Geometric Landscape Annealing as an Optimization Principle Underlying the Coherent Ising Machine

Atsushi Yamamura, Hideo Mabuchi, and Surya Ganguli
Department of Applied Physics, Stanford University, Stanford, California 94305, USA

(Received 18 October 2023; revised 27 June 2024; accepted 14 August 2024; published 27 September 2024)

Given the fundamental importance of combinatorial optimization across many diverse domains, there has been widespread interest in the development of unconventional physical computing architectures that can deliver better solutions with lower resource costs. However, a theoretical understanding of their performance remains elusive. We develop such understanding for the case of the coherent Ising machine (CIM), a network of optical parametric oscillators that can be applied to any quadratic unconstrained binary optimization problem. We focus on how the CIM finds low-energy solutions of the Sherrington-Kirkpatrick spin glass. As the laser gain of this system is annealed, the CIM interpolates between gradient descent on coupled soft spins to descent on coupled binary spins. By combining the Kac-Rice formula, the replica method, and supersymmetry breaking, we develop a detailed understanding of the evolving geometry of the high-dimensional energy landscape of the CIM as the laser gain increases, finding several phase transitions in the landscape, from flat to rough to rigid. Additionally, we develop a novel cavity method that provides a geometric interpretation of supersymmetry breaking in terms of the reactivity of a rough landscape to specific external perturbations. Our energy landscape theory successfully matches numerical experiments, provides geometric insights into the principles of CIM operation, and yields optimal annealing schedules.

DOI: 10.1103/PhysRevX.14.031054

Subject Areas: Complex Systems, Optics, Statistical Physics

I. INTRODUCTION

Combinatorial optimization [1] is a key enabler of performance in diverse application domains, including, for example, machine learning, robotics, chip design, operations research, and manufacturing. Thus, the codevelopment of algorithms and hardware that can provide better solutions with lower consumption of resources such as time and energy could substantially impact many fields. Promising recent demonstrations of unconventional hardware architectures have ignited broad interest in physics-based approaches to solving NP-hard problems, in which combinatorial optimization over discrete variables is embedded in the analog evolution of nonlinear dynamical systems [2–6]. This interplay between discrete optimization and analog evolution spawns a rich new field of research based on fresh foundations to complement more traditional approaches. While benchmarking experiments have established high-performance scaling of physics-based approaches up to as many as 10^5 optimization variables [7], scant theory exists for extrapolating future prospects for unconventional architectures or analyzing their strengths and weaknesses relative to mainstream heuristics.

In this article, we develop substantial components of a theoretical framework for how the coherent Ising machine (CIM) [8–14], an unconventional physical optimization architecture based on coupled optical parametric oscillators (OPOs) solves a generic class of Sherrington-Kirkpatrick (SK) spin-glass optimization problems [15]. As described in more detail below, the CIM solves an optimization problem by performing gradient descent on a high-dimensional energy landscape whose geometry anneals over time from an initial simple landscape to a final complex landscape corresponding to the optimization problem of interest. Understanding how the CIM solves optimization problems, therefore, requires understanding the evolving high-dimensional geometry of its energy landscape.

Our main contributions are as follows. First, we elucidate mechanisms underlying how the CIM performs well on the generic class of SK spin-glass problems by developing a statistical mechanics based analysis of the evolving geometry of the landscape, thereby deriving the first theory of how this unconventional physical computing device operates in an ensemble of problems. In particular, we combine the Kac-Rice formula, the replica method, and supersymmetry breaking to reveal that, as the annealing process proceeds, the landscape undergoes a sequence of geometric

phase transitions. Our quantitative understanding of these phase transitions allows us to derive an optimal annealing schedule for the CIM, thereby showing for the first time how to rationally design this important schedule using principles of high-dimensional geometry. Our theoretical analysis is also of general interest to the field of statistical mechanics of random energy landscapes, because we introduce a novel cavity method that provides considerable geometric insights into the mysterious nature of supersymmetry-breaking calculations in this field. In particular, our cavity analysis provides a geometric interpretation of the supersymmetry-breaking order parameters in terms of the exponential reactivity of a complex energy landscape, with exponentially many critical points, to specific small perturbations.

In the rest of the introduction, we introduce the CIM as a novel physical computing device and then review the theory of random landscapes, which sets the context for our new theory. Then, we discuss how the confluence of these two disparate fields provide new ways to think about how to solve optimization problems using unconventional analog dynamics in evolving high-dimensional landscapes.

A. Review of coherent Ising machines

The architecture of a CIM [8–14] typically involves a closed-loop optical fiber, wherein pulses of degenerate OPOs circulate. The phase of electromagnetic oscillations of the pulse's optical modes encodes individual binary variables. This binary encoding is due to the bistable phase induced by phase-sensitive amplification dynamics which forces each optical oscillator to oscillate either in or out of phase with respect to its pump light. These binary variables can be coupled via either delay lines or measurement feedback mechanisms, as their corresponding pulses of light traverse the closed-loop optical fiber. The coupling can be arbitrarily programmed to correspond to any symmetric connectivity matrix between N such pulses or OPOs. The combined phase dynamics of this network of OPOs can be thought of as an Ising network of soft spins with arbitrary programmable connectivity, undergoing energy-minimizing dynamics, with some annealing, as described in more detail below. Indeed, large-scale physical implementations exist with $N = O(10^5)$ OPOs with $O(10^{10})$ connections [7].

Overall, the CIM may be understood as a heuristic solver for the Ising ground-state problem, which is to identify the spin configuration $\{s_i\}_{i=1,2,\ldots,N}$ with spin variables $s_i = \pm 1$ that minimizes the Ising Hamiltonian $\mathcal{H} = -\frac{1}{2}\sum_{i,j} J_{ij} s_i s_j$, where J_{ij} is an $N \times N$ symmetric matrix. This problem, also known as quadratic unconstrained binary optimization (QUBO), is known to be NP-hard [16]. Indeed, many optimization problems, including partitioning, covering, packing, matching, clique finding, graph coloring, minimum spanning trees, and the traveling salesman problem, can be mapped to a corresponding QUBO problem with only polynomial overhead [17]. Thus, solving QUBO or Ising optimization problems is of wide interest and applicability in combinatorial optimization, yet there is no theoretical understanding of when and how the CIM successfully solves such optimization problems.

Indeed, the central role of OPOs as building blocks makes the CIM architecture especially interesting within the broader field of physics-based optimization, as comprehensive quantitative models for OPO networks can be constructed in ways that interpolate between classical and quantum operating regimes (as a function, e.g., of linear decoherence rates relative to coherent nonlinear dynamical rates [18]). This makes CIM theory a fertile setting for exploring how novel information dynamics that emerge in the classical-quantum crossover [19] may impact optimization performance. But the first step in this program must be to establish a baseline understanding of classical CIM mechanics, against which quantum differences can be highlighted. In this article, we begin to draw this classical baseline.

The CIM approaches QUBO by relaxing the binary Ising spins to continuous soft spins. Each OPO functions as a relaxed analog state (soft spin) with a continuous state variable x, subject to a double-well energy potential $E_I(x, a) := \frac{1}{4}x^4 - \frac{a}{2}x^2$. Here, x can be thought of roughly as the phase of a single OPO, and the minima correspond to this phase being 0 or π relative to its pump field, as described above. The important laser gain parameter a controls the depth of the two wells. As the gain parameter increases, each OPO becomes strongly confined in one of the wells, effectively functioning as a binary spin s_i. At very large gain, there are 2^N minima in the energy landscape, and the global minimum corresponds to the ground state of the Ising Hamiltonian [12]. To locate a global minimum, the CIM anneals the gain, by first minimizing the energy of the soft-spin network at a low gain, where the energy landscape is convex, and then adiabatically increasing the gain parameter until each soft spin starts to exhibit behavior akin to a binary spin. Such optimization mechanisms, which continually reshape the energy landscape starting from a trivial form, have also been suggested in various other contexts such as mean-field annealing [20] (a deterministic approximation of simulated annealing), annealed stochastic gradient descent [21] in the context of deep neural networks, and topology trivialization [22] in the random landscape literature.

Numerous numerical and experimental benchmarks have shown that this landscape annealing approach can achieve high performance [7–9,23,24]. However, to the best of our knowledge, no theoretical analysis for this performance has been established. This is in stark contrast to other well-recognized annealing algorithms, such as simulated annealing and quantum annealing, which are known to successfully find the optimum given a sufficiently slow annealing schedule [25,26]. Interestingly, the CIM may fail

to find the ground state for certain frustrated instances, even if the annealing speed is appropriately slow [12]. This is believed to stem from the amplitude heterogeneity of the soft spins, which makes the mapping from Ising energy to the soft-spin network's energy less precise when the gain is not substantial enough. Indeed, right after the landscape becomes nonconvex, the global minimum of the energy landscape lies along the eigenvector of the J matrix with the minimum eigenvalue, generally different from the true Ising ground-state configuration [27]. This amplitude heterogeneity issue has been discussed, and a few methods have been proposed to mitigate its effect [27–30]. While the amplitude heterogeneity initially compels the Ising machine to find the eigenvector rather than a global minimum, as we further ramp up the gain, the signs of soft-spin variables x_i successively flip, leading to a continuous decrease in Ising energy. These configuration adjustments enhance the Ising machine, making it a robust Ising optimizer rather than just a simple linear solver. Indeed, when optimizing the SK energy function, a previous work reported that the CIM finds SK ground states in finite-size systems with a finite probability [31]. To understand how the CIM state evolves with the landscape annealing process, we need to understand the changes in the energy landscape as the gain increases.

