

Photonic Recurrent Ising Sampler

Charles Roques-Carmes,¹ Yichen Shen,² Cristian Zanoci,² Mihika Prabhu,¹ Fadi Atieh,² Li Jing,² Tena Dubček,² Vladimir Čeperić,² John D. Joannopoulos,^{2,3} Dirk Englund,¹ and Marin Soljačić^{1,2}

¹*Research Laboratory of Electronics, Massachusetts Institute of Technology, 50 Vassar Street, Cambridge MA 02139, USA**

²*Department of Physics, Massachusetts Institute of Technology,*

77 Massachusetts Avenue, Cambridge, MA 02139, USA

³*Institute for Soldier Nanotechnologies, 500 Technology Square, Cambridge, MA 02139, USA*

The failure of conventional electronic architectures to solve large combinatorial problems motivates the development of novel computational architectures. There has been much effort recently in developing photonic networks which can exploit fundamental properties enshrined in the wave nature of light and of its interaction with matter: high-speed, low-power, optical passivity, and parallelization. However, unleashing the true potential of photonic architectures requires the development of featured algorithms which largely exploit these fundamental properties. We here present the Photonic Recurrent Ising Sampler (PRIS), an algorithm tailored for photonic parallel networks that can sample combinatorially hard distributions of Ising problems, in a fast and efficient manner. The PRIS finds the ground state of general Ising problems and is able to probe critical behaviors of universality classes and their critical exponents. In addition to the attractive features of photonic networks, the PRIS relies on intrinsic dynamic noise to find ground states more efficiently.

Probabilistic algorithms are the cornerstone of both numerical methods in statistical physics¹ and combinatorial optimization². A broad class of problems in statistical physics, such as growth patterns in clusters³, percolation⁴, heterogeneity in lipid membranes⁵, and complex networks⁶, can be described by Monte Carlo (MC) methods. These methods have played an instrumental role in predicting phase transitions and the critical exponents of various universality classes – families of physical systems exhibiting similar scaling properties around their critical temperature¹. Additionally, stochastic techniques, such as simulated annealing⁷, can find approximate solutions of NP-hard problems for which the number of candidate solutions grows exponentially. These algorithms are usually tailored for conventional electronic hardware; however, a number of optical machines have recently been shown to solve the well-known Ising^{8,9} and Traveling Salesman problems^{10,11}. For computationally demanding problems, MC methods can benefit from parallelization speedups^{1,12}, but the determination of an efficient parallelization approach is highly problem-specific¹. In this Letter, we propose a fast and efficient probabilistic technique relying essentially on iterative matrix multiplications which can benefit from photonic analog computing platforms^{13,14}, and in particular from recently demonstrated high-speed, low-power, parallel photonic computing platforms¹⁵. Our probabilistic approach also takes advantage of optical passivity and dynamic noise to find ground states of arbitrary Ising problems more efficiently and probe their critical behaviors, yielding accurate predictions of critical exponents of the universality classes of conventional Ising models. Our algorithm presents attractive scaling properties when benchmarked versus conventional sequential sampling algorithms. Our findings suggest a new paradigm in probabilistic computing for efficient optimization and sampling by leveraging the potential of parallel photonic computing.

Half a century before the contemporary Machine Learning

Renaissance¹⁶, the Little¹⁷ and then the Hopfield^{18,19} networks were considered as early architectures of recurrent neural networks (RNN). The latter was suggested as an algorithm to solve combinatorially hard problems, as it was shown to deterministically converge to local minima of arbitrary quadratic Hamiltonians of the form

$$H^{(K)} = -\frac{1}{2} \sum_{1 \leq i,j \leq N} K_{ij} \sigma_i \sigma_j, \quad (1)$$

which is the most general form of an Ising Hamiltonian in the absence of an external magnetic field²⁰. In Equation (1), we equivalently denote the set of spins as $\sigma \in \{-1, 1\}^N$ or $S \in \{0, 1\}^N$ (with $\sigma = 2S - 1$), and K is a $N \times N$ real symmetric matrix, with $K_{ii} = 0$ (as diagonal terms only contribute to a global offset of the Hamiltonian, see Supplementary Information (SI), section I). In the following, we will refer to the eigenvalue decomposition of K as $K = UDU^\dagger$.

In the context of physics, Ising models describe the interaction of many particles in terms of the coupling matrix K . These systems are observed in a particular spin configuration S with a probability given by the Gibbs distribution $p(S) \propto \exp(-\beta H^{(K)}(S))$, where $\beta = 1/(k_B T)$, with k_B the Boltzmann constant and T the temperature. At low temperature, when $\beta \rightarrow \infty$, the Gibbs probability of observing the system in its ground state is close to 1, thus naturally minimizing the quadratic function in Equation (1). Similar optimization problems are often encountered in computer science. A natural idea is to engineer physical systems so that their energy duplicates a function to be optimized and will output the state minimizing this energy with high probability. This analogy between statistical physics and information theory has nurtured a great variety of concepts in both fields²¹, for instance, the analogy between neural networks and spin glasses^{18,22}.

Many complex systems fall into the Ising universality class²³ — such as ferromagnets^{20,24}, liquid-vapor

transitions²⁵, lipid membranes⁵, brain functions²⁶, and strongly-interacting systems in Quantum Chromodynamics²⁷ — and can thus be described by this century-old model. From the perspective of optimization, finding the spin distribution minimizing $H^{(K)}$ for an arbitrary matrix K belongs to the class of NP-hard problems²⁸.

The deterministic convergence of the Hopfield network to a *local* minimum makes it impossible to scale such networks to *deterministically* find the *global* minimum²⁹ — thus jeopardizing any electronic¹⁹ or optical³⁰ implementation of these algorithms. Therefore, these early RNN architectures were soon superseded by probabilistic optimization algorithms, such as simulated annealing⁷, usually tailored for conventional electronic hardware. Nonetheless, optical hardware can benefit from parallel, low-energy, high-speed computations¹⁵. Here, we propose a *probabilistic* photonic implementation of a passive RNN, able to model the arbitrary Ising-type Hamiltonian shown in Equation (1).

The proposed architecture of our photonic network is shown in Figure 1. The spin state at time step t , encoded in the phase and amplitude of N parallel photonic signals $S^{(t)} \in \{0, 1\}^N$, first goes through a linear photonic domain decomposed in its eigenvalue form $2J = USq_\alpha(D)U^\dagger$, where $Sq_\alpha(D)$ is a diagonal matrix derived from D . The signal is then fed into nonlinear optoelectronic domain, where it is perturbed by a Gaussian distribution of standard deviation ϕ (simulating noise present in the photonic implementation) and is imparted a nonlinear threshold function Th_θ ($Th_\theta(x) = 1$ if $x > \theta$, 0 otherwise). The signal is then recurrently fed back to the linear photonic domain, and the process repeats. The static unit transformation between two time steps t and $t + 1$ of this RNN can be summarized as

$$\begin{aligned} S^{(t+1)} &= Th_\theta(X^{(t)}) \\ X^{(t)} &\sim \mathcal{N}(2JS^{(t)}|\phi), \end{aligned} \quad (2)$$

where $\mathcal{N}(x|\phi)$ denotes a Gaussian distribution of mean x and standard deviation ϕ . We call this algorithm, which is tailored for a photonic implementation, the Photonic Recurrent Ising Sampler (PRIS). The detailed choice of algorithm parameters is described in the SI, section II.

This simple recurrent loop can be readily implemented in the photonic domain. For example, the linear photonic interference unit can be realized with MZI networks^{15,31–33}, diffractive optics^{34,35}, ring resonator filter banks^{36–38}, and free space lens-SLM-lens systems³⁹; the diagonal matrix multiplication $Sq_\alpha(D)$ can be implemented with a simple electro-optical absorber, a modulator or a single MZI^{15,40,41}; the nonlinear optoelectronic unit can be implemented with an optical nonlinearity^{40–44}, or simple analog/digital electronics^{45–48}.

The long-time dynamics of the PRIS is described by an effective Hamiltonian H_L (see Refs.^{22,49} and SI, section II), which can be expanded, in the large noise approximation

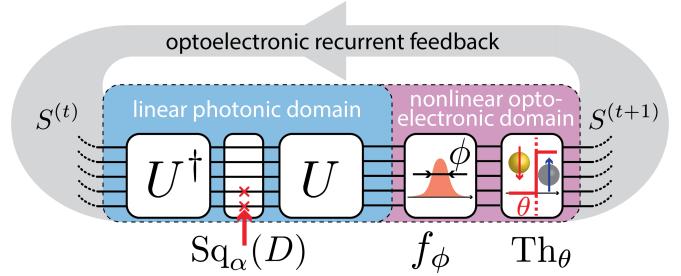


FIG. 1. **Operation principle of the PRIS.** A photonic analog signal, encoding the current spin state, goes through transformations in linear photonic and nonlinear optoelectronic domains. The result of this transformation is recurrently fed back to the input of this passive photonic system.

$(\phi \gg 1)$, into H_2 :

$$H_L = -\frac{1}{\beta} \sum_i \log \cosh \left(\beta \sum_j J_{ij} \sigma_j \right), \quad (3)$$

$$H_2 = -\frac{\beta}{2} \sum_{1 \leq i, j \leq N} [J^2]_{ij} \sigma_i \sigma_j. \quad (4)$$

Here, $\beta = 1/(k\phi)$ is analogous to the inverse temperature from statistical mechanics, where k is a constant, only depending on the noise distribution (see SI, section II). Examining Equation (4), we can deduce a mapping of the PRIS to the general Ising model shown in Equation (1). We set the PRIS matrix J to be a modified square-root of K by imposing the following condition on the PRIS

$$Sq_\alpha(D) = 2\text{Re}(\sqrt{D + \alpha\Delta}), \quad (5)$$

where $\alpha \in [-1, 1]$ is the eigenvalue dropout level and Δ is a diagonal offset matrix defined as the sum of the off-diagonal term of the Ising coupling matrix $\Delta_{ii} = \sum_{j \neq i} |K_{ij}|$. The addition of Δ only results in a global offset on the Hamiltonian. The purpose of the Δ offset is to make the matrix in the square root symmetric positive definite when $\alpha = 1$. Thus, other definitions of the diagonal offset could be proposed. When $\alpha < 1$, some lower eigenvalues are then dropped out by taking the real part (see SI, section III); we show below that this improves the performance of the PRIS. When choosing this definition of $Sq_\alpha(D)$ and operating the PRIS in the large noise limit, we can implement any general Ising model (Equation (1)) on the PRIS (Equation (4)).

It has been noted that by setting $Sq_\alpha(D) = D$ (i.e. the linear photonic domain matrix amounts to the Ising coupling matrix $2J = K$), the free energy of the system equals the Ising free energy at any finite temperature (up to a factor of 2) in the particular case of associative memory couplings²². In this regime of operation, the PRIS can also benefit from computational speed-ups, if implemented on a conventional architecture, for instance if the coupling matrix is sparse. However, as has been pointed out in theory²² and by our simulations (see SI, section IV, Figure 7), some additional considerations should be taken into account in order to eliminate non-ergodic

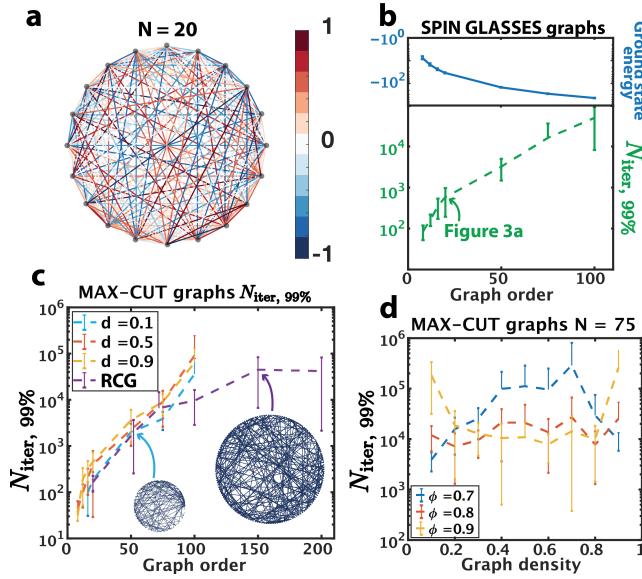


FIG. 2. Scaling performance of the PRIS. **a:** A fully-connected spin glass with uniformly-distributed continuous couplings in $[-1, 1]$. **(b, top)** Ground state energy versus graph order of these graphs. $N_{\text{iter}, 99\%}$ versus graph size for spin glasses **(b, bottom)** and MAX-CUT graphs **(c)**. **d:** $N_{\text{iter}, 99\%}$ versus graph density for MAX-CUT graphs and $N = 75$. The graph density is defined as $d = 2|E|/(N(N - 1))$, $|E|$ being the number of undirected edges.

behaviors in this system. As the regime of operation described by Equation (54) is general to any coupling, we will use it in the following demonstrations.

We investigate the performance of the PRIS on finding the ground state of general Ising problems (Equation (1)) with two types of Ising models: MAX-CUT graphs, which can be mapped to an instance of the unweighted MAX-CUT problem⁹ and spin glasses, whose connections are uniformly distributed in $[-1, 1]$ (an example illustration of the latter is shown in Figure 2(a)). Both families of models are computationally NP-hard problems²⁸, thus their computational complexity grows exponentially with the graph order N .

The number of steps necessary to find the ground state with 99% probability, $N_{\text{iter}, 99\%}$ is shown in Figure 2(b-c) for these two types of graphs (see definition in SI, section IV). As the PRIS can be implemented with high-speed parallel photonic networks, the on-chip real time of a unit step can be less than a nanosecond^{15,50} (and the initial setup time for a given Ising model is typically of the order of microseconds with thermal phase shifters⁵¹). In such architectures, the PRIS would thus find ground states of arbitrary Ising problems with graph orders $N \sim 100$ within less than a millisecond. Interestingly, both classical and quantum optical Ising machines have exhibited limitations in their performance related to the graph density^{9,52}. We observe that the PRIS is roughly insensitive to the graph density, when optimizing the noise level ϕ (see Figure 2(d)). A more comprehensive comparison should take into account the static fabrication error in integrated photonic networks¹⁵, even though careful calibration of their control

electronics can significantly reduce its impact on the computation.

