




Scaling advantage of chaotic amplitude control for high-performance combinatorial optimization

Timothée Leleu ^{1,2}✉, Farad Khoystatee^{1,3,4,5}, Timothée Levi ^{1,5,6}, Ryan Hamerly ^{7,8}, Takashi Kohno^{1,2,6} & Kazuyuki Aihara^{1,2}

The development of physical simulators, called Ising machines, that sample from low energy states of the Ising Hamiltonian has the potential to transform our ability to understand and control complex systems. However, most of the physical implementations of such machines have been based on a similar concept that is closely related to relaxational dynamics such as in simulated, mean-field, chaotic, and quantum annealing. Here we show that dynamics that includes a nonrelaxational component and is associated with a finite positive Gibbs entropy production rate can accelerate the sampling of low energy states compared to that of conventional methods. By implementing such dynamics on field programmable gate array, we show that the addition of nonrelaxational dynamics that we propose, called chaotic amplitude control, exhibits exponents of the scaling with problem size of the time to find optimal solutions and its variance that are smaller than those of relaxational schemes recently implemented on Ising machines.

¹Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro City, Tokyo 153-8505, Japan. ²International Research Center for Neurointelligence, University of Tokyo, 7-3 Hongō, Bunkyo City, Tokyo 113-0033, Japan. ³Theoretical Quantum Physics Laboratory, RIKEN Cluster for Pioneering Research, Saitama, 2-1, Hirosawa, Wako, Saitama 351-0198, Japan. ⁴Physics Department, The University of Michigan, 500 South State Street, Ann Arbor, MI 48109, USA. ⁵IMS laboratory, CNRS UMR 5218, University of Bordeaux, 33405 Talence, France. ⁶LIMMS/CNRS-IIS, University of Tokyo, 4-6-1 Komaba, Meguro City, Tokyo 153-8505, Japan. ⁷Research Laboratory of Electronics, MIT, 50 Vassar Street, Cambridge, MA 02139, USA. ⁸PHI (Physics & Informatics) Laboratories, NTT Research Inc., 940 Stewart Drive, Sunnyvale, CA 94085, USA. ✉email: timothee.leleu@ircn.jp

Many complex systems such as spin glasses, interacting proteins, large-scale hardware, and financial portfolios, can be described as ensembles of disordered elements that have competing frustrated interactions¹ and rugged energy landscapes. There has been a growing interest in using physical simulators called Ising machines in order to reduce time and resources needed to identify configurations that minimize their total interaction energy, notably that of the Ising Hamiltonians \mathcal{H} with $\mathcal{H}(\sigma) = -\frac{1}{2}\sum_{ij}\omega_{ij}\sigma_i\sigma_j$ (with ω_{ij} the symmetric Ising couplings, i.e., $\omega_{ij} = \omega_{ji}$, and $\sigma_i = \pm 1$) that is related to many non-deterministic polynomial-time hard (NP-hard) combinatorial optimization problems and various real-world applications² (see ref. ³ for a review). Recently proposed implementations include memresistor networks⁴, micro- or nano-electromechanical systems⁵, micro-magnets^{6,7}, coherent optical systems⁸, hybrid opto-electronic hardware^{9–11}, integrated photonics^{12–14}, flux qubits¹⁵, and Bose-Einstein condensates¹⁶. In principle, these physical systems often possess unique properties, such as coherent superposition in flux qubits¹⁷ and energy efficiency of memresistors^{4,18}, which could lead to a distinctive advantage compared to conventional computers (see Fig. 1(a) and (b)) for the sampling of low energy states. In practice, the difficulty in constructing connections between constituting elements of the hardware is often the main limiting factor to scalability and performance for these systems^{15,19}. Moreover, these devices often implement schemes that are directly related to the concept of annealing (either simulated^{20,21}, mean-field^{22,23}, chaotic^{18,24}, and quantum^{17,25}) in which the escape from the numerous local minima²⁶ and saddle points²⁷ of the free energy function can only

be achieved under very slow modulation of the control parameter (see Fig. 1(c)). These methods are dependent on non-equilibrium dynamics called aging that, according to recent numerical studies²⁸, is strongly non-ergodic and seems to explore only a confined subspace determined by the initial condition rather than wander in the entire configurational space²⁹ for mean-field spin glass models. In general, such systems find the solutions of minimal energy only after many repetitions of the relaxation process.

Alternative dynamics that is not based on the concepts of annealing and relaxation may perform better for solving hard combinatorial optimization problems^{30–32}. Various kinds of dynamics have been proposed^{3,33–36}, notably chaotic dynamics^{18,37–40}, but have either not been implemented onto specialized hardware^{37,41} or use chaotic dynamics merely as a replacement to random fluctuations^{18,38}. It was recently shown that the control of amplitude in mean-field dynamics can improve the performance of Ising machines by introducing error-correction terms (see Fig. 1(d)), effectively doubling the dimensionality of the system, whose role is to correct the amplitude heterogeneity³⁰. Because of the similarity of such dynamics with that of a neural network, it can be implemented especially efficiently in electronic neuromorphic hardware where memory is distributed with the processing^{42–44}.

In this paper, we show that the addition of the nonrelaxational part to the relaxation process makes the dynamics able to escape at a much faster rate than relaxational ones from local minima and saddles of the corresponding energy function. The exponential scaling factor with respect to system size of the time needed to reach optimal configurations of spin glasses that can be

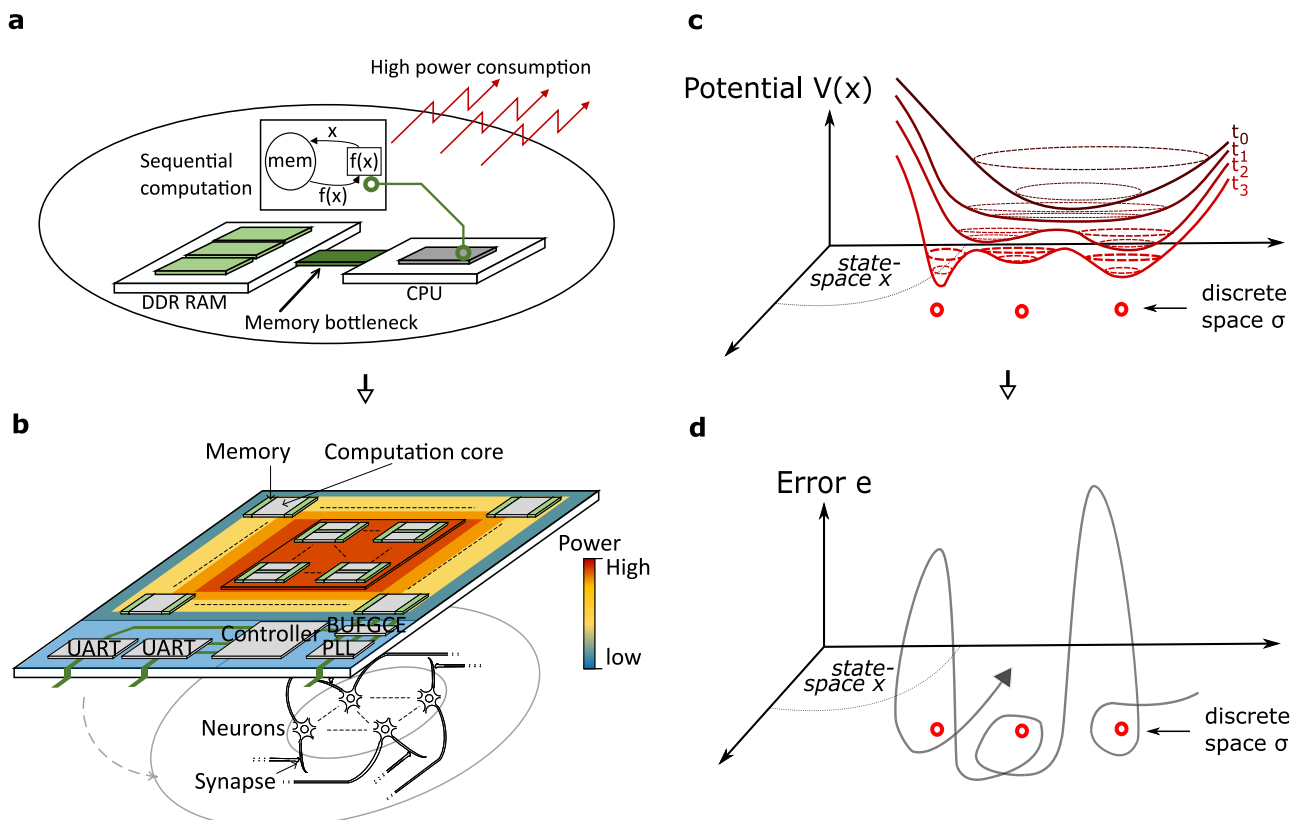


Fig. 1 Schematic representation of the proposed chip. Schematic representation of **a** conventional central processing unit (CPU) and separate memory with the von Neumann bottleneck problem and **b** the proposed neuromorphic chip for combinatorial optimization. Universal Asynchronous Receiver/Transmitter (UART), phase-locked loops (PLL), and global clock buffer with clock enable (BUFGE) are used for input/output of data, clock management, and clock gating, respectively. Neurons and synapses are shown in the bottom to illustrate the analogy between the organization of the chip and biological neural networks. Schema of dynamics in analog state-space x of algorithms based on **c** annealing on a potential function shown at different times t_i and **d** a trajectory of proposed chaotic amplitude control scheme shown by the curved gray line with an arrow. Red circles show the discrete space σ .

