

Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity

Timothée Leleu,¹ Yoshihisa Yamamoto,^{2,3} Peter L. McMahon,^{3,4,5} and Kazuyuki Aihara^{1,6}

¹*Institute of Industrial Science, The University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*

²*ImPACT program, The Japan Science and Technology Agency,
Gobancho 7, Chiyoda-ku, Tokyo 102-0076, Japan,*

³*E. L. Ginzton Laboratory, Stanford University, Stanford, CA94305, USA.*

⁴*National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan.*

⁵*School of Applied and Engineering Physics, Cornell University, Ithaca, NY14853, USA*

⁶*International Research Center for Neurointelligence (IRCIN),
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan.*

The relaxation of binary spins to analog values has been the subject of much debate in the field of statistical physics, neural networks, and more recently quantum computing, notably because the benefits of using an analog state for finding lower energy spin configurations are usually offset by the negative impact of the improper mapping of the energy function that results from the relaxation. We show that it is possible to destabilize trapping sets of analog states that correspond to local minima of the binary spin Hamiltonian by extending the phase space to include error signals that correct amplitude inhomogeneity of the analog spin states and controlling the divergence of their velocity. Performance of the proposed analog spin system in finding lower energy states is competitive against state-of-the-art heuristics.

Many algorithms and hardware dedicated to solving hard combinatorial optimization problems utilize a mapping of the cost function to the energy landscape of simple physical systems such as classical spins [1–5], quantum spins [6, 7], optical oscillators [8–12], solid-state oscillators [13, 14], and neural networks [15] (see [16–18] for reviews). From a random initial state, the probability to find the lowest energy states, or ground states, depends critically on whether the non-equilibrium dynamics of such systems can escape efficiently from local minima. The difficulty in solving hard problems stems from the fact that the number of these local minima, for which energy is not decreased by any single spin flip at zero temperature, generally grows exponentially with the problem size. By coupling the system to a (Markovian) heat bath, transitions over energy barriers are allowed, and convergence to the ground state is assured for a slow enough decrease of the temperature [1, 19]. Such approaches have been improved upon in recent years [2], and generalized to the case of quantum spins in quantum annealing [6, 7].

Alternatively, it has been suggested in various systems (including soft spins [20], analog neural networks [15], coherent Ising machines [8–11]) that relaxing the binary spins $\sigma_i = \pm 1$ to analog values x_i with $x_i \in \mathbb{R}$ may increase the probability of finding lower energy states. In these gain-dissipative systems, combinatorial optimization is classically achieved by mapping local minima to fixed point attractors [8, 15]. These attractors are created because of dissipation that induces phase-space contraction under the action of the dynamics [21]. By tuning the gain, i.e., the average energy supplied to the system, dissipation can be compensated and the rate and directions in which phase-space volumes are contracted can

be controlled. Earlier studies on the Hopfield neural network suggest that reduction of the gain (or steepness of the neuron transfer function) results in improved quality of solutions of traveling salesman problems [15] and in the exponential decrease of the number of fixed points [22, 23]. Moreover, gradual reduction of the gain, which can be related to temperature via the naive mean field of Thouless, Anderson, and Palmer, can serve as “mean field annealing” [24, 25]. In the framework of the coherent Ising machine, it has been shown numerically [8] and experimentally [10, 11] that such machines can be used as an efficient heuristic solver for hard combinatorial optimization problems such as MAX-CUT (which is equivalent to Ising problems). These schemes rely on setting the gain to a minimum at which (most) suboptimal configurations, or excited states, cannot be stable fixed points [8, 26]. In the field of combinatorial optimization, such analog systems have been relatively less studied than their binary spins counterparts, and it is believed that they do not perform as well as state-of-the-art heuristics [27] notably because it is difficult to map low energy configurations of the binary system to analog states with smaller loss [26, 28], although these systems are ideal for efficient implementation on dedicated hardware [10, 15, 29].

In this Letter, we argue that analog bistable systems, even when simulated on a classical computer, can in fact find low energy states at least as efficiently as current state-of-the-art heuristics. Our model is based on the observation that the improper mapping of the objective function by the loss landscape results from the fact that the amplitudes of analog variables are in general heterogeneous (i.e., not all equal). We propose to correct the amplitude heterogeneity by extending the phase space

[30] using auxiliary degrees of freedom, which we call error variables. The proposed system utilizes the fact that the local stability (the Jacobian matrix) of fixed points can be controlled *a priori* in analog systems in order to reduce the number of stable local minima. Moreover, we show that periodic and chaotic attractors can be avoided by controlling the gain such that the divergence of the error signal velocity is close to zero but positive, which ensures that phase-space volumes in the auxiliary subspace never contract and, in turn, forbids the creation of attractors. The error variables play the role of a non-Markovian reservoir that guarantees positive entropy production in the system despite dissipation. We have performed numerical simulations demonstrating that this effect generalizes well by finding the ground states of spin glasses and solving Ising problems from standard benchmark sets against state-of-the-art heuristics, and suggest that orders of magnitude decrease of the time-to-solution can be obtained in the case of an implementation on an analog Ising machine.

We consider a network of analog bistable units x_i , $i \in \{1, \dots, N\}$, for which time-evolution is given as follows:

$$\frac{dx_i}{dt} = f_i = \phi(x_i) + e_i I_i, \quad (1)$$

where $\phi(x_i)$ represents the time-evolution of isolated units with $\phi(x_i) = -\frac{\partial V_b}{\partial x_i}$ and $V_b = -(-1+p)\frac{x_i^2}{2} + \frac{x_i^4}{4}$ the paradigmatic bistable potential. The function $\phi(x_i)$ can be written as $\phi(x_i) = -x_i + px_i - x_i^3$ in which the first, second, and third terms can be interpreted as the loss, linear gain with rate p , and nonlinear saturation. These dynamics can be used to describe various systems such as soft spins [20], open-dissipative quantum systems such as degenerate parametric oscillators [26, 34–36], and weakly coupled neural networks near pitchfork bifurcations [37]. For solving the combinatorial optimization problem that is defined by the cost function $V(\boldsymbol{\sigma})$, the coupling I_i is chosen as $I_i \propto -\frac{\partial V(\boldsymbol{\sigma})}{\partial x_i}$, i.e., I_i is the gradient of the potential V . In particular, Ising problems with the cost function $V(\boldsymbol{\sigma}) = \mathcal{H} = -\frac{1}{2} \sum_{ij} \omega_{ij} \sigma_i \sigma_j$ can be solved using the injection term $I_i = \epsilon \sum_j \omega_{ij} x_j$ with ϵ the coupling strength ($\epsilon > 0$, $\omega_{ii} = 0$, and $\omega_{ij} = \omega_{ji}$ for $j \neq i$).

In order to destabilize states that correspond to local minima of the Ising Hamiltonian, we propose to control the target amplitude, noted a with $a > 0$, of the variables x_i independently of the linear gain p by considering the following error signal induced by amplitude heterogeneity:

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

where e_i and β are the error variables and the rate of change of error variables, respectively, with $e_i > 0$ and $\beta > 0$ [43].