For a given Ising problem, there remain two degrees of freedom in the execution of the PRIS: the noise and eigenvalue dropout levels. The noise level ϕ determines the level of entropy in the Gibbs distribution probed by the PRIS $p(E) \propto \exp(-\beta(E - \phi S(E)))$, where $S(E)$ is the Boltzmann entropy associated with the energy level E . On the one hand, increasing ϕ will result in an exponential decay of the probability of finding the ground state $p(H_{\min}, \phi)$. On the other hand, too small a noise level will not satisfy the large noise approximation $H_L \sim H_2$ and result in large autocorrelation times (as the spin state could get stuck in a local minimum of the Hamiltonian). Figure 3(e) demonstrates the existence of an optimal noise level ϕ , minimizing the number of iterations required to find the ground state of a given Ising problem, for various graph sizes, densities, and eigenvalue dropout levels. This optimal noise value can be approximated upon evaluation of the probability of finding the ground state $p(H_{\min}, \phi)$ and the energy autocorrelation time τ_{auto}^E , as the minimum of the following heuristic

$$N_{\text{iter},q} \sim \tau_{\text{eq}}^E(\phi) + \tau_{\text{auto}}^E(\phi) \frac{\log(1 - q)}{\log(1 - p(H_{\min}, \phi))}, \quad (6)$$

which approximates the number of iterations required to find the ground state with probability q (see Figure 3(a-e)). In this expression, $\tau_{\text{eq}}^E(\phi)$ is the energy equilibrium time. As can be seen in Figure 3(e), decreasing α (and thus dropping more eigenvalues, with the lowest eigenvalues being dropped out first) will result in a smaller optimal noise level ϕ . Comparing the energy landscape for various eigenvalue dropout levels (Figure 3(h)) confirms this statement: as α is reduced, the energy landscape is perturbed. However, for the random spin glass studied in Figure 3(f-g), the ground state remains the same down to $\alpha = 0$. This hints at a general observation: as lower eigenvalues tend to increase the energy, the Ising ground state will in general be contained in the span of eigenvectors associated with higher eigenvalues. Nonetheless, the global picture is more complex, as the solution of this optimization problem should also enforce the constraint $\sigma \in \{-1, 1\}^N$. We observe in our simulations that $\alpha = 0$ yields a higher ground state probability and lower autocorrelation times than $\alpha > 0$ for all the Ising problems we used in our benchmark. In some sparse models, the optimal value can even be $\alpha < 0$. The eigenvalue dropout is thus a parameter that defines the dimensionality of the ground state search.

The influence of eigenvalue dropout can also be seen from the perspective of the transition matrix. Figure 3(f-g) shows the eigenvalue distribution of the transition matrix for various noise and eigenvalue dropout levels. As the PRIS matrix eigenvalues are dropped out, the transition matrix eigenvalues become more nonuniform, as in the case of large noise (Figure 3(g)). Overall, the eigenvalue dropout can be understood as a means of pushing the PRIS to operate in the large noise approximation, without perturbing the Hamiltonian in such a way that would prevent it from finding the ground state. The

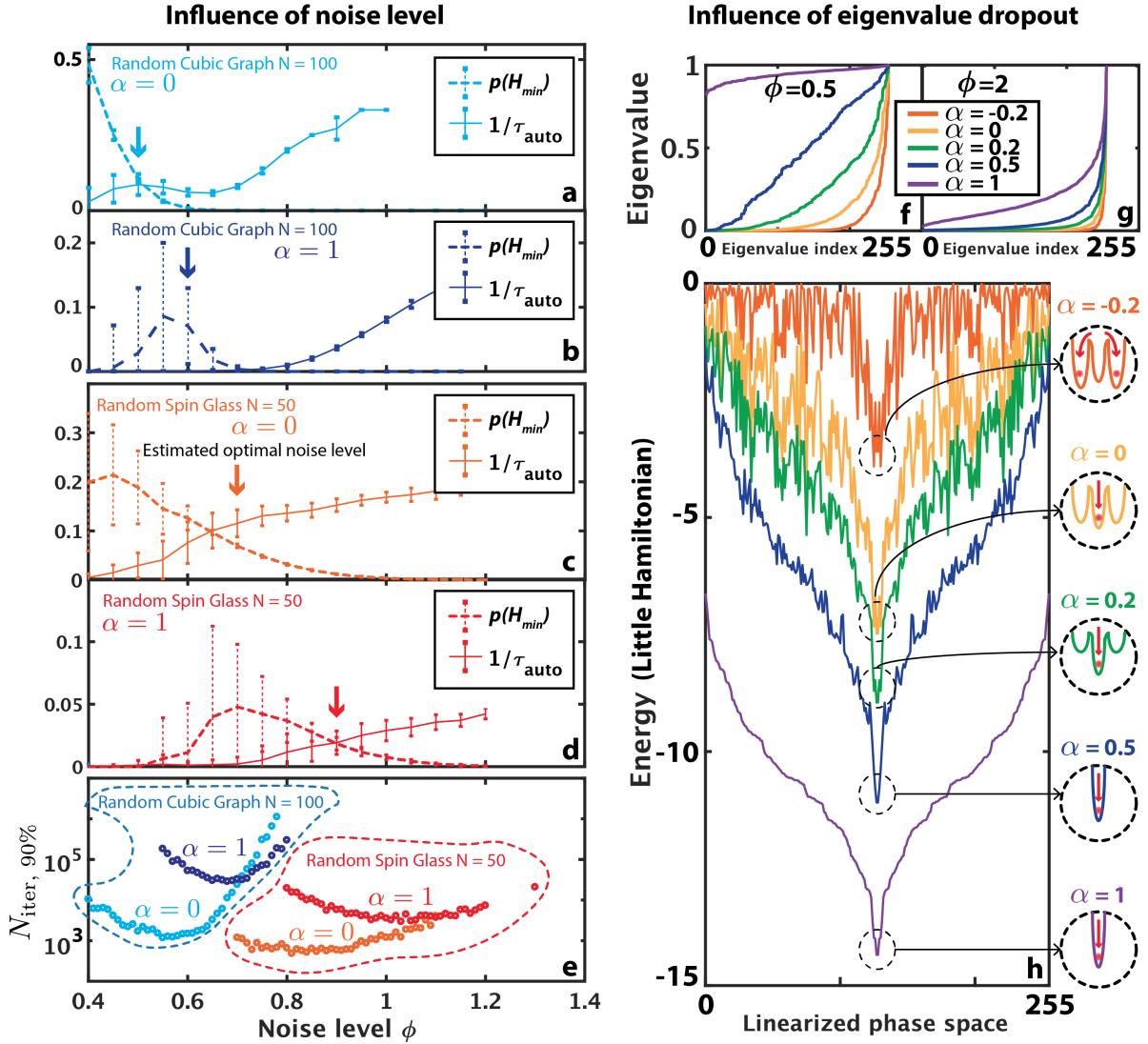


FIG. 3. Influence of noise and eigenvalue dropout levels. **a-d:** Probability of finding ground state, and the inverse of autocorrelation time as a function of noise level ϕ for a sample Random Cubic Graph⁹ ($N = 100$, $\alpha = 0$ (a), $\alpha = 1$ (b)) and a sample spin glass ($N = 50$, $\alpha = 0$ (c), $\alpha = 1$ (d)). The arrows indicate the estimated optimal noise level, from Equation (6), taking τ_{eq}^E to be constant. **e:** $N_{\text{iter},90\%}$ versus noise level ϕ for these same graphs and eigenvalue dropout levels. **f-g:** Eigenvalues of the transition matrix of a sample spin glass ($N = 8$) with $\phi = 0.5$ (f) and $\phi = 2$ (g). **h:** The corresponding energy is plotted for various eigenvalue dropout levels α . The inset is a schematic of the relative position of the global minimum when $\alpha = 1$ with respect to nearly local minima when $\alpha < 1$.

existence of these two degrees of freedom suggests a realm of algorithmic techniques to optimize the PRIS operation. One could suggest, for instance, setting $\alpha \approx 0$, and then letting the PRIS run for a given number of iterations, before increasing the dimensionality of the search. This class of algorithms could rely on the development of high-speed, low-loss integrated modulators^{50,53–55}.

The existence of an effective Hamiltonian describing the PRIS dynamics (Equation (4)) further suggests the ability to generate samples of the associated Gibbs distribution at *any finite temperature*. This is particularly interesting considering the various ways in which noise can be added in integrated photonic circuits by tuning the operating temperature, laser

power, photodiode regimes of operation, etc.⁴⁵. This alludes to the possibility of detecting phase transitions and characterizing critical exponents of universality classes, leveraging the high speed at which photonic systems can generate uncorrelated MC samples of the Gibbs distribution associated with Equations (3,4). In this part, we operate the PRIS in the regime where the linear photonic matrix is equal to the Ising coupling matrix²². This allows us to speedup the computation on a CPU by leveraging symmetry and sparsity. We show that the regime of operation described by Equation (54) also probes the expected phase transition (see SI, section IV).

A standard way of locating the critical temperature of a system is through the use of the Binder cumulant¹ $U_4(L) = 1 -$

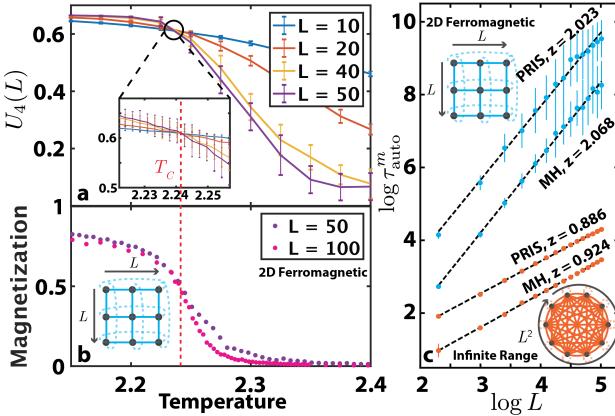


FIG. 4. Detecting and characterizing critical behaviors. **a:** Binder cumulants $U_4(L)$ for various graph sizes L on the 2D Ferromagnetic Ising model. Their intersection determines the critical temperature of the model T_C (denoted by a dotted line). **b:** Magnetization estimated from the PRIS for various L . **c:** Scaling of the PRIS magnetization autocorrelation time for various Ising models, benchmarked versus the Metropolis-Hastings algorithm (MH). The complexity of a single time step scales like $N^2 = L^4$ for MH on a CPU and like $N = L^2$ for the PRIS on a photonic platform. For readability, error bars in (b) are not shown (see SI).

$\langle m^4 \rangle / (3 \langle m^2 \rangle^2)$, where $m = \sum_{i=1}^N \sigma_i / N$ is the magnetization and $\langle \cdot \rangle$ denotes the ensemble average. As shown in Figure 4(a), the Binder cumulants intersect for various graph sizes $L^2 = N$ at the critical temperature of $T_C = 2.241$ (compared to the theoretical value of 2.269 for the two-dimensional Ferromagnetic Ising model, i.e. within 1.3%). The MC samples generated by the PRIS can be used to compute physical observables of the modeled system, which exhibit the emblematic order-disorder phase transition of the two-dimensional Ising model^{1,24} (Figure 4(b)). In addition, critical parameters describing the scaling of the magnetization and susceptibility at the critical temperature can be extracted from the PRIS to within 10% of the theoretical value (see SI, section IV).

In Figure 4(c), we benchmark the performance of the PRIS against the well-known Metropolis-Hastings (MH) algorithm^{1,56,57}. In the context of Markov Chain MC, one should compare the autocorrelation time of a given observable. The scaling of the magnetization autocorrelation time $\tau_m^{\text{auto}} = \mathcal{O}(L^z) = \mathcal{O}(N^{z/2})$ at the critical temperature is shown in Figure 4(c) for two analytically-solvable models: the two-dimensional ferromagnetic and the infinite-range Ising models. Both algorithms yield autocorrelation time critical exponents close to the theoretical value ($z \sim 2.1$)¹ for the two-dimensional Ising model. However, the PRIS seems to perform better on denser models such as the infinite-range Ising model, where it yields a smaller autocorrelation time critical exponent. More significantly, the advantage of the PRIS resides in its possible implementation with parallel photonic networks, so that the computational complexity of a single step is $\mathcal{O}(N)^{15,32,33}$. Thus, the computational complexity of generating an uncorrelated sample scales like $\mathcal{O}(N^{1+z_{\text{PRIS}}/2})$

for the PRIS on a parallel architecture, while it scales like $\mathcal{O}(N^{2+z_{\text{MH}}/2})$ for a sequential implementation of MH, on a CPU for instance. Implementing the PRIS on a photonic parallel architecture also ensures that the prefactor in this order of magnitude estimate is small (and only limited by the clock rate of a single recurrent step of this high-speed network). Thus, as long as $z_{\text{PRIS}} < z_{\text{MH}} + 2$, the PRIS exhibits a clear advantage over MH implemented on a sequential architecture.

To conclude, we presented in this Letter the PRIS, a photonic-based algorithm able to efficiently find ground states of general Ising problems. Essential to our approach are the high-speed, low-energy, and parallel computations enabled by recently-demonstrated photonic networks¹⁵. Moreover, our system requires some amount of noise to perform better, which is an unusual behavior only observed in very few physical systems. For instance, neuroscientists have conjectured that this could be a feature of the brain⁵⁸. The PRIS also performs a static transformation (and the state evolves to find the ground state). This kind of computation, quite general in Markov Chain MC sampling, can rely on a fundamental property of photonics — passivity — and thus reach even higher efficiencies.

The ability of the PRIS to detect phase transitions and probe critical exponents is particularly promising for the study of universality classes, as numerical simulations suffer from critical slowing down: the autocorrelation time grows exponentially at the critical point, thus making most samples too correlated to yield accurate estimates of physical observables. Our study suggests that this fundamental issue could be bypassed with the PRIS, which can generate a very large number of samples per unit time – only limited by the bandwidth of active silicon photonics components.