found after a very long computation time using state-of-the-art heuristics, called time to solution, is smaller in the former case, which raises the question whether the nonrelaxational component makes the dynamics qualitatively different from the very slow relaxation observed in classic Monte-Carlo simulations of spin glasses. In order to extend numerical analysis to large problem sizes and limit finite-size effects, we implement a scheme that we name chaotic amplitude control (CAC) on a field programmable gate array (FPGA, see Fig. 1(b)) and show that the developed hardware can be faster for finding these optimal configurations in the limit of large problem sizes than many state-of-the-art algorithms and Ising machines for some reference benchmarks with enhanced energy efficiency.

Results

Mean-field dynamics. For the sake of simplicity, we consider the classical limit of Ising machines for which the state space is often naturally described by analog variables (i.e., real numbers) noted x_i in the following. The variables x_i represent measured physical quantities such as voltage⁴ or optical field amplitude^{8–13} and these systems can often be simplified to networks of interacting nonlinear elements whose time evolution can be written as follows:

$$\frac{dx_i}{dt} = f_i(x_i) + \beta_i(t) \sum_j \omega_{ij} g_j(x_j) + \sigma_0 \eta_i, \quad (1)$$

where f_i and g_i represent the nonlinear gain and interaction, respectively, and are assumed to be monotonic, odd, and invertible “sigmoidal” functions for the sake of simplicity; η_i , experimental white noise of standard deviation σ_0 with $\langle \eta_i(t) \eta_j(t') \rangle = \delta_{ij} \delta(t - t')$; and N , the number of spins. δ_{ij} and $\delta(t)$ are the Kronecker delta symbol and Dirac delta function, respectively. Ordinary differential equations similar to eq. (1) have been used in various computational models that are applied to NP-hard combinatorial optimization problems such as Hopfield-Tank neural networks⁴⁵, coherent Ising machines⁴⁶, and correspond to the “soft” spin description of frustrated spin systems⁴⁷. Moreover, the steady states of eq. (1) correspond to the solutions of the “naive” Thouless-Anderson-Palmer (nTAP) equations that arise from the mean-field description of Sherrington-Kirkpatrick spin glasses when the Onsager reaction term has been discarded⁴⁸. In the case of neural networks in particular, the variables x_i and constant parameters ω_{ij} correspond to firing rates of neurons and synaptic coupling weights, respectively.

It is well known that, when $\beta_i = \beta$ for all i and the noise is not taken into account ($\sigma_0 = 0$), the time evolution of this system is motion in the state space that seeks out minima of a potential function⁴⁹ (or Lyapunov function) V given as $V = \beta \mathcal{H}(\mathbf{y}) + \sum_i V_b(y_i)$, where V_b is a bistable potential with $V_b(y_i) = -\int_0^{y_i} f_i(g_i^{-1}(y)) dy$ and $\mathcal{H}(\mathbf{y}) = -\frac{1}{2} \sum_{ij} \omega_{ij} y_i y_j$ is the extension of the Ising Hamiltonian in the real space with $y_i = g_i(x_i)$ (see Supplementary Note 1.1). The bifurcation parameter β , which can be interpreted as the inverse temperature of the naive TAP equations⁴⁸, the steepness of the neuronal transfer function in Hopfield-Tank neural networks⁴⁵, or to the coupling strength in coherent Ising machines^{8–10}, is usually decreased gradually in order to improve the quality of solutions found. This procedure has been called mean-field annealing²³, and can be interpreted as a quasi-static deformation of the potential function V (see Fig. 1(c)). There is, however, no guarantee that a sufficiently slow deformation of the landscape V will ensure convergence to the lowest energy state contrary to the quantum adiabatic theorem⁵⁰ or the convergence theorem of simulated annealing⁵¹. At fixed β , global convergence to the minimum of

the potential V can be assured if σ_0 is gradually decreased with $\sigma_0(t)^2 \sim \frac{c}{\log(2+t)}$ and c sufficiently large⁵². The parameter σ_0^2 is analogous to the temperature in simulated annealing in this case. The global minimum of the potential V does not, however, generally correspond to that of the Ising Hamiltonians \mathcal{H} at finite β . Moreover, the statistical analysis of spin glasses suggests that the potential V is highly non-convex at low temperature and that simple gradient descent very unlikely reaches the global minimum of $\mathcal{H}(\boldsymbol{\sigma})$ because of the presence of exponentially numerous local minima²⁶ and saddle points²⁷ as the size of the system increases. The slow relaxation time of Monte-Carlo simulations of spin glasses, such as when using simulated annealing, might also be explained by similar trapping dynamics during the descent of the free energy landscape obtained from the TAP equations²⁷. In the following, we consider in particular the soft spin description obtained by taking $f_i(x_i) = (-1 + p)x_i - x_i^3$ and $y_i = g(x_i) = x_i$, where p is the gain parameter, which is the canonical model of the system described in eq. (1) at proximity of a pitchfork bifurcation with respect to the parameter p (see theory of weakly connected neural networks⁵³). In this case, the potential function V_b is given as $V_b(x_i) = (1 - p)\frac{x_i^2}{2} + \frac{x_i^4}{4}$ and eq. (1) can be written as $\frac{dx_i}{dt} = -\frac{\partial V}{\partial x_i}, \forall i$.

Chaotic amplitude control. In order to define deterministic dynamics that is inclined to visit spin configurations associated with lower Ising Hamiltonian without relying entirely on the descent of a potential function, we introduce error signals, noted $e_i \in \mathbb{R}$, that modulate the strength of coupling β_i to the i th nonlinear element such that $\beta_i(t)$ defined in eq. (1) is expressed as $\beta_i(t) = \beta e_i(t)$ with $\beta > 0$. The time evolution of the error signals e_i are given as follows³⁰:

$$\frac{de_i}{dt} = -\xi(g(x_i)^2 - a)e_i, \quad (2)$$

where a and ξ are the target amplitude and the rate of change of error variables, respectively, with $a > 0$ and $\xi > 0$. If the system settles to a steady state, the values $y_i^* = g(x_i^*)$ become exactly binary with $y_i^* = \pm\sqrt{a}$. When $p < 1$, the internal fields h_i at the steady state, defined as $h_i = \sum_j \omega_{ij} \sigma_j$ with $\sigma_j = y_j^*/|y_j^*|$, are such that $h_i \sigma_i > 0, \forall i$ ³⁰. Thus, each equilibrium point of the analog system corresponds to that of a zero-temperature local minimum of the binary spin system.

