s. As detailed in Supplement 1, Section 4.B: we estimate the power consumption of the photonic system to be $(18N)$ mW. These values correspond to an energy consumption per FLOP of $\sim(9/N)$ pJ/FLOP. At the time of writing, a state-of-the-art NVIDIA V100 GPU operating at a comparable precision of 16-bit consumes 2.2 pJ/FLOP [51], suggesting an energy scaling advantage for a passive photonic architecture as the graph size increases.

This work also paves the way to larger-scale INPRISs by identifying key tradeoffs in their design. While homodyne detection allows the reconstruction of the spin state phase and amplitude, an increased number of idler signals will increase the PNP footprint for a given number of spins and determine the overall accuracy of this operation. Other tradeoffs, such as optimal values of eigenvalue dropout [17], should be taken into account when designing quasi-passive photonic GHz INPRISs on ASICs which could outperform current optical and electronic Ising machines by several orders of magnitude [20,26,28].

Funding. Semiconductor Research Corporation (2016-EP-2693-B); National Science Foundation (CCF-1640012, GRFP 1122374, and MRSEC Program DMR-1419807); Air Force Office of Scientific Research (FA9550-14-1-0052); Army Research Laboratory and the U.S. Army Research Office; Institute for Soldier Nanotechnologies (W911NF-18-2-0048); H2020 Marie Skłodowska-Curie Actions (751016); Intelligence Community Postdoctoral Research Fellowship Program.

Acknowledgment. The authors would like to acknowledge Aram Harrow, Mehran Kardar, Ido Kaminer, Manan Raval, and Jamison Sloan for helpful discussions.

Disclosures. C. R.-C., Y.S., L. J., and M. S.: U. S. Patent Application No. 16/032,737 (P), Y.S.: Lightelligence (I, E, R), L.J., J.D.J., and M.S.: Lightelligence (I), R.H.: NTT Research (E), T. B.-J. and M. H.: Nokia Corporation (E).

See Supplement 1 for supporting content.

[†]These authors contributed equally to this work.