The experimental realization of the PRIS on a photonic platform would require additional work compared to the demonstration of deep learning with nanophotonic circuits¹⁵. The noise level can be dynamically induced by several well-known sources of noise in photonic and electronic systems⁴⁵. However, controlling and/or characterizing the noise level can be quite challenging. Moreover, the PRIS requires an additional homodyne detection unit, in order to detect both the amplitude and the phase of the output signal from the linear photonic domain. Nonetheless, these experimental challenges do not impact the promising scaling properties of the PRIS.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Aram Harrow, Mehran Kardar, Ido Kaminer, Miriam Farber, Theodor Misiakiewicz, Manan Raval, Nicholas Rivera, Nicolas Romeo, Jamison Sloan, and Can Knaut for helpful discussions. The authors would also like to thank Angelika Wiegele (Alpen-Adria-Universität Klagenfurt) for providing solutions of the Ising models considered in this work with $N \geq 50$ (computed with BiqMac⁵⁹). This work was supported in part by the Semiconductor Research Corporation (SRC) under SRC

contract #2016-EP-2693-B (Energy Efficient Computing with Chip-Based Photonics MIT). This work was supported in part by the National Science Foundation (NSF) with NSF Award #CCF-1640012 (E2DCA: Type I: Collaborative Research: Energy Efficient Computing with Chip-Based Photonics). This material is based upon work supported in part by the U. S. Army Research Laboratory and the U.S. Army Research Office through the Institute for Soldier Nanotechnologies, under contract number W911NF-18-2-0048. C. Z. was financially supported by the Whiteman Fellowship. M. P. was financially supported by NSF Graduate Research Fellowship grant number 1122374.

AUTHOR CONTRIBUTIONS

C. R.-C., Y. S. and M.S. conceived the project. C. R.-C. and Y. S. developed the analytical models and numerical calculations, with contributions from C. Z., M. P., L. J. and T. D.; C. R.-C. and C. Z. performed the benchmarking of the PRIS on analytically solvable Ising models. C. R.-C. and F.A. developed the analytics for various noise distributions. Y. S., J. D. J., D. E. and M.S. supervised the project. C. R.-C. wrote the manuscript with input from all authors.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to C. R.-C.

COMPETING FINANCIAL INTERESTS

The authors declare the following patent application: U.S. Patent Application No.: 16/032,737.

* Corresponding author e-mail: chrc@mit.edu

- [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 Berlin Heidelberg, 2013.
- [3] M. Kardar, G. Parisi, and Y.-C. Zhang, “Dynamic Scaling of Growing Interfaces,” *Physical Review Letters*, vol. 56, pp. 889–892, mar 1986.
- [4] M. B. Isichenko, “Percolation, statistical topography, and transport in random media,” *Reviews of Modern Physics*, vol. 64, pp. 961–1043, oct 1992.
- [5] 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 Biophysica Acta (BBA) - Biomembranes*, vol. 1788, pp. 53–63, jan 2009.
- [6] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of Modern Physics*, vol. 74, pp. 47–97, jan 2002.
- [7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” *Science (New York, N.Y.)*, vol. 220, pp. 671–80, may 1983.
- [8] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, “Coherent Ising machine based on degenerate optical parametric oscillators,” *Physical Review A*, vol. 88, p. 063853, dec 2013.
- [9] 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 (New York, N.Y.)*, vol. 354, pp. 614–617, nov 2016.
- [10] 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: Science & Applications*, vol. 3, pp. e147–e147, feb 2014.
- [11] 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,” *Optics Express*, vol. 26, p. 702, jan 2018.
- [12] W. M. Macready, A. G. Siapas, and S. A. Kauffman, “Criticality and Parallelism in Combinatorial Optimization,” *Science*, vol. 271, pp. 56–59, jan 1996.
- [13] A. Silva, F. Monticone, G. Castaldi, V. Galdi, A. Alù, and N. Engheta, “Performing mathematical operations with metamaterials,” *Science*, vol. 343, no. 6167, pp. 160–163, 2014.
- [14] A. F. Koenderink, A. Alù, and A. Polman, “Nanophotonics: Shrinking light-based technology,” may 2015.
- [15] 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,” *Nature Photonics*, vol. 11, pp. 441–446, jun 2017.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, may 2015.
- [17] W. A. Little, “The existence of persistent states in the brain,” *Mathematical Biosciences*, vol. 19, no. 1-2, pp. 101–120, 1974.
- [18] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, pp. 2554–8, apr 1982.
- [19] J. J. Hopfield and D. W. Tank, “Neural computation of decisions in optimization problems,” *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152.
- [20] E. Ising, “Beitrag zur Theorie des Ferromagnetismus,” *Z. Phys.*, 1925.
- [21] M. Mézard and A. Montanari, *Information, Physics, and Computation*. 2009.
- [22] D. J. Amit, H. Gutfreund, and H. Sompolinsky, “Spin-glass models of neural networks,” *Physical Review A*, vol. 32, pp. 1007–1018, aug 1985.
- [23] A. Pelissetto and E. Vicari, “Critical phenomena and renormalization-group theory,” *Physics Reports*, vol. 368, pp. 549–727, oct 2002.
- [24] L. Onsager, “Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition,” *Physical Review*, vol. 65, pp. 117–149, feb 1944.
- [25] N. V. Brilliantov, “Effective magnetic Hamiltonian and Ginzburg criterion for fluids,” *Physical Review E*, vol. 58, pp. 2628–2631, aug 1998.

- [26] D. J. Amit, *Modeling brain function : the world of attractor neural networks*. Cambridge University Press, 1989.
- [27] M. A. Halasz, A. D. Jackson, R. E. Shrock, M. A. Stephanov, and J. J. M. Verbaarschot, “Phase diagram of QCD,” *Physical Review D*, vol. 58, p. 096007, sep 1998.
- [28] F. Barahona, “On the computational complexity of Ising spin glass models,” *Journal of Physics A: Mathematical and General*, vol. 15, pp. 3241–3253, oct 1982.
- [29] J. Bruck and J. W. Goodman, “On the power of neural networks for solving hard problems,” *Journal of Complexity*, vol. 6, pp. 129–135, jun 1990.
- [30] N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, “Optical implementation of the Hopfield model,” *Applied Optics*, vol. 24, p. 1469, may 1985.
- [31] 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*, vol. 349, pp. 711–716, aug 2015.
- [32] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, “Experimental realization of any discrete unitary operator,” *Physical Review Letters*, vol. 73, pp. 58–61, jul 1994.
- [33] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, “Optimal design for universal multiport interferometers,” *Optica*, vol. 3, p. 1460, dec 2016.
- [34] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, M. Jarrahi, and A. Ozcan, “All-Optical Machine Learning Using Diffractive Deep Neural Networks,” apr 2018.
- [35] M. Gruber, J. Jahns, and S. Sinzinger, “Planar-integrated optical vector-matrix multiplier,” *Applied Optics*, vol. 39, p. 5367, oct 2000.
- [36] A. N. Tait, M. A. Nahmias, Y. Tian, B. J. Shastri, and P. R. Prucnal, “Photonic Neuromorphic Signal Processing and Computing,” pp. 183–222, Springer, Berlin, Heidelberg, 2014.
- [37] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, “Broadcast and weight: An integrated network for scalable photonic spike processing,” *Journal of Lightwave Technology*, vol. 32, no. 21, pp. 3427–3439, 2014.
- [38] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, “Experimental demonstration of reservoir computing on a silicon photonics chip,” *Nature Communications*, vol. 5, p. 3541, dec 2014.
- [39] A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Dremeau, S. Gigan, and F. Krzakala, “Random projections through multiple optical scattering: Approximating Kernels at the speed of light,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6215–6219, IEEE, mar 2016.
- [40] 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 Journal of Selected Topics in Quantum Electronics*, vol. 20, pp. 43–48, jan 2014.
- [41] 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 Research*, vol. 4, pp. 297–307, mar 2011.
- [42] A. C. Selden, “Pulse transmission through a saturable absorber,” *British Journal of Applied Physics*, vol. 18, pp. 743–748, jun 1967.
- [43] M. Soljačić, M. Ibanescu, S. G. Johnson, Y. Fink, and J. D. Joannopoulos, “Optimal bistable switching in nonlinear photonic crystals,” *Physical Review E*, vol. 66, p. 055601, nov 2002.
- [44] R. W. Schirmer and A. L. Gaeta, “Nonlinear mirror based on two-photon absorption,” *Journal of the Optical Society of America B*, vol. 14, p. 2865, nov 1997.
- [45] P. Horowitz and H. Winfield, “The Art of Electronics,” *American Journal of Physics*, 1990.
- [46] B. Boser, E. Sackinger, J. Bromley, Y. Le Cun, and L. Jackel, “An analog neural network processor with programmable topology,” *IEEE Journal of Solid-State Circuits*, vol. 26, no. 12, pp. 2017–2025, 1991.
- [47] J. Misra and I. Saha, “Artificial neural networks in hardware: A survey of two decades of progress,” *Neurocomputing*, vol. 74, pp. 239–255, dec 2010.
- [48] D. Vrataric, V. Ceperic, and A. Baric, “Area-efficient differential Gaussian circuit for dedicated hardware implementations of Gaussian function based machine learning algorithms,” *Neurocomputing*, vol. 118, pp. 329–333, oct 2013.
- [49] P. Peretto, “Collective properties of neural networks: A statistical physics approach,” *Biological Cybernetics*, vol. 50, pp. 51–62, feb 1984.
- [50] M. Lipson, “Guiding, Modulating, and Emitting Light on Silicon-Challenges and Opportunities,” *Journal of Lightwave Technology*, Vol. 23, Issue 12, pp. 4222-, vol. 23, p. 4222, dec 2005.
- [51] 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,” *Optics Express*, vol. 22, p. 10487, may 2014.
- [52] 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, E. Rieffel, H. Takesue, and Y. Yamamoto, “Scaling advantages of all-to-all connectivity in physical annealers: the Coherent Ising Machine vs. D-Wave 2000Q,” *arXiv preprints*, may 2018.
- [53] V. R. Almeida, C. A. Barrios, R. R. Panepucci, and M. Lipson, “All-optical control of light on a silicon chip,” *Nature*, vol. 431, pp. 1081–1084, oct 2004.
- [54] C. T. Phare, Y. H. Daniel Lee, J. Cardenas, and M. Lipson, “Graphene electro-optic modulator with 30 GHz bandwidth,” *Nature Photonics*, 2015.
- [55] C. Haffner, D. Chelladurai, Y. Fedoryshyn, A. Josten, B. Baeuerle, W. Heni, T. Watanabe, T. Cui, B. Cheng, S. Saha, D. L. Elder, L. R. Dalton, A. Boltasseva, V. M. Shalaev, N. Kinsey, and J. Leuthold, “Low-loss plasmon-assisted electro-optic modulator,” *Nature*, 2018.
- [56] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The Journal of Chemical Physics*, 1953.
- [57] W. K. Hastings, “Monte carlo sampling methods using Markov chains and their applications,” *Biometrika*, 1970.
- [58] D. C. Knill and A. Pouget, “The Bayesian brain: the role of uncertainty in neural coding and computation,” *Trends in Neurosciences*, vol. 27, pp. 712–719, dec 2004.
- [59] F. Rendl, G. Rinaldi, and A. Wiegele, “Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations,” *Mathematical Programming*, vol. 121, pp. 307–335, feb 2010.
- [60] “In this part, we operate the PRIS in the regime where the linear photonic matrix is equal to the Ising coupling matrix. This allows us to speedup the computation on our CPU by leveraging symmetry and sparsity. We show the other regime of o,”
- [61] S. Friedli and Y. Velenik, *Statistical Mechanics of Lattice Systems*. Cambridge University Press, nov 2017.

- [62] S. Kirkpatrick and B. Selman, “Critical behavior in the satisfiability of random boolean expressions,” *Science (New York, N.Y.)*, 1994.
- [63] D. W. Heermann and A. N. Burkitt, “System size dependence of the autocorrelation time for the Swendsen-Wang Ising model,” *Physica A: Statistical Mechanics and its Applications*, vol. 162, pp. 210–214, jan 1990.
- [64] A. C. Selden, “Pulse transmission through a saturable absorber,” *British Journal of Applied Physics*, vol. 18, pp. 743–748, jun 1967.
- [65] A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Dremeau, S. Gigan, and F. Krzakala, “Random projections through multiple optical scattering: Approximating Kernels at the speed of light,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6215–6219, IEEE, mar 2016.
- [66] A. N. Tait, M. A. Nahmias, Y. Tian, B. J. Shastri, and P. R. Prucnal, “Photonic Neuromorphic Signal Processing and Computing,” pp. 183–222, Springer, Berlin, Heidelberg, 2014.
- [67] M. Gruber, J. Jahns, and S. Sinzinger, “Planar-integrated optical vector-matrix multiplier,” *Applied Optics*, vol. 39, p. 5367, oct 2000.
- [68] A. Azzalini and A. D. VALLE, “The multivariate skew-normal distribution,” *Biometrika*, vol. 83, pp. 715–726, dec 1996.
- [69] C.-K. Looi, “Neural network methods in combinatorial optimization,” *Computers & Operations Research*, vol. 19, pp. 191–208, apr 1992.
- [70] R. M. Karp, “Reducibility among Combinatorial Problems,” in *Complexity of Computer Computations*, pp. 85–103, Boston, MA: Springer US, 1972.

PHOTONIC RECURRENT ISING SAMPLER – SUPPLEMENTARY INFORMATION

In this work, we first provide a comprehensive theoretical analysis of the dynamics of the Photonic Recurrent Ising Sampler (PRIS), with an emphasis on how the weight matrix in the algorithm should be chosen. We then detail the parameters of our numerical simulations carried to estimate the performance of the PRIS on finding the ground state of arbitrary Ising problems and to sample critical behaviors.

I – THEORETICAL PRELIMINARIES

Definition of an Ising problem

The general definition of an Ising problem is the following: given a matrix $(K_{ij})_{(i,j) \in I^2}$, and a vector $\{b_i\}_{i \in I}$ where I is a finite set with cardinality $|I| = N$, we want to find the spin distribution $\{\sigma_i\}_{i \in I} \in \{-1, 1\}^N$ minimizing the following Hamiltonian function:

$$H_{K,b}(\{\sigma_i\}) = - \sum_{1 \leq i < j \leq N} K_{ij} \sigma_i \sigma_j - \sum_i b_i \sigma_i \quad (7)$$

$$= -\frac{1}{2} \sum_{1 \leq i, j \leq N} K_{ij} \sigma_i \sigma_j - \sum_i b_i \sigma_i. \quad (8)$$

In statistical mechanics, this model represents the energy of an interacting set of spins. The coupling term K_{ij} represents the coupling between spins i and j . In the general definition of the Ising model, the coupling matrix K is assumed to be symmetric. We could also assume that it has a null diagonal, as it would only add a constant offset to the Hamiltonian. We are not making this extra assumption, as it will prove later to be useful. An external magnetic field b_i can be applied, which breaks the symmetry of the problem. This general class of Ising problems is NP-hard⁷⁰ and many subclasses of this problem exhibit a similar complexity. In this work, we will characterize our optical algorithms on two subclasses of problems:

- ▷ General antiferromagnets, for which the coupling K can only take two discrete values $K_{ij} \in \{0, -1\}$, and $b_i = 0$ for all i . As every problem in this subclass is equivalent to an unweighted MAX-CUT problem, and that MAX-CUT is known to be NP-hard⁷⁰, this subclass is already NP-hard.
- ▷ Spin glasses, for which the coupling K can take on continuous values in the range $[-1, 1]$, and $b_i = 0$.