The dynamics described by the coupled equations (1) and (2) is not derived from a potential function because error signals e_i introduce asymmetric interactions between the x_i and the computational principle is not related to a gradient descent. Rather, the addition of error variables results in additional dimensions in the phase space via which the dynamics can escape local minima. The mechanism of this escape can be summarized as follows. It can be shown (see the Supplementary Note 1.2) that the dimension of the unstable manifold at equilibrium points corresponding to local minima $\boldsymbol{\sigma}$ of the Ising Hamiltonian depends on the number of eigenvalues $\mu(\boldsymbol{\sigma})$ with $\mu(\boldsymbol{\sigma}) > F(a)$ where $\mu(\boldsymbol{\sigma})$ are the eigenvalues of the matrix $\left\{ \frac{\omega_{ij}}{|h_i|} \right\}_{ij}$ (with internal field h_i) and F a function given as

$F(y) = \frac{\psi(y)}{\varphi(y)} y$ and $\psi(y) = \frac{f(g^{-1}(y))}{(g^{-1})'(y)}$. Thus, there exists a value of a such that all local minima (including the ground state) are unstable and for which the system exhibits chaotic dynamics that explores successively candidate boolean configurations. The energy is evaluated at each step and the best configuration visited is kept as the solution of a run. This chaotic search is particularly efficient for sampling configurations of the Ising Hamiltonian close to that of the ground state using a single run although the distribution of sampled states is not necessarily given by the Boltzmann distribution. Note

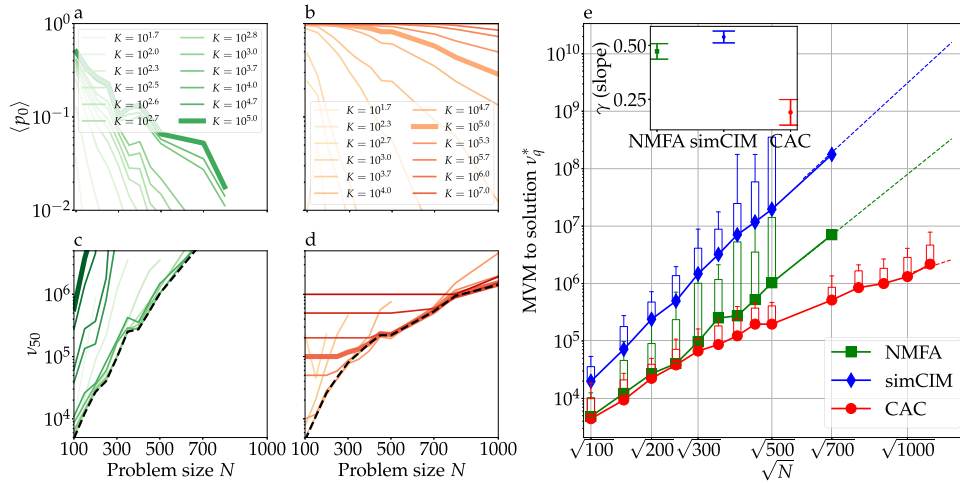


Fig. 2 Matrix-vector multiplication to solution for Sherrington-Kirkpatrick instances. **a, b** Average success probability $\langle p_0 \rangle$ of finding the solution state configuration of 100 Sherrington-Kirkpatrick spin glass instances and **c, d** 50th percentile (median) of the matrix-vector multiplication (MVM) to solution distribution ν_{50} vs. system size N . **a, c** noisy mean-field annealing (NMFA). **b, d** CAC. Thicker lines correspond to the duration of simulation $K = 10^{5.0}$. Dotted black lines show the minimum of the median MVM to solution ν_{50} with respect to the number of MVMs of a single run K . **e** MVM to solution $\nu_q^*(N)$ distribution. Lower, higher, and upper whisker of boxes show the 50th, 80th, and 90th percentiles of the distribution. The upper right inset shows the exponential scaling factor γ of the 50th percentile with $\nu_{50} \sim e^{\gamma\sqrt{N}}$ for CAC, NMFA, and simCIM.

that the use of chaotic dynamics for solving Ising problems has been discussed previously^{18,54}, notably in the context of neural networks, and it has been argued that chaotic fluctuations may possess better properties than Brownian noise for escaping from local minima traps. In the case of the proposed scheme, the chaotic dynamics is not merely used as a replacement for noise. Rather, the interaction between nonlinear gain and error-correction results in the destabilization of states associated with lower Ising Hamiltonian. Increasing the amplitude of additive noise σ_0 in eq. (1) does not, in fact, significantly increase the efficiency of the system and σ_0 is thus set to zero.

Ensuring that fixed points are locally unstable does not guarantee that the system does not relax to periodic and chaotic attractors. We have previously proposed that non-trivial attractors can also be destabilized by ensuring the positive rate of Gibbs entropy production using a modulation of the target amplitude³⁰. In this paper, we propose an alternative heuristic modulation of the target amplitude that is more suitable for a digital implementation on FPGA than the one proposed in³⁰ without significant decrease in performance for most problem instances (see Supplementary Note 1.3 for a comparison of the two schemes). Because the value of a for which all local minima is unstable is not known a priori, we propose instead to destabilize the local minima traps by dynamically modulating a depending on the visited configurations σ as follows:

$$a(t) = \alpha - \rho \tanh(\delta\Delta\mathcal{H}(t)), \quad (3)$$

where $\Delta\mathcal{H}(t) = \mathcal{H}_{\text{opt}} - \mathcal{H}(t)$; $\mathcal{H}(t)$, the Ising Hamiltonian of the configuration visited at time t ; and \mathcal{H}_{opt} , a given target energy. In practice, we set \mathcal{H}_{opt} to the lowest energy visited during the current run, i.e., $\mathcal{H}_{\text{opt}}(t) = \min_{t' \leq t} \mathcal{H}(t')$. The function \tanh is the tangent hyperbolic. ρ and δ are positive real constants. Lastly, the parameter ξ (see eq. (2)) is modulated as follows: $\frac{d\xi}{dt} = \Gamma$ when $t - t_r < \Delta t$, where t_r is the last time for which either the best known energy \mathcal{H}_{opt} was updated or ξ was reset. Otherwise, ξ is reset to 0 if $t - t_r \geq \Delta t$ and t_r is set to t . Numerical simulations shown in the following suggest that this modulation results in the destabilization of non-trivial attractors (periodic, chaotic, etc.) for typical problem instances.

Hardware independent time-to-solution scaling. In order to test if the nonrelaxational dynamics of chaotic amplitude control might be able to accelerate the search of mean-field dynamics for finding the ground state of typical frustrated systems, we look for the ground states of Sherrington-Kirkpatrick (SK) spin glass instances using the numerical simulation of eqs. (1) to (3) and compare time to solutions with those of two closely related relaxational schemes: noisy mean-field annealing (NMFA)²² and the simulation of the coherent Ising machine (simCIM, see Supplementary Note 4.1-2). Because finding the ground states of SK spin glasses is a NP-hard problem, it cannot be assured that the states found by heuristic solvers, even after a very long computation time, are the real ground states. In order to compare the performance of a set of heuristic solvers, we can define for each instance the “optimal” energy as the one equal to the lowest energy found repeatedly ($>10^6$ times) by these heuristics after a very long computation time. States that have this optimal energy are called solution states in the following. Because the arithmetic complexity of calculating one step of these three schemes is dominated by the matrix-vector multiplication (MVM), it is sufficient for the sake of comparison to count the number of MVM, noted ν , to find the solution state energy of a given instance with 99% success probability, with $\nu(K) = K \frac{\ln(1-0.99)}{\ln(1-p_0(K))}$ and $p_0(K)$ the probability of visiting a solution state configuration at least once after a number K of MVMs in a single run. In Fig. 2, NMFA (a) and the CAC (b) are compared using the averaged success probability $\langle p_0 \rangle$ of finding the solution state for 100 randomly generated SK spin glass instances per problem size N (see Supplementary Note 3 for details about the benchmark set). Note that the success probability of the mean-field annealing method does not seem to converge to 1 even for large annealing time (see Fig. 2(a)). Because the success probability of NMFA and simCIM remains low at larger problem size, its correct estimation requires simulating a larger number of runs, which we achieved by using GPU implementations of these methods. On the other hand, the average success probability $\langle p_0 \rangle$ of CAC is of order 1 when the maximal number of MVM is large enough, suggesting that the system rarely gets trapped in local minima of the Ising Hamiltonian or non-trivial attractors. In Fig. 2(c) and (d) are shown the q^{th} percentile (with $q = 50$, i.e., the median) of the

MVM to solution distribution, noted $\nu_q(K; N)$, for various duration of simulation K , where K is the number of MVMs of a single run. The minimum of these curves, noted $\nu_q^*(N)$ with $\nu_q^*(N) = \min_K \nu_q(K; N)$, represents the optimal scaling of MVM to solution vs. problem size N^{15} . Using the hypothesis of an exponential scaling with the square root of problem size N , CAC exhibits significantly smaller scaling exponent ($\gamma = 0.18 \pm 0.06$) than the NMFA ($\gamma = 0.47 \pm 0.04$) and simCIM ($\gamma = 0.54 \pm 0.03$, see inset in Fig. 2(e)). We have verified that this scaling advantage holds for various parameters of the mean-field annealing (see Supplementary Note 4.1). Note that a root-exponential scaling behavior of the median time to solution has been previously reported for SK spin glass problems^{15,55} and other NP-Hard problems⁵⁶. We consider both cases of a root-exponential and exponential scaling as possible hypotheses in the following. Note that a similar root-exponential scaling at finite size N is observed when we apply the CAC algorithm to solving problems from the recently proposed Wishart planted instances^{57,58} in the “easy” regime (see Supplementary Note 3.3).