This type of question has been extensively investigated for simulated annealing and quantum annealing, especially with purely random instances. In the former case, we generally observe a phase transition from the paramagnetic phase to the spin-glass phase as we cool down the system [32]. In the spin-glass phase, free energies of different thermodynamic states are generally crossing successively, and the low free energy states at two slightly different temperatures can be dramatically different [33–35]. This phenomenon, known as temperature chaos, is related to exponentially long thermalization times in the system size [36], which can dramatically slow down simulated annealing or parallel tempering algorithms that attempt to find low-energy states. Interestingly, for the SK model, an efficient message-passing algorithm [37] was derived to find low-energy states, though the time it takes to find a low-energy state is thought to diverge as the fractional energy gap of that state relative to the ground state goes to 0. The goal of our paper is not to find the best possible algorithm for the SK spin glass but rather to understand how an unconventional physical computing device solves this problem.

Quantum annealing via a transverse magnetic field exhibits similar level crossing properties [26]; it undergoes a phase transition from a quantum paramagnetic phase to a spin-glass or many-body localized phase as the transverse field is reduced [38]. In systems with local interactions, energy level crossings of low-energy states occur in the localized phase, and it takes exponential time in system size to follow the ground state due to the small overlap of those localized states [39,40] (see Ref. [41] for reviews on this topic.)

To our knowledge, such analysis has not been applied to soft-spin networks and hardware like the coherent Ising machine. In this paper, we focus on purely random instances, corresponding in the Ising setting to the SK spin glass [32], and we examine phase transitions in the geometry of the CIM energy landscape. We discover significant phase transitions in the energy landscape as well as evidence for potential level crossings within a particular phase. Furthermore, we demonstrate that these phase transitions are intimately tied to the annealing schedule and optimization performance. In addition to contributing to a type of baseline theory that can eventually be used to study the impact of increasingly quantum OPO behavior in CIM-type architectures, our analysis may also be useful for exploring the potential utility of nondegenerate oscillatory OPO dynamics [42] for evading landscape obstacles within the QUBO setting (see Sec. X). Such studies will be the subject of future work, but our results here provide essential foundations.

The structure of this paper is as follows: After discussing how our work on CIM theory connects with statistical physics results based on related technical approaches, we review in Sec. II the classical formulation of the CIM as a soft-spin network as well as the structure of the energy landscape in both the small- and large-gain regimes, in the case of random connectivity matrices corresponding to the SK spin glass. We furthermore derive a theory delineating the dependence of the curvature of the landscape, quantified through the Hessian eigenspectrum, on where one is located in the landscape. This dependence is critical in all following sections, given the CIM energy landscape possesses no special symmetries. In Sec. III, we demonstrate numerically that the CIM performs well in finding a near-ground-state solution of the SK spin glass, using an optimal annealing schedule for the laser gain that we derive using our subsequent landscape analysis; it significantly outperforms a spectral algorithm and finds a solution that is within about 1% of the true intensive ground-state energy [43].

In Sec. IV, we begin our geometric landscape annealing analysis by performing a supersymmetry-breaking replica calculation to derive detailed predictions about the structure and organization of critical points of the CIM energy landscape and how it evolves as the laser gain is increased. The detailed derivation of the result in this section can be found in Appendix A. In Sec. V, we rederive these results by developing a novel supersymmetry (SUSY)-breaking cavity method, thereby providing considerable geometric insight into the meaning of SUSY breaking in terms of extreme landscape reactivity to external perturbations. Readers who are more interested in the implications of our work for CIM performance, rather than general theory about landscape geometry, can skip the technical details of this cavity derivation. In Sec. VI, we further analyze our replica and cavity theory predictions and compare them to

numerical explorations of the CIM energy landscape, finding an excellent match between theory and numerical experiments. In Sec. VII, we derive a supersymmetric but full replica symmetry-breaking theory of global minima of the CIM energy landscape and further confirm the predictions of this theory in numerical experiments. Together, Secs. VI and VII provide matching theory and experiments for the typical energy, distance from the origin, and Hessian eigenspectra of saddle points, local minima, and global minima as a function of laser gain and reveal a sequence of important phase transitions in the landscape geometry which we summarize in a phase diagram in Sec. VIII.

In Sec. IX, we relate the phase transitions in the landscape geometry to the performance of the CIM as a function of the annealing schedule and explain how these phase transitions suggest the optimal annealing schedule employed earlier in Sec. III to obtain good CIM performance for the SK spin glass. We end with a discussion and future directions in Sec. X. Finally, we provide self-contained derivations in Appendix A, as well as detailed explanations of our numerical experiments in Appendix B. Readers unfamiliar with the spin-glass theory may consult Ref. [44], where we provide detailed step-by-step derivations for pedagogical purposes.

B. Review of theory of random landscapes

The fundamental problem of understanding how the high-dimensional geometry of even the classical CIM energy landscape evolves with increasing laser gain poses several interesting challenges from the perspective of random landscape theory, which has a rich history involving the analysis of several models, including, for example, Thouless-Anderson-Palmer (TAP) free energy landscapes [45–54], random Gaussian fields [55–59], and spherical spin glasses [60–69]. Here, we situate our work within this prior context.

To describe the evolving CIM landscape geometry, we seek to describe changes in the number, location, energy, Hessian eigenspectrum, and local susceptibility of various critical points, including typical saddle points, local minima, and global minima. We apply a combination of the Kac-Rice method [55,70], replica theory [32], random matrix theory [71], and supersymmetry [49,51,62,72–76] to derive an analytic theory of the organization of critical points in the CIM energy landscape as a function of laser gain. Prior theoretical studies of the geometry of critical points in continuous high-dimensional random landscapes focused on simplified settings in which symmetry played a crucial role in carrying forward calculations. For example, in the case of random Gaussian fields [55–59] and spherical spin glasses [60–69], translational and spherical symmetry, respectively, were crucial. The reason symmetry has greatly simplified past calculations is that, as we see below, the combination of the Kac-Rice and replica methods require an analysis of how the Hessian eigenspectrum of the energy landscape depends on the location $\mathbf{x}$ within the landscape. When strong translational or spherical symmetries are present, the Hessian eigenspectrum becomes independent of location and the problem of computing properties of critical points can be reduced to computing properties of the spectrum of a single random Hessian matrix. The TAP free energy landscape, on the other hand, does not possess such simple symmetries but does have a nongeneric property, namely, that the Hessian eigenspectrum of typical critical points has a bulk that is gapped away from the origin, apart from a single zero eigenvalue [47,51], which again simplifies certain analyses as described below.

In contrast, as we see below, the CIM energy landscape possesses neither translational nor spherical symmetry, and its Hessian eigenspectra extend continuously to zero, even for local minima. All of this necessitates a more involved analysis of the relationship between the Hessian eigenspectra of critical points and their location in the CIM energy landscape. One of the contributions of this article from the perspective of random landscape theory is to provide an analysis of how the Hessian eigenspectrum depends on location in a scenario in which no strong symmetries are present. Intriguingly, in the case of the CIM, we find a simple connection from location to Hessian eigenspectrum through Dyson's Brownian motion [77]. We furthermore provide a framework for incorporating this dependence into the combined Kac-Rice and replica methods to analytically derive the organization of critical points in the CIM energy landscape for arbitrary laser gains. Such a framework could be broadly useful for other random landscape problems.

Our work also sheds new light on the geometric meaning of SUSY breaking, which is one approach to analyzing random landscape geometries [49,51,62,72–76]. The reason SUSY can emerge in random landscape analysis is that the Kac-Rice formula can be expressed in terms of a partition function integral over bosonic degrees of freedom related to the location $\mathbf{x}$ as well as fermionic degrees of freedom, which, when integrated alone, yield the determinant of the Hessian of the energy landscape. This integral possesses a SUSY that exchanges bosonic and fermionic degrees of freedom. When the integral is computed via the saddle point method, the correct saddle point can sometimes break SUSY and, therefore, yield nonzero SUSY-breaking order parameters. Given the abstract nature of this calculation, the fundamental geometric meaning of SUSY breaking and the resultant nonzero order parameters has often remained mysterious in general settings.

Prior work has derived geometric interpretations of SUSY breaking in limited settings [54] using modifications of the cavity method [51,76] that take into account the possibility that critical points may have Hessian eigenspectra with a *single* zero mode corresponding to a single flat direction in the energy landscape, with the rest of the bulk spectrum gapped away from the origin. Indeed, Ref. [54] showed that the presence of this single flat

direction indicates SUSY breaking, and the SUSY-breaking order parameters for local minima are related to the inner product between the location of the minimum and the flat direction. This analysis suffices for the TAP free energy landscape of the SK model, which is known to have such an isolated single flat direction, or soft mode, around local minima [47,51]. However, as we see below, this is not the case for the CIM energy landscape, in which typical critical points can have a continuous Hessian spectral density extending to 0, indicating an extensive number of near-flat directions about such critical points.