However, our approach can be easily extended to any coupling matrix K . The possible extension to non-zero external magnetic fields will also be discussed. A complete study of the hardness of the various subclasses of Ising problems can be found in Ref.²⁸.

II – GENERAL THEORY OF PHOTONIC RECURRENT ISING SAMPLER

Optical Parallel Recurrent MCMC algorithm

The Little and Hopfield networks^{17,18} are two early forms of recurrent neural networks, originally designed to understand and model associative memory processes. In their general acceptance, the Little network can be understood as a synchronous version of the Hopfield network. Using the latter to solve the NP-hard Ising problem was previously investigated^{19,69}. The behavior of both networks was later studied from a statistical mechanical perspective by P. Peretto⁴⁹ and Daniel J. Amit and colleagues²². We here generalize their study to the case of noisy, synchronous Hopfield network, or noisy Little network⁴⁹ and investigate the possibility of sampling hard Ising problems with a parallel, recurrent optical system. In the following, we equivalently use the term “neuron” and “spin” in the framework of spin glass models of neural networks.

Our machine is applying the following algorithm:

- ▷ Initialize assembly of neurons $\{\sigma_i\}$ with random values. The signal is coded into optical domain with the reduced spins $S_i = (\sigma_i + 1)/2 \in \{0, 1\}$.
- ▷ At each step of the algorithm, each neuron is applied a potential v_i (random variable) whose expected value is given by a matrix multiplication with $C = 2J$:

$$\bar{v}_i = \sum_j C_{ij} S_j = \sum_{j=1}^N J_{ij} \sigma_j + \sum_{j=1}^N C_{ij}/2 \quad (9)$$

and whose probability density is given by the function f_ϕ with mean \bar{v}_i and standard deviation ϕ :

$$\int f_\phi(x) x dx = \bar{v}_i, \quad (10)$$

$$\int f_\phi(x)(x - \bar{v}_i)^2 dx = \phi^2. \quad (11)$$

- ▷ A non-linear threshold Th_θ is then applied to \bar{v}_i to yield the neural state at the next time step:

$$\text{Th}_{\theta_i}(S_i) = \begin{cases} 0, & \text{if } v_i < \theta_i, \\ 1, & \text{otherwise.} \end{cases} \quad (12)$$

To summarize, the sequential transformation of the neuron state is:

$$S^{(t+1)} = \text{Th}_\theta(X^{(t)}), \quad (13)$$

$$X^{(t)} \sim \mathcal{N}(CS^t|\phi), \quad (14)$$

where $\mathcal{N}(x|\phi)$ is a gaussian distribution with mean x and standard deviation ϕ . The algorithm is fully determined by the following set of transformation variables: (C, f_φ, θ) .

Determination of the Temperature factor

In this section, we compute the stationary distribution of the spin variable S^t for the general case with variables (C, f_φ, θ) . We first determine the probability of a single spin-flip, knowing that the current spin state is J :

$$W(\sigma_i(I)|J) = G(h_i(J)\sigma_i(I)), \quad (15)$$

$$h_i(J) = -\sum_{j=1}^N J_{ij} \sigma_j(J) - h_i^0, \quad (16)$$

$$h_i^0 = \sum_j C_{ij}/2 - \theta_i, \quad (17)$$

Noise distribution	$G(x)$	$\frac{k}{2}$	$\epsilon_0 \equiv \max_{X \in \mathbb{R}} \epsilon_{\gamma_{\text{opt}}}(X) $
Logistic	$\frac{1}{1+\exp\left(\frac{\pi x}{\sqrt{3}\phi}\right)}$	$\frac{\sqrt{3}}{\pi}$	0
Gaussian	$\frac{1}{2}(1 - \text{erf}(\frac{x}{\sqrt{2}\phi}))$	0.5877	0.0095
Cauchy*	$\frac{1}{\pi} \arctan\left(\frac{x}{\phi}\right) + \frac{1}{2}$	1.16	0.0495
Laplace	$\begin{cases} \frac{1}{2} \exp(-\frac{\sqrt{2}x}{\phi}) & x \leq 0 \\ 1 - \frac{1}{2} \exp(\frac{\sqrt{2}x}{\phi}) & x \geq 0 \end{cases}$	0.4735	0.0199
Uniform	$\begin{cases} 0 & x \leq -\sqrt{3}\phi \\ \frac{x+\sqrt{3}\phi}{2\sqrt{3}\phi} & x \in [-\sqrt{3}\phi, \sqrt{3}\phi[\\ 1 & x \geq \sqrt{3}\phi \end{cases}$	0.6136	0.0561

TABLE I. **Summary of temperature factors for various noise distributions.** Each noise distribution is defined so that its standard deviation, when applicable, is equal to ϕ . (*the standard deviation of the Cauchy distribution is not defined, the linear dependence is measured as a function of the scaling parameter ϕ in this particular case.)

where we define G as a rescaled version of the noise cumulative density function:

$$F(x) = \int_{-\infty}^x f_\phi(u) du, \quad (18)$$

$$F_0(x) = F(x + \bar{v}_i), \quad (19)$$

$$G(x) = 1 - F_0(x). \quad (20)$$

In the following, we only assume that f_ϕ is symmetric and has a standard deviation ϕ . The symmetry assumption is not constraining as most noise distributions found in nature have a symmetric (and, in many cases, gaussian) distribution⁴⁵. Our analysis can also be extended to noise distributions whose variance is infinite by parameterizing the distribution with another parameter, usually referred to as a “scaling” parameter (see Table I). The case of f_ϕ being a gaussian distribution is discussed in Ref.⁴⁹, where the function G can be approximated by a sigmoid function when rescaling it by the adequate factor. In the general case, in order to minimize this approximation error, we choose the following rescaling factor

$$\begin{aligned} G_\gamma(x) &= G(\gamma x), \\ 1 - s(x) &= \frac{1}{1 + e^{-x}}, \\ \gamma_{\text{opt}} &= \operatorname{argmin}_{\gamma'} \max_{X \in \mathbb{R}} |G_{\gamma'}(X) - (1 - s(X))| = \operatorname{argmin}_{\gamma'} \max_{X \in \mathbb{R}} |\epsilon_\gamma(X)|, \end{aligned}$$

where $s(x) = 1/(1 + e^{-x})$ is the sigmoid function.

Naturally, the optimal scaling factor will depend on the standard deviation ϕ of the original probability distribution f . We numerically observe that this dependence is linear

$$\gamma_{\text{opt}} = \frac{k}{2} \phi. \quad (21)$$

Optimal scaling factors and corresponding errors are reported in Table I. Identifying the transition probability to a sigmoid function is a necessary step to compute the effective Hamiltonian of the spin distribution⁴⁹.

Transition probability

By choosing adequately the temperature factor $\frac{k}{2}$, we minimize the error when approximating the transition probability by a sigmoid function (in the following, $h_i = h_i(J)$ and $\sigma_i = \sigma_i(I)$, for readability):

$$W(\sigma_i|J) = G_{\gamma_{\text{opt}}} \left(\frac{h_i \sigma_i}{\gamma_{\text{opt}}} \right) \quad (22)$$

$$= s \left(\frac{h_i \sigma_i}{\gamma_{\text{opt}}} \right) + \epsilon_{\gamma_{\text{opt}}} \left(\frac{h_i \sigma_i}{\gamma_{\text{opt}}} \right) \quad (23)$$

$$\approx s \left(\frac{h_i \sigma_i}{\gamma_{\text{opt}}} \right). \quad (24)$$

In the following, we will drop the subscript in γ_{opt} and will assume that its value is known (it can be determined numerically, given the noise distribution). The error function ϵ can be used as an expansion parameter, in order to estimate the accuracy of the approximation performed when neglecting this term (as is done in Ref.⁴⁹).

Expansion of the transition probability in terms of ϵ

The transition probability can be derived by multiplying the probabilities of single spin-flips, these events being independent:

$$W(I|J) = \prod_i W(\sigma_i|J) \quad (25)$$

$$= \prod_i \left(s \left(\frac{h_i \sigma_i}{k\phi/2} \right) + \epsilon \left(\frac{h_i \sigma_i}{k\phi/2} \right) \right) \quad (26)$$

$$= W^0(I|J) + \sum_{k=1}^N W^k(I|J), \quad (27)$$

where $W^k(I|J)$ is the k -th order term in the expansion, assuming the error function ϵ is small ($\epsilon \ll s$). The zero-th order term of this expansion gives us back the result from⁴⁹:

$$W^0(I|J) = \frac{e^{-\beta H^0(I|J)}}{\sum_K e^{-\beta H^0(K|J)}}, \quad (28)$$

$$H^0(I|J) = - \sum_{ij} J_{ij} \sigma_i(I) \sigma_j(J) - \sum_i h_i^0 \sigma_i(I), \quad (29)$$

$$\beta = \frac{1}{k\phi}. \quad (30)$$

Another way to write this zero-th order term⁴⁹ will prove to be useful later in our derivations:

$$W^0(I|J) = \frac{e^{-\beta H^0(I|J)}}{\prod_i 2 \cosh(\beta h_i(J))}. \quad (31)$$

The general expression for the k -th order term can be derived:

$$W^k(I|J) = W^0(I|J) \sum_{j_1} \dots \sum_{j_k \notin \{j_1, \dots, j_{k-1}\}} \prod_l \epsilon(2\beta h_{j_l} \sigma_{j_l}) (1 + \exp(2\beta h_{j_l} \sigma_{j_l})). \quad (32)$$

The N -th order term scales, in the worst case scenario, like ϵ_0^N :

$$|W^N(I|J)| = \left| \prod_k \epsilon(2\beta h_{j_k} \sigma_{j_k}) \right| \leq \epsilon_0^N. \quad (33)$$

For $k < N$, analyzing the scaling of the k -th order term requires more assumptions. Let's look at the case $k = 1$, to verify that we can safely neglect higher-order terms in this expansion:

$$\left| \frac{W^1(I|J)}{W^0(I|J)} \right| \leq \sum_j |\epsilon(2\beta h_j \sigma_j) (1 + \exp(2\beta h_j \sigma_j))|. \quad (34)$$

This ratio scales exponentially with $\beta h_j \sigma_j / 2$. However, in this case, the error function ϵ also goes to zero. In addition, increasing the value of $\beta h_j \sigma_j / 2$ also increases the value of the Hamiltonian $H^0(I|J)$, and thus reduces the likelihood of the transition $W(I|J)$. **Thus, the larger the ratio $|W^1(I|J)|/|W^0(I|J)|$, the smaller the transition probability.** In the following, we will neglect all high order terms $k \geq 1$.

We do not extend our derivations to a general non-symmetric noise distribution in this work. We still suggest the following ideas to treat this extension:

- ▷ We could first treat the non-symmetric case as a perturbation of the symmetric case and thus derive the error on the effective Hamiltonian of the spin distribution.
- ▷ In this view, we suggest the parametrization of the skewness using the skew normal distribution⁶⁸, which is a natural extension of the toy-model symmetric noise distribution analyzed in Ref.⁴⁹.

Detailed balance condition

To determine the stationary distribution of the spin state, we first need to verify that the transition probability satisfies the triangular equality⁴⁹, equivalent to the detailed balance condition in the context of Markov Chains¹:

$$W(I|K)W(K|J)W(J|I) = W(J|K)W(K|I)W(I|J). \quad (35)$$

This is equivalent to:

$$H(I|K) + H(K|J) + H(J|I) = H(J|K) + H(K|I) + H(I|J), \quad (36)$$

which is satisfied if J is symmetric (there is no condition on the linear term h_i^0 in the Hamiltonian). We can then deduce the effective Hamiltonian by decoupling the ratio of transition probabilities into the ratio of two terms depending on distributions I and J (by using Eq. (31))⁴⁹

$$\frac{W^0(I|J)}{W^0(J|I)} = \frac{\exp(\beta \sum_i h_i^0 \sigma_i(I))}{\exp(\beta \sum_i h_i^0 \sigma_i(J))} \frac{\prod_i \cosh(\beta(\sum_j J_{ij} \sigma_j(I) + h_i^0))}{\prod_i \cosh(\beta(\sum_j J_{ij} \sigma_j(J) + h_i^0))} \quad (37)$$

$$= \frac{F(I)}{F(J)}, \quad (38)$$

and $H_L(I) = -\frac{1}{\beta} \ln F(I)$ gives the effective Hamiltonian describing the dynamics of the spin distribution:

$$H_L(I) = -\frac{1}{\beta} \sum_i \log \cosh(\beta(\sum_j J_{ij} \sigma_j(I) + h_i^0)) - \sum_i h_i^0 \sigma_i(I). \quad (39)$$

Effective Gibbs distribution

Assuming J is symmetric, the stationary distribution of the spin state is a Gibbs distribution given by the effective Hamiltonian $H_L(I)$ ^{22,49}:

$$P(\{\sigma_i\}) \propto \exp(-\beta H_L(\{\sigma_i\})). \quad (40)$$

In the following, we define $\|\cdot\|_k$ as the usual k -norm (where k is an integer).