First, we examine the existence of fixed point attractors. The fixed points of the dynamical system described by eqs. (1) and (2) are given as follows:

$$\begin{cases} \frac{de_i}{dt} = 0, \\ \frac{dx_i}{dt} = 0, \end{cases} \implies \begin{cases} x_i^2 = a, \\ e_i = \frac{1-p+a}{\epsilon h_i \sigma_i}, \end{cases} \quad \forall i, \quad (3)$$

where h_i are the elements of the vector $\mathbf{h} = \Omega \boldsymbol{\sigma}$ and Ω the matrix of couplings with $\Omega = \{\omega_{ij}\}_{ij}$. The configuration $\boldsymbol{\sigma}$ corresponds to the sign of the state \mathbf{x} at the fixed point. Note that the analog states x_i are exactly binary at the steady state with $x_i = \sigma_i \sqrt{a}$. Moreover, the internal fields h_i are such that $h_i \sigma_i > 0$, $\forall i$, at equilibrium when $p < 1$ because $e_i > 0$, $\forall i$. Thus, all fixed points of the analog system correspond to local minima of the binary spin system at $T = 0$.

The linear stability of these fixed points can be examined by analyzing the following Jacobian matrix:

$$J = \begin{bmatrix} J_{xx} & J_{xe} \\ J_{ex} & J_{ee} \end{bmatrix}, \quad (4)$$

with $J_{xx} = (-1 + p - 3a)I + \epsilon D[\mathbf{e}]\Omega$, $J_{xe} = \epsilon \sqrt{a} D[\mathbf{h}]$, $J_{ex} = -2\beta \sqrt{a} D[\boldsymbol{\sigma} \cdot \mathbf{e}]$, and $J_{ee} = 0$. Moreover, $D[\mathbf{X}]$ is the diagonal matrix with elements $D_{ii}[\mathbf{X}] = X_i$ and $D_{ij}[\mathbf{X}] = 0$ for $i \neq j$, where X_i expresses the components of a vector such as e_i or h_i . Note that $J_{ex} J_{xe} = bI$, with $b = -2\beta a(1 - p + a)$. The eigenvalues λ_j^\pm of the Jacobian matrix can be explicitly calculated by considering its characteristic polynomial, and are given as follows:

$$\lambda_j^\pm = \begin{cases} \frac{1}{2}(-2a + (1 - p + a)\mu_j \pm \sqrt{\Delta_j}) & \text{if } \Delta_j > 0, \\ \frac{1}{2}(-2a + (1 - p + a)\mu_j \pm i\sqrt{-\Delta_j}) & \text{otherwise,} \end{cases} \quad (5)$$

with $\Delta_j = (-2a + (1 - p + a)\mu_j)^2 + 4b$, $i^2 = -1$, and μ_j the j^{th} eigenvalues of the matrix $\tilde{\Omega} = D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]\Omega - I$ with $(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1} = \{(h_i \sigma_i)^{-1}\}_i$. Because the vector $\boldsymbol{\sigma} \cdot \mathbf{h}$ has positive components at local minima, the eigenvalues of $D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]\Omega$ are the same as the ones of $D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]^{\frac{1}{2}} \Omega D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]^{\frac{1}{2}}$ (Sylvester's law of inertia), which is a symmetric real matrix. Thus, the eigenvalues μ_j are always real.

The $2N$ eigenvalues of the system are always pairs λ_j^+ and λ_j^- with $j \in \{1, \dots, N\}$. Each pair become the same real value when $\Delta_j = 0$, i.e., under the condition given as follows:

$$\theta = G_\pm(\beta, \mu_j), \quad (6)$$

with $\theta = \frac{a}{1-p}$ and $G_\pm(\beta, \mu) = \frac{4\beta - \mu(\mu-2) \pm 4\sqrt{\beta(\beta+\mu)}}{(\mu-2)^2 - 8\beta}$.

The stability of fixed points depends on the sign of the real part of the eigenvalues of the Jacobian matrix. It

can be shown that $\text{Re}[\lambda_j^+] = 0$ in the following set (note that $\text{Re}[\lambda_j^+] \geq \text{Re}[\lambda_j^-]$, $\forall j$):

$$\text{Re}[\lambda_j^+] = 0 \Leftrightarrow \begin{cases} \theta = 0 \text{ or } \theta = -1, & \text{if } \Delta_j > 0, \\ \theta = F(\mu_j(\boldsymbol{\sigma})), & \text{if } \Delta_j < 0, \end{cases} \quad (7)$$

with $F(\mu) = -\frac{\mu}{(\mu-2)}$. The parameter θ can be interpreted as the ratio between the target amplitude a and the effective loss (the loss minus the linear gain p) of the analog system. The stability of fixed points is illustrated in Fig. 1 in the space $\{\theta = \frac{a}{1-p}, \mu_j(\boldsymbol{\sigma})\}$ where θ depend on the controllable parameters of the analog system whereas $\mu_j(\boldsymbol{\sigma})$ are determined by the spin configuration $\boldsymbol{\sigma}$ and couplings Ω . We denote $\mu_0(\boldsymbol{\sigma})$ the maximum eigenvalue of $\tilde{\Omega}$ calculated at the fixed point \boldsymbol{x} corresponding to the configuration $\boldsymbol{\sigma}$. The state \boldsymbol{x} becomes unstable at a supercritical Andronov-Hopf bifurcation when the real part of the dominant eigenvalue of the Jacobian matrix J , noted $\lambda_0(\boldsymbol{\sigma})$, becomes positive, i.e., $\mu_0(\boldsymbol{\sigma}) > F^{-1}(\theta)$ for $p < 1$ (see eq. 7).

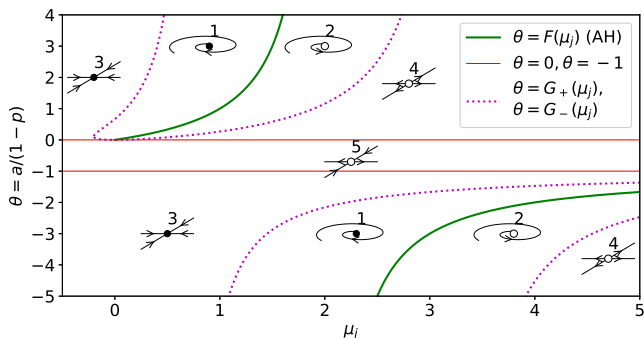


FIG. 1: Bifurcation diagram in the space $\{\theta = \frac{a}{1-p}, \mu_j(\boldsymbol{\sigma})\}$ at configuration $\boldsymbol{\sigma}$. Thick lines correspond to the sets where $\text{Re}[\lambda_j^\pm] = 0$, i.e., $\theta = 0, -1$, or $F(\mu_j)$ with $\Delta_j < 0$, and dotted lines to the sets where $\theta = G_\pm(\beta, \mu_j)$. In the regions marked 1 and 2, the pair of eigenvalues λ_j^+ and λ_j^- are complex conjugate with negative real parts and positive real parts, respectively, whereas in regions 3, 4, and 5, they are both real with both, none, and one of their real parts being negative, respectively. $\beta = 0.2$. AH: Andronov-Hopf bifurcation.

