

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, California 94305, 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, New York 14853, USA*

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



(Received 11 May 2018; revised manuscript received 29 October 2018; published 1 February 2019)

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.

DOI: [10.1103/PhysRevLett.122.040607](https://doi.org/10.1103/PhysRevLett.122.040607)

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 nonequilibrium 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 maximum cut (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 spin 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, heterogenous (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 $\phi(x_i) = -(\partial V_b / \partial x_i)$ and $V_b = -(-1 + p)(x_i^2/2) + (x_i^4/4)$ 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 [26,31–33], and weakly coupled neural networks near pitchfork bifurcations [34]. For solving the combinatorial optimization problem that is defined by the cost function $V(\boldsymbol{\sigma})$, the coupling I_i is chosen as $I_i \propto -[\partial V(\mathbf{x}) / \partial x_i]$; i.e., \mathbf{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 as 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 by other means than 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 $\beta > 0$ [35].

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:

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

where h_i are the elements of the vector $\mathbf{h} = \boldsymbol{\Omega} \boldsymbol{\sigma}$ and $\boldsymbol{\Omega}$ is the matrix of couplings with $\boldsymbol{\Omega} = \{\omega_{ij}\}_{ij}$. 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 $J_{xx} = (-1 + p - 3a)I + \epsilon D[\mathbf{e}]\boldsymbol{\Omega}$, $J_{xe} = \epsilon \sqrt{a} D[\mathbf{h}]$, $J_{ex} = -2\beta \sqrt{a} D[\boldsymbol{\sigma} \cdot \mathbf{e}]$, 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 $\tilde{\boldsymbol{\Omega}} = D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]\boldsymbol{\Omega} - I$ 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 $D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]\boldsymbol{\Omega}$ are the same as the ones of $D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]^{\frac{1}{2}} \boldsymbol{\Omega} D[(\boldsymbol{\sigma} \cdot \mathbf{h})^{-1}]^{\frac{1}{2}}$ (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 becomes the same real value when $\Delta_j = 0$, i.e., under the condition given as follows:

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

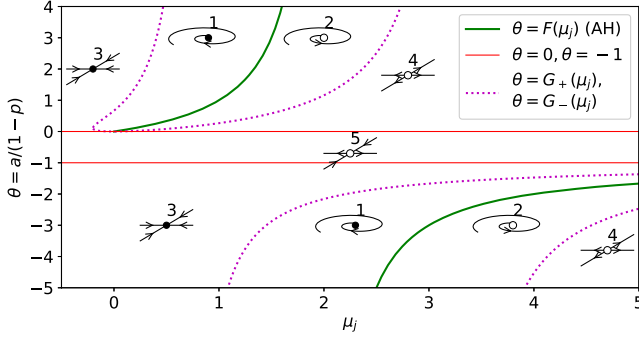


FIG. 1. Bifurcation diagram in the space $\{\theta = [a/(1-p)], \mu_j(\sigma)\}$ at configuration σ . Thick lines correspond to the sets where $\text{Re}[\lambda_j^\pm] = 0$, i.e., $\theta = 0, -1$, or $F(\mu_j)$ with $\Delta_j < 0$, and dotted lines correspond to the sets where $\theta = G_\pm(\beta, \mu_j)$. 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 3–5, they are both real with both, none, and one of their real parts being negative, respectively. $\beta = 0.2$. Andronov-Hopf bifurcation (AH).

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

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 $\text{Re}[\lambda_j^+] \geq \text{Re}[\lambda_j^-]$, $\forall j$):

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

with $F(\mu) = -[\mu/(\mu - 2)]$. 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 $\{\theta = [a/(1-p)], \mu_j(\sigma)\}$, where θ depend on the controllable parameters of the analog system, whereas $\mu_j(\sigma)$ are determined by the spin configuration σ and couplings Ω . We denote $\mu_0(\sigma)$ the maximum eigenvalue of $\tilde{\Omega}$ calculated at the fixed point \mathbf{x} corresponding to the configuration σ . The state \mathbf{x} becomes unstable at a supercritical Andronov-Hopf bifurcation when the real part of the dominant eigenvalue of the Jacobian matrix J , noted $\lambda_0(\sigma)$, becomes positive, i.e., $\mu_0(\sigma) > F^{-1}(\theta)$ for $p < 1$ (see Eq. (7)).