Small noise approximation

In the small noise approximation ($\beta \gg 1$), the effective Hamiltonian can be simplified using the following Taylor expansion : $\log \cosh x \approx |x| - \log 2$:

$$H_L(\{\sigma_i\}, \beta) \approx -\sum_i \left| \sum_j J_{ij} \sigma_j + h_i^0 \right| + \frac{N}{\beta} \log 2 - \sum_i h_i^0 \sigma_i \quad (41)$$

$$= -\|J\vec{\sigma} + \vec{h}^0\|_1 + \frac{N}{\beta} \log 2 - \vec{h}^0 \cdot \vec{\sigma} := H_1(\{\sigma_i\}, \beta). \quad (42)$$

Large noise approximation

In the large noise approximation ($\beta \ll 1$), the effective Hamiltonian can be simplified using the following Taylor expansion : $\log \cosh x \approx \log(1 + x^2/2) \approx x^2/2$,

$$H_L(\{\sigma_i\}, \beta) \approx -\underbrace{\frac{\beta}{2} \sum_{ij} \tilde{K}_{ij} \sigma_i \sigma_j}_{\text{Quadratic term}} - \underbrace{\sum_i \sigma_i \left(\beta \sum_k J_{ki} h_k^0 + h_i^0 \right)}_{\text{Linear term}} \quad (43)$$

$$- \underbrace{\frac{\beta}{2} \sum_j (h_j^0)^2}_{\text{Constant (independent of } \sigma\text{)}} \quad (44)$$

$$= -\frac{\beta}{2} \|J\vec{\sigma} + \vec{h}^0\|_2^2 - \vec{h}^0 \cdot \vec{\sigma} := H_2(\{\sigma_i\}, \beta), \quad (45)$$

with $\tilde{K} = J^2$.

Discussion on h_i^0

Identifying the Hamiltonian in Eq. (45) to the general Ising Hamiltonian in Eq. (8) requires solving the following system of equations:

$$\left\{ \begin{array}{l} K_{ij} + \delta_{ij}\Delta_i = \beta \sum_k J_{ik}J_{kj} \\ b_i = \beta \sum_j J_{ki}h_k^0 + h_i^0 \end{array} \right. \quad (46)$$

$$(47)$$

$$J_{ij} = J_{ji}, \quad (48)$$

with unknown variables (h_i^0, J) . δ_{ij} is the Kronecker symbol and Δ_i represents a degree of freedom on the diagonal of the Hamiltonian (which results in an additional constant term). Studying the set of general Ising Hamiltonians (K, b) (Eq. (8)) which can be mapped to an Ising network (J, h^0) in the large noise approximation (Eq. (45)) would require an extensive study that we do not carry in this work. However, one can notice that:

- ▷ In the case where the external field satisfies $J \cdot h^0 = 0$, the system has a trivial solution $h_i^0 = b_i$ and $J = \sqrt{K}$ (the conditions for which a square root of K can be found are discussed in the next section).
- ▷ In the more general case, one is given N “free” degrees of freedoms Δ_i in finding solutions to Eq. (46), while there are N constraints to solve in Eq. (47). Thus, it seems likely that, in non-degenerate cases, this system will have a non-trivial solution. However, one should make sure that while tuning the degrees of freedom Δ_i , the matrix J remains a real square root of the matrix K (see discussion in the next section on finding a square root). If not, this algorithm should be generalized to using complex-valued matrices J . We do not discuss this generalization of the algorithm in this work. However, considering that arbitrary (complex) unitary transformations can be implemented with our photonic architecture¹⁵, this would be an interesting extension to this work.

In the following, for the sake of simplicity, we will assume $h_i^0 \equiv 0$. In this case, the large noise approximation Hamiltonian gives:

$$H_L(\{\sigma_i\}, \beta) \approx -\frac{\beta}{2} \sum_{ij} \tilde{K}_{ij} \sigma_i \sigma_j, \quad (49)$$

with $\tilde{K} = J^2$. Thus, the implementation of the Little network to model a general Ising Hamiltonian (Eq. (8)) strongly relies **on the possibility of finding a symmetric, real square root to the matrix K** .

Let us remind the reader of the assumptions we have made so far:

- ▷ The model is *synchronous*, or as stated in Little’s original paper¹⁷ “we shall suppose that the neurons are not permitted to fire at any random time but rather that they are synchronized such that they can only fire at some integral multiple of a period τ ”.
- ▷ The neurons have no memory of states older than their previous state (Markov process).
- ▷ The connections do not evolve in time (there is no learning involved).
- ▷ J is symmetric and $h_i^0 = 0$, which results in

$$\theta_i = \sum_j C_{ij}/2, \quad (50)$$

from Eq. (17).

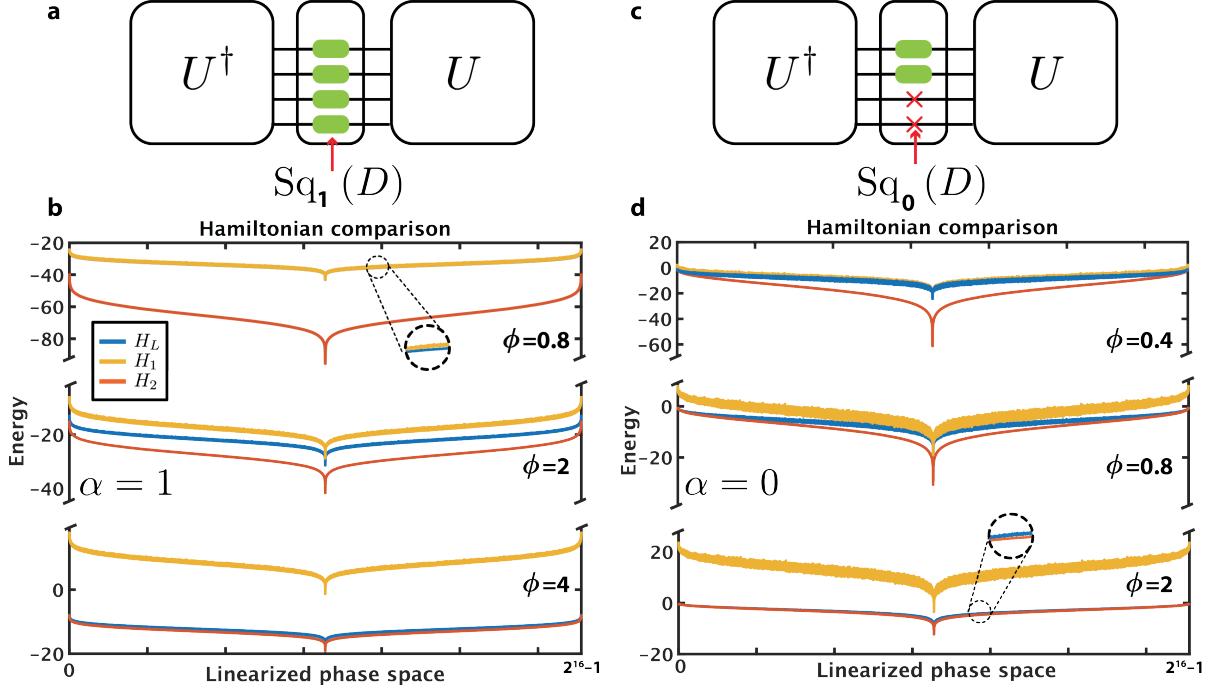


FIG. 1. Transitioning between Hamiltonians. **a:** To approximate a given Ising model in the large noise limit, one should take into account all eigenvalues. **b:** Effective Hamiltonian and its large/low-noise limit approximation when $\alpha = 1$ (no eigenvalue dropout). **c:** However, to reduce the noise threshold of the large noise expansion, one should drop out some eigenvalues. **d:** Effective Hamiltonian and its large/low-noise limit approximation when $\alpha = 0$ (all negative eigenvalues are dropped out).

Inequalities between Hamiltonians

Using basic algebra and functional analysis, we can show that $\log \cosh x \leq x^2/2$ and that $\log \cosh x \geq |x| - \log 2$, which results in the following inequality on the Hamiltonians

$$H_2(\{\sigma_i\}, \beta) \leq H_L(\{\sigma_i\}, \beta) \leq H_1(\{\sigma_i\}, \beta). \quad (51)$$

By summing over spin configurations, we get the reversed inequality for the partition functions

$$Z_2(\beta) \geq Z_L(\beta) \geq Z_1(\beta). \quad (52)$$

Another inequality can be derived by using the equivalence of norms $\|\cdot\|_1$ and $\|\cdot\|_2$ in finite dimensions ($\|\cdot\|_2 \leq \|\cdot\|_1 \leq \sqrt{N}\|\cdot\|_2$):

$$0 \leq \sqrt{-\frac{2}{\beta}H_2(\{\sigma_i\}, \beta)} \leq -H_1(\{\sigma_i\}, \beta) + \frac{N}{\beta} \log 2 \leq \sqrt{\frac{-2N}{\beta}H_2(\{\sigma_i\})}. \quad (53)$$

These Hamiltonians are plotted for a sample random Ising problem in Figure 1.

Suggested algorithms

As suggested above, from the large noise expansion of the effective Hamiltonian, a natural way to probe the Gibbs distribution of some Ising model defined by coupling matrix K (Eq. (8)) is to set the matrix of the recurrent loop C to be a modified square root of K :

$$\text{Sq}_\alpha(J) = \text{Re}(\sqrt{K + \alpha\Delta}), \quad (54)$$

where $\alpha \in [-1, 1]$ is the eigenvalue dropout level, a parameter we will discuss below, and Δ is a diagonal offset matrix defined as the sum of the off-diagonal term of the Ising coupling matrix $\Delta_{ii} = \sum_{j \neq i} |K_{ij}|$. An alternative solution would be to set $C = K$. This solution has been studied in the particular case of associative memory couplings²², where it was shown that the PRIS in this configuration would be described by a free energy function equal to the Ising free energy (up to a factor of 2). We will discuss the pros and cons of both algorithms in section IV.

III - CONSTRUCTION OF THE WEIGHT MATRIX

We here propose a technique to build a square root of K in order to find an approximate solution to *any* Ising problem. We start with the general K Ising weight matrix K from Eq. (8). We notice that as K is symmetric and $K_{ii} = 0$, K will never obey the condition of Lemma 1.

Lemma 1 (*Diagonal dominance.*) *If a $A \in S_N(\mathbb{R})$ is a real symmetric matrix such that for all i , $|A_{ii}| \geq \sum_{j \neq i} |A_{ij}|$ and $A_{ii} > 0$, there exists $B \in S_N(\mathbb{R})$ such that $B^2 = A$.*

Proof 1 Let X be a vector of size N and norm 1 such that $AX = \lambda X$ with $\lambda < 0$. We assume i_0 is the coordinate of X with maximum absolute value (and $x_{i_0} > 0$): $x_{i_0} = \|x\|_\infty$. We have $\sum_j A_{i_0 j} x_j = \lambda x_{i_0}$. We thus get :

$$\begin{aligned} \|x\|_\infty |A_{i_0 i_0}| &< |A_{i_0 i_0} - \lambda| \|x\|_\infty \\ &= \left| \sum_{j \neq i_0} A_{i_0 j} x_j \right| \\ &\leq \sum_{j \neq i_0} |A_{i_0 j}| \|x\|_\infty. \end{aligned}$$

By simplifying this inequality, we get $|A_{i_0 i_0}| < \sum_{j \neq i_0} |A_{i_0 j}|$ which contradicts our assumption. Thus, for all X , $X^T A X \geq 0$, which means that A is positive semidefinite.

We can thus write $A = UDU^T$ where U is unitary and $D = \text{Diag}(\lambda_1, \dots, \lambda_N)$ with $\lambda_i \geq 0$. The result is given by $B = U\sqrt{D}U^T \in S_N(\mathbb{R})$ with $\sqrt{D} = \text{Diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_N})$

To be able to apply Lemma 1, we need to add diagonal terms to the Ising matrix K from Eq. 8 by defining $\tilde{K} = K + \Delta$ where Δ is a diagonal matrix such that \tilde{K} verifies the assumptions of theorem 1. From there, we can define $J = \sqrt{\tilde{K}}$. In the large noise approximation, we thus get the following Hamiltonian describing the PRIS:

$$\begin{aligned} H_L(\{\sigma_i\}, \beta) &\approx -\frac{\beta}{2} \sum_{ij} \tilde{K}_{ij} \sigma_i \sigma_j = -\beta \sum_{1 \leq i < j \leq N} \tilde{K}_{ij} \sigma_i \sigma_j - \frac{\beta}{2} \sum_{1 \leq i \leq N} \tilde{K}_{ii} \sigma_i^2 \\ &= -\beta \sum_{1 \leq i < j \leq N} K_{ij} \sigma_i \sigma_j - \frac{\beta}{2} \sum_{1 \leq i \leq N} \Delta_{ii}. \end{aligned}$$

Since the Δ matrix is fixed, the second term in this equation is a constant (independent of the spin distribution, given an Ising problem to minimize). The Gibbs distribution is thus $P(\{\sigma_i\}) \propto \exp(-\beta H_L(\{\sigma_i\})) \approx C_{te} \exp(-\beta H_{K,0})$ where $H_{K,0}$ is the Ising Hamiltonian defined in Eq. (8).

Discussion on the diagonal offset

The diagonal offset that is added to the original Ising matrix can be considered as a supplementary degree of freedom. To verify Lemma 1, we can choose $\Delta_i = \sum_{j \neq i} |K_{ij}|$, to make sure \tilde{K} is positive definite and has a real symmetric square root $J^2 = \tilde{K}$.

However, there is no reciprocal to Lemma 1. We here show that even some partial or weak reciprocals are usually wrong:

▷ The direct reciprocal of Lemma 1 is generally wrong. We consider the following counter-example:

$$A = \begin{pmatrix} 1 & -2 \\ -2 & 8 \end{pmatrix}$$

. A is symmetric, has positive diagonal elements but is not diagonally dominant (the inequality does not hold for only one diagonal element). However, A is symmetric positive definite: its determinant is 4 and its trace 9, so the sum and product of eigenvalues is positive, which means both are positive.

▷ Even a weak reciprocal of Lemma 1 is wrong:

If $A \in S_N(\mathbb{R})$, A has positive diagonal elements, such that for all i , $|A_{ii}| < \sum_{j \neq i} |A_{ij}|$, then A has at least one negative eigenvalue.

For example, one can consider :

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

. A is symmetric, has positive diagonal elements, verifies: $\forall i, |A_{ii}| < \sum_{j \neq i} |A_{ij}|$, but still, its eigenvalues are 0, 0, and 3.

▷ One (very) weak reciprocal of Lemma 1 is the following Lemma 2. It means that any Ising matrix (that has a zero diagonal) will never have a real symmetric square root, unless we add an offset to its diagonal.