Next, we consider the existence of limit cycles and chaotic attractors. A necessary condition for the existence of these attractors is that the divergence of the velocity, or equivalently, the rate of change of phase-space volumes, is negative at the proximity of the trapping sets[21]. In order to estimate the divergence at configuration $\boldsymbol{\sigma}$, we utilize the fact that the fast subsystem on \boldsymbol{x} converges to the slow manifold characterized by $\frac{dx_i}{dt} \approx 0$, $\forall i$, in the limit $\beta \ll 1$. Moreover, the analog states become approximately binary in the limit of small coupling strength ($\epsilon \ll 1$) with x_i^2 given as $x_i^2 = (x_i^{(0)})^2 + 2\epsilon x_i^{(0)} x_i^{(1)} + \mathcal{O}(\epsilon^2)$ with $(x_i^{(0)})^2 = -1+p$ and

$2\epsilon x_i^{(0)} x_i^{(1)} = \epsilon e^{(0)} h_i \sigma_i$ for $p > 1$ (see supplementary material). The term $e^{(0)}(t)$ is the zeroth order approximation of $e_i(t)$ with $e_i(t) = e^{(0)}(t) + \epsilon e_i^{(1)}(t) + \mathcal{O}(\epsilon^2)$, which is valid for $e^{(0)}(t)\epsilon \ll -1+p$, and $e^{(0)}(t) = e^{-\beta \int (-1+p-a) dt'}$ for $e_i(0) = 1$, $\forall i$. Moreover, the first order term $e_i^{(1)}$ is given as $e_i^{(1)} = -\beta \epsilon e^{(0)} [\int_0^t h_i(t') \sigma_i(t') dt']$ and depends on the previous history of states visited. Although this Taylor approximation is valid only for $p > 1$, numerical simulations show that it is still accurate by continuity for $p < 1$ and $|1-p| \ll 1$ for a finite value of ϵ . Then, the divergence of the vector field \boldsymbol{g} , defined as $\text{div } \boldsymbol{g} = \sum_i \frac{\partial g_i}{\partial e_i}$ can thus be approximated as follows (see supplementary material):

$$\text{div } \boldsymbol{g} = \beta [N(1-p+a) + 2\epsilon e^{(0)}(t) \mathcal{H}(t) + \mathcal{O}(\frac{(e^{(0)})^2 \epsilon^2}{p-1})]. \quad (8)$$

In order to prevent the system from being trapped in limit cycles and chaotic attractors, we propose to set the divergence of velocity in the auxiliary subspace on \boldsymbol{e} to be always positive along the trajectories of the system. This can be achieved by modulating the target amplitude a as follows:

$$a = \alpha + \epsilon < e_i(t_c) h_i(t) \sigma_i(t) >, \quad (9)$$

where α is the target amplitude baseline and t_c , with $t_c < t$, is the time of the last change of configuration $\boldsymbol{\sigma}$, i.e., the time at which one of the x_i has changed its sign. Moreover, $\langle X_i \rangle$ denotes the ensemble average of X_i with $\langle X_i \rangle = \frac{1}{N} \sum_i X_i$. The control scheme described in eq. (9) implies, using eq. (8), that the divergence is approximately given as $\text{div } \boldsymbol{g} \approx \kappa$ with $\kappa = N\beta(1-p+a)$ as long as the configuration $\boldsymbol{\sigma}(t)$ switches rapidly before the error variables \boldsymbol{e} blow up. The parameter κ can be set to $\kappa > 0$, $\forall t$, by modulating a and p such that the two conditions given in eqs. (7) and (9) are respected, which ensures that there are no stable fixed points at suboptimal configurations and that the divergence in the auxiliary subspace is always positive, respectively. Note that, when β is very small, the fast variables \boldsymbol{x} are slaved to the slow error signals \boldsymbol{e} . Thus, positive divergence in the auxiliary subspace should imply that the whole system cannot be trapped in a region of negative divergence at the proximity of an attractor. In order to verify this claim, we have simulated the proposed system using the weights ω_{ij} from 5000, 5000, 3000, and 2000 instances of Sherrington-Kirkpatrick (SK) spin-glass problems (an NP-hard problem) of size $N = 80$, $N = 100$, $N = 150$, and $N = 200$ spins, respectively[44]. For an instance of size $N = 100$, Figure 2 (a) shows that the system exhibits constant phase-space volume expansion or contraction in the auxiliary subspace \boldsymbol{e} when $\kappa > 0$ and $\kappa < 0$, respectively. Moreover, the system does not become trapped in

an undesirable attractor and the ground-state configuration of the Ising Hamiltonian is visited with probability $P_0 = 1$ from any random initial condition when $\kappa \gg 0$ (see Figs. 2 (b) and 3). As the constant divergence in the auxiliary subspace is increased, the system always visits a ground-state configuration, but the time-to-solution increases, which is likely a consequence of the fact that the dynamics becomes more complex[32]. Therefore, the optimal regime for finding lower energy states is for positive but small divergence of the auxiliary subspace velocity. In this case, the probability of finding the ground state from a randomly chosen set of initial conditions is superior to 99.9 percent even for $N = 200$ (see Fig. 3). Thus, numerical simulations suggest that the system does not become bounded within a subspace. Further analysis is necessary, however, for clarifying the conditions, such as the maximal value of β , under which this is the case. The proposed dynamics finds ground-state configurations more reliably than the state-of-the-art algorithm, called BLS[38], for harder problem instances. Figure 3 shows that the proportion of unsolved instances $p_0(t)$ is well fitted by a power law such that $p_0 \sim t^{-\gamma(N)}$ for $t \gg 0$. The positive divergence implies that error variables eventually become very large. However, the parameter κ can always be chosen sufficiently small such that a ground-state configuration is visited before the error variables blow up, which is confirmed by our numerical simulations. In practice, the error variables are rescaled such that $e_i(t) = \frac{e_i(t^-)}{\langle e_i(t^-) \rangle}$, with $t^- < t$, whenever the mean error signal $\langle e_i(t^-) \rangle$ is superior to the threshold, noted Γ , in order to insure the stability of numerical simulations. Note that the proposed scheme does not guarantee that the best solution found after t is the optimal one (but see [39] for a heuristic approach to predicting the likelihood of optimality for MaxSAT problems).

The increase of dynamical complexity that results from the addition of the amplitude-heterogeneity error correction scheme can be interpreted in terms of entropy. With the adiabatic elimination of the fast variables \mathbf{x} , the sign of the entropy production rate $\frac{dS}{dt}$ [40, 41], defined as $\frac{dS}{dt} = \int \rho(\mathbf{e}) \text{div } \mathbf{g} \, d\mathbf{e}$ with S the Gibbs entropy $S = - \int \rho(\mathbf{e}) \ln \rho(\mathbf{e}) \, d\mathbf{e}$ and $\rho(\mathbf{e})$ the probability density of states, is equal to the sign of κ when using the control of divergence proposed in eq. (9). Thus, the coupling to the auxiliary subspace implies that the system has always a positive production of entropy despite dissipation when $\kappa > 0$.