Another main contribution of our work is to not only derive the properties of critical points using the Kac-Rice formula combined with the SUSY-breaking replica method, but also derive a generalized cavity method for the SUSY-breaking phase. We demonstrate that the generalized cavity and SUSY-breaking replica methods yield identical results, but our novel general cavity method yields important geometric insights into the meaning of SUSY breaking in more general scenarios than previously derived. Importantly, unlike prior work, our cavity method can handle Hessian eigenspectra whose spectral density extends continuously to zero, indicating a critical point that is marginally stable, with extensively many soft modes, corresponding to the small eigenvalues. These soft modes are highly susceptible to perturbations of the landscape. Our cavity method shows that SUSY breaking coincides with the presence of exponentially many such marginally stable, soft critical points with high susceptibility to perturbations. In such a scenario, a small change in the landscape can induce bifurcations in these exponentially many critical points, resulting in exponentially more or fewer critical points. Moreover, we show that the nonzero SUSY-breaking order parameters quantitatively reflect the exponential reactivity of the number of critical points of the energy landscape to specific perturbations. Thus, our work provides a new, general, and quantitative geometric interpretation of SUSY breaking in terms of the extreme reactivity of the landscape stemming from exponentially many marginally stable critical points.

Thus, overall, we see that the general analysis of a physical analog computing device for solving random discrete combinatorial optimization problems, even in the classical limit, yields an incredibly rich theoretical picture that interfaces with numerous branches of physics and mathematics, including the replica method, the cavity method, supersymmetry breaking, random matrix theory, Dyson's Brownian motion, and the geometry of random landscapes. This rich picture serves as an interesting foundational baseline for analyzing how the classical to quantum transition may aid in optimization, in a physically implementable device.

II. THE OVERVIEW OF THE CIM AND ITS ADIABATIC EVOLUTION

Our fundamental problem of interest is to find ground states of the Ising energy function, given by

$$E_{\text{Ising}}(\mathbf{s}) = \frac{1}{2}\sum_{i,j=1}^{N} J_{ij} s_i s_j, \tag{1}$$

where each $s_i = \pm 1$ is a binary spin. The reason for this is that many optimization problems can be cast as Ising optimization problems for a given choice of spin connectivity J_{ij} [17]. However, we focus, in particular, on one generic ensemble of optimization problems in which J_{ij} are chosen to be independent identically distributed (i.i.d.) zero mean random Gaussian variables with variance $1/N$. This is known as the Sherrington-Kirkpatrick spin glass [32].

A. A model of the coherent Ising machine

We consider a model of the CIM as a network of N soft spins, each of which is described by a scalar $x_i \in \mathbb{R}$ $(i = 1, 2, \ldots, N)$, corresponding to the x quadrature of a degenerate OPO. The total energy of the network is given by

$$E_{\text{tot}}(\mathbf{x}) = \sum_{i=1}^{N} E_I(x_i) + \frac{1}{2}\sum_{i,j=1}^{N} J_{ij} x_i x_j, \tag{2}$$

where $E_I(x)$ is a single-site energy function governing the dynamics of a single OPO and J_{ij} reflects the symmetric network connectivity between the OPOs.

While many of our derivations apply to arbitrary internal energy functions $E_I(x)$ that are bounded from below, we focus our comparisons to numerics using the particular internal energy function

$$E_I(x) = \frac{1}{4}x^4 - \frac{a}{2}x^2, \tag{3}$$

which governs the dynamics of each individual OPO in the CIM. Here, a is an important effective laser gain parameter that controls the overall shape of the internal energy of individual OPOs. Note that a reflects a balancing between the linear dissipation and the gain of the CIM system. Therefore, it can be negative when the dissipation is stronger. For $a < 0$, $E_I(x)$ is convex with a single minimum at $x = 0$. But, as a increases beyond 0 to become positive, the single OPO energy landscape undergoes a pitchfork bifurcation wherein the minimum at $x = 0$ becomes a local maximum and two new minima appear at $x = \pm\sqrt{a}$, both with energy $E_I = -\frac{1}{4}a^2$. This corresponds to a symmetric double-well potential in which the wells move further out and become deeper and sharper as a increases, leading to stronger confinement of the soft spins around $x = \pm\sqrt{a}$.

The simplified dynamics of the CIM at zero temperature and fixed gain a can be described as gradient descent dynamics [12,27]

$$\tau\frac{dx_i}{dt} = -\frac{dE_{\text{tot}}(\mathbf{x})}{dx_i}. \tag{4}$$

We work in units of time in which the intrinsic CIM timescale $\tau = 1$. The CIM is typically operated by annealing the gain a as follows [11]. First, the gain parameter a is large and negative, so that the initial CIM state is prepared near the origin $\mathbf{x} = 0$, corresponding to all OPOs approximately in their vacuum state. Then, the gain a is slowly increased over time, while the OPOs simultaneously undergo their natural gradient descent dynamics in Eq. (4). Finally, at a large enough gain a, the OPO states x_i are measured and their signs $s_i = \mathrm{sgn}x_i$ are interpreted as a binary spin configuration, which ideally would achieve a very low Ising energy in the original Ising energy minimization problem of interest in Eq. (1).

This typical annealing of the gain a leads to several questions. First, how and why does annealing a lead to a final answer with low Ising energy? Second, what determines a good annealing schedule, and at what value of a should we stop annealing? In this work, we take a high-dimensional geometric perspective to these questions, by seeking to understand the changing structure of $E_{\mathrm{tot}}(\mathbf{x})$ in Eq. (2) as a increases.

In particular, as a increases, a sequence of bifurcations in the geometry of the high-dimensional energy landscape $E_{\mathrm{tot}}(\mathbf{x})$ takes place. In each such bifurcation, new critical points [i.e., points where the gradient $\nabla E_{\mathrm{tot}}(\mathbf{x})$ vanishes] are either created or destroyed. Additionally, at bifurcations, the index of a critical point can change, where the index is defined to be the number of negative eigenvalues of the Hessian matrix of second derivatives of $E_{\mathrm{tot}}(\mathbf{x})$, evaluated at the critical point. We seek to understand, at each value of a, the high-dimensional geometry of $E_{\mathrm{tot}}(\mathbf{x})$ by analyzing where critical points of a given index lie in terms of their typical energies and their typical locations in $\mathbf{x}$ space. An elucidation of this changing high-dimensional geometry provides insights into the functional optimization advantage gained by annealing the laser gain in the CIM. Furthermore, it suggests properties of good annealing schedules for a.

B. Energy landscape geometry at extremal gains

As a warmup to understanding the high-dimensional geometry of $E_{\mathrm{tot}}(\mathbf{x})$ for arbitrary a, we first focus on two extremal regimes: small $a \ll 0$ and large $a \gg 0$.

1. The small laser gain regime: The CIM computes a spectral approximation to the Ising problem

For $a \ll 0$, we expect the energy landscape to be convex, with the only minimum occurring at $\mathbf{x} = 0$. As a increases, the landscape first becomes nonconvex, by definition, when the Hessian matrix $H(\mathbf{x})$ at any location $\mathbf{x}$ first acquires a negative eigenvalue. The elements of this N by N Hessian matrix are given by

$$H(\mathbf{x})_{ij} = \frac{\partial^2 E_{\mathrm{tot}}}{\partial x_i \partial x_j} = H^I(\mathbf{x})_{ij} + J_{ij}, \tag{5}$$

where

$$H^I(\mathbf{x})_{ij} = \partial^2 E_I(x_i)\delta_{ij} \tag{6}$$

is the diagonal contribution to the Hessian coming from the internal single-site OPO energy function $E_I(x)$ alone. To determine both the smallest a and the location $\mathbf{x}$ at which the first negative eigenvalue of $H(\mathbf{x})$ can occur, we lower bound the eigenvalues of $H(\mathbf{x})$ for all $\mathbf{x}$ as follows.

First, note that, since $H(\mathbf{x}) = H^I(\mathbf{x}) + J$ and the minimum eigenvalue $\lambda_{\min}$ of a symmetric matrix is a concave function of its matrix elements, we have, by Jensen's inequality,

$$\begin{aligned}\lambda_{\min}[H(\mathbf{x})] &\geq \lambda_{\min}(H^I) + \lambda_{\min}(J) \\ &= \min_i \partial^2 E_I(x_i) + \lambda_{\min}(J) \\ &= \min_i 3x_i^2 - a + \lambda_{\min}(J).\end{aligned} \tag{7}$$

In the last line, we use the specific form of the single OPO energy function in Eq. (3). Then, a sufficient condition for $\lambda_{\min}[H(\mathbf{x})]$ to be non-negative is that its lower bound (7) is also non-negative. This yields the sufficient (but not necessary) condition that if $a \leq \min_i 3x_i^2 + \lambda_{\min}(J)$ at any spin configuration $\mathbf{x}$, then $E_{\mathrm{tot}}(\mathbf{x})$ is convex at $\mathbf{x}$. The contrapositive then implies that if $E_{\mathrm{tot}}(\mathbf{x})$ violates convexity at any fixed location $\mathbf{x}$, because the Hessian obeys $\lambda_{\min}[H(\mathbf{x})] < 0$, then we must have $a > \min_i 3x_i^2 + \lambda_{\min}(J)$. This is a necessary (but not sufficient) condition for $E_{\mathrm{tot}}(\mathbf{x})$ to be nonconvex at $\mathbf{x}$.