Benchmark of the FPGA implementation. Although comparison of algebraic complexity indicates that CAC has a scaling advantage over mean-field annealing, it is in practice necessary to compare its physical implementation against other state-of-the-art methods because the performance of hardware depends on other factors such as memory access and information propagation delays. To this end, CAC is implemented into a FPGA because its similarity with neural networks makes it well-fitted for a design where memory is distributed with processing (see Supplementary Note 2 for the details of the FPGA implementation). The organization of the electronic circuit can be understood using the following analogy. Pairs of analog values x_i and e_i , which represent averaged activity of two types of neurons, are encoded within neighboring circuits. This microstructure is repeated N times on the whole surface of the chip, which resembles the columnar organization of the brain. The nonlinear processes $f_i(x_i)$, which model the local-population activation functions and are independent for $i \neq j$, are calculated in parallel. The coupling between elements i and $j \in \{1, \dots, N\}$ that is achieved by the dot product in eq. (1) is implemented by circuits that are at the periphery of the chip and are organized in a summation tree reminiscent of dendritic branching (see Fig. 1(b)). The power consumption of the developed hardware never exceeds 5W because of limitations of the development board that we have used.

First, we compare the FPGA implementation of CAC against state-of-the-art CPU algorithms: break-out local search⁵⁹ (BLS) that has been used to find many of the best known maximum-cuts (equivalently, Ising energies) from the GSET benchmark set (<https://web.stanford.edu/~yyye/yyye/Gset/>), a well-optimized single-core CPU implementation of parallel tempering (or random replica exchange Monte-Carlo Markov chain sampling)^{60,61} (PT, courtesy of S. Mandrà), simulated annealing (SA)⁶². Figure 3(a) shows that the CAC on FPGA has the smallest real time to solution τ_q^* against most other state-of-the-art algorithms despite just 5W power-consumption where $\tau_q^*(N)$ is the optimal q^{th} percentile of time to solution with 99% success probability and is given as $\tau_q^*(N) = \min_T \tau_q(T; N)$ where $\tau(T)$ of a given instance is $\tau(T) = T \frac{\ln(1-0.99)}{\ln(1-p_0(T))}$ and T is the duration in seconds of a run. The probability $p_0(T)$ is evaluated using 100 runs per instance. The results of the CAC algorithm run on a CPU are also included in Fig. 3. The CPU implementation of CAC written in python for this work is not optimized and is consequently slower than other algorithms. However, its scaling of time to solution with problem size is consistent with that of

CAC on FPGA. Figure 3 shows that CAC implemented on either CPU or FPGA has a significantly smaller increase of time to solution with problem size than SA run on CPU. Note that the power consumption of transistors in the FPGA and CPU scales proportionally to their clock frequencies. In order to compare different hardware despite the heterogeneity in their power consumption, the q^{th} percentile of energy-to-solution E_q^* , i.e., the energy E_q^* required to solve SK instances with $E_q^* = P\tau_q^*$ and P the power consumption, is plotted in Fig. 3(b). For the sake of simplicity, we assume a 20 watts power consumption for the CPU. These numbers represent typical orders to magnitude for contemporary digital systems. CAC on FPGA is 10^2 to 10^3 times more energy efficient than state-of-the-art algorithms running on classical computers.

The Monte-Carlo methods SA and PT have moreover been recently implemented on a special-purpose electronic chip called Digital Annealer (DA)²⁰. In Fig. 4, we show the scaling exponents of 50th and 80th percentiles of the MVM and time to solution distribution for problem sizes $N = 800$ to $N = 1100$ based on the hypothesis of scaling in $e^{\gamma N}$ obtained by fitting data shown in Figs. 2 and 3 (see “CAC fully parallel” and “CAC 100×100 ”, respectively, in Fig. 4 for the numerical values of the scaling exponents), and compare them to that reported for SA and PT implemented on CPU and DA. Note that we replicated the benchmark method of ref. ²⁰ for the sake of the comparison by fitting time to solution from $N = 800$ up to $N = 1100$. The scaling obtained from fitting the MVM to solution in Fig. 2 is based on the assumption that the matrix-vector multiplication can be calculated fully in parallel in a time that scales as $\log(N)$ instead of N^2 (see Methods section) at least up to $N = 1100$. We include this hypothesis because many other Ising machines exploit the parallelization of matrix-vector multiplication for speed up^{20,63}, whereas the current implementation of CAC iterates on block matrices of size 100 by 100 and is thus only partially parallel because of resource limitations specific to the downscale FPGA used in this work. Note that the number of matrix-vector multiplications necessary to find the solution state dominates the exponential scaling rather than the time to compute one matrix-vector multiplication. The scaling of time to solution for chaotic amplitude control observed is significantly smaller than the ones of standard Monte-Carlo methods SA and PT²⁰, especially in the case of a fully parallel implementation. The scaling exponents of fully parallel CAC is smaller than that of DA and on par with that of PT on DA (PTDA), although CAC does not require simulating replica of the system and is thus faster in absolute time than PTDA²⁰.

Next, the proposed implementation of chaotic amplitude control is compared to other recently developed Ising machines (see Fig. 5). The relatively slow increase of time to solution with respect to the number of spins N when solving larger SK problems using CAC suggests that our FPGA implementation is faster than the Hopfield neural network implemented using memresistors (mem-HNN)⁴, the restricted Boltzmann machine using a FPGA (FPGA-RBM)⁵⁵ at large N . Extrapolations are based on the hypotheses of scaling in $e^{\gamma N}$ and $e^{\gamma\sqrt{N}}$ by fitting the available experimental data up to $N = 100$ for mem-HNN and FPGA-RBM, $N = 150$ for NTT CIM, and $N = 1100$ for FPGA-CAC. Figure 5 shows that mem-HNN, FPGA-RBM, and NTT CIM have similar scaling exponents, although FPGA-RBM tends to exhibit a scaling in $e^{\gamma N}$ rather than $e^{\gamma\sqrt{N}}$ for $N \approx 100$ ⁵⁵. It can be nonetheless expected that the algorithm implemented in mem-HNN, which is similar to mean-field annealing, has the same scaling behavior as simCIM and NMFA (see Fig. 2). Note that the lines showing the scaling in $e^{\gamma\sqrt{N}}$ in Fig. 5 can be seen as a lower bound of the TTS if the sub-exponential scaling is finite-size effect.

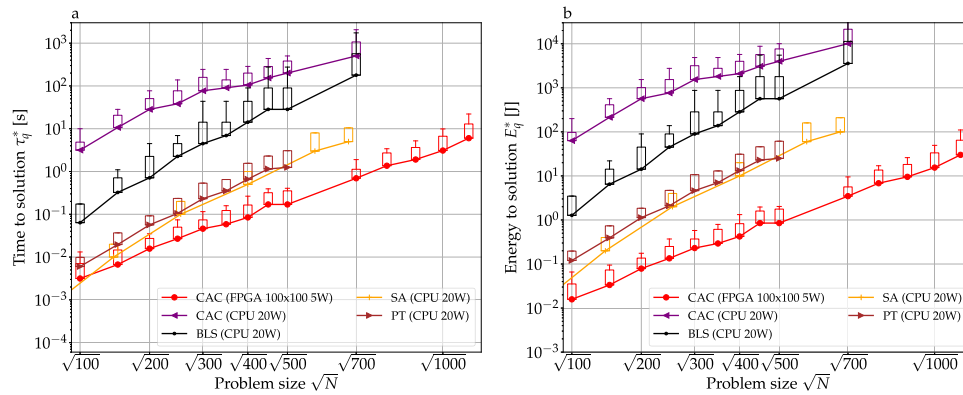


Fig. 3 Time to solution of the field programmable gate array (FPGA) implementation for Sherrington-Kirkpatrick instances. **a** Lower, higher, and upper whisker of boxes show the 50th, 80th, and 90th percentiles of the time to solution τ_{99}^* distribution in seconds vs. the square root of problem size \sqrt{N} for the FPGA implementation of chaotic amplitude control (CAC) with a maximum of 5W power consumption and the following algorithms running on a CPU (20W): CAC, simulated annealing (SA), and parallel tempering (PT). **b** The same as **a** for the energy-to-solution E_{99}^* . Power consumption of the FPGA is considered constant with respect to N because of the pipelined implementation.