For further comparison against state-of-the-art algorithms, the proposed scheme is adapted for solving MAX-CUT problems of the G-set [38], which is a reference benchmark used in the community of combinatorial optimization. When simulated on a desktop computer, the scheme can achieve performance in terms of solution quality and time-to-solution that is qualitatively similar[45] to that of BLS[38], which itself outperforms

other recent heuristics on MAX-CUT problems. For several instances, the proposed scheme finds solutions of better quality than previously known from [38] (see the supplementary material for the details of the benchmark on the G-set). Importantly, the advantage of the proposed scheme relies on the ability to implement it on hardware dedicated to solving Ising problems in the analog domain [31]. If we assume that the presented scheme is realized in hardware that has the same experimental parameters as existing coherent-Ising-machine implementations [46], then we predict that such a hardware solver will feature a time-to-solution for solving Ising problems that is a factor of 100-1000 smaller than that of the state-of-the-art classical heuristic algorithms running on a conventional desktop computer, even when the problem size is not large, i.e., $N \ll 1000$.

This research was supported by ImPACT Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan). We thank Ryan Hamerly and the anonymous reviewers for comments on the manuscript. We were informed of related work about amplitude heterogeneity error correction (see [42]) during the preparation of this manuscript.

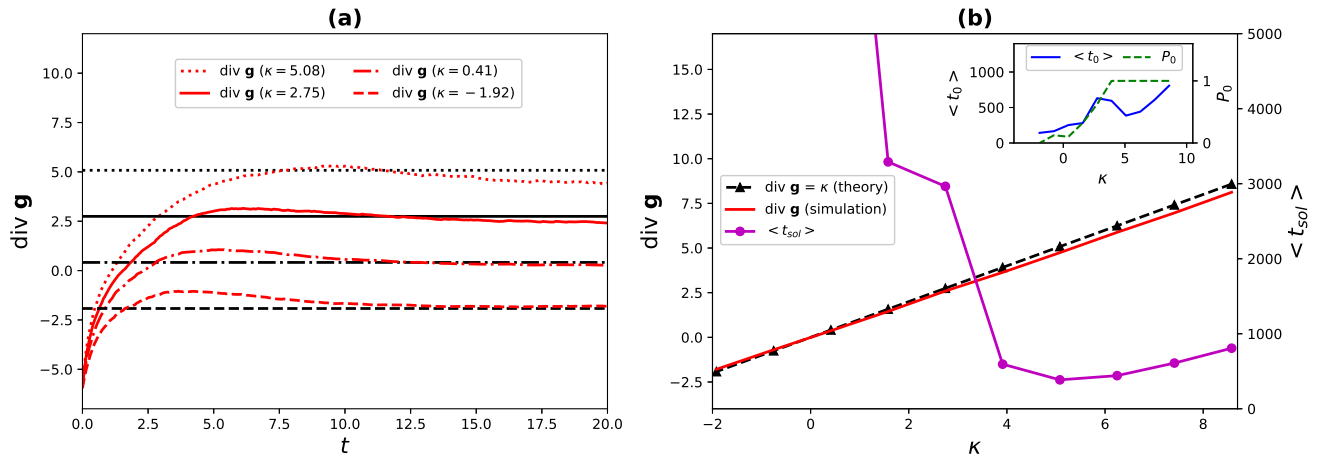


FIG. 2: Control of the divergence in the auxiliary subspace. (a) Divergence $\text{div } \mathbf{g}$ in the auxiliary subspace vs. normalized simulation time t . After some transient behavior, the divergence is constant and approximately described by its analytical value $\text{div } \mathbf{g} \approx \kappa$. (b) Divergence $\text{div } \mathbf{g}$ and average time $\langle t_{sol} \rangle$ for finding the ground state of the Ising Hamiltonian with a probability superior to 99% with $t_{sol} = \langle t_0 \rangle \frac{\log(0.1)}{\log(1-P_0)}$ for $P_0 < 0.99$, and $t_{sol} = \langle t_0 \rangle$ otherwise, with P_0 and $\langle t_0 \rangle$ (see inset) the probability of finding the ground state during a single run and averaged time-to-solution for successful runs, respectively.

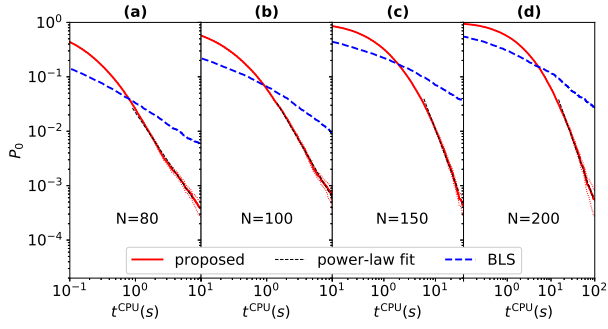


FIG. 3: Proportion $p_0(t)$ of instances unsolved, i.e., for which a ground-state configuration has not been found, after the CPU time t in seconds of the SK problems of size $N = 80$, $N = 100$, $N = 150$, and $N = 200$ shown in (a), (b), (c), and (d), respectively, for the proposed scheme and the breakout local search (BLS) algorithm[38] in the logarithmic scale. The 95% confidence interval shown by dotted lines is calculated using 20 runs per instance.