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 σ , we utilize the fact that the fast subsystem on \mathbf{x} converges to the slow manifold characterized by $(dx_i/dt) \approx 0$, $\forall i$, 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 Supplemental Material [37]). 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 \mathbf{g} , defined as $\text{div} \mathbf{g} = \sum_i (\partial g_i / \partial e_i)$ can thus be approximated as follows (see Supplemental Material [37]):

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

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 \mathbf{e} to be always positive along the trajectories of the system. This can be achieved by modulating the target amplitude a as follows:

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

where α is the target amplitude baseline and t_c , with $t_c < t$, is the time of the last change of configuration σ , 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 $\langle X_i \rangle = (1/N) \sum_i X_i$. The control scheme described in Eq. (9) implies, using Eq. (8), that the divergence is approximately given as $\text{div} \mathbf{g} \approx \kappa$ with $\kappa = N\beta(1-p+a)$ as long as the configuration $\sigma(t)$ switches rapidly before the error variables \mathbf{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 \mathbf{x} are slaved to the slow error signals \mathbf{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 (a nondeterministic polynomial-time hard problem) of size $N = 80, 100, 150$, and 200 spins, respectively [37]. For an instance of size $N = 100$, Fig. 2(a) shows that the

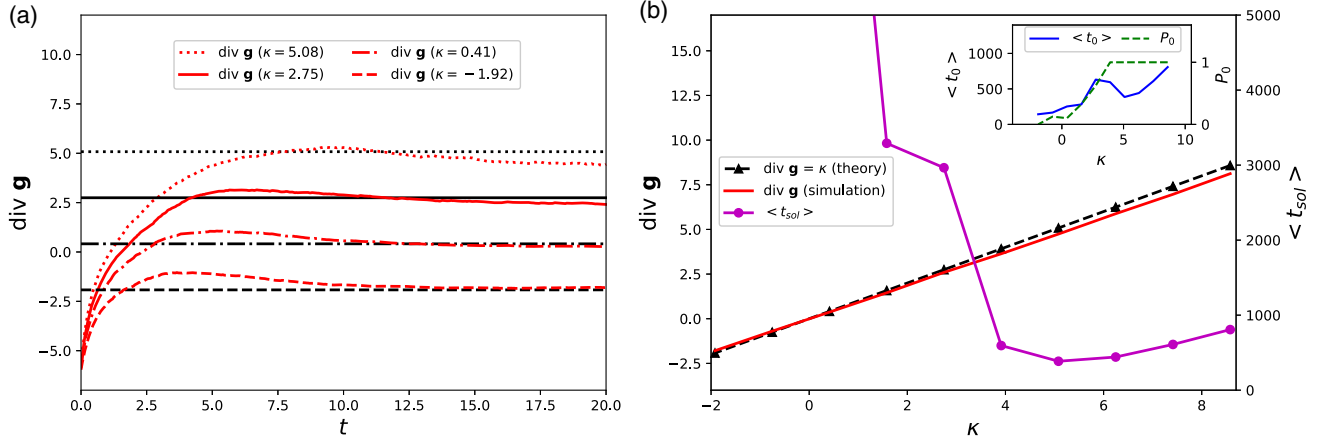


FIG. 2. Control of the divergence in the auxiliary subspace. (a) Divergence $\text{div } \mathbf{g}$ in the auxiliary subspace vs normalized simulation time t . After some transient behavior, the divergence is constant and approximately described by its analytical value $\text{div } \mathbf{g} \approx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

system exhibits constant phase-space volume expansion or contraction in the auxiliary subspace \mathbf{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 Fig. 2(b)]. 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 [36]. 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% 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 the breakout local search (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 $e_i(t) = [e_i(t^-) / \langle e_i(t^-) \rangle]$, 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 maximum satisfiability 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 (dS/dt) [40,41], defined as $(dS/dt) = \int \rho(\mathbf{e}) \text{div } \mathbf{g} d\mathbf{e}$, with as 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 $\kappa > 0$.