Lemma 2 *If $A \in S_N(\mathbb{R})$ (real symmetric matrix) such that for all i , $A_{ii} = 0$, then A has at least one negative (and one positive) eigenvalue.*

Proof 2 We have $\sum_i A_{ii} = \sum_i \lambda_i = 0$ so there is at least one negative and one positive eigenvalue.

▷ **Lemma 3 (Weak reciprocal: unweighted MAX-CUT case.)** *If K is equal to minus the adjacency matrix of a graph, then the matrix $\tilde{K} = K + \alpha\Delta$ is symmetric positive semi-definite if and only if $\alpha = 1$.*

Proof 3 Let's first notice that this case still holds for a NP-hard subclass of Ising problems. In this case, \tilde{K} 's elements are -1 if there is a spin-spin connection and 0 otherwise. Thus, $\tilde{K} = K + \Delta$ has a zero eigenvalue: $KX = 0$ with $X = (1, 1, \dots, 1)^T$ (because the sum of each row is equal to 0). We notice that for $\alpha < 1$:

$$K + \alpha\Delta = (K + \Delta) + (\alpha - 1)\Delta. \quad (55)$$

As seen before, $K + \Delta$ has a zero eigenvalue, and $(\alpha - 1)\Delta$ only has negative eigenvalues. Using Weyl's inequalities, we can deduce that $K + \alpha\Delta$ has at least one strictly negative eigenvalue.

Studying the reciprocal of Lemma 1 gives us a better insight on what is the optimal diagonal offset Δ to add to the original Ising matrix. We can consider Δ as a supplementary degree of freedom in our algorithm.

Tuning the search dimensionality with eigenvalue dropout

Tuning the eigenvalue dropout gives us a way of tuning the search dimensionality. If $K = UDU^\dagger$ where U is real unitary and $D = \text{Diag}(\lambda_1, \dots, \lambda_N)$, we can parametrize the phase space in terms of the eigenvectors of K , which we denote as e^j (associated to eigenvalue λ_j):

$$\vec{\sigma} = \sum_j \mu_j e^j, \quad (56)$$

$$H(\vec{\sigma}) = -\frac{1}{2} \sigma^T K \sigma \quad (57)$$

$$= -\frac{1}{2} \sum_j \mu_j^2 \lambda_j. \quad (58)$$

We can rephrase the optimization problem as follows:

$$\text{Find } \begin{cases} \underset{\mu}{\text{argmin}} -\frac{1}{2} \sum_j \mu_j^2 \lambda_j \\ \sum_j \mu_j e_i^j = \pm 1 \end{cases} \quad (59)$$

As only the positive eigenvalues can decrease the energy, we would like to conclude that only the eigenvectors associated with positive eigenvalues will contribute to minimizing the energy. However, one should make sure that the hard constraint in Eq. (59) remains satisfied. We can only conclude on the heuristic result that the ground state of the optimization probably will prefer having components in eigenspaces associated with positive eigenvalues. As reducing the eigenvalue dropout level results in dropping out the negative eigenvalues, this explains why we usually observe a better performance for a certain level around $\alpha = 0$.

Modified algorithm with tunable offset

We define Δ^0 as the minimum offset to verify the assumption of Lemma 1 and make sure the modified Ising matrix is symmetric positive semi-definite. Δ^0 is defined as:

$$\begin{aligned}\Delta_{ii}^0 &= \sum_{j \neq i} |A_{ij}|, \\ \Delta_{ij}^0 &= 0 \text{ for } i \neq j.\end{aligned}$$

We define the modified algorithm as follows:

- ▷ $K = \tilde{K} + \alpha\Delta^0$ where $\alpha \in [0, 1]$
- ▷ $J = \text{Re}\sqrt{K}$, defined as follows: as K is symmetric, there is a unitary (and real) U such that $K = U\text{Diag}(\lambda_1, \dots, \lambda_N)U^\dagger$. We define $\sqrt{\lambda_i} = i\sqrt{|\lambda_i|}$ if $\lambda_i < 0$. Then

$$J = U \left(\text{Diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_N}) \right) U^\dagger.$$

This modified algorithm only corresponds to a particular parametrization of the weight matrix and can thus be implemented with the PRIS. The writing of this problem as a function of the coupling matrix eigenvalues (see above) proves that, as long as $\alpha \geq 0$ in the modified algorithm, the ground state of the Ising problem will likely remain unchanged (see Figure 4).

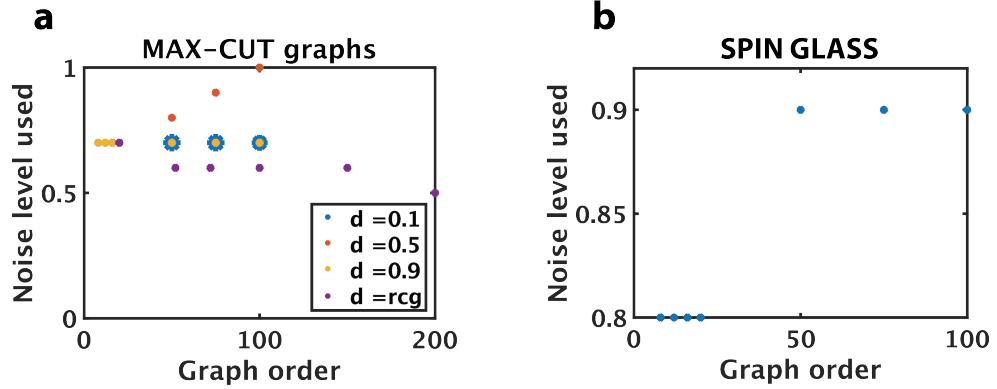


FIG. 2. Optimized noise level used in Figure 2 (main text). **a:** for MAX-CUT graphs (Figure 2c in the main text). **b:** for spin glasses (Figure 2b in the main text).

IV – NUMERICAL SIMULATIONS

Evaluating performance on finding ground state

To evaluate the performance of the algorithm on several Ising problems, we simulate the execution of an ideal photonic system, performing computations without static error. The noise is artificially added after the matrix multiplication unit and follows a Gaussian distribution, as discussed above. This results in an algorithm similar to the one described in the section II of this work.

For each randomly generated graph, we first compute its ground state with the online platform BiqMac⁵⁹. We then make 100 measurements of the number of steps required (with a random initial state) to get to this ground state. From these 100 runs, we define the estimate of finding the ground state of the problem with q percent probability $N_{\text{iter},q\%}$ as the q -th 100-quantile.

In the main text, we present the scaling performance of the PRIS as a function of the graph order. For each graph order and density, we generate 10 random samples with these properties. We then optimize the noise level (minimizing $N_{\text{iter},99\%}$) on a random sample graph and generate a total of 10 samples for each pair of graph order/density. The optimal value of ϕ is shown in Figure 2.

Also in the main text, we study the influence of eigenvalue dropout and of the noise level on the PRIS performance. We show that the optimal level of eigenvalue dropout is usually $\alpha < 1$, and usually around $\alpha = 0$. In some cases, it can even be $\alpha < 0$ as we show in Figure 3 where the optimal $(\alpha, \phi) = (-0.15, 0.55)$ for a random cubic graph with $N = 52$. In addition to Figure 3(f-h) from the main text showing the influence of eigenvalue dropout on a random spin glass, the influence of dropout on a random cubic graph is shown in Figure 4. Similar observations can be made, but random cubic graphs, which show highly degenerated hamiltonian landscapes, are more robust to eigenvalue dropout. Even with $\alpha = -0.8$, in the case shown in Figure 4(f) the ground state remains unaffected.

Evaluating sampling performance and critical parameters on two-dimensional ferromagnetic and infinite-range Ising models

In the main text, we evaluate the performance of the PRIS on sampling the Gibbs distribution of analytically solvable Ising models and estimating their critical exponents. We chose the following two problems:

- ▷ Two-dimensional ferromagnetic Ising model on a square lattice with periodic boundary conditions. This model was first analytically solved by Onsager²⁴. Its energy is defined as

$$H = -\frac{1}{2} \sum_{\langle ij \rangle} \sigma_i \sigma_j, \quad (60)$$

where $\langle ij \rangle$ refers to nearest neighbors (i, j) under periodic boundary conditions (see Figure 5).

- ▷ Infinite-range Ising model, where each node is connected with a positive coupling of $1/N$ to all its neighbors, including

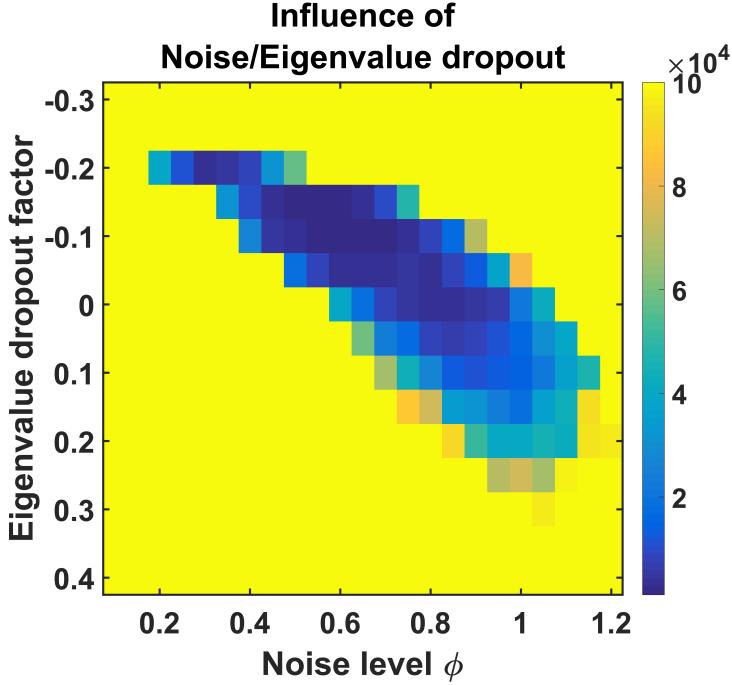


FIG. 3. Conjugated influence of eigenvalue dropout and noise levels for finding the ground state.

itself. This model can be analytically solved by mean field theory⁶¹. Its energy is defined as

$$H = -\frac{1}{2N} \sum_{ij} \sigma_i \sigma_j = -\frac{1}{2N} (\sum_i \sigma_i)^2. \quad (61)$$

In order to measure the critical exponents, we need to make sure the free energy at all temperatures is equivalent to that of the corresponding Ising model *at all temperatures*. It has been shown that the free energy of a Little network with coupling matrix K equals that of an Ising model defined by the same coupling matrix K , when the coupling weights are configured to learn a set of stable configurations (associative memory)²². However, this network has been discarded because of the possible existence of loops between some of its degenerate states, which can result in odd dynamics of the system, if no additional precautions are taken (see, for instance, Figure 7, where this simple algorithm actually converges to the *maximal* energy). We observe that adding a diagonal offset to the coupling, as we do in the square-root version of this algorithm, suppresses these odd dynamics (Figure 7(b)).

The algorithm described earlier to find the ground state of Ising problems could also be used to measure critical exponents, as can be seen in Figure 6. However, there are some complications that arise:

- ▷ Taking the square root of the coupling matrix prevents the use of symmetry and sparsity to reduce the algorithm complexity. Thus, for large graphs, the time needed to make a single matrix multiplication becomes quite large on a CPU.
- ▷ Taking the square root also modifies the coupling amplitude. It is now unclear how the temperature of the system should be defined (which is a problem if we want to estimate the critical temperature with the PRIS). From the large-noise expansion of the Hamiltonian, we should define the effective temperature as $T = 2k^2\phi^2$. However, we observe this definition does not match the theoretical value of the critical temperature of the 2D Ferromagnetic Ising model.

For both Ising problems studied, we perform the following analysis:

- ▷ We first estimate the Binder cumulant U_4 (see definition in the main text) as a function of the system temperature, for various graph sizes $N = L^2$. The cumulants U_4 plotted for different L intersect at the critical temperature¹ (see Figure 4(a) in the main text).

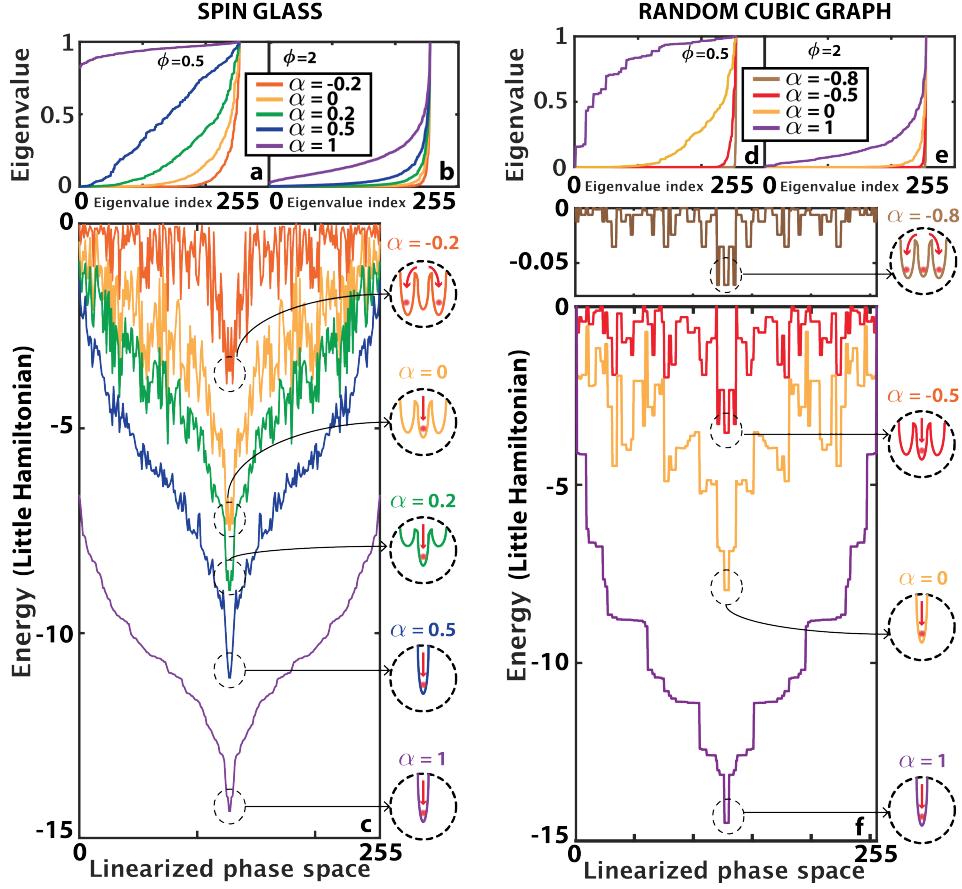


FIG. 4. **Influence of dropout** (a-c) for a random spin glass and (d-f) for a random cubic graph with $N = 8$. (a, b, d, e): Eigenvalue distribution for various noise levels. (c, f): Energy landscape (Hamiltonian) plotted over linearized phase space for various eigenvalue dropout levels α .