-
- [1] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).
- [2] W. Wang, J. Machta, and H. G. Katzgraber, *Phys. Rev. E* **92**, 013303 (2015).
- [3] G. Zaránd, F. Pázmándi, K. F. Pál, and G. T. Zimányi, *Phys. Rev. Lett.* **89**, 150201 (2002).
- [4] S. Boettcher and A. G. Percus, *Phys. Rev. Lett.* **86**, 5211 (2001).
- [5] B. Sutton, K. Y. Camsari, B. Behin-Aein, and S. Datta, *Sci. Rep.* **7**, 44370 (2017).
- [6] T. Kadowaki and H. Nishimori, *Phys. Rev. E* **58**, 5355 (1998).
- [7] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, arXiv preprint quant-ph/0001106 (2000).
- [8] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, *Phys. Rev. A* **88**, 063853 (2013).
- [9] T. Inagaki, K. Inaba, R. Hamerly, K. Inoue, Y. Yamamoto, and H. Takesue, *Nature Photonics* **10**, 415 (2016).
- [10] A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto, *Nature Photonics* **8**, 937 (2014).
- [11] 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, *Science* **354**, 614 (2016).
- [12] W. R. Clements, J. J. Renema, Y. H. Wen, H. M. Chrzanowski, W. S. Kolthammer, and I. A. Walmsley, *Phys. Rev. A* **96**, 043850 (2017).
- [13] T. Wang and J. Roychowdhury, arXiv preprint arXiv:1709.08102 (2017).
- [14] G. Csaba, A. Raychowdhury, S. Datta, and W. Porod, in *2018 IEEE International Symposium on Circuits and Systems* (2018) pp. 1–5.
- [15] J. J. Hopfield and D. W. Tank, *Biol. Cybern.* **52**, 141 (1985).
- [16] A. Percus, G. Istrate, and C. Moore, *Computational complexity and statistical physics* (OUP USA, 2006).
- [17] M. Mezard and A. Montanari, *Information, physics, and computation* (Oxford University Press, 2009).
- [18] A. Lucas, *Front. Phys.* **2**, 5 (2014).
- [19] S. Geman and D. Geman, *IEEE Trans. Pattern Anal. Mach. Intell.* , 721 (1984).
- [20] H. Sompolinsky and A. Zippelius, *Phys. Rev. Lett.* **47**, 359 (1981).
- [21] S. Strogatz, *Nonlinear Dynamics And Chaos*, Studies in nonlinearity (Sarat Book House, 2007).
- [22] T. Fukai and M. Shiino, *Phys. Rev. A* **42**, 7459 (1990).
- [23] F. R. Waugh, C. M. Marcus, and R. M. Westervelt, *Phys. Rev. Lett.* **64**, 1986 (1990).
- [24] G. Bilbro, R. Mann, T. K. Miller, W. E. Snyder, D. E. Van den Bout, and M. White, in *Advances in neural information processing systems* (1989) pp. 91–98.
- [25] C. Peterson and B. Söderberg, *Int. J. Neural Syst.* **1**, 3 (1989).
- [26] T. Leleu, Y. Yamamoto, S. Utsunomiya, and K. Aihara, *Phys. Rev. E* **95**, 022118 (2017).
- [27] G. Wilson and G. Pawley, *Biol. Cybern.* **58**, 63 (1988).
- [28] E. B. Baum, *Phys. Rev. Lett.* **57**, 2764 (1986).
- [29] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, arXiv preprint arXiv:1705.06963 (2017).
- [30] N. Surlas, *Phys. Rev. Lett.* **94**, 070601 (2005).
- [31] Y. Yamamoto, K. Aihara, T. Leleu, K.-i. Kawarabayashi, S. Kako, M. Fejer, K. Inoue, and H. Takesue, *npj Quantum Inf.* **3**, 49 (2017).
- [32] M. Ercsey-Ravasz and Z. Toroczkai, *Nat. Phys.* **7**, 966 (2011).
- [33] A. Franz, K. H. Hoffmann, and P. Salamon, *Phys. Rev. Lett.* **86**, 5219 (2001).
- [34] T. Shoji, K. Aihara, and Y. Yamamoto, *Phys. Rev. A* **96**, 053833 (2017).
- [35] A. Yamamura, K. Aihara, and Y. Yamamoto, *Phys. Rev. A* **96**, 053834 (2017).
- [36] W. Wustmann and V. Shumeiko, *Phys. Rev. B* **87**, 184501 (2013).
- [37] F. C. Hoppensteadt and E. M. Izhikevich, *Weakly connected neural networks*, Vol. 126 (Springer Science & Business Media, 2012).
- [38] U. Benlic and J.-K. Hao, *Eng. Appl. Artif. Intell.* **26**, 1162 (2013).
- [39] B. Molnár, F. Molnár, M. Varga, Z. Toroczkai, and M. Ercsey-Ravasz, *Nature communications* **9**, 4864 (2018).
- [40] L. Andrey, *Physics Letters A* **111**, 45 (1985).
- [41] D. Ruelle, “What physical quantities make sense in non-equilibrium statistical mechanics,” in *Boltzmann’s Legacy* (European Mathematical Society, 2008) pp. 89–97.
- [42] K. P. Kalinin and N. G. Berloff, arXiv preprint arXiv:1807.00699 (2018).
- [43] Note that exponentially increasing error variables have been previously proposed in [32] for solving constraint satisfaction problems.
- [44] See the supplementary material for more details about the simulation scheme.
- [45] Note that computer simulations are obtained using CPUs with similar clock rates (Xeon E5440 2.83 GHz and Xeon E5-2680 2.70 GHz for BLS, a Xeon X5690 3.47 GHz for the proposed scheme), but using different programming languages (C++ for BLS and Matlab for the proposed scheme). Therefore, the averaged time-to-solutions are presented as indicative purposes for comparison of their orders of magnitudes.
- [46] Estimation of the time-to-solution $\langle t^{\text{CIM}} \rangle$ in the case of an implementation on the coherent Ising machine is calculated using the normalized time-to-solution $\langle t \rangle$ obtained from numerical simulations of the proposed scheme and the value of the time normalization constant (the cavity photon lifetime) that has been obtained experimentally [9–11], i.e., $\langle t^{\text{CIM}} \rangle \approx 10^{-5} \langle t \rangle$. This approximation is justified by the fact that the computational effort related to the amplitude heterogeneity error correction scheme (see eqs. (2) and (9)) is negligible compared to the one related to the coherent Ising machine (see eq. (1)).

Supplementary material

Approximation of the state-space velocity divergence in the limit $\beta \ll 1$ and $\epsilon \ll 1$

The time-evolution of the proposed system, described using eqs. (1) and (2), can be rewritten as follows:

$$\frac{dx_i}{dt} = (-1 + p)x_i - x_i^3 + \epsilon e_i \sum_j \omega_{ij} x_j, \quad (\text{s1})$$

with $e_i(t)$ given as $e_i(t) = e_i(0)e^{-\beta \int_0^t (x_i(t')^2 - a(t')) dt'}$ after integrating eq. (2). Then, the state $x_i(t)$ of the system can be approximated at the proximity of the steady-state $\frac{dx_i}{dt} = 0$ using the Taylor approximation $x_i = x_i^{(0)} + \epsilon x_i^{(1)} + \mathcal{O}(\epsilon^2)$ when $p > 1$. At the zeroth order, the state can be described as follows:

$$(x_i^{(0)})^2 = (-1 + p), \quad (\text{s2})$$

$$e_i^{(0)} = e_i(0)e^{-\beta \int_0^t (-1+p(t')-a(t')) dt'}. \quad (\text{s3})$$

In the following, $e_i^{(0)}$ is simply written $e^{(0)}$ because it is independent of the index i when $e_i(0) = 1, \forall i$. At the first order, the state is described as follows:

$$x_i^{(1)} = \frac{e^{(0)} \sum_j \omega_{ij} x_j^{(0)}}{2(x_i^{(0)})^2}, \quad (\text{s4})$$

$$e_i^{(1)} = -\beta e^{(0)} \int_0^t 2x_i^{(0)}(t')x_i^{(1)}(t') dt'. \quad (\text{s5})$$

Moreover, the divergence of state-space velocity \mathbf{f} and \mathbf{g} are given as follows (see eqs. (1) and (2)):

$$\text{div } \mathbf{f} = \sum_i \frac{\partial f_i}{\partial x_i} = N(-1 + p) - 3 \sum_i x_i^2, \quad (\text{s6})$$

$$\text{div } \mathbf{g} = \sum_i \frac{\partial g_i}{\partial e_i} = -\beta \left(\sum_i x_i^2 - aN \right). \quad (\text{s7})$$

Using the Taylor approximation of x_i^2 , equal to $x_i^2 = (x_i^{(0)})^2 + 2\epsilon x_i^{(0)} x_i^{(1)} + \mathcal{O}(\epsilon^2)$, the divergence is given as follows:

$$\text{div } \mathbf{f} = -2N(-1 + p) + 6\epsilon e^{(0)}(t)\mathcal{H} + \mathcal{O}\left(\frac{(e^{(0)})^2 \epsilon^2}{p-1}\right), \quad (\text{s8})$$

$$\text{div } \mathbf{g} = \beta [N(1 - p + a) + 2\epsilon e^{(0)}(t)\mathcal{H} + \mathcal{O}\left(\frac{(e^{(0)})^2 \epsilon^2}{p-1}\right)]. \quad (\text{s9})$$

Although the Taylor approximations given in eqs. (s2) to (s9) are only valid for $p > 1$, numerical simulations show that these are still accurate by continuity for $p < 1$ and $|1 - p| \ll 1$ for a finite value of ϵ (see supplementary Fig. s1 in the case $e_i = 1, \forall i$).