As a increases, this inequality is first satisfied at the origin $\mathbf{x} = 0$, yielding the result that the origin is the first place where the Hessian $H(\mathbf{x})$ acquires a negative eigenvalue. Moreover, this occurs when a crosses $\lambda_{\min}(J)$. Since the Hessian at the origin is simply $H(\mathbf{0}) = -aI + J$, the associated eigenvector of this Hessian is simply the minimal eigenvector $\mathbf{v}_{\min}$ of J which solves the variational problem

$$\mathbf{v}_{\min} = \mathrm{argmin}_{\{\mathbf{v}|\mathbf{v}^T\mathbf{v}=1\}} \mathbf{v}^T J \mathbf{v}. \tag{8}$$

As a increases beyond $\lambda_{\min}(J)$, the first nonconvex behavior of $E_{\mathrm{tot}}(\mathbf{x})$ is a pitchfork bifurcation where the minimum at $\mathbf{x} = 0$ becomes an index 1 saddle with a single negative curvature direction along $\mathbf{v}_{\min}$ and two new minima appearing that are closely aligned to $\pm\mathbf{v}_{\min}$. If one simply computes the signs of the spin configuration $\mathbf{x}$ in these minima, then one obtains an Ising configuration given by $s_i = \mathrm{sgn}(\mathbf{v}_{\min})$, where $\mathbf{v}_{\min}$ is the solution to Eq. (8). This is known as the spectral approximation to the Ising energy minimization problem in Eq. (1). Thus, for small a just above $\lambda_{\min}(J)$, the CIM computes the spectral approximation. We see below that increasing a can improve upon this spectral solution by finding Ising spin configurations with energy lower than that of the spectral solution.

In summary, our analysis above yields the following picture. For any fixed value of a, $E_{\rm tot}(\mathbf{x})$ can be only nonconvex in the region obeying $\min_i 3x_i^2 < a - \lambda_{\min}(J)$ (a necessary condition for nonconvexity). Contrapositively, if $\min_i 3x_i^2 \geq a - \lambda_{\min}(J)$, then $E_{\rm tot}(\mathbf{x})$ must be convex at $\mathbf{x}$ (a sufficient condition for convexity).

2. The large laser gain regime: The CIM global minimum coincides with the Ising global minimum

In the absence of the connectivity J, the N spins decouple and the energy landscape of Eqs. (2) and (3) has 3^N critical points given by

$$x_i = \sqrt{a} s_i, \quad \text{where } s_i \in \{-1, 0, +1\}. \tag{9}$$

Thus, in the absence of connectivity J, the scale of the soft spins x_i grows as the square root of gain a. If we work with rescaled variables $x_i' := a^{-1/2} x_i$ which remain $O(1)$ as a becomes large, the total energy in Eqs. (2) and (3) can be written as

$$a^{-2} E_{\rm tot} = \sum_i \frac{1}{4} x_i'^4 - \frac{1}{2} x_i'^2 + \frac{1}{2a} \sum_{i,j} J_{ij} x_i' x_j'. \tag{10}$$

This shows that, for large $a \gg \lambda_{\max}(J)$, the effect of the connectivity J on the geometry of the energy landscape can be treated as a weak perturbation of the decoupled landscape in which $J = 0$. Therefore, it is useful to first understand this simple decoupled energy landscape.

In this landscape with 3^N critical points given by Eq. (9), the Hessian matrix H of each critical point is diagonal, with each diagonal element either (i) taking the value $-a$ for every "uncommitted" spin sitting at the saddle point $x_i = 0$ of the double-well potential in Eq. (3) or (ii) taking the value $2a$ for every "committed" spin sitting at a minimum $x_i = \pm\sqrt{a}$ of the double-well potential. Thus, the intensive index r of each critical point, defined as the fraction of negative eigenvalues of H, simply corresponds to the fraction of uncommitted spins in the critical point. Since each uncommitted (committed) spin contributes internal energy $E_I = 0$ ($E_I = -\frac{1}{4}a^2$) in Eq. (3), the energy of every critical point is determined by its index r via

$$E_{\rm tot} = -\frac{N}{4}(1 - r)a^2. \tag{11}$$

Thus, a saddle point's energy decreases linearly with its index.

However, the introduction of the connectivity J breaks the energy degeneracy between all critical points of the same index. Applying perturbation theory in the small parameter $1/a$ to Eq. (10) shows that each critical point of the decoupled landscape in Eq. (9) moves to

$$x_i = \sqrt{a} s_i - (3s_i^2 - 1)^{-1} a^{-1/2} h_i^0 + O(a^{-3/2}), \tag{12}$$

where $h_i^0 = \sum_j J_{ij} s_j$ is the field on spin i before the perturbation. Inserting Eq. (12) into Eqs. (2) and (3) shows that the energy of each critical point at large a is given by

$$E_{\rm tot} = -\frac{N}{4}(1 - r)a^2 + \frac{a}{2} \sum_{i,j} J_{ij} s_i s_j + O(a^0). \tag{13}$$

Thus, to leading order in a, the term breaking the degeneracy of critical points in the decoupled landscape is proportional to the Ising energy in Eq. (1). This implies that, at large a, the sign configuration of the global minimum of the CIM energy function in Eqs. (2) and (3) is equal to that of the global minimum of the Ising energy function in Eq. (1).

Additionally, the Hessian $H(\mathbf{x})$ in Eq. (5) at a critical point $\mathbf{x}$ in Eq. (12) takes the form $H = H^I(\mathbf{x}) + J$, where $H^I(\mathbf{x})$ is diagonal with elements

$$H_{ii}^I = \begin{cases} 2a - 3h_i^0 + O(1/a) & \text{for } s_i = \pm 1, \\ -a + O(1/a) & \text{for } s_i = 0. \end{cases} \tag{14}$$

The eigenvalue spectrum of this Hessian, in the case where J is the random Gaussian connectivity of the SK model, can be understood using the random matrix theory of the next subsection, which also forms a basis for many subsequent analyses.

C. A theory of Hessian eigenspectra in the CIM with an SK spin-glass connectivity

The eigenvalue distribution of the Hessian $H(\mathbf{x})$ in Eqs. (5) and (6) plays a key role in this work. Here, we provide a theory for the spectrum of $H(\mathbf{x})$, at any spin configuration $\mathbf{x}$, when J_{ij} is a rotationally invariant symmetric Wigner random matrix with i.i.d. elements distributed as

$$J_{ij} = J_{ji} \sim \begin{cases} \mathcal{N}(0, g^2/N) & \text{for } i \neq j \\ \mathcal{N}(0, 2g^2/N) & \text{for } i = j, \end{cases} \tag{15}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a Gaussian distribution with mean μ and variance σ^2. This connectivity corresponds to the SK spin glass in Eq. (1). Because of the fundamental importance of the eigenvalue distribution of $H(\mathbf{x})$ in understanding the high-dimensional geometry of the CIM energy landscape, we discuss this spectral distribution in the next two subsections in two different ways: first, in a conceptual way, as the outcome of a Dyson's Brownian motion with initial condition determined by $\mathbf{x}$ and, second, in a computationally tractable manner in terms of a self-consistent formula involving the resolvent of $H(\mathbf{x})$. Finally, in the third subsection, we apply this random matrix theory to analytically calculate the Hessian eigenspectra of CIM

critical points at large a and verify our formula by comparing to numerics. In the following, we set the connectivity variance parameter g in Eq. (15) to 1 without loss of generality, because the case of $g \neq 1$ can be reduced to $g = 1$ through the rescaling $\mathbf{x} \to \sqrt{g}\mathbf{x}$ and $a \to ga$. Note that, for $g = 1$, the eigenvalue spectrum of J follows the well-known Wigner semicircular law with minimum and maximum eigenvalues given by $\lambda_{\min}(J) \approx -2$ and $\lambda_{\max}(J) \approx +2$ [78], respectively.

1. From the distribution of spins to Hessian eigenspectra through Dyson's Brownian motion

Now, at any spin configuration $\mathbf{x}$ for which the diagonal elements $H^I_{ii}(x_i) = \partial^2 E_I(x_i)$ are large relative to the elements of J_{ij}, one can compute the eigenvalues of $H(\mathbf{x})$ through first-order perturbation theory, treating J as perturbation to H^I in Eq. (5). This yields an approximate expression for the eigenvalues λ_i of $H(\mathbf{x})$ given by

$$\lambda_i = H^I_{ii} + J_{ii} + \sum_{j \neq i} \frac{J^2_{ij}}{H^I_{ii} - H^I_{ij}}. \tag{16}$$

This expression is applicable, for example, when $\mathbf{x}$ corresponds to a critical point of the CIM energy landscape at large a, where each x_i in Eq. (12) is $O(\sqrt{a})$, and, therefore, each H^I_{ii} in Eq. (14) is $O(a)$.