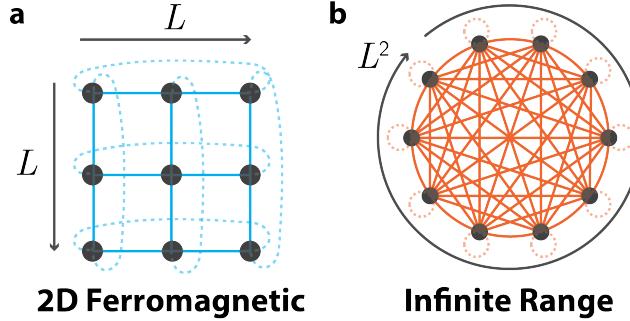


FIG. 5. **Examples of analytically solvable Ising models studied in our benchmark.** (a) Two-dimensional ferromagnetic Ising model and (b) infinite-range (mean-field) Ising model.

▷ We then estimate the dependence on the linear dimension L of various observables at the critical temperature and deduce the corresponding critical exponent. This is enabled by the scaling law of observables at the critical temperature¹:

$$m \sim L^{-\beta_C/\nu_C}, \quad (62)$$

$$\chi \sim L^{\gamma_C/\nu_C}, \quad (63)$$

$$\tau_{\text{auto}}^E \sim L^{z_C^E}, \quad (64)$$

$$\tau_{\text{auto}}^m \sim L^{z_C^m}, \quad (65)$$

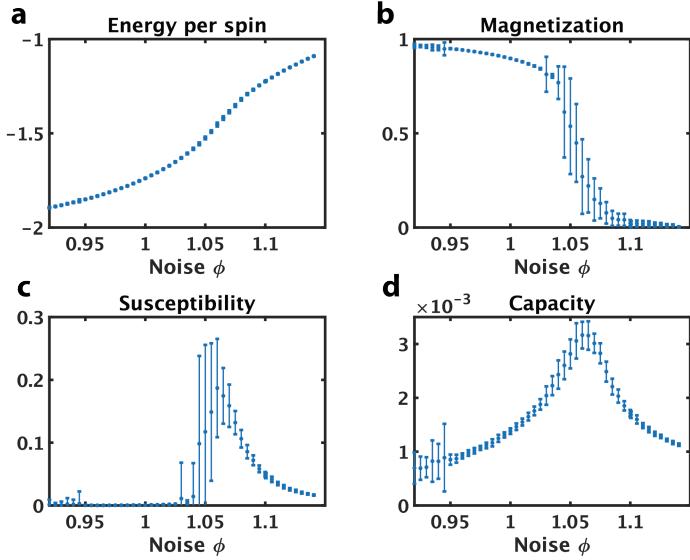


FIG. 6. Physical observables of the two-dimensional Ising model obtained with $L = 36$ square-rooted PRIS.

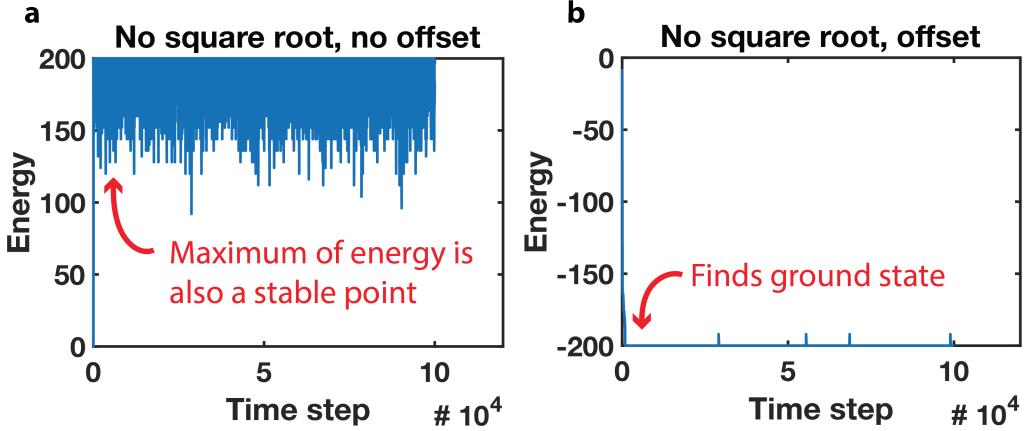


FIG. 7. Comparison of different versions of the algorithms on the 2D Ferromagnetic Ising Model. (a) With no offset, the no square root algorithm (defined by the following regime of operation of the PRIS: $C = K$) results in convergence to the state with maximal energy, for some runs (with random initial conditions). (b) Adding an offset to the same regime of operation cancels this behavior and the algorithm always converges to the ground state after some time.

where m , χ , τ_{auto}^E , and τ_{auto}^m are respectively the magnetization, the magnetic susceptibility, the energy autocorrelation time and the magnetization autocorrelation time¹. To estimate the critical exponents, we run the algorithm 120 times for 10^5 iterations, with random initial conditions, at the critical temperature for each L and fit these observables with a power law in L .

The estimates of all observables are obtained by taking every τ_{auto}^E generated samples, where τ_{auto}^E is the energy autocorrelation time, and dropping the first 10% of the data (arbitrary burn-in or equilibrium time). Our findings are summarized in Table II, Figures 8 and 9, and are benchmarked versus the Metropolis-Hastings algorithm, which is summarized in Refs.^{1,56,57}.

Algorithm	β_C	R^2	γ_C	R^2	z_C^E	R^2	z_C^m	R^2
2D ferromagnetic Ising model								
MH	0.1257	0.992	1.735	0.998	1.393	0.993	2.068	0.998
PRIS	0.1194	0.972	1.867	0.990	1.860	0.977	2.023	0.994
Infinite-range Ising model								
MH	0.5245	0.997	1.011	0.999	0.920	0.999	0.924	0.999
PRIS	0.5650	0.999	0.922	0.999	0.914	0.999	0.886	0.999

TABLE II. **Summary of critical exponents measured with the PRIS and MH.** R^2 is the coefficient of determination of each power law fitting.

V – OTHER CONSIDERATIONS

Mapping to the MAX-CUT problem

The MAX-CUT problem of a weighted undirected graph can be phrased in terms of its adjacency matrix A :

$$\text{Find } \underset{\sigma}{\operatorname{argmax}} \frac{1}{4} \sum_{ij} A_{ij} (1 - \sigma_i \sigma_j) := \underset{\sigma}{\operatorname{argmax}} C_A(\sigma), \quad (66)$$

where $\sigma \in \{-1, 1\}^N$ is a spin vector. The value of this MAX-CUT is called C_{\max} . More intuitively, this problem can be interpreted as finding a subset of vertices of the graph, such that the number of edges connecting this subset to its complementary is maximized. The vertices of this subset (resp. of its complementary) will have spin up (resp. spin down).

We can map the general Ising problem (spin glass) (Equation (8)) to any *weighted* MAX-CUT problem. This mapping is useful because the publicly-available solver we use to find the Ising ground state works in terms of the MAX-CUT problem⁵⁹. The energy of the Ising ground state and the MAX-CUT are related as follows:

$$H_{\min} = -\frac{1}{2} \sum_{ij} K_{ij} - 2C_{\max}, \quad (67)$$

where $K = -A$. From this linear relation, we also deduce that the solution of the weighted MAX-CUT problem and of the general Ising model (spin glass) are the same:

$$\underset{\sigma}{\operatorname{argmax}} C_A(\sigma) = \underset{\sigma}{\operatorname{argmin}} H_K(\sigma). \quad (68)$$

* Corresponding author e-mail: chrc@mit.edu

-
- [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 Berlin Heidelberg, 2013.
- [3] M. Kardar, G. Parisi, and Y.-C. Zhang, “Dynamic Scaling of Growing Interfaces,” *Physical Review Letters*, vol. 56, pp. 889–892, mar 1986.
- [4] M. B. Isichenko, “Percolation, statistical topography, and transport in random media,” *Reviews of Modern Physics*, vol. 64, pp. 961–1043, oct 1992.
- [5] 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 Biophysica Acta (BBA) - Biomembranes*, vol. 1788, pp. 53–63, jan 2009.
- [6] R. Albert and A.-L. Barabási, “Statistical mechanics of complex networks,” *Reviews of Modern Physics*, vol. 74, pp. 47–97, jan 2002.
- [7] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing.,” *Science (New York, N.Y.)*, vol. 220, pp. 671–80, may 1983.
- [8] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, “Coherent Ising machine based on degenerate optical parametric oscillators,” *Physical Review A*, vol. 88, p. 063853, dec 2013.
- [9] 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 (New York, N.Y.)*, vol. 354, pp. 614–617, nov 2016.
- [10] 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: Science & Applications*, vol. 3, pp. e147–e147, feb 2014.
- [11] 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,” *Optics Express*, vol. 26, p. 702, jan 2018.
- [12] W. M. Macready, A. G. Siapas, and S. A. Kauffman, “Criticality and Parallelism in Combinatorial Optimization,” *Science*, vol. 271, pp. 56–59, jan 1996.
- [13] A. Silva, F. Monticone, G. Castaldi, V. Galdi, A. Alù, and N. Engheta, “Performing mathematical operations with metamaterials,” *Science*, vol. 343, no. 6167, pp. 160–163, 2014.
- [14] A. F. Koenderink, A. Alù, and A. Polman, “Nanophotonics: Shrinking light-based technology,” may 2015.
- [15] 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,” *Nature Photonics*, vol. 11, pp. 441–446, jun 2017.
- [16] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning.” *Nature*, vol. 521, pp. 436–444, may 2015.
- [17] W. A. Little, “The existence of persistent states in the brain,” *Mathematical Biosciences*, vol. 19, no. 1-2, pp. 101–120, 1974.
- [18] J. J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities.,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, pp. 2554–8, apr 1982.
- [19] J. J. Hopfield and D. W. Tank, “Neural computation of decisions in optimization problems,” *Biological Cybernetics*, vol. 52, no. 3, pp. 141–152.
- [20] E. Ising, “Beitrag zur Theorie des Ferromagnetismus,” *Z. Phys.*, 1925.
- [21] M. Mézard and A. Montanari, *Information, Physics, and Computation*. 2009.
- [22] D. J. Amit, H. Gutfreund, and H. Sompolinsky, “Spin-glass models of neural networks,” *Physical Review A*, vol. 32, pp. 1007–1018, aug 1985.
- [23] A. Pelissetto and E. Vicari, “Critical phenomena and renormalization-group theory,” *Physics Reports*, vol. 368, pp. 549–727, oct 2002.
- [24] L. Onsager, “Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition,” *Physical Review*, vol. 65, pp. 117–149, feb 1944.
- [25] N. V. Brilliantov, “Effective magnetic Hamiltonian and Ginzburg criterion for fluids,” *Physical Review E*, vol. 58, pp. 2628–2631, aug 1998.
- [26] D. J. Amit, *Modeling brain function : the world of attractor neural networks*. Cambridge University Press, 1989.
- [27] M. A. Halasz, A. D. Jackson, R. E. Shrock, M. A. Stephanov, and J. J. M. Verbaarschot, “Phase diagram of QCD,” *Physical Review D*, vol. 58, p. 096007, sep 1998.
- [28] F. Barahona, “On the computational complexity of Ising spin glass models,” *Journal of Physics A: Mathematical and General*, vol. 15, pp. 3241–3253, oct 1982.
- [29] J. Bruck and J. W. Goodman, “On the power of neural networks for solving hard problems,” *Journal of Complexity*, vol. 6, pp. 129–135, jun 1990.
- [30] N. H. Farhat, D. Psaltis, A. Prata, and E. Paek, “Optical implementation of the Hopfield model,” *Applied Optics*, vol. 24, p. 1469, may 1985.
- [31] 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*, vol. 349, pp. 711–716, aug 2015.
- [32] M. Reck, A. Zeilinger, H. J. Bernstein, and P. Bertani, “Experimental realization of any discrete unitary operator,” *Physical Review Letters*, vol. 73, pp. 58–61, jul 1994.
- [33] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, “Optimal design for universal multiport interferometers,” *Optica*, vol. 3, p. 1460, dec 2016.
- [34] X. Lin, Y. Rivenson, N. T. Yardimci, M. Veli, M. Jarrahi, and A. Ozcan, “All-Optical Machine Learning Using Diffractive Deep Neural