Numerical simulation scheme and parameter values

In the following, the scheme used for numerical simulations is detailed for the sake of reproducibility. The ODEs describing the time-evolution of the system are approximated using a Euler approximation given as follows (see eqs. (1) and (2)):

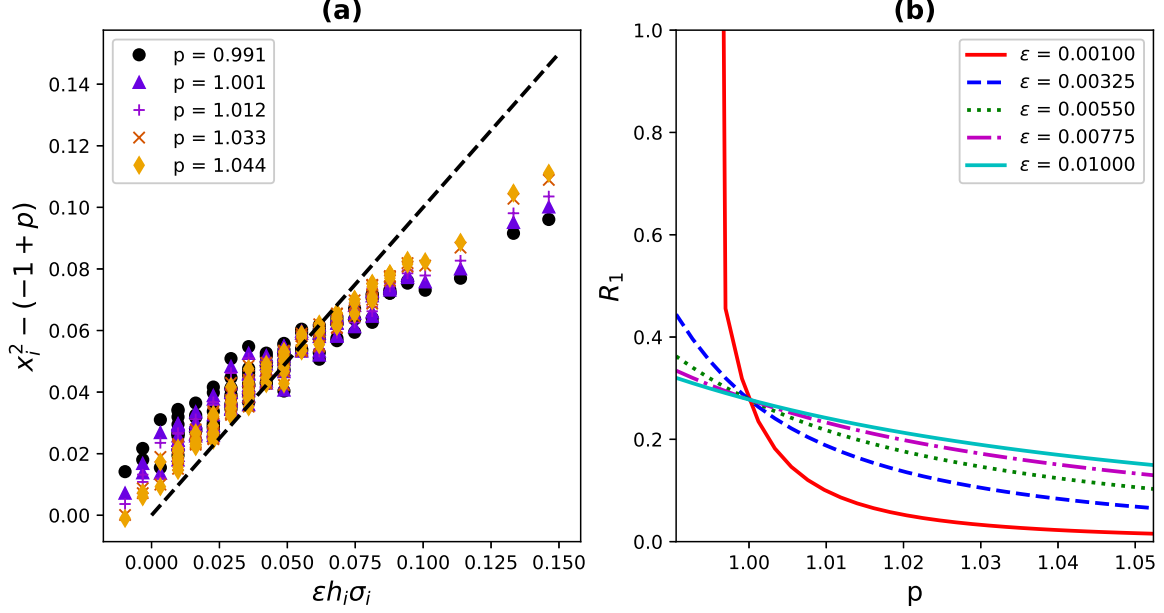


FIG. s1: Accuracy of the Taylor approximation of x_i^2 with $x_i^2 = (x_i^{(0)})^2 + 2x_i^{(0)}x_i^{(1)}\epsilon + \mathcal{O}(\epsilon^2)$ when $e_i(0) = 1, \forall i$. (a) First order term $x_i^2 - (x_i^{(0)})^2 = x_i^2 - (-1+p)$ vs. $2x_i^{(0)}x_i^{(1)} = \epsilon h_i \sigma_i$ obtained using numerical simulations of a Sherrington-Kirkpatrick spin glass problem of size $N = 100$. Each marker corresponds to an index i with $i \in \{1, \dots, N\}$. (b) Normalized averaged remainder of the Taylor approximation R_1 with $R_1 = \frac{1}{N} \sum_i \frac{|x_i^2 - (-1+p) - \epsilon h_i \sigma_i|}{x_i^2}$ vs. the parameter value p . The remainder of the Taylor approximation is smoothly varying at the proximity of $p = 1$ for a finite value of ϵ .

$$x_i(t + \Delta t) = x_i(t) + \Delta x_i(t) \Delta t, \forall i, \quad (\text{s10})$$

$$e_i(t + \Delta t) = \begin{cases} e_i(t) + \Delta e_i(t) \Delta t, \forall i, & \text{if } \langle e_i(t) + \Delta e_i(t) \Delta t \rangle < \Gamma, \\ \frac{e_i(t) + \Delta e_i(t) \Delta t}{\langle e_i(t) + \Delta e_i(t) \Delta t \rangle} & \text{otherwise.} \end{cases} \quad (\text{s11})$$

with $\langle X_i \rangle = \frac{1}{N} \sum_i X_i$ and Γ is the maximum mean error. Moreover, $\Delta x_i(t)$ and $\Delta e_i(t)$ are given as follows:

$$\Delta x_i(t) = (-1 + p(t))x_i(t) - x_i(t)^3 + \epsilon e_i(t) \sum_j \omega_{ij} x_j(t), \quad (\text{s12})$$

$$\Delta e_i(t) = -\beta(x_i(t)^2 - a(t))e_i(t), \quad (\text{s13})$$

If the configuration $\sigma(t)$, defined as $\sigma_i(t) = \frac{x_i(t)}{|x_i(t)|}$, $\forall i$, is not a stable local minimum, i.e., there exists an index j such that $\sigma_j(t)h_j(t) < 0$ with $\mathbf{h}(t) = \Omega \sigma(t)$, then the target amplitude $a(t)$ and linear gain $p(t)$ are given as follows (see eq. (9)):

$$a(t) = \alpha + \epsilon \langle e_i(t_c) h_i(t) \sigma_i(t) \rangle, \quad (\text{s14})$$

$$p(t) = \pi, \quad (\text{s15})$$

where π is the linear gain baseline with $\pi < 1$; α , the target amplitude baseline; and t_c , with $t_c < t$, is the time of the last change of configuration σ , i.e., the time at which one of the $x_i(t)$ has changed its sign.

If the configuration $\sigma(t)$ is a local minimum, i.e., $\sigma_i(t)h_i(t) > 0, \forall i$, and $\theta(t) > \eta F(\mu_0(\sigma(t)))$ (with $\mu_0(\sigma(t)) < 2$, $\theta(t)$ given as $\theta(t) = \frac{a(t)}{1-p(t)}$, and $0 < \eta < 1$) then $a(t)$ and $p(t)$ are given as follows (see eq. (7)):

Symbol	meaning	value
Δt	Euler time-step	0.05
ϵ	coupling strength	0.07
α	target amplitude baseline	0.2
π	linear gain baseline	0.2
β	relative time-scale of error variables time-evolution	0.05
η	destabilization strength from local minima	0.5
Γ	maximal mean error	5.0

TABLE I: Parameters description in the case of the Sherrington-Kirkpatrick spin glass problems.

$$a(t) = \eta F(\mu_0(\boldsymbol{\sigma}(t)))(1 - p(t)), \quad (\text{s16})$$

$$p(t) = \pi + a(t) - a(t - 1), \quad (\text{s17})$$

with $F(X) = -\frac{X}{X-2}$. Note that $\mu_0(\boldsymbol{\sigma}(t))$ is the dominant eigenvalue of the matrix $\tilde{\Omega} = D[(\boldsymbol{\sigma}(t) \cdot \mathbf{h}(t))^{-1}]\Omega - I$.