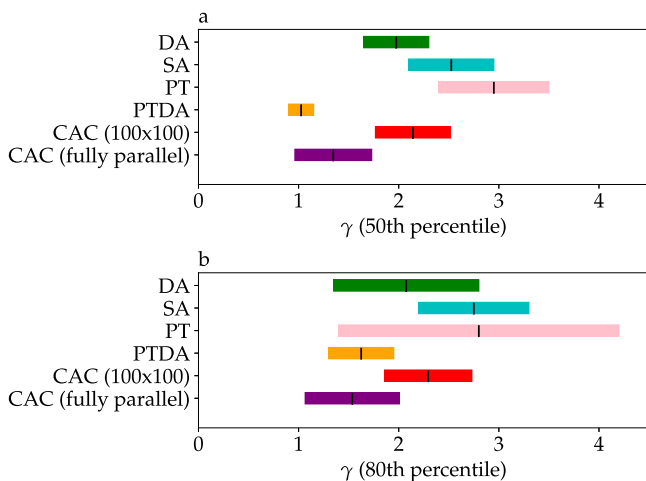


Fig. 4 Scaling exponents of the proposed method. Scaling exponents $\frac{y}{\log(10)}$ of the 50th (**a**) and 80th (**b**) percentiles of the time to solution distribution based on the hypotheses of scaling in $e^{\gamma N}$ obtained by fitting data of the number of matrix-vector multiplication (MVM) to solution vs. problem size N shown in Fig. 2 of the proposed chaotic amplitude control dynamics and the scaling exponents reported in ref. 20. Colored boxes show the 90% confidence interval in the scaling exponents. CAC: chaotic amplitude control; SA: simulated annealing; PT: parallel tempering; DA: digital annealer; PTDA: parallel tempering on DA. Exponents of SA, PT, DA, and PTDA are taken from ref. 20. CAC (100 × 100) and CAC (fully parallel) are estimated using time (see Fig. 3) and MVM (see Fig. 2) to solution, respectively.

It is noteworthy to mention that a recent implementation of the simulated bifurcation machine⁶³ (SBM), which is not based on the gradient descent, similarly to CAC, but based on adiabatic evolutions of energy conservative systems performs well in solving SK problems. Both SBM and CAC exhibit smaller time to solution than other gradient based methods. SBM has been implemented on a FPGA (the Intel Stratix 10 GX) that has approximately 5 to 10 times more adaptive logic modules than the KU040 FPGA used to implement CAC. In order to compare SBM and CAC if implemented on an equivalent FPGA hardware, we plot in Fig. 5 the estimation of the time to solution for a fully parallel implementation of CAC using the hypothesis that one matrix-vector multiplication of size 1100×1100 can be achieved in $0.3 \mu\text{s}$. This is the same time to compute a MVM that we can infer from time to solution reported in ref. 63 for SBM

with binary connectivity given that problems of size $N = 100$ ($N = 700$) are solved in $29 \mu\text{s}$ (55ms) and 94 (81000) MVMs, respectively. Note that SBM can reach the solution states after approximately 20 times less MVMs than CAC at $N = 100$ but only 5 times less MVM at $N = 1000$, suggesting that the speed of SBM depends largely on hardware rather than an algorithmic scaling advantage. Moreover, the simulated bifurcation machine⁶³ does not perform significantly better than our current implementation of CAC for solving instances of the reference MAXCUT benchmark set called GSET (<https://web.stanford.edu/~yyye/yyye/Gset/>) (see Table 1) because CAC can find better cuts and exhibits smaller time to solution for multiple GSET instances even compared to the case of the implementation on the smaller KU040 FPGA with the probability of finding maximum cuts of the GSET in a single run that is much smaller with SBM. Comparison of the scaling behavior of time to solution between CAC and SBM is unfortunately not possible based on available data⁶³.

Estimation of time to solution distribution. Next, we consider the whole distribution of time to solution in order to compare the ability of various methods to solve harder instances. As shown in Fig. 6(a), the cumulative distribution function (CDF) $P(\tau; T)$ of time to solution with 99% success probability τ is not uniquely defined as it depends on the duration T of the runs. We can define an optimal CDF $P^*(\tau)$ that is independent of the runtime T as follows: $P^*(\tau) = \max_T P(\tau; T)$. Numerical simulations show that this optimal CDF is well described by lognormal distribution, that is $P^*(\log(\tau)) \sim \mathcal{N}(\mu, \sqrt{\nu})$ where $\sqrt{\nu}$ is the standard deviation of $\log(\tau)$ (see Fig. 6(b), (c), and (d) for the cases of CAC, SA, and NMFA, respectively). In Fig. 6(e), it is shown that the scaling of the standard deviation $\sqrt{\nu}(N)$ with the problem size N is significantly smaller for CAC, which implies that harder instances can be solved relatively more rapidly than using other methods. This result confirms the advantageous scaling of higher percentiles for CAC that was observed in Figs. 2 and 3.

Conclusions

The framework described in this paper can be extended to solve other types of constrained combinatorial optimization problems such as the traveling salesman⁴⁵, vehicle routing, and lead optimization problems. Moreover, it can be adapted to a variety of recently proposed Ising machines^{4-13,15}, which would benefit from implementing a scheme that does not rely solely on the descent of a potential function. In particular, the performance of CIM^{9,10}, mem-HNN⁴, and chip-scale photonic Ising machine¹⁴, which have small

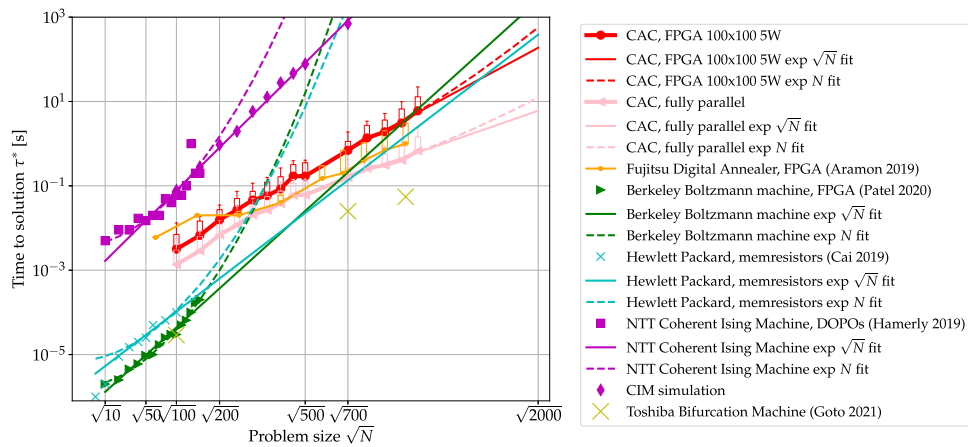


Fig. 5 Median time to solution and extrapolations for the exponential and sub-exponential scaling hypotheses on different Ising machines. For chaotic amplitude control on field programmable gate array (CAC, FPGA) and digital annealer (DA), the lower, higher, and upper whisker of boxes show the 50th, 80th, and 90th percentiles of the real time to solution distribution. Markers show the experimental and simulated data. Dashed and full straight lines show the fit of the logarithm of median time to solution vs. problem size N and its square root \sqrt{N} , respectively, and extrapolations up to $N = 2000$.