However, at smaller a, when critical points are closer to the origin, the perturbative expression in Eq. (16) may not be accurate. One can go beyond this perturbation theory by exploiting the fact that $H(\mathbf{x})$ is the sum of a fixed matrix H^I and a Wigner matrix. This sum can be thought of as the outcome of a white-noise-driven diffusion process in the space symmetric matrices running from time $t = 0$ to $t = g$ starting from the initial condition $H^I(\mathbf{x})$ and ending at $H(\mathbf{x})$. This diffusion process on symmetric matrices, in turn, induces the well-known Dyson's Brownian motion on the corresponding eigenvalues [71,77], described by the stochastic differential equation

$$d\lambda_i = \sqrt{\frac{2}{N}} dW_{ii} + \frac{1}{N} \sum_{j \neq i} \frac{dt}{\lambda_i - \lambda_j}, \tag{17}$$

where dW_{ii} is a standard white noise process. This stochastic evolution has a physical interpretation in which each λ_i can be thought of as a Coulomb charge in the complex plane, confined to the real axis, feeling a deterministic, repulsive 2D Coulomb force from all the other charges λ_j, in addition to an independent stochastic drive. If this Brownian motion is initialized at $t = 0$ so that $\lambda_i(0) = H^I_{ii}(x_i)$ and is run up to time $t = g$, then the resulting eigenvalue distribution

$$\rho_H(\lambda) \equiv \frac{1}{N} \sum_{i=1}^{N} \delta[\lambda - \lambda_i(g)] \tag{18}$$

will, at large N, converge to the eigenvalue distribution of $H(\mathbf{x})$ in Eq. (5) with J_{ij} distributed as in Eq. (15).

Thus, Dyson's Brownian motion provides an elegant and intuitive understanding of the relationship between a spin configuration $\mathbf{x}$ and the eigenvalue distribution of the Hessian $H(\mathbf{x})$: Simply initialize a set of N charges at the positions $H^I_{ii} = \partial^2 E_I(x_i)$ and allow them to diffuse under Eq. (17) for a time g. However, this does not by itself provide an analytic method for computing the final outcome of the diffusion in Eq. (18).

2. From the distribution of spins to the Hessian eigenspectra through the resolvent

In Appendix A, we provide a calculation of the Hessian eigenspectrum $\rho_H(\lambda)$ of $H(\mathbf{x})$ as a function of the distribution of spins at $\mathbf{x}$, defined as

$$P_x(x) \equiv \frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i). \tag{19}$$

Our replica calculation yields a self-consistent equation for the resolvent of $H(\mathbf{x})$. A more detailed step-by-step derivation can be found in Ref. [44]. In general, the resolvent of any N-by-N symmetric matrix H is defined as

$$R(z) = \frac{1}{N} \mathrm{Tr} \frac{1}{H - z}, \tag{20}$$

where $z \in \mathbb{C}$ is a complex scalar. One can recover the eigenvalue density $\rho_H(\lambda)$ from the resolvent $R(z)$ via the inversion formula

$$\rho_H(\lambda) = \lim_{\varepsilon \to 0^+} \frac{R(\lambda - i\varepsilon) - R(\lambda + i\varepsilon)}{2\pi i}. \tag{21}$$

For $H(\mathbf{x})$, our replica-based self-consistent equation for its resolvent, when $g = 1$, is given by (see Ref. [44], Sec. A, for a derivation)

$$R(z) = \int \frac{P_x(x)}{\partial^2 E_I(x) - z - R(z)} dx, \tag{22}$$

where $P_x(x)$ is the distribution of spins in Eq. (19). This result agrees with Pastur's self-consistent equation for the resolvent of the sum of a fixed matrix and a Wigner matrix [79].

Thus, we obtain a simple calculational framework to obtain the Hessian eigenspectrum at any spin configuration $\mathbf{x}$: (i) Insert the distribution of spins $P_x(x)$ in Eq. (19) into the self-consistent equation for the resolvent $R(z)$ in Eq. (22), (ii) solve this equation to find $R(z)$, and (iii) insert

this solution into the inversion formula in Eq. (21) to obtain the Hessian eigenvalue distribution $\rho_H(\lambda)$. The result will be equivalent to the distribution in Eq. (18) at time $t = 1$ obtained by running Dyson's Brownian motion in Eq. (17) starting from the initial distribution of H^I_{ii} induced by the distribution of $P_x(x)$ under the map $x \to \partial^2 E_I(x)$.

3. Hessian eigenspectra of critical points at laser gain

Given this random matrix theory, we now return to the large-gain regime in Sec. II B 2 to compute the Hessian eigenspectra of critical points of the form in Eq. (12). In a typical index r critical point, a fraction r of the spins (before the perturbation by J) take the uncommitted value $s_i = 0$, while the remaining fraction $1 - r$ takes the committed values $s_i = \pm 1$ with equal probability. Moreover, the field $h_i^0 = \sum_j J_{ij} s_j$ in Eq. (12), which perturbs the critical point after introducing the SK connectivity in Eq. (15), is, at large N, a zero mean Gaussian random variable with variance $(1 - r)$, originating from the fraction $1 - r$ of nonzero committed spins. Thus, the distribution of the diagonal elements in $H^I(\mathbf{x})$ in Eq. (14) is given by

$$p_{H^I}(h) = r\delta(h + a) + (1 - r)\mathcal{N}[2a, 9(1 - r)]. \quad (23)$$

This corresponds to a mixture of a δ function at $-a$ with weight r coming from the uncommitted spins and a Gaussian centered at $2a$ with weight $1 - r$ coming from the committed spins. The variance of $9(1 - r)$ arises from the amplification of h_i^0 by a factor of 3 in Eq. (14).

This initial distribution then undergoes Dyson's Brownian motion in Eq. (17) to yield the full distribution $\rho_H(\lambda)$ of $H(\mathbf{x})$. Alternatively, we can make the change of variables from x to $h = \partial^2 E_I(x)$ in Eq. (22) to obtain a self-consistent equation $R_H(z)$ in terms of $p_{H^I}(h)$:

$$R(z) = \int \frac{p_{H^I}(h)}{h - z - R(z)} dh. \quad (24)$$

We can then solve this equation (numerically) and insert the solution into Eq. (21) to obtain $\rho_H(\lambda)$.

We calculate the Hessian eigenspectrum in this fashion both for typical critical points with index $r = 1/3$ and for typical minima with index $r = 0$, finding an excellent match with direct numerical searches for such critical points at a finite system size of $N = 10^3$ and at large laser gain $a = 9$ (Fig. 1). Some features of the outcome of Dyson's Brownian motion in going from $p_{H^I}(h)$ in Eq. (23) to $\rho_H(\lambda)$ are readily apparent in Fig. 1. For example, at large a for a typical critical point with index $r = 1/3$, the charges start in two far apart clumps in Eq. (23), with a delta function at $-a$ and a Gaussian at $2a$. Thus, these two distant charge clumps do not interact strongly with each other in the diffusion. However, each clump itself expands under the repulsive diffusion. The delta function expands into a Wigner semicircle, still centered at $-a$, while the Gaussian expands a bit more, largely retaining its shape and remaining centered at $2a$ (Fig. 1, top).

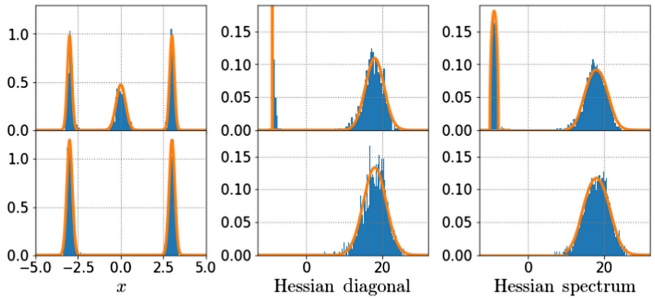

FIG. 1. Distribution of spins, Hessian diagonal elements, and Hessian spectrum. The upper (or lower) panels showcase the distribution of OPO amplitudes x (left), Hessian diagonal elements (middle), and Hessian eigenvalues (right) corresponding to a typical critical point (or a typical local minimum) with a large gain $a = 9$. The empirical distributions portrayed as blue histograms are obtained with a system size of $N = 10^3$. The orange curves in the left and middle figures are obtained with the perturbation theory in Eqs. (12) and (14). The distributions of Hessian diagonal elements in the middle panels diffuse via Dyson's Brownian motion in Eq. (17) to generate the Hessian eigenspectrum in the right panels. The orange curves in the right panel are obtained from solutions of Eq. (24).

III. THE PERFORMANCE OF GEOMETRIC LANDSCAPE ANNEALING FOR THE SK SPIN GLASS

We have seen in Sec. II B 1 that, at small gains a just above $\lambda_{\min}(J)$, the CIM global minimum computes the spectral solution in Eq. (8), which is not of direct interest. On the other hand, in Sec. II B 2, at large gain $a \gg \lambda_{\max}(J)$, we have seen that the CIM global minimum computes the Ising energy minimization, which *is* of direct interest. However, our analysis of the energy landscape at large laser gain in Sec. II B 2 reveals a complex landscape with exponentially many local minima and saddle points of all indices. Thus, direct gradient descent in the large laser gain energy landscape of the CIM is unlikely to find the CIM global minimum (as we verify below in Sec. IX). Therefore, to understand how the CIM solves optimization problems by annealing the laser gain, a key first step is to understand how the geometry of the landscape changes from small to large gain.

In particular, we would like to understand, in general, how the first local minimum to occur, which is aligned along the lowest eigenvector $\mathbf{v}_{\min}$ in Eq. (8), changes as the laser gain is increased. There are several possibilities.

The first is that this local minimum is continuously connected to one of the CIM global minima as we increase the gain to large values. In this case, annealing will find the global minimum. The second possibility is that the first

local minimum to appear as the gain increases is continuously connected to a higher-energy CIM local minimum at large gain. In this case, annealing will not find the CIM global minimum. A third possibility is that this first minimum may disappear through a saddle-node bifurcation and then slowly annealed gradient descent will flow to another nearby minimum, which, in turn, can exhibit these same three possibilities.