Lastly, the parameters of the system are described as given in Table 1 and the proposed scheme is simulated using Python on a Xeon E5-2680 2.70 GHz when solving the Sherrington-Kirkpatrick spin glass problems.

The time-evolution of the system is given as shown in the Supplementary Figure s2 when the parameter values from Table I are used.

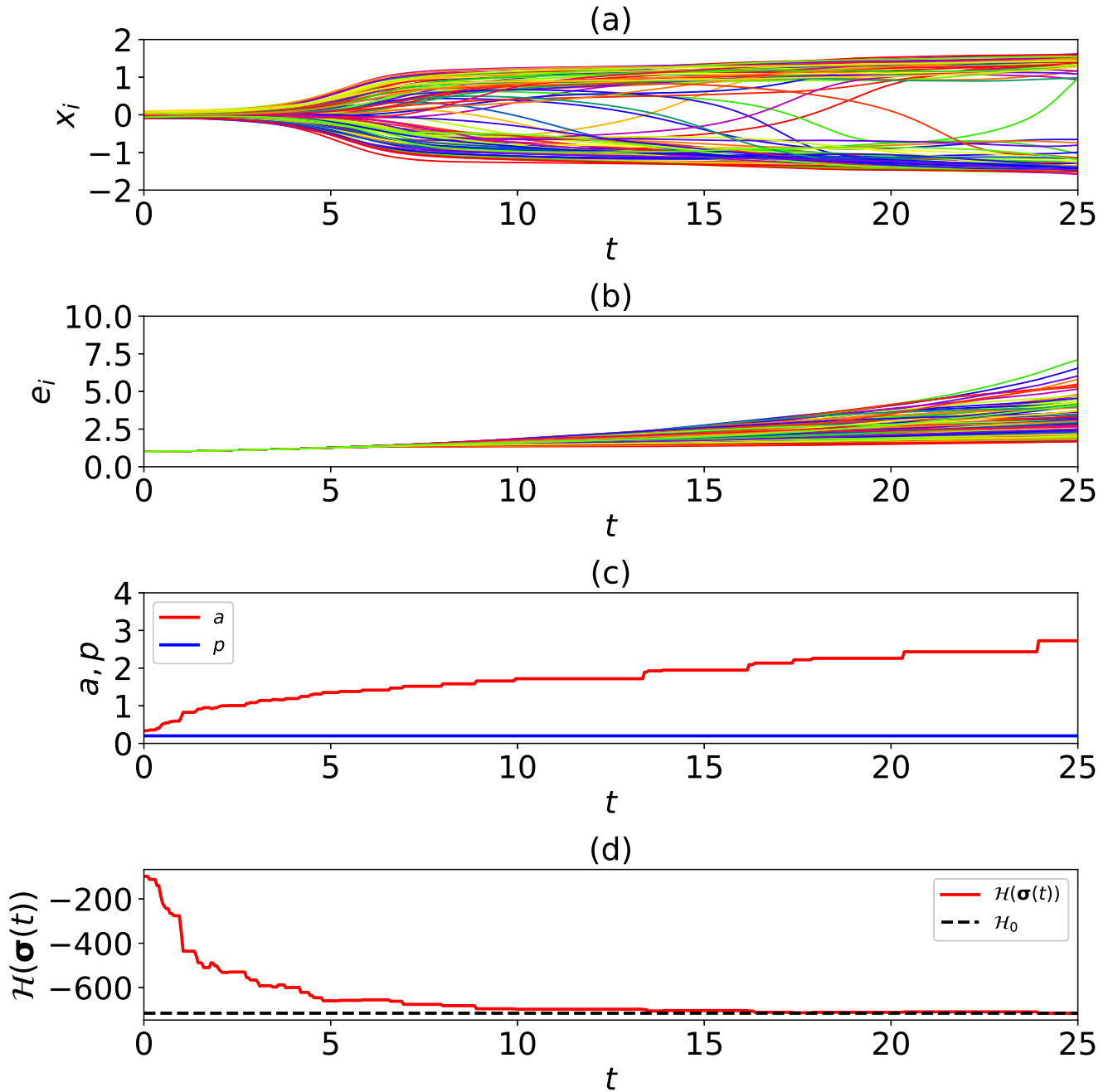


FIG. s2: Time-evolution of the states x_i , error variables e_i , target amplitude a , linear gain p , and Ining Hamiltonian \mathcal{H} (cost function) are shown in (a), (b), (c), and (d), respectively, when solving a Sherrington-Kirkpatrick spin glass problem of size $N = 100$. In (a) and (b), the different colors correspond different indices of i with $i \in \{1, \dots, N\}$. At $t = 25$, a ground-state configuration with energy $\mathcal{H} = \mathcal{H}_0$ has been found.

Symbol	meaning	value
ϵ	coupling strength	$3 / \langle \sum_i \omega_{ij} \rangle$
α	target amplitude baseline	1
π	linear gain baseline	$1 - \epsilon g(\nu_1, \nu_2, \nu_3, \nu_4)$
ρ	amplitude and gain variation	1
δ	sensitivity to energy variations	7
γ	rate of increase of β	0.065 / N
τ	max. time w/o energy change	7 N

TABLE II: Parameters description in the case of MAX-CUT on the G-set.

Benchmark results on the G-set

In the case of the benchmark against BLS on the G-set, the proposed scheme is adapted as described in this section. First, the ODEs describing the system are simulated using a Euler approximation with time-step $\Delta t = 0.05$ (see eqs. [s10] and [s11]). Contrarily to the case of the Sherrington-Kirkpatrick problems, the error variables are not rescaled, i.e., $\Gamma = +\infty$. Note that the objective function in the case of the MAX-CUT problems is given as follows:

$$\mathcal{C} = -\frac{1}{2}(\mathcal{H} + \frac{1}{2} \sum_{ij} \omega_{ij}), \quad (\text{s17})$$

where \mathcal{C} is the value of the cut and \mathcal{H} is the Ising Hamiltonian given as follows:

$$\mathcal{H} = -\frac{1}{2} \sum_{ij} \omega_{ij} \sigma_i \sigma_j. \quad (\text{s18})$$