REFERENCES

1. D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (2009).
2. J. Hromkovič, *Algorithmics for Hard Problems: Introduction to Combinatorial Optimization, Randomization, Approximation, and Heuristics* (Springer, 2013).
3. E. Aarts and J. Korst, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing* (Wiley, 1989).
4. R. M. Karp, "Reducibility among combinatorial problems," in *Complexity of Computer Computations* (Springer, 1972), pp. 85–103.
5. A. Lucas, "Ising formulations of many NP problems," Front. Phys. **2**, 5 (2014).
6. F. Glover and G. Kochenberger, *Handbook of Metaheuristics* (Springer, 2006).
7. E. Ising, "Beitrag zur Theorie des Ferromagnetismus," in *Z. Phys.* (1925).
8. A. R. Honerkamp-Smith, S. L. Veatch, and S. L. Keller, "An introduction to critical points for biophysicists; observations of compositional heterogeneity in lipid membranes," Biochimica et Biophys. Acta (BBA) – Biomembr. **1788**, 53–63 (2009).
9. N. Ghofraniha, I. Viola, F. Di Maria, G. Barbarella, G. Gigli, L. Leuzzi, and C. Conti, "Experimental evidence of replica symmetry breaking in random lasers," Nat. Commun. **6**, 6058 (2015).
10. W. M. Macready, A. G. Siapas, and S. A. Kauffman, "Criticality and parallelism in combinatorial optimization," Science **271**, 56–59 (1996).
11. S. Tsukamoto, M. Takatsu, S. Matsubara, and H. Tamura, "An accelerator architecture for combinatorial optimization problems," Fujitsu Sci. Tech. J. **53**, 8–13 (2017).
12. D. J. Earl and M. W. Deem, "Parallel tempering: theory, applications, and new perspectives," Phys. Chem. Chem. Phys. **7**, 3910 (2005).
13. J. Carolan, C. Harrold, C. Sparrow, E. Martín-López, N. J. Russell, J. W. Silverstone, P. J. Shadbolt, N. Matsuda, M. Oguma, M. Itoh, G. D. Marshall, M. G. Thompson, J. C. Matthews, T. Hashimoto, J. L. O'Brien, and A. Laing, "Universal linear optics," Science **349**, 711–716 (2015).
14. Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," Nat. Photonics **11**, 441–446 (2017).
15. N. C. Harris, G. R. Steinbrecher, M. Prabhu, Y. Lahini, J. Mower, D. Bunandar, C. Chen, F. N. C. Wong, T. Baehr-Jones, M. Hochberg, S. Lloyd, and D. Englund, "Quantum transport simulations in a programmable nanophotonic processor," Nat. Photonics **11**, 447–452 (2017).
16. N. C. Harris, J. Carolan, D. Bunandar, M. Prabhu, M. Hochberg, T. Baehr-Jones, M. L. Fanto, A. M. Smith, C. C. Tison, P. M. Alsing, and D. Englund, "Linear programmable nanophotonic processors," Optica **5**, 1623 (2018).
17. C. Roques-Carmes, Y. Shen, C. Zanoci, M. Prabhu, F. Atieh, L. Jing, T. Dubček, C. Mao, M. R. Johnson, V. Čeperić, and J. D. Joannopoulos, "Heuristic recurrent algorithms for photonic Ising machines," Nat. Commun. **11**, 1–8 (2020).
18. A. Vergis, K. Steiglitz, and B. Dickinson, "The complexity of analog computation," Math. Comput. Simul. **28**, 91–113 (1986).
19. Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, "Coherent Ising machine based on degenerate optical parametric oscillators," Phys. Rev. A **88**, 063853 (2013).
20. P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, "A fully programmable 100-spin coherent Ising machine with all-to-all connections," Science **354**, 614–617 (2016).
21. K. Wu, J. García de Abajo, C. Soci, P. Ping Shum, and N. I. Zheludev, "An optical fiber network oracle for NP-complete problems," Light. Sci. Appl. **3**, 147 (2014).
22. M. R. Vázquez, V. Bharadwaj, B. Sotillo, S.-Z. A. Lo, R. Ramponi, N. I. Zheludev, G. Lanzani, S. M. Eaton, and C. Soci, "Optical NP problem solver on laser-written waveguide platform," Opt. Express **26**, 702–710 (2018).
23. S. Utsunomiya, N. Namekata, K. Takata, D. Akamatsu, S. Inoue, and Y. Yamamoto, "Binary phase oscillation of two mutually coupled semiconductor lasers," Opt. Express **23**, 6029–6040 (2015).
24. M. Babaeian, D. T. Nguyen, V. Demir, M. Akbulut, P.-A. Blanche, Y. Kaneda, S. Guha, M. A. Neifeld, and N. Peyghambarian, "A single shot coherent Ising machine based on a network of injection-locked multicore fiber lasers," Nat. Commun. **10**, 3516 (2019).
25. D. Pierangeli, G. Marcucci, and C. Conti, "Large-scale photonic Ising machine by spatial light modulation," Phys. Rev. Lett. **122**, 213902 (2019).
26. R. Hamerly, T. Inagaki, P. L. McMahon, D. Venturelli, A. Marandi, T. Onodera, E. Ng, C. Langrock, K. Inaba, T. Honjo, K. Enbutsu, T. Umeki, R. Kasahara, S. Utsunomiya, S. Kako, K.-I. Kawarabayashi, R. L. Byer, M. M. Fejer, H. Mabuchi, D. Englund, E. Rieffel, H. Takesue, and Y. Yamamoto, "Experimental investigation of performance differences between coherent Ising machines and a quantum annealer," Sci. Adv. **5**, eaau0823 (2019).
27. R. Hamerly, L. Bernstein, A. Sludds, M. Soljačić, and D. Englund, "Large-scale optical neural networks based on photoelectric multiplication," Phys. Rev. X **9**, 021032 (2019).
28. T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K.-I. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, "A coherent Ising machine for 2000-node optimization problems," Science **354**, 603–606 (2016).
29. Y. Okawachi, M. Yu, K. Luke, D. O. Carvalho, S. Ramelow, A. Farsi, M. Lipson, and A. L. Gaeta, "Dual-pumped degenerate Kerr oscillator in a silicon nitride microresonator," Opt. Lett. **40**, 5267–5270 (2015).
30. T. Leleu, Y. Yamamoto, P. L. McMahon, and K. Aihara, "Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity," Phys. Rev. Lett. **122**, 040607 (2019).
31. A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto, "Network of time-multiplexed optical parametric oscillators as a coherent Ising machine," Nat. Photonics **8**, 937–942 (2014).
32. W. A. Little, "The existence of persistent states in the brain," Math. Biosci. **19**, 101–120 (1974).
33. J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc. Natl. Acad. Sci. USA **79**, 2554–2558 (1982).
34. P. Peretto, "Collective properties of neural networks: a statistical physics approach," Biol. Cybern. **50**, 51–62 (1984).
35. W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walsmley, "Optimal design for universal multiport interferometers," Optica **3**, 1460 (2016).
36. M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, "Experimental realization of any discrete unitary operator," Phys. Rev. Lett. **73**, 58–61 (1994).
37. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," J. Mach. Learn. Res. **15**, 1929–1958 (2014).