It is an exceedingly difficult problem to analytically predict, in advance of geometric landscape annealing, which of these possibilities will occur for any large, fixed connectivity matrix J. One would have to map out the entire bifurcation structure of critical points as a increases. Moreover, one would have to analytically derive the CIM ground-state energy for that connectivity J at large a and compare it to the energy of all critical points that are continuously connected through bifurcations to the first minimum to appear along $\mathbf{v}_{\min}$ near the origin. All of this is more complex than simply performing geometric landscape annealing itself.

We circumvent these difficulties by not analyzing any fixed connectivity J but rather analyzing typical CIM behavior in random Gaussian connectivities J in Eq. (15) corresponding to an SK spin glass. For this problem, we can use techniques from the statistical mechanics of quenched disorder to analytically calculate the CIM ground-state energy at arbitrary laser gain a, as well as the location, Hessian eigenstructure, and energy levels of critical points of any index. We can compare these quantities to numerical simulations of geometric landscape annealing to provide insights into its operation.

First, we assess the performance of CIM geometric landscape annealing as we increase the laser gain a. We perform numerical simulations of geometric landscape annealing with several system sizes N by the integration of Eq. (4). During the integration, we slowly increase the gain parameter a starting from $\lambda_{\min}(J)$ to achieve the best performance for each system size N. In Fig. 2, the blue dots represent the medians of the final Ising energy obtained by the simulations with several instances, and the blue dotted line is the linear regression of those points against $N^{-2/3}$. This scaling comes from the finite-size scaling of the SK model's ground-state energy [80,81]. The y intercept of this blue line represents the reachable lowest energy by the annealing dynamics in the large-N limit. We call this energy E_{anneal}. The horizontal red dotted line is the theoretically obtained ground-state energy $E_g \sim -0.763$ in the large-N limit [82,83], and the horizontal green dashed line is the energy obtained from the Ising spin configuration by rounding the principal eigenvector, which yields the known value $E_{sp} = -2/\pi$ [37]. We can see that the energy $E_{\text{anneal}} \sim -0.75$ obtained by the annealing process in the large-N limit is much lower than E_{sp}. This means that the first minimum to appear along the eigenvector $\mathbf{v}_{\min}$ must undergo multiple sign flips as a increases.

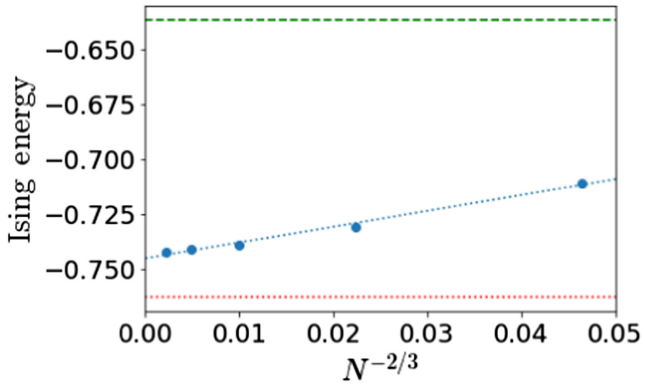

FIG. 2. Ising energy of the final state obtained by the annealing process for random instances. We simulate the annealing process of the soft-spin network with many instances for each system size N and plot the medians as blue dots. The blue dotted line is the linear regression against $N^{-2/3}$. This scaling comes from the SK model's finite scaling [80,81]. The y intercept of this line represents the reachable lowest Ising energy by the dynamics under the large-N limit. The horizontal red dotted line is the SK model's ground-state energy in the large-N limit $E_g \sim -0.763$. The horizontal green dashed line is $E_{sp} = -2/\pi$, the energy of spin configuration obtained by rounding the principal eigenvector of J. The annealing schedule used here is given by Eq. (81) with $\tau = 10^2$, $a_{\max} = 0.0$, and $a(0) = \lambda_{\min}$. The number of sampled instances is 300, 300, 100, 100, 20, and 5 for $N = 10^2$, 3×10^2, 10^3, 3×10^3, and 10^4, respectively.

Remarkably, these bifurcations substantially lower the Ising energy found by the CIM, making it very close to the actual ground-state Ising energy of the SK model, as reported numerically in a previous work [31]. In the remainder of this paper, we study the changing geometry of the CIM energy landscape to understand how geometric annealing of this landscape empowers the performance of the CIM in finding low Ising energy solutions.

IV. THE EVOLVING ENERGY LANDSCAPE GEOMETRY

To address the questions raised above, we first analytically derive a formula for the typical number $\mathcal{N}(r, e|J)$ of critical points of a given intensive index r and energy e. In order to average over the connectivity J, we work with the complexity $\Sigma(r, e|J)$ of critical points, which is defined via the relation

$$\mathcal{N}(r, e|J) = e^{N\Sigma(r,e|J)}. \tag{25}$$

This complexity can be formally written as a sum over all critical points:

$$\Sigma(r, e|J) = \frac{1}{N} \log \sum_{\alpha \in \mathrm{Crt}(E)} \delta[\mathcal{I}(\mathbf{x}^\alpha) - r]\delta[E(\mathbf{x}^\alpha) - e], \tag{26}$$

where $E(\mathbf{x}) = (1/N)E_{\text{tot}}(\mathbf{x})$ is the intensive energy and $\text{Crt}(E)$ denotes the set of all critical points of $E(\mathbf{x})$.

Unlike the potentially exponentially large number $\mathcal{N}(r, e|J)$ itself, which could fluctuate across random samples of J, we expect the complexity function $\Sigma(r, e|J)$ to be self-averaging with respect to J. This means that typical values of $\Sigma(r, e|J)$ for random samples of J concentrate closely around the sample average $\Sigma(r, e) \equiv \langle \Sigma(r, e|J) \rangle_J$, where $\langle \cdot \rangle_J$ denotes an average over J. Furthermore, in order to compute this sample-averaged complexity $\Sigma(r, e)$, we first compute the sample average of the grand potential $\Omega(\beta, \mu|J)$, defined as

$$-\beta\Omega(\beta, \mu|J) = \frac{1}{N} \log \sum_{\alpha \in \text{Crt}(E)} e^{-\beta E(\mathbf{x}^\alpha) + \mu \mathcal{I}(\mathbf{x}^\alpha)}, \quad (27)$$

where $\mathcal{I}(x^\alpha) \in [0, 1]$ denotes the intensive index of the critical point $\mathbf{x}^\alpha$, i.e., $\mathcal{I}(\mathbf{x}^\alpha) = I(\mathbf{x}^\alpha)/N$. If one can compute the sample-averaged grand potential $\Omega(\beta, \mu) \equiv \langle \Omega(\beta, \mu|J) \rangle_J$, then one can recover the average complexity $\Sigma(e, r)$ via Legendre transform:

$$\Sigma(e, r) = \inf_{\beta, \mu} [\beta e - \mu r - \beta\Omega(\beta, \mu)]. \quad (28)$$

Here, the effective inverse temperature β and energy density e form a Legendre dual pair, as do the chemical potential μ and the intensive index r. Indeed, the typical values of e and r that dominate in the sum in Eq. (27) are those that achieve the infimum in Eq. (28). If the infimum does not occur at a boundary, the typical e and r are related to β and μ through

$$e = \frac{\partial}{\partial \beta} [\beta\Omega(\beta, \mu)], \qquad r = -\frac{\partial}{\partial \mu} [\beta\Omega(\beta, \mu)]. \quad (29)$$

Finally, we compute the sample-averaged grand potential $\Omega(\beta, \mu)$ via the replica trick [84], i.e.,

$$\begin{aligned} -\beta\Omega(\beta, \mu) &= \frac{1}{N} \langle \log(Z) \rangle_J \\ &= \frac{1}{N} \lim_{n \to 0} \frac{1}{n} \log \langle Z^n \rangle_J, \end{aligned} \quad (30)$$

where Z is the partition function

$$Z = \sum_{\alpha \in \text{Crt}(E)} e^{-\beta E(\mathbf{x}^\alpha) + \mu \mathcal{I}(\mathbf{x}^\alpha)}. \quad (31)$$

Below, we apply the Kac-Rice formula to Eq. (30) to compute $\Omega(\beta, \mu)$. This replica-based calculation is given in Appendix A. Step-by-step pedagogical derivations are presented in Ref. [44], Sec. S-II. This method involves introducing both bosonic degrees of freedom (replicated soft spins $\mathbf{x}^a$ for $a = 1, \ldots, n$) as well as fermionic degrees of freedom whose integral computes the determinant of the Hessian which arises in the Kac-Rice formula below. The resulting integrals possess both replica symmetry, involving permutations of the replicas, as well as supersymmetry, involving exchanges of bosonic and fermionic degrees of freedom. Such a SUSY-based framework has also been used in a variety of works [49,51,62,72–76]. These integrals can be solved via a saddle point approximation, and the order parameters, whose extremal values determine the saddle point, can either exhibit or break replica symmetry or supersymmetry. Which pattern of symmetry breaking occurs or not depends on the particular values of the inverse temperature β, chemical potential μ, and gain a considered.