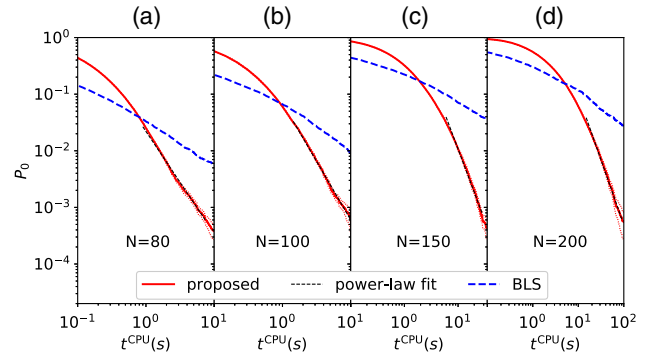


FIG. 3. Proportion $p_0(t)$ of instances unsolved, i.e., for which a ground-state configuration has not been found, after the CPU time t in seconds of the SK problems of size $N = 80, 100, 150$, and 200 shown in (a)–(d), respectively, for the proposed scheme and the BLS algorithm [38] in the logarithmic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

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 [42] to that of the 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 Supplemental Material [37] for 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 [43]. If we assume that the presented scheme is realized in hardware that has the same experimental parameters as existing coherent-Ising-machine implementations [44], 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 the 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.

Note added.—Recently, we were informed of related work about amplitude heterogeneity error correction (see [45]).

*timothee@sat.t.u-tokyo.ac.jp

- [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:quant-ph/0001106](https://arxiv.org/abs/quant-ph/0001106).
- [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, *Nat. Photonics* **10**, 415 (2016).
- [10] A. Marandi, Z. Wang, K. Takata, R. L. Byer, and Y. Yamamoto, *Nat. 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:1709.08102](https://arxiv.org/abs/1709.08102).
- [14] G. Csaba, A. Raychowdhury, S. Datta, and W. Porod, in *IEEE International Symposium on Circuits and Systems (ISCAS), Florence, Italy, 2018* (IEEE, 2018).
- [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* (Oxford University Press, New York, 2006).
- [17] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, New York, 2009).
- [18] A. Lucas, *Front. Phys.* **2**, 5 (2014).
- [19] S. Geman and D. Geman, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984* 721 (IEEE, 1984), Vol. PAMI-6, pp. 721–741.
- [20] H. Sompolinsky and A. Zippelius, *Phys. Rev. Lett.* **47**, 359 (1981).
- [21] S. H. Strogatz, *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*, Studies in Nonlinearity (Westview Press, Perseus Books Group, Cambridge, 2001).
- [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* (Morgan Kaufman Publisher Inc., San Francisco, 1989), pp. 91–98.
- [25] C. Peterson and B. Söderberg, *International Journal of Neural Systems* **01**, 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:1705.06963](https://arxiv.org/abs/1705.06963).
- [30] N. Sourlas, *Phys. Rev. Lett.* **94**, 070601 (2005).
- [31] T. Shoji, K. Aihara, and Y. Yamamoto, *Phys. Rev. A* **96**, 053833 (2017).
- [32] A. Yamamura, K. Aihara, and Y. Yamamoto, *Phys. Rev. A* **96**, 053834 (2017).
- [33] W. Wustmann and V. Shumeiko, *Phys. Rev. B* **87**, 184501 (2013).
- [34] F. C. Hoppensteadt and E. M. Izhikevich, *Weakly Connected Neural Networks* (Springer Science & Business Media, New York, 2012), Vol. 126.
- [35] Note that exponentially increasing error variables have been previously proposed in [36] for solving constraint satisfaction problems.
- [36] M. Ercsey-Ravasz and Z. Toroczkai, *Nat. Phys.* **7**, 966 (2011).
- [37] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.122.040607> for more details about the simulation scheme.

- [38] U. Benlic and J.-K. Hao, *Engineering Applications of Artificial Intelligence* **26**, 1162 (2013).
- [39] B. Molnár, F. Molnár, M. Varga, Z. Toroczkai, and M. Ercsey-Ravasz, *Nat. Commun.* **9**, 4864 (2018).
- [40] L. Andrey, *Phys. Lett.* **111A**, 45 (1985).
- [41] D. Ruelle, *Boltzmann's Legacy* (European Mathematical Society, Zurich, 2008) pp. 89–97.
- [42] 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.
- [43] Y. Yamamoto, K. Aihara, T. Leleu, K.-i. Kawarabayashi, S. Kako, M. Fejer, K. Inoue, and H. Takesue, *npj Quantum Inf.* **3**, 49 (2017).
- [44] Estimation of the time to solution $\langle t^{\text{CIM}} \rangle$ in the case of an implementation on the coherent Ising machine (CIM) 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)].
- [45] K. P. Kalinin and N. G. Berloff, [arXiv:1807.00699](https://arxiv.org/abs/1807.00699).