38. Z. Cheng, H. K. Tsang, X. Wang, K. Xu, and J.-B. Xu, "In-plane optical absorption and free carrier absorption in graphene-on-silicon waveguides," IEEE J. Sel. Top. Quantum Electron. **20**, 43–48 (2014).
39. Q. Bao, H. Zhang, Z. Ni, Y. Wang, L. Polavarapu, Z. Shen, Q.-H. Xu, D. Tang, and K. P. Loh, "Monolayer graphene as a saturable absorber in a mode-locked laser," Nano Res. **4**, 297–307 (2011).
40. Q. Xu, B. Schmidt, S. Pradhan, and M. Lipson, "Micrometre-scale silicon electro-optic modulator," Nature **435**, 325–327 (2005).
41. A. H. Atabaki, S. Moazeni, F. Pavanello, H. Gevorgyan, J. Notaros, L. Alloatti, M. T. Wade, C. Sun, S. A. Kruger, H. Meng, K. Al Qubaisi, I. Wang, B. Zhang, A. Khilo, C. V. Baiocco, M. A. Popović, V. M. Stojanović, and R. J. Ram, "Integrating photonics with silicon nanoelectronics for the next generation of systems on a chip," Nature **556**, 349–354 (2018).
42. N. C. Harris, Y. Ma, J. Mower, T. Baehr-Jones, D. Englund, M. Hochberg, and C. Galland, "Efficient, compact and low loss thermo-optic phase shifter in silicon," Opt. Express **22**, 10487 (2014).
43. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220**, 671–680 (1983).
44. F. Böhm, T. Inagaki, K. Inaba, T. Honjo, K. Enbutsu, T. Umeki, R. Kasahara, and H. Takesue, "Understanding dynamics of coherent Ising machines through simulation of large-scale 2D Ising models," Nat. Commun. **9**, 5020 (2018).
45. D. A. B. Miller, "Perfect optics with imperfect components," Optica **2**, 747 (2015).
46. R. Burgwal, W. R. Clements, D. H. Smith, J. C. Gates, W. S. Kolthammer, J. J. Renema, and I. A. Walmsley, "Using an imperfect photonic network to implement random unitaries," Opt. Express **25**, 28236 (2017).
47. X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, Y. Luo, M. Jarrahi, and A. Ozcan, "All-optical machine learning using diffractive deep neural networks," Science **361**, 1004–1008 (2018).
48. Q. Wang, E. T. F. Rogers, B. Gholipour, C.-M. Wang, G. Yuan, J. Teng, and N. I. Zheludev, "Optically reconfigurable metasurfaces and photonic devices based on phase change materials," Nat. Photonics **10**, 60–65 (2016).
49. B. J. Eggleton, B. Luther-Davies, and K. Richardson, "Chalcogenide photonics," Nat. Photonics **5**, 141 (2011).
50. C. R. de Galarreta, A. M. Alexeev, Y.-Y. Au, M. Lopez-Garcia, M. Klemm, M. Cryan, J. Bertolotti, and C. D. Wright, "Nonvolatile reconfigurable phase-change metadevices for beam steering in the near infrared," Adv. Funct. Mater. **28**, 1704993 (2018).
51. NVIDIA Corporation, "NVIDIA V100 TENSOR CORE GPU," Tech. rep. (2020).