In the following, we consider three regimes in detail. First, we consider $\beta = \mu = 0$, corresponding to a white average in Eq. (31) in which all critical points are equally weighted. This white average yields information about the typical behavior of a randomly chosen critical point, regardless of its energy or index. We find that in this regime, at the saddle point order parameters, replica symmetry always holds, but SUSY is preserved at low laser gain a, while it is broken at large laser gain. The order parameters at the saddle point yield information about the distribution of spins and Hessian eigenvalues at typical critical points.

The second regime we consider is $\beta = 0$ and $\mu \to -\infty$. This concentrates the sum over critical points onto those with vanishing intensive index, independent of their energy. This corresponds to a sum over all minima (we refer to critical points with zero *intensive* index as minima). We find a similar pattern for typical minima as we do for typical critical points: Order parameters at the saddle point exhibit replica symmetry but can break SUSY depending on the laser gain. The order parameters, distribution of spins, and Hessian eigenvalues for $\beta = 0$ and arbitrary μ are given in Sec. IV A in the case of replica symmetry and broken SUSY.

Unfortunately, the geometric interpretation of these replica calculations, and, in particular, the geometric meaning of broken SUSY in terms of the original energy landscape, is unclear. Because the geometric interpretation of SUSY breaking is a subject of considerable interest, we provide in Sec. V a completely different derivation of the results in Sec. IV A using the cavity method instead of the supersymmetric method (see Appendix A or Ref. [44], Sec. S-IV, for a detailed cavity derivation). This derivation yields a new interpretation of nonzero supersymmetry-breaking order parameters as signaling a high sensitivity of the complex energy landscape to small changes in external fields. We discuss, in particular, the case of typical critical points in Sec. VI A and the typical minima in Sec. VI B and successfully match our theoretical predictions with numerical experiments.

In Sec. VII A, we move on to the case of the global minima, corresponding to the regime $\beta \to \infty$ in Eq. (31). We find that the global minima of the energy landscape occur at significantly lower energies than that of typical local minima for large values of the laser gain. Therefore, as

in other spin-glass problems, replica symmetry is broken. We analyze the global minima through two methods: (i) a replica-based calculation of the grand potential (see Appendix A and Ref. [44], Sec. S-II.G) or (ii) a calculation of the free energy (see Ref. [44], Sec. S-III), and we demonstrate their equivalence.

Finally, in Sec. VIII, we summarize and describe the significant phase transitions we can observe in the energy landscape due to successive SUSY and replica symmetry breaking and their geometric consequences.

A. The replica-based calculation

In our setting, the Kac-Rice formula (see Ref. [70] for an introduction) enables us to convert the sum of any function $F(\mathbf{x})$ over all critical points $\mathbf{x}^\alpha$ for $\alpha \in \mathrm{Crt}(E)$ of a landscape $E(\mathbf{x})$ into an integral over the entire domain $\mathbf{x} \in \mathbb{R}^N$ of the landscape. It is given by

$$\sum_{\alpha \in \mathrm{Crt}(E)} F(\mathbf{x}^\alpha) = \int \prod_{i=1}^{N} [dx_i \delta[\partial_i E(\mathbf{x})]] |\det H(\mathbf{x})| F(\mathbf{x}), \quad (32)$$

where $H(\mathbf{x})$ is the Hessian of E at $\mathbf{x}$. Here, the δ functions in Eq. (32) localize the integral to critical points of $E(\mathbf{x})$ as desired, while the absolute value of the Hessian determinant $|\det H(\mathbf{x})|$ corresponds to the Jacobian of the change of variables from x_i to $y_i = \partial_i E(\mathbf{x})$. Indeed, performing this change of variables on the right-hand side of Eq. (32) and then integrating recovers the left-hand side.

Now applying the Kac-Rice formula in Eq. (32) to the partition function in Eq. (31) yields

$$Z = \int \prod_{i=1}^{N} [dx_i \delta[\partial_i E(\mathbf{x})]] |\det H(\mathbf{x})| e^{-\beta E(\mathbf{x}) + \mu \mathcal{I}(\mathbf{x})}. \quad (33)$$

Then inserting Eq. (33) into Eq. (30) provides the starting point for the replica- and supersymmetry-based calculation of the sample-averaged grand potential. A detailed derivation is given in Appendix A and Ref. [44], Sec. S-II. The final answer at a replica symmetric, annealed level with $\beta = 0$, but with broken supersymmetry, is given by

$$-\beta\Omega(0,\mu) = -\frac{1}{2}(Cq + A^2) - At + \log \frac{1}{\sqrt{2\pi q}} \int dx |\partial^2 E_I(x) - t| \exp\left(-\frac{1}{2q} h(x)^2 + \frac{A}{q} x h(x) + \frac{1}{2}\frac{qC - A^2}{q} x^2 + \mu \bar{\mathcal{I}}(x)\right). \quad (34)$$

Additionally, beyond the sample-averaged grand potential, we consider the sample-averaged distribution of spins in an ensemble of critical points, defined as

$$P(x) = \left\langle Z^{-1} \sum\nolimits_{\alpha \in \mathrm{Crt}(E)} e^{-\beta E_I(\mathbf{x}^\alpha) + \mu \mathcal{I}(\mathbf{x}^\alpha)} \left(\frac{1}{N} \sum_{i=1}^{N} \delta(x - x_i^\alpha)\right) \right\rangle_J, \quad (35)$$

where Z is given in Eq. (31). We derive a formula for this distribution in Appendix A and Ref. [44], Sec. S-II.B. In the case of $\beta = 0$ and arbitrary μ, which is relevant for typical critical points ($\mu = 0$) and typical minima ($\mu \to -\infty$), the answer is

$$P(x) \propto |\partial^2 E_I(x) - t| \exp\left(-\frac{1}{2q} h(x)^2 + \frac{A}{q} x h(x) + \frac{1}{2}\frac{qC - A^2}{q} x^2 + \mu \bar{\mathcal{I}}(x)\right). \quad (36)$$

Here, $\bar{\mathcal{I}}(x)$ is given by

$$\bar{\mathcal{I}}(x) = \Theta\{-[\partial^2 E_I(x) - t]\}, \quad (37)$$

where Θ is the Heaviside step function, and

$$h(x) = \partial E_I(x) - tx. \quad (38)$$

The formulas for the sample-averaged grand potential in Eq. (34) and distribution of spins in Eq. (36) depend on four order parameters q, t, A, and C which satisfy the following self-consistent equations arising from extremizing the grand potential in Eq. (34):

$$\begin{aligned} q &= \langle x \rangle^2, \\ t &= \left\langle \frac{1}{\partial^2 E_I(x) - t} \right\rangle, \\ A &= \frac{\langle x h(x) \rangle}{2q} - \frac{t}{2}, \\ C &= -q^{-1} + q^{-2} \langle h^2(x) \rangle - 2q^{-2} A \langle x h(x) \rangle + q^{-1} A^2. \end{aligned} \quad (39)$$

Here, $\langle\cdot\rangle$ denotes an average with respect to the distribution $P(x)$ in Eq. (36). We note that self-consistent solutions with nonzero values for the order parameters A and C correspond to broken supersymmetry [48,62].

Finally, with knowledge of the typical distribution of spins $P(x)$ in an ensemble of critical points, we can obtain the typical distribution of Hessian eigenvalues by inserting $P(x)$ in Eq. (36) into Eq. (22), solving for the resolvent $R(z)$, and inserting this solution into Eq. (21) to obtain $\rho_H(\lambda)$ for any $\beta = 0$, μ, and a. Note that we here assume that the correlation between $H^I(\mathbf{x})$ and J is negligible for any critical point $\mathbf{x}$.

Importantly, we note that the last self-consistent equation for t in Eq. (22) is equivalent to the self-consistent equation for $R(z)$ at $z = 0$ in Eq. (22). But, more precisely, while the resolvent $R(z)$ of a large random symmetric matrix H is not well defined at any point λ on the real axis where the eigenvalue distribution $\rho_H(\lambda)$ is nonzero, $R(z)$ *is* defined on the complex plane near the real axis for $z = \lambda + i\varepsilon$ with arbitrarily small ε. Thus, we can define the complex number $t_R + it_I = R(0 + i\varepsilon)$ for a small ε. By the inversion formula in Eq. (21), t_I is nonzero if and only if the Hessian eigenvalue density $\rho_H(0)$ is nonzero. We see empirically that $\rho_H(0)$ is very close to 0. Therefore, assuming $t_I = 0$, t in Eq. (39) should be properly be thought of as $t_R = R(0 + i\varepsilon)$. On the other hand, if the Hessian eigenvalue density $\rho_H(\lambda)$ at the origin $\lambda = 0$ were nonzero, one would have to self-consistently solve for another order parameter t_I. The full self-consistent equations for all five order parameters q, t_R, t_I, A, and C are given in Ref. [44], Sec. S-II.F. However, to match the numerics below, we need only to find approximate self-consistent solutions to Eq. (39) assuming that $t_I = 0$ or, equivalently, $\rho_H(0) = 0$.

In summary, the replica analysis provides an efficient calculational framework to obtain key information about the number and properties of typical critical points ($\mu = 0$) and typical minima ($\mu \to -\infty$), as well as critical points of any index r related to μ through Legendre duality in Eq. (29). The procedure is as follows: (i) Solve the self-consistent equations for the order parameters in Eq. (39); (ii) insert them into Eq. (36) to obtain the typical distribution of spins $P(x)$ at a critical point; (iii) insert $P(x)$ into Eqs. (22) and (21) to obtain the typical distribution of Hessian eigenvalues $\rho_H(\lambda)$; (iv) insert the formula for the grand potential $\Omega(\beta, \mu)$ in Eq. (30) into the Legendre transform in Eq. (28) to obtain the complexity $\Sigma(e, r)$ at typical energy e and index r given by Eq. (29).