Table 1 Performance of the field programmable gate array (FPGA) implementation of chaotic amplitude control (CAC) on the GSET. Performance of the FPGA implementation of CAC in finding the maximum cuts known, i.e., lowest Ising Hamiltonian known, of graphs in the GSET benchmark. id , C^{opt} , C^{CAC} , C^{SBM} are the name of instances, best maximum cuts known from ref. 73, the proposed method after 100 runs, and Toshiba bifurcation machine on FPGA⁶³ (FPGA-SBM), respectively. C^{SBM} is evaluated using >100 runs but shorter time per run. The difference in cuts between CAC and SMB is given by $\Delta C_{CAC/SBM} = C^{CAC} - C^{SBM}$. p_0^{CAC} and p_0^{SBM} are the probabilities that FPGA-CAC and FPGA-SBM find the cut C^{CAC} and C^{SBM} in a single run, respectively. Moreover, $\langle t^{BLS} \rangle$ and $\langle t^{CAC} \rangle$ are the averaged time over 20 and 100 runs, respectively, to find the corresponding best cut in successful runs using BLS written C++ and running on a Xeon E5440 2.83 GHz⁵⁹ and the proposed scheme implemented on the KU040 FPGA. Moreover, TTS^{CAC} and TTS^{SBM} are the time to find the best cut with 99% success probability using CAC and SBM, respectively, on FPGA. Times are written with parentheses around them if the cuts found are not the best known ones. The times are in seconds. For each instance, the smaller TTS from the two last columns of the table is shown in bold when the comparison is applicable.

id	N	C^{opt}	C^{CAC}	C^{SBM}	$\Delta C_{CAC/SBM}$	p_0^{CAC}	p_0^{SBM}	$\langle t^{BLS} \rangle$	$\langle t^{CAC} \rangle$	TTS^{CAC}	TTS^{SBM}
22	2000	13359	13359	13359	0	0.11	0.018	560	3.33	157.11	2.7
23	2000	13344	13342	13342	0	0.25	0.0490	(278)	(6.52)	(45.26)	(0.94)
24	2000	13337	13337	13337	0	0.34	0.0077	311	7.27	20.78	3.1
25	2000	13340	13340	13340	0	0.38	0.0022	148	10.43	33.61	15
26	2000	13328	13328	13328	0	0.42	0.0050	429	15.32	40.95	3.2
27	2000	3341	3341	3341	0	0.91	0.0479	449	7.69	4.1	0.26
28	2000	3298	3298	3298	0	0.18	0.0690	432	10.12	63.32	0.45
29	2000	3405	3405	3405	0	0.47	0.0039	17	11.72	2732.62	1.2
30	2000	3413	3413	3413	0	0.01	0.0045	(283)	7.06	21.86	2.8
31	2000	3310	3309	3310	-1	0.18	0.0012	(285)	(9.98)	(4067.07)	5.3
32	2000	1410	1410	1410	0	0.01	0.0012	336	17.67	2323.94	33
33	2000	1382	1380	1382	-2	0.03	0.0002	402	(22.93)	(749.47)	120
34	2000	1384	1384	1384	0	0.04	0.0015	170	24.85	255.62	27
35	2000	7687	7685	7685	0	0.26	0.0002	(442)	(20.67)	(437.53)	(479)
36	2000	7680	7679	7677	2	0.09	0.0001	(604)	(20.97)	(159.94)	(1597)
37	2000	7691	7690	7691	-1	0.03	0.0001	(444)	(5.24)	(5.24)	1278
38	2000	7688	7688	7688	0	0.01	0.0003	(461)	19.05	5872.14	213
39	2000	2408	2408	2408	0	0.44	0.0006	251	11.23	16.39	266
40	2000	2400	2396	2400	-4	0.03	0.0001	431	(15.35)	(1703.22)	48
41	2000	2405	2405	2405	0	0.02	0.0020	73	3.04	623.12	48
42	2000	2481	2481	2479	2	0.39	0.0002	183	17.89	178.53	(240)

time to solution for small problem sizes ($N \approx 100$) for which it may be sufficient to do rapid sampling based on convex optimization⁶⁴ but with a relatively large scaling exponent, which limits their scalability, could be improved by adapting the implementation we propose if these hardware can be shown to be able to simulate larger numbers of spins experimentally. Rapid progress in the growing field of Ising machines may allow to verify scaling behaviors of the

various methods at larger problem sizes and, thus, limit further finite-size effects.

The scaling exponents we have reported in this paper are based on the integration of the chaotic amplitude control dynamics using a Euler approximation of its ODEs. We have noted that the scaling of MVMs to solution of SK spin glass problems is reduced when the Euler time step is decreased (see Supplementary

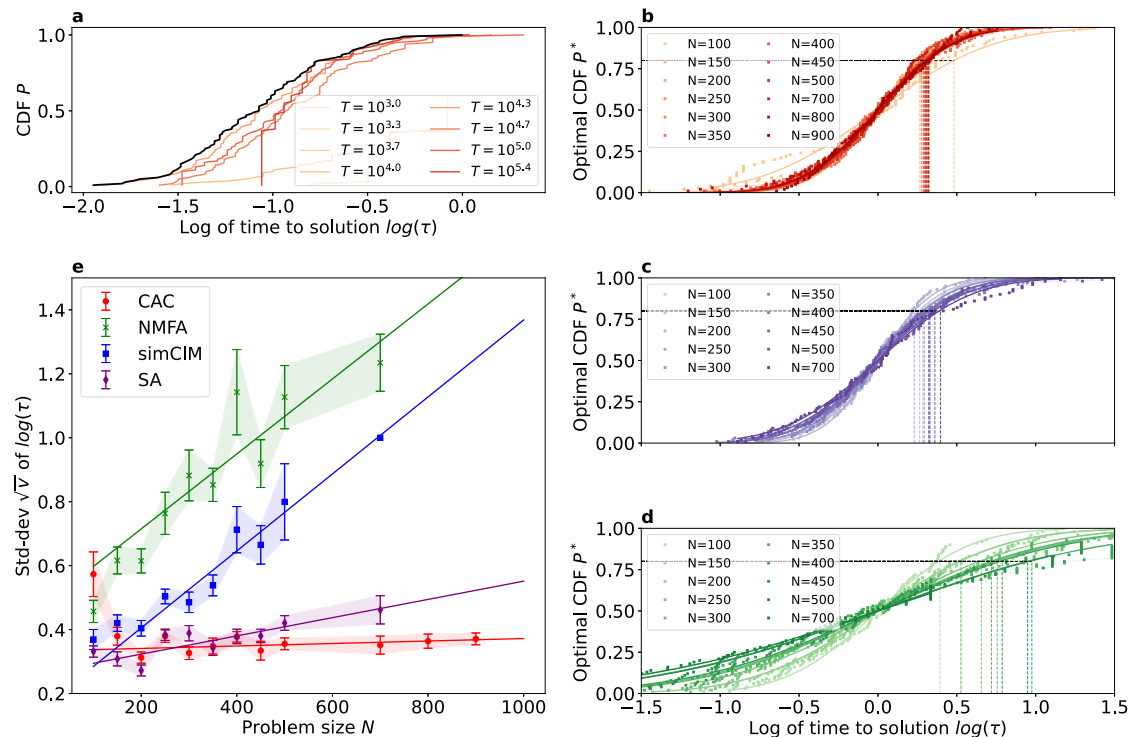


Fig. 6 Estimation of time to solution distribution of chaotic amplitude control for Sherrington-Kirkpatrick (SK) instances. **a** Cumulative distribution of the time to solution $P(\tau)$ for $N = 400$ SK problems. The black line corresponds to the maximum of $P(\tau, T)$ with respect to the duration of the runs T . **b, c, d** Optimal cumulative distribution $P^*(\tau)$ with $P^*(\log(\tau)) \sim \mathcal{N}(\mu(N), \sqrt{v(N)})$ for chaotic amplitude control (CAC) (**b**), simulated annealing (SA) (**c**), and noisy mean-field annealing (NMFA) (**d**), respectively. **e** Standard deviation \sqrt{v} of the logarithm of time to solution distribution vs. problem size N . Shaded regions show the 99% confidence interval in the standard deviation. simCIM: simulation of the coherent Ising machine.

Note 2.8). The scaling exponents of CAC might thus be smaller than reported in this paper in the limit of a more accurate numerical integration over continuous time. It is therefore important future work to evaluate the scaling for $N \gg 1000$ using a faster numerical simulation method. Such numerical calculations require a careful analysis of the integration method of ODEs, numerical precision, and tuning of parameters. It is also of considerable interest to implement CAC on multiple FPGAs for very high parallelization⁶⁵ and on analog physical systems for further reduction of power consumption.

Nonrelaxational dynamics described herein is not limited to artificial simulators and likely also emerge in natural complex systems. In particular, it has been hypothesized that the brain operates out of equilibrium at large scales and produces more entropy when performing physically and cognitively demanding tasks^{66,67} in particular for the ones involving creativity for maximizing reward rather than memory retrieval. Such neural processes cannot be explained simply by the relaxation to fixed point attractors⁴⁹, periodic limit cycles⁶⁸, or even low-dimensional chaotic attractors whose self-similarity may not be equivalent to complexity but set the conditions for its emergence^{67,69}. Similarly, evolutionary dynamics that is characterized as non-ergodic when the time required for representative genome exploration is longer than available evolutionary time⁷⁰ may benefit from nonrelaxational dynamics rather than slow glassy relaxation for faster identification of high-fitness solutions. A detailed analytic comparison between the slow relaxation dynamics observed in Monte-Carlo simulations of spin glasses with the one proposed in this paper is needed in order to explain the apparent difference in their scaling of time to solution exhibited by our numerical results.