For MAX-CUT, the modulation of θ is achieved by modulating both the target amplitude a and the pump rate p . First, the target amplitude is modulated as follows:

$$a(t) = \alpha - \rho \phi(\delta \Delta \mathcal{C}(t)), \quad (\text{s19})$$

where $\Delta \mathcal{C}(t) = \mathcal{C}_{\text{opt}} - \mathcal{C}(t)$ with $\mathcal{C}_{\text{opt}} = -\frac{1}{2}(\mathcal{H}_{\text{opt}} + \frac{1}{2} \sum_{ij} \omega_{ij})$. Moreover, $\mathcal{C}(t)$ is the cut of the currently visited configuration and \mathcal{H}_{opt} the target energy. In practice, we set \mathcal{H}_{opt} to the lowest energy visited, i.e., $\mathcal{H}_{\text{opt}}(t) = \min_{t' \leq t} \mathcal{H}(t')$. Lastly, the function ϕ is the tangent hyperbolic function. Next, the pump rate is modulated as follows:

$$p(t) = \pi + \rho \phi(\delta \Delta \mathcal{C}(t)). \quad (\text{s20})$$

In this case, the parameter β is time-dependent. It is linearly increased with a rate equal to γ during the simulation, and reset to zero if the energy does not decrease during a duration τ with $\tau > 0$. The dynamics of $\beta(t)$ is given as follows ($\beta(0) = 0$):

$$\partial_t \beta = \gamma, \quad (\text{s21})$$

when $t - t_c < \tau$, with t_c the last time the best known cut \mathcal{C}_{opt} increased or β was reset. Otherwise, β is reset to 0 if $t - t_c \geq \tau$ and t_c is set to t . The parameters used for the numerical simulations of the proposed scheme are summarized in Tab. II. Moreover, the function g is given as $g(\nu_1, \nu_2, \nu_3, \nu_4) = -6.3674 - 0.2579\nu_1 - 1.0548\nu_2 - 4.2597\nu_3 + 6.1727\nu_4$ where ν_1, ν_2, ν_3 , and ν_4 are the largest, second, third, and fourth largest eigenvalues, respectively, of the weight matrix $\Omega = \{\omega_{ij}\}_{ij}$.

Results of the benchmark on the G-set are shown in Tables III, IV, and V.

id	N	C^{BLS}	C^*	$C^* - C^{\text{BLS}}$	$\langle t^{\text{BLS}} \rangle$ (s)	$\langle t^{\text{CPU}} \rangle$ (s)	$\langle t^{\text{CIM}} \rangle$ (s)
1	800	11624(9)	11624(20)	0	13.00	12.69	0.03
2	800	11620(11)	11620(20)	0	41.00	49.07	0.13
3	800	11622(19)	11622(20)	0	83.00	24.16	0.06
4	800	11646(16)	11646(20)	0	214.00	38.26	0.10
5	800	11631(20)	11631(19)	0	14.00	25.80	0.06
6	800	2178(20)	2178(20)	0	18.00	14.74	0.04
7	800	2006(12)	2006(20)	0	317.00	32.19	0.08
8	800	2005(18)	2005(18)	0	195.00	75.20	0.18
9	800	2054(14)	2054(20)	0	97.00	34.26	0.09
10	800	2000(13)	2000(19)	0	79.00	134.81	0.30
11	800	564(20)	564(20)	0	1.00	21.19	0.07
12	800	556(20)	556(20)	0	2.00	12.40	0.04
13	800	582(20)	582(20)	0	2.00	31.10	0.10
14	800	3064(6)	3064(9)	0	119.00	107.89	0.34
15	800	3050(20)	3050(6)	0	43.00	37.72	0.10
16	800	3052(8)	3052(20)	0	70.00	89.18	0.27
17	800	3047(16)	3047(17)	0	96.00	131.32	0.39
18	800	992(14)	992(7)	0	106.00	72.14	0.19
19	800	906(6)	906(15)	0	20.00	101.04	0.29
20	800	941(20)	941(19)	0	9.00	37.86	0.10
21	800	931(14)	931(9)	0	42.00	109.44	0.32

TABLE III: Performance of proposed method in finding ground-states, i.e., maximum cuts, of graphs in the G-set. id , C^{BLS} , and C^* are the name of instances, maximum cuts found by BLS[s1] and the proposed method, respectively, after 20 runs. The number of runs that found the optimal cut are indicated in parenthesis. Moreover, $\langle t^{\text{BLS}} \rangle$, $\langle t^{\text{CPU}} \rangle$, and $\langle t^{\text{CIM}} \rangle$ are the averaged time-to-solution using BLS written C++ and running on a Xeon E5440 2.83 GHz[s1], the proposed scheme simulated using Matlab on a Xeon X5690 3.47 GHz, and the expected time-to-solution for an implementation of the proposed scheme using the coherent Ising machine[s2] (see footnote [45]), respectively. Part I.

id	N	C^{BLS}	C^*	$C^* - C^{\text{BLS}}$	$\langle t^{\text{BLS}} \rangle$ (s)	$\langle t^{\text{CPU}} \rangle$ (s)	$\langle t^{\text{CIM}} \rangle$ (s)
22	2000	13359(1)	13359(9)	0	560.00	229.53	0.43
23	2000	13344(10)	13342(11)	-2	278.00	383.30	0.71
24	2000	13337(5)	13337(17)	0	311.00	493.25	0.90
25	2000	13340(1)	13340(10)	0	148.00	414.78	0.78
26	2000	13328(3)	13328(8)	0	429.00	534.15	0.98
27	2000	3341(10)	3341(20)	0	449.00	228.86	0.43
28	2000	3298(8)	3298(16)	0	432.00	434.56	0.78
29	2000	3405(1)	3405(19)	0	17.00	526.89	0.90
30	2000	3412(10)	3413(1)	1	283.00	164.02	0.29
31	2000	3309(8)	3310(2)	1	285.00	483.15	0.87
32	2000	1410(13)	1408(7)	-2	336.00	422.86	0.97
33	2000	1382(1)	1380(3)	-2	402.00	320.80	0.75
34	2000	1384(20)	1384(10)	0	170.00	410.30	0.96
35	2000	7684(2)	7681(1)	-3	442.00	655.11	1.47
36	2000	7678(1)	7676(8)	-2	604.00	251.63	0.48
37	2000	7689(1)	7690(1)	1	444.00	285.74	0.54
38	2000	7687(4)	7688(1)	1	461.00	199.44	0.38
39	2000	2408(13)	2408(5)	0	251.00	438.54	0.84
40	2000	2400(1)	2400(1)	0	431.00	528.28	1.10
41	2000	2405(17)	2405(3)	0	73.00	132.46	0.24
42	2000	2481(2)	2478(1)	-3	183.00	199.26	0.35

TABLE IV: Same as Tab. III, part II.

id	N	C^{BLS}	C^*	$C^* - C^{\text{BLS}}$	$\langle t^{\text{BLS}} \rangle$ (s)	$\langle t^{\text{CPU}} \rangle$ (s)	$\langle t^{\text{CIM}} \rangle$ (s)
43	1000	6660(18)	6660(20)	0	26.00	30.01	0.08
44	1000	6650(14)	6650(20)	0	43.00	46.35	0.12
45	1000	6654(15)	6654(20)	0	104.00	104.78	0.27
46	1000	6649(13)	6649(20)	0	67.00	110.58	0.30
47	1000	6657(11)	6657(20)	0	102.00	102.36	0.28
48	3000	6000(20)	6000(19)	0	0.00	4.76	0.01
49	3000	6000(20)	6000(12)	0	0.00	4.99	0.01
50	3000	5880(19)	5880(6)	0	169.00	18.40	0.04
51	1000	3848(17)	3848(8)	0	81.00	181.31	0.49
52	1000	3851(19)	3851(3)	0	78.00	199.90	0.62
53	1000	3850(13)	3850(6)	0	117.00	187.95	0.51
54	1000	3852(10)	3851(19)	-1	131.00	201.95	0.60

TABLE V: Same as Tab. III, part III.

K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, *Science* **354**, 614 (2016).