- Networks," apr 2018.
- [35] M. Gruber, J. Jahns, and S. Sinzinger, "Planar-integrated optical vector-matrix multiplier," *Applied Optics*, vol. 39, p. 5367, oct 2000.
- [36] A. N. Tait, M. A. Nahmias, Y. Tian, B. J. Shastri, and P. R. Prucnal, "Photonic Neuromorphic Signal Processing and Computing," pp. 183–222, Springer, Berlin, Heidelberg, 2014.
- [37] A. N. Tait, M. A. Nahmias, B. J. Shastri, and P. R. Prucnal, "Broadcast and weight: An integrated network for scalable photonic spike processing," *Journal of Lightwave Technology*, vol. 32, no. 21, pp. 3427–3439, 2014.
- [38] K. Vandoorne, P. Mechet, T. Van Vaerenbergh, M. Fiers, G. Morthier, D. Verstraeten, B. Schrauwen, J. Dambre, and P. Bienstman, "Experimental demonstration of reservoir computing on a silicon photonics chip," *Nature Communications*, vol. 5, p. 3541, dec 2014.
- [39] A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Dremeau, S. Gigan, and F. Krzakala, "Random projections through multiple optical scattering: Approximating Kernels at the speed of light," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6215–6219, IEEE, mar 2016.
- [40] 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 Journal of Selected Topics in Quantum Electronics*, vol. 20, pp. 43–48, jan 2014.
- [41] 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 Research*, vol. 4, pp. 297–307, mar 2011.
- [42] A. C. Selden, "Pulse transmission through a saturable absorber," *British Journal of Applied Physics*, vol. 18, pp. 743–748, jun 1967.
- [43] M. Soljačić, M. Ibanescu, S. G. Johnson, Y. Fink, and J. D. Joannopoulos, "Optimal bistable switching in nonlinear photonic crystals," *Physical Review E*, vol. 66, p. 055601, nov 2002.
- [44] R. W. Schirmer and A. L. Gaeta, "Nonlinear mirror based on two-photon absorption," *Journal of the Optical Society of America B*, vol. 14, p. 2865, nov 1997.
- [45] P. Horowitz and H. Winfield, "The Art of Electronics," *American Journal of Physics*, 1990.
- [46] B. Boser, E. Sackinger, J. Bromley, Y. Le Cun, and L. Jackel, "An analog neural network processor with programmable topology," *IEEE Journal of Solid-State Circuits*, vol. 26, no. 12, pp. 2017–2025, 1991.
- [47] J. Misra and I. Saha, "Artificial neural networks in hardware: A survey of two decades of progress," *Neurocomputing*, vol. 74, pp. 239–255, dec 2010.
- [48] D. Vrataric, V. Ceperic, and A. Baric, "Area-efficient differential Gaussian circuit for dedicated hardware implementations of Gaussian function based machine learning algorithms," *Neurocomputing*, vol. 118, pp. 329–333, oct 2013.
- [49] P. Peretto, "Collective properties of neural networks: A statistical physics approach," *Biological Cybernetics*, vol. 50, pp. 51–62, feb 1984.
- [50] M. Lipson, "Guiding, Modulating, and Emitting Light on Silicon-Challenges and Opportunities," *Journal of Lightwave Technology, Vol. 23, Issue 12, pp. 4222-*, vol. 23, p. 4222, dec 2005.
- [51] 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," *Optics Express*, vol. 22, p. 10487, may 2014.
- [52] R. Hamerly, T. Inagaki, P. L. McMahon, D. Venturelli, A. Marandi, T. Onodera, E. Ng, C. Langrock, K. Inaba, T. Honjo, K. Enbusu, T. Umeki, R. Kasahara, S. Utsunomiya, S. Kako, K.-i. Kawarabayashi, R. L. Byer, M. M. Fejer, H. Mabuchi, E. Rieffel, H. Takesue, and Y. Yamamoto, "Scaling advantages of all-to-all connectivity in physical annealers: the Coherent Ising Machine vs. D-Wave 2000Q," *arXiv preprints*, may 2018.
- [53] V. R. Almeida, C. A. Barrios, R. R. Panepucci, and M. Lipson, "All-optical control of light on a silicon chip," *Nature*, vol. 431, pp. 1081–1084, oct 2004.
- [54] C. T. Phare, Y. H. Daniel Lee, J. Cardenas, and M. Lipson, "Graphene electro-optic modulator with 30 GHz bandwidth," *Nature Photonics*, 2015.
- [55] C. Haffner, D. Chelladurai, Y. Fedoryshyn, A. Josten, B. Baeuerle, W. Heni, T. Watanabe, T. Cui, B. Cheng, S. Saha, D. L. Elder, L. R. Dalton, A. Boltasseva, V. M. Shalaev, N. Kinsey, and J. Leuthold, "Low-loss plasmon-assisted electro-optic modulator," *Nature*, 2018.
- [56] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The Journal of Chemical Physics*, 1953.
- [57] W. K. Hastings, "Monte carlo sampling methods using Markov chains and their applications," *Biometrika*, 1970.
- [58] D. C. Knill and A. Pouget, "The Bayesian brain: the role of uncertainty in neural coding and computation," *Trends in Neurosciences*, vol. 27, pp. 712–719, dec 2004.
- [59] F. Rendl, G. Rinaldi, and A. Wiegele, "Solving Max-Cut to optimality by intersecting semidefinite and polyhedral relaxations," *Mathematical Programming*, vol. 121, pp. 307–335, feb 2010.
- [60] "In this part, we operate the PRIS in the regime where the linear photonic matrix is equal to the Ising coupling matrix. This allows us to speedup the computation on our CPU by leveraging symmetry and sparsity. We show the other regime of o,"
- [61] S. Friedli and Y. Velenik, *Statistical Mechanics of Lattice Systems*. Cambridge University Press, nov 2017.
- [62] S. Kirkpatrick and B. Selman, "Critical behavior in the satisfiability of random boolean expressions," *Science (New York, N.Y.)*, 1994.
- [63] D. W. Heermann and A. N. Burkitt, "System size dependence of the autocorrelation time for the Swendsen-Wang Ising model," *Physica A: Statistical Mechanics and its Applications*, vol. 162, pp. 210–214, jan 1990.
- [64] A. C. Selden, "Pulse transmission through a saturable absorber," *British Journal of Applied Physics*, vol. 18, pp. 743–748, jun 1967.
- [65] A. Saade, F. Caltagirone, I. Carron, L. Daudet, A. Dremeau, S. Gigan, and F. Krzakala, "Random projections through multiple optical scattering: Approximating Kernels at the speed of light," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6215–6219, IEEE, mar 2016.
- [66] A. N. Tait, M. A. Nahmias, Y. Tian, B. J. Shastri, and P. R. Prucnal, "Photonic Neuromorphic Signal Processing and Computing," pp. 183–222, Springer, Berlin, Heidelberg, 2014.
- [67] M. Gruber, J. Jahns, and S. Sinzinger, "Planar-integrated optical vector-matrix multiplier," *Applied Optics*, vol. 39, p. 5367, oct 2000.
- [68] A. Azzalini and A. D. VALLE, "The multivariate skew-normal distribution," *Biometrika*, vol. 83, pp. 715–726, dec 1996.

- [69] C.-K. Looi, “Neural network methods in combinatorial optimization,” *Computers & Operations Research*, vol. 19, pp. 191–208, apr 1992.
- [70] R. M. Karp, “Reducibility among Combinatorial Problems,” in *Complexity of Computer Computations*, pp. 85–103, Boston, MA: Springer US, 1972.

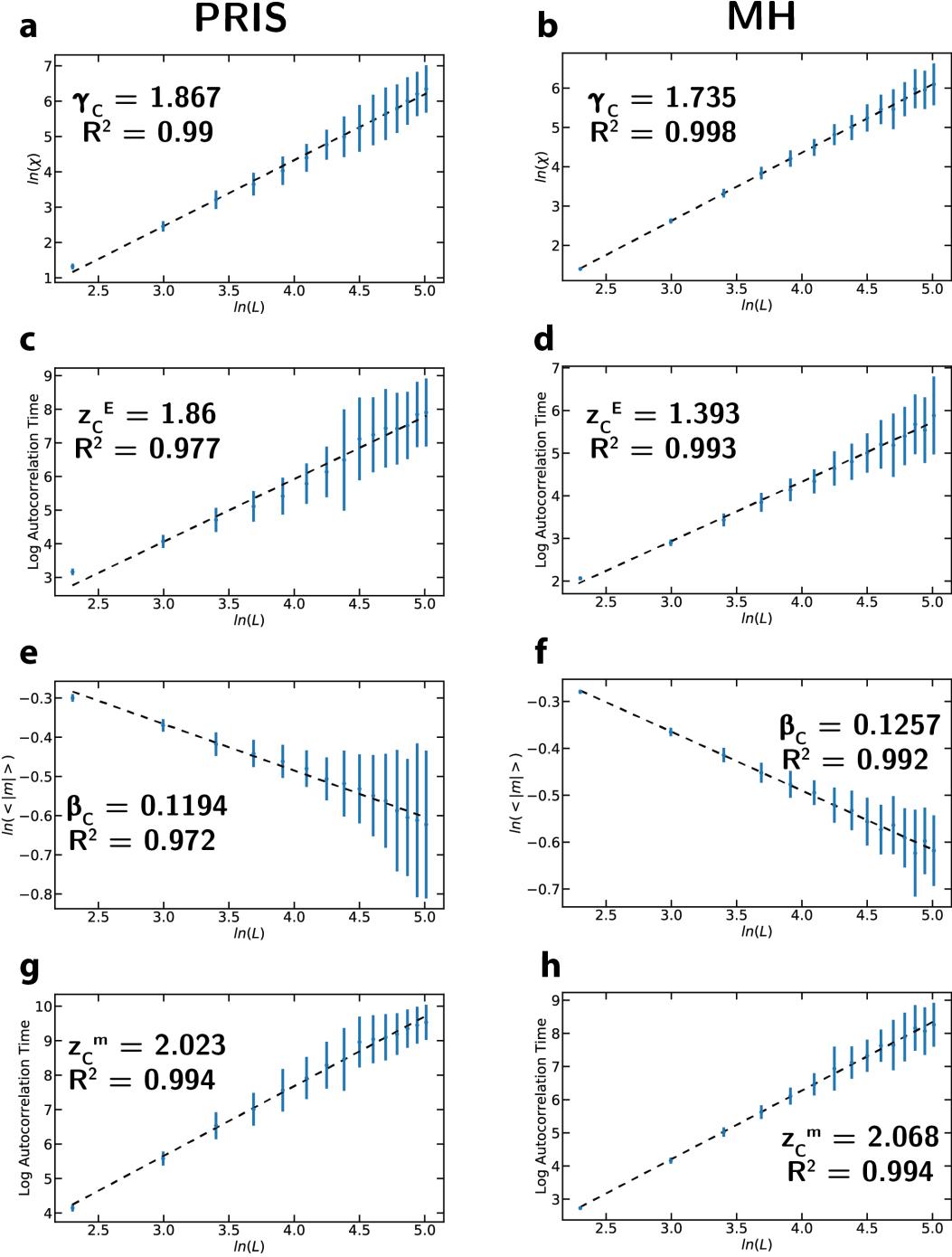


FIG. 8. Probing the critical exponents of the 2D Ferromagnetic Ising model. Fits are shown with the resulting critical exponent for the PRIS (**a, c, e, g**) and the MH (**b, d, f, h**) algorithms for the susceptibility (**a-b**), energy autocorrelation time (**c-d**), magnetization (**e-f**), and magnetization autocorrelation time (**g-h**).

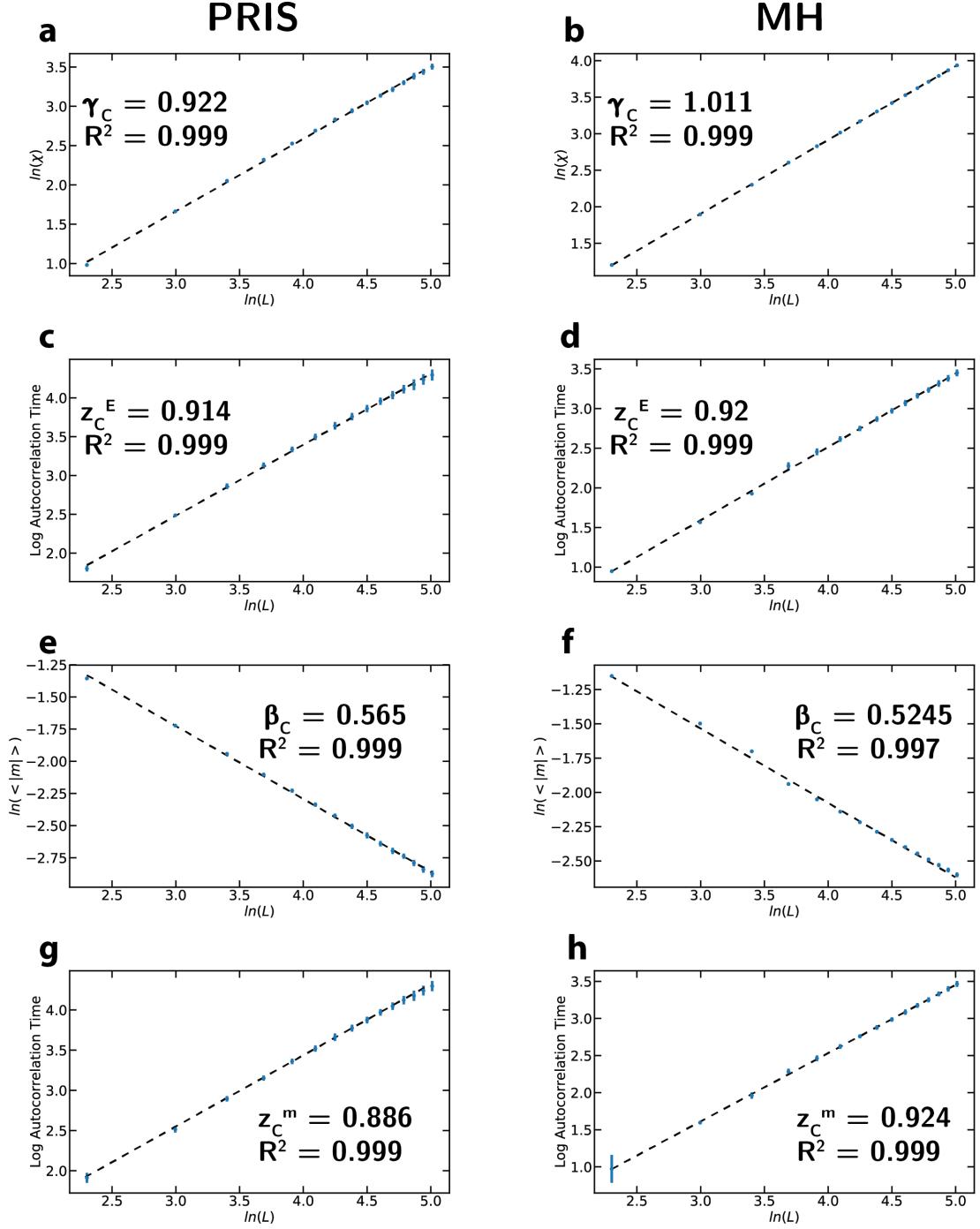


FIG. 9. **Probing the critical exponents of the infinite-range Ising model.** Fits are shown with the resulting critical exponent for the PRIS (a, c, e, g) and the MH (b, d, f, h) algorithms for the susceptibility (a-b), energy autocorrelation time (c-d), magnetization (e-f), and magnetization autocorrelation time (g-h).