Methods

We target the implementation of a low-power system with maximum power supply of 5W using a XCKU040 Kintex Ultrascale Xilinx FPGA integrated on an Avnet

board. The implemented circuit can process Ising problems of up to at least 1100 spins fully connected and of >2000 spins sparsely connected within the 5W power supply limit. Data is encoded into 18 bits fixed point vectors with 1 sign, 5 integer and 12 decimal bits to optimize computation time and power consumption. An important feature of our FPGA implementation of CAC is the use of several clock frequencies to concentrate the electrical power on the circuits that are the bottleneck of computation and require a high-speed clock. For the realization of the matrix-vector multiplication, each element of the matrix is encoded with 2 bits precision (w_{ij} is $-1, 0$ or 1). An approximation based on the combination of logic equations describing the behavior of a multiplexer allows to achieve 10^4 multiplications within one clock cycle. The results of these multiplications are summed using cascading DSP and CARRY8 connected in a tree structure. Using pipelining, a matrix-vector multiplication for a squared matrix of size N is computed in $2 + 5 \frac{\log(N-4)}{\log(5)} + \left(\frac{N}{u}\right)^2$ clock cycles (see Supplementary Note 2.4) at a clock frequency of 50MHz with $u = 100$, which is determined by the limitation of the number of available electronic component of the XCKU040 FPGA. The block size u can be made at least 3 times larger using commercially available FPGAs, which implies that the number of clock cycles needed to compute a dot product can scale almost logarithmically for problems of size $N \approx 1000$ (see Supplementary Note 2.4 for discussions of scalability) and that the calculation time can be further significantly decreased using a higher-end FPGA. The calculation of the nonlinearity f_i and error terms is achieved at higher frequency (300MHz and 100Mhz) using DSP in $8 + (N/u)$ and $9 + (N/u)$ clock cycles, respectively. In order to minimize energy resources and maximize speed, the nonlinear and error terms are calculated multiple times during the calculation of a single matrix-vector multiplication (see Supplementary Note 2).

Prior to computing the benchmark on the Sherrington-Kirkpatrick instances, the parameters of the system (see Supplementary Note 2.8) are optimized automatically using Bayesian optimization and bandit-based methods⁷¹. The automatic tuning of parameters for some previously unseen instances is out of the scope of this work but can be achieved to some extent using machine learning techniques⁷².

Data availability

Sherrington-Kirkpatrick instances used in this paper are available upon request to T. Leleu. The GSET instances are available at <https://web.stanford.edu/~yyye/yyye/Gset/>.

Code availability

Requests for code availability should be addressed to T. Leleu.

Received: 21 March 2021; Accepted: 23 November 2021;

Published online: 15 December 2021

References

- Parisi, G., Mézard, M. & Virasoro, M. *Spin Glass Theory and Beyond*. Vol 187, 202 (World Scientific, 1987).
- Kochenberger, G. et al. The unconstrained binary quadratic programming problem: a survey. *J. Comb. Optim.* **28**, 58–81 (2014).
- Vadlamani, S. K., Xiao, T. P. & Yablonoitch, E. Physics successfully implements Lagrange multiplier optimization. *Proc. Natl Acad. Sci. USA* **117**, 26639–26650 (2020).
- Cai, F. et al. Power-efficient combinatorial optimization using intrinsic noise in memristor Hopfield neural networks. *Nat. Electron.* **3**, 1–10 (2020).
- Mahboob, I., Okamoto, H. & Yamaguchi, H. An electromechanical Ising Hamiltonian. *Sci. Adv.* **2**, e1600236 (2016).
- Camsari, K. Y., Faria, R., Sutton, B. M. & Datta, S. Stochastic p-bits for invertible logic. *Phys. Rev. X* **7**, 031014 (2017).
- Camsari, K. Y., Sutton, B. M. & Datta, S. p-bits for probabilistic spin logic. *Appl. Phys. Rev.* **6**, 011305 (2019).
- Marandi, A., Wang, Z., Takata, K., Byer, R. L. & Yamamoto, Y. Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. *Nat. Photon.* **8**, 937–942 (2014).
- McMahon, P. L. et al. A fully programmable 100-spin coherent Ising machine with all-to-all connections. *Science* **354**, 614–617 (2016).
- Inagaki, T. et al. Large-scale Ising spin network based on degenerate optical parametric oscillators. *Nat. Photon.* **10**, 415–419 (2016).
- Pierangeli, D., Marcucci, G. & Conti, C. Large-scale photonic Ising machine by spatial light modulation. *Phys. Rev. Lett.* **122**, 213902 (2019).
- Roques-Carmes, C. et al. Heuristic recurrent algorithms for photonic Ising machines. *Nat. Commun.* **11**, 1–8 (2020).
- Prabhu, M. et al. Accelerating recurrent Ising machines in photonic integrated circuits. *Optica* **7**, 551–558 (2020).
- Okawachi, Y. et al. Demonstration of chip-based coupled degenerate optical parametric oscillators for realizing a nanophotonic spin-glass. *Nat. Commun.* **11**, 1–7 (2020).
- Hamerly, R. et al. Experimental investigation of performance differences between coherent Ising machines and a quantum annealer. *Sci. Adv.* **5**, eaau0823 (2019).
- Kalinin, K. P. & Berloff, N. G. Global optimization of spin Hamiltonians with gain-dissipative systems. *Sci. Rep.* **8**, 17791 (2018).
- Johnson, M. W. et al. Quantum annealing with manufactured spins. *Nature* **473**, 194 (2011).
- Kumar, S., Strachan, J. P. & Williams, R. S. Chaotic dynamics in nanoscale NbO₂ Mott memristors for analogue computing. *Nature* **548**, 318 (2017).
- Heim, B., Rønnow, T. F., Isakov, S. V. & Troyer, M. Quantum versus classical annealing of Ising spin glasses. *Science* **348**, 215–217 (2015).
- Aramon, M. et al. Physics-inspired optimization for quadratic unconstrained problems using a digital annealer. *Front. Phys.* **7**, 48 (2019).
- Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science* **220**, 671–680 (1983).
- King, A. D., Bernoudy, W., King, J., Berkley, A. J. & Lanting, T. Emulating the coherent Ising machine with a mean-field algorithm. Preprint at <https://arxiv.org/abs/1806.08422> (2018).
- Bilbro, G. et al. in *Advances in Neural Information Processing Systems*, (ed. Touretzky, D. S.), 91–98 (Morgan Kaufmann Publishers Inc., 1989).
- Chen, L. & Aihara, K. Chaotic simulated annealing by a neural network model with transient chaos. *Neural Netw.* **8**, 915–930 (1995).
- Kadowaki, T. & Nishimori, H. Quantum annealing in the transverse Ising model. *Phys. Rev. E* **58**, 5355 (1998).
- Tanaka, F. & Edwards, S. Analytic theory of the ground state properties of a spin glass. i. Ising spin glass. *J. Phys. F: Metal Phys.* **10**, 2769 (1980).
- Biroli, G. Dynamical TAP approach to mean field glassy systems. *J. Phys.* **32**, 8365 (1999).
- Bernaschi, M., Billoire, A., Maiorano, A., Parisi, G. & Ricci-Tersenghi, F. Strong ergodicity breaking in aging of mean-field spin glasses. *Proc. Natl Acad. Sci. USA* **117**, 17522–17527 (2020).
- Cugliandolo, L. F. & Kurchan, J. On the out-of-equilibrium relaxation of the Sherrington-Kirkpatrick model. *J. Phys. A* **27**, 5749 (1994).
- Leleu, T., Yamamoto, Y., McMahon, P. L. & Aihara, K. Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity. *Phys. Rev. Lett.* **122**, 040607 (2019).
- Ercsey-Ravasz, M. & Toroczkai, Z. Optimization hardness as transient chaos in an analog approach to constraint satisfaction. *Nat. Phys.* **7**, 966–970 (2011).
- Molnár, B., Molnár, F., Varga, M., Toroczkai, Z. & Ercsey-Ravasz, M. A continuous-time MaxSat solver with high analog performance. *Nat. Commun.* **9**, 4864 (2018).
- Aspelmeier, T. & Moore, M. Realizable solutions of the Thouless-Anderson-Palmer equations. *Phys. Rev. E* **100**, 032127 (2019).
- Boettcher, S. & Percus, A. G. Optimization with extremal dynamics. *Phys. Rev. Lett.* **86**, 5211–5214 (2001).
- Zarand, G., Pazmandi, F., Pal, K. & Zimanyi, G. Using hysteresis for optimization. *Phys. Rev. Lett.* **89**, 150201 (2002).
- Leleu, T., Yamamoto, Y., Utsunomiya, S. & Aihara, K. Combinatorial optimization using dynamical phase transitions in driven-dissipative systems. *Phys. Rev. E* **95**, 022118 (2017).
- Aspelmeier, T., Blythe, R., Bray, A. J. & Moore, M. A. Free-energy landscapes, dynamics, and the edge of chaos in mean-field models of spin glasses. *Phys. Rev. B* **74**, 184411 (2006).
- Hasegawa, M., Ikeguchi, T. & Aihara, K. Combination of chaotic neurodynamics with the 2-opt algorithm to solve traveling salesman problems. *Phys. Rev. Lett.* **79**, 2344 (1997).
- Horio, Y. & Aihara, K. Analog computation through high-dimensional physical chaotic neuro-dynamics. *Phys. D: Nonlin. Phenom.* **237**, 1215–1225 (2008).
- Aihara, K. Chaos engineering and its application to parallel distributed processing with chaotic neural networks. *Proc. IEEE* **90**, 919–930 (2002).
- Montanari, A. Optimization of the Sherrington-Kirkpatrick Hamiltonian. *SIAM J. Comput.* FOCS19-1–FOCS19-38 <https://doi.org/10.1137/20M132016X> (2018).
- Furber, S. B., Galluppi, F., Temple, S. & Plana, L. A. The spinnaker project. *Proc. IEEE* **102**, 652–665 (2014).
- Davies, M. et al. Loihi: A neuromorphic manycore processor with on-chip learning. *IEEE Micro* **38**, 82–99 (2018).
- Benjamin, B. V. et al. Neurogrid: A mixed-analog-digital multichip system for large-scale neural simulations. *Proc. IEEE* **102**, 699–716 (2014).
- Hopfield, J. J. & Tank, D. W. "Neural" computation of decisions in optimization problems. *Biol. Cybern.* **52**, 141–152 (1985).
- Wang, Z., Marandi, A., Wen, K., Byer, R. L. & Yamamoto, Y. Coherent Ising machine based on degenerate optical parametric oscillators. *Phys. Rev. A* **88**, 063853 (2013).
- Sompolinsky, H. & Zippelius, A. Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. *Phys. Rev. B* **25**, 6860 (1982).
- Thouless, D. J., Anderson, P. W. & Palmer, R. G. Solution of 'solvable model of a spin glass'. *Philos. Magaz.* **35**, 593–601 (1977).
- Hopfield, J. J. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Natl Acad. Sci. USA* **81**, 3088–3092 (1984).
- Farhi, E., Goldstone, J., Gutmann, S. & Sipser, M. Quantum computation by adiabatic evolution. Preprint at <https://arxiv.org/abs/quant-ph/0001106> (2000).
- Granville, V., Krivánek, M. & Rasson, J.-P. Simulated annealing: a proof of convergence. *IEEE Trans. Pattern Anal. Mach. Intell.* **16**, 652–656 (1994).
- Geman, S. & Hwang, C.-R. Diffusions for global optimization. *SIAM J. Contr. Optimiz.* **24**, 1031–1043 (1986).
- Hoppensteadt, F. C. & Izhikevich, E. M. Weakly connected neural networks, Vol. 126 (Springer Science & Business Media, 2012).
- Goto, H. Bifurcation-based adiabatic quantum computation with a nonlinear oscillator network. *Sci. Rep.* **6**, 21686 (2016).
- Patel, S., Chen, L., Canozo, P. & Salahuddin, S. Ising model optimization problems on a FPGA accelerated restricted boltzmann machine. Preprint at <https://arxiv.org/abs/2008.04436> (2020).
- Hoos, H. H. & Stützle, T. On the empirical scaling of run-time for finding optimal solutions to the travelling salesman problem. *Eur. J. Operat. Res.* **238**, 87–94 (2014).
- Hamze, F., Raymond, J., Pattison, C. A., Biswas, K. & Katzgraber, H. G. Wishart planted ensemble: A tunably rugged pairwise Ising model with a first-order phase transition. *Phys. Rev. E* **101**, 052102 (2020).
- Perera, D. et al. Chook—a comprehensive suite for generating binary optimization problems with planted solutions. Preprint at <https://arxiv.org/abs/2005.14344> (2021).
- Benlic, U. & Hao, J.-K. Breakout local search for the Max-Cut Problem. *Eng. Appl. Artif. Intell.* **26**, 1162–1173 (2013).
- Hukushima, K. & Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Japan* **65**, 1604–1608 (1996).
- Mandra, S., Villalonga, B., Boixo, S., Katzgraber, H. & Rieffel, E. State-of-the-art classical tools to benchmark NISQ devices. (APS Meeting Abstracts, 2019).
- Isakov, S. V., Zintchenko, I. N., Rønnow, T. F. & Troyer, M. Optimised simulated annealing for Ising spin glasses. *Comput. Phys. Commun.* **192**, 265–271 (2015).

63. Goto, H. et al. High-performance combinatorial optimization based on classical mechanics. *Sci. Adv.* **7**, eabe7953 (2021).
64. Ma, Y.-A., Chen, Y., Jin, C., Flammarion, N. & Jordan, M. I. Sampling can be faster than optimization. *Proc. Natl Acad. Sci. USA* **116**, 20881–20885 (2019).
65. Tatsumura, K., Yamasaki, M. & Goto, H. Scaling out Ising machines using a multi-chip architecture for simulated bifurcation. *Nat. Electron.* **4**, 208–217 (2021).
66. Lynn, C. W., Papadopoulos, L., Kahn, A. E. & Bassett, D. S. Human information processing in complex networks. *Nat. Phys.* **16**, 965–973 (2020).
67. Wolf, Y. I., Katsnelson, M. I. & Koonin, E. V. Physical foundations of biological complexity. *Proc. Natl Acad. Sci. USA* **115**, E8678–E8687 (2018).
68. Yan, H. et al. Nonequilibrium landscape theory of neural networks. *Proc. Natl Acad. Sci. USA* **110**, E4185–E4194 (2013).
69. Ageev, D., Arefeva, I., Bagrov, A. & Katsnelson, M. I. Holographic local quench and effective complexity. *J. High Energy Phys.* **2018**, 71 (2018).
70. McLeish, T. C. Are there ergodic limits to evolution? ergodic exploration of genome space and convergence. *Interf. Focus* **5**, 20150041 (2015).
71. Falkner, S., Klein, A. & Hutter, F. Robust and efficient hyperparameter optimization at scale. In: *International Conference on Machine Learning*, 1437–1446 (PMLR, 2018).
72. Dunning, I., Gupta, S. & Silberholz, J. What works best when? a systematic evaluation of heuristics for Max-Cut and QUBO. *INFORMS J. Comput.* **30**, 608–624 (2018).
73. Ma, F. & Hao, J.-K. A multiple search operator heuristic for the max-k-cut problem. *Ann. Oper. Res.* **248**, 365–403 (2017).

Acknowledgements

The authors thank Salvatore Mandrà for providing results of the PT algorithm, Davide Venturelli, David Bernal, Sam Reifstein, and the anonymous referee for their suggestion of applying CAC on the Wishart planted instances. This research is partially supported by the Brain-Morphic AI to Resolve Social Issues (BMAI) project (NEC corporation), the University of Tokyo GAP Fund Program, AMED under Grant Number JP21dm0307009, UTokyo Center for Integrative Science of Human Behavior (CiSHuB), and Japan Science and Technology Agency Moonshot R&D Grant Number JPMJMS2021

Author contributions

T. Leleu proposed the project. F.K. and T. Leleu performed the FPGA implementation, numerical simulations, and data analysis. R.H. and T. Levi supported theoretical analysis

and experimental data. T. Leleu, F.K., T. Levi, R.H., T.K., and K.A. wrote the manuscript with inputs from all authors.

Competing interests

T. Leleu and K.A. are inventors on patent WO-2020027785-A1 submitted by the University of Tokyo that covers amplitude heterogeneity error-correction scheme. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-021-00768-0>.

Correspondence and requests for materials should be addressed to Timothée Leleu.

Peer review information *Communications Physics* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021