Finally, we note that for typical critical points and typical minima, if we wish only to compute the grand potential at $\beta = 0$, we can still compute the typical energy e of critical points without using the first Legendre dual relation in Eq. (29). We do this by noting that any critical point $\mathbf{x}$ of Eq. (2) obeys $\partial E_I(x_i) + h_i = 0$, where $h_i \equiv \sum_j J_{ij}x_j$. This implies that, at any critical point $\mathbf{x}$, the normalized intensive energy obeys the special relation

$$E(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\left[E_I(x_i) + \frac{1}{2}x_i h_i\right] = \frac{1}{N}\sum_{i=1}^{N}\left[E_I(x_i) - \frac{1}{2}x_i \partial E_I(x_i)\right]. \tag{40}$$

This site-decoupled expression for the energy allows us to calculate the typical energy e at critical points directly from the typical distribution of spins $P(x)$ in Eq. (36) via

$$e = \int dx P(x)\left[E_I(x) - \frac{1}{2}x\partial E_I(x)\right]. \tag{41}$$

Similarly, the typical intensive index r can be calculated, without resorting to the second Legendre dual relation in Eq. (29), by directly using the typical Hessian eigenvalue distribution $\rho_H(\lambda)$ obtained from Eq. (21) using the distribution of $P(x)$ in Eq. (36) inserted into the formula for $R(z)$ in Eq. (22). In terms of this $\rho_H(\lambda)$, r is simply

$$r = \int_{-\infty}^{0} d\lambda \rho_H(\lambda). \tag{42}$$

Overall, these results yield a complete characterization of the typical energy e, index r, grand potential $\Omega(0, \mu)$, complexity $\Sigma(e, r)$, distribution of spins $P(x)$, and distribution of Hessian eigenvalues $\rho_H(\lambda)$ of both typical critical points and typical minima. We successfully confirm these theoretical predictions with numerical simulations below in Sec. VI. But first, we provide an alternate derivation of these results by developing a novel cavity method.

V. A GEOMETRIC INTERPRETATION OF SUPERSYMMETRY BREAKING VIA A GENERALIZED CAVITY METHOD

While the replica-based calculation above provides detailed information about critical points, the form of the answers is difficult to understand. For example, why do the grand potential $\beta\Omega$ in Eq. (30), the distribution of spins $P(x)$ in Eq. (36), and the self-consistent equations for the order parameters in Eq. (39) take the forms that they do? Moreover, what is the geometric meaning of the order parameters, especially the SUSY-breaking order parameters A and C? In essence, what is the qualitative difference between high-dimensional energy landscapes described by broken SUSY versus preserved SUSY? To obtain answers to these questions, we develop a new generalized version of the cavity method and demonstrate the equivalence between our generalized cavity method and replica derivations (see Appendix A and Ref. [44], Sec. S-IV, for detailed derivations). Our generalized cavity method yields considerable conceptual insights into the replica results as well as a geometric interpretation of SUSY breaking.

Note that readers who are less interested in the technical details may skip this section.

A. The naive cavity method

We first take a naive approach to the cavity method, which we see is appropriate when SUSY is preserved. The cavity method, in general, for many mean-field systems involves (i) analyzing the effect of adding a single new degree of freedom to a system (called a cavity system because it excludes the new degree of freedom), (ii) describing how the cavity system responds to the new degree of freedom, often using simple perturbation theory under the assumption that the single new degree of freedom exerts a small effect on the large cavity, and (iii) quantifying how the response of the cavity exerts a backreaction onto the new degree of freedom as it comes to equilibrium with the cavity system. The backreaction of the cavity onto the new degree of freedom depends on certain order parameters associated with the cavity. The cavity method then yields self-consistent equations for these order parameters, assuming the cavity system without the new degree of freedom and the full system with the new degree of freedom have the same order parameters, due to the existence of a thermodynamic limit.

For example, in the context of the CIM, critical points of any index (not necessarily energy minima) obey the gradient equations

$$\partial E_I(x_i) + \sum_{j=1}^{N-1} J_{ij} x_j = 0 \quad \text{for } i = 1, \ldots, N-1. \tag{43}$$

Here, this corresponds to a cavity system with only $N-1$ spins. Next, we introduce a new spin x_0 coupled to the cavity system via new random coupling constants $\{J_{0i}\}_{i=1,\ldots,N-1}$. The gradient equations for $x_1, \ldots x_N$ in the presence of the new spin in the full system become

$$\partial E_I(x_i) + \sum_{j=1}^{N-1} J_{ij} x_j + J_{i0} x_0 = 0, \tag{44}$$

while the new spin, after it equilibrates with the cavity, must obey

$$\partial E_I(x_0) + \sum_{i=1}^{N-1} J_{0i} x_i = 0. \tag{45}$$

The cavity method relates the critical point solutions of the full system in Eqs. (44) and (45) to the critical point solutions of the cavity system in Eq. (43). In particular, let $x_i^{/0}$ for $i = 1, \ldots, N-1$ be a critical point of the cavity system in the *absence* of spin 0. Thus, $x_i^{/0}$ is a solution to Eq. (43) for all $i = 1, \ldots, N-1$. Now, when the new spin 0 is brought into contact with the cavity and held at a *fixed* value x_0, the cavity will react to the new spin so as to solve the modified equations (44), which are simply equivalent to the original cavity equations (43) plus a small perturbative term $J_{i0}x_0$ that is $O(1/\sqrt{N})$.

Assuming the effect of the new spin x_0 on the cavity is small, one can solve Eq. (44) using perturbative linear response theory, by Taylor expanding the first two terms about $x_i = x_i^{/0}$ and using the fact that $x_i^{/0}$ satisfies Eq. (43). The resulting approximate linear response of the cavity to the new spin x_0 [i.e., approximate solution to Eq. (44)] is

$$x_i = x_i^{/0} - \sum_{j=1}^{N} H_{ij}^{-1}(\mathbf{x}^{/0}) J_{j0} x_0. \tag{46}$$

Here, $H_{ij}^{-1}(\mathbf{x}^{/0})$ is the inverse Hessian of the cavity system evaluated at its critical point $\mathbf{x}^{/0}$ *before* the new spin x_0 is introduced. As usual, this inverse Hessian acts as a linear susceptibility matrix $\chi = H^{-1}(\mathbf{x}^{/0})$ that translates the force $J_{j0}x_0$ exerted by the new spin into the response of the cavity from $x_i^{/0}$ to x_i in Eq. (46).

Now, with Eq. (44) solved perturbatively via the cavity response in Eq. (46) for arbitrary x_0, we must next find the equilibrium value of x_0 that generates an approximate critical point of the full system by inserting Eq. (46) into Eq. (45), obtaining

$$\partial E_I(x_0) - \sum_{i,j=1}^{N} J_{0i} H_{ij}^{-1}(\mathbf{x}^{/0}) J_{j0} x_0 + \sum_{i=1}^{N} J_{0i} x_i^{/0} = 0. \tag{47}$$

Here, the final term

$$h^{/0} \equiv -\sum_{i=1}^{N} J_{0i} x_i^{/0} \tag{48}$$

is the cavity field that the cavity would have exerted on the new spin had it not reacted to the new spin at all and remained at configuration $\mathbf{x}^{/0}$. The second term takes into account the reaction of the cavity to x^0 through the force $J_{j0}x^0$ and its resultant backreaction on the new spin through the connections J_{0i}. This is an example of an Onsager backreaction-type term [32].

Now both the cavity field and the backreaction term depend on the cavity system through two simple order parameters. First, note that $x_i^{/0}$, a critical point of the cavity system in the absence of the new spin x_0, is necessarily independent of the new connectivity J_{0i}, which is not a part of the cavity system. Thus, we can apply the central limit theorem to conclude that $h^{/0}$ in Eq. (48) is a random Gaussian variable distributed as $\mathcal{N}(0, q)$, where the variance q is an order parameter given by

$$q = \frac{1}{N-1}\sum_{i=1}^{N-1}(x_i^{/0})^2 = \frac{1}{N}\sum_{i=0}^{N-1}(x_i)^2. \tag{49}$$

Here, we assume that the order parameter q is self-averaging and is the same in both the cavity system and the full system, at large N. Similarly, we assume that the Onsager backreaction term is self-averaging, and we replace it with its average over the connectivity in Eq. (15), yielding a second-order parameter t which we assume is the same both in the cavity and the full system:

$$t = \frac{1}{N-1}\mathrm{Tr}H^{-1}(\mathbf{x}^{/0}) = \frac{1}{N}\mathrm{Tr}H^{-1}(\mathbf{x}). \tag{50}$$

While q is the squared length of a critical point, t is the trace of a critical point's linear susceptibility matrix to small external forces.

With the definition of the cavity field $h^{/0}$ in Eq. (48) and the order parameters q in Eq. (49) and t in Eq. (50), the solution(s) of x_0 in Eq. (47) is in one-to-one correspondence with critical points of a mean-field energy function

$$E_{\mathrm{MF}}[h](x) \equiv E_I(x) - \frac{1}{2}tx^2 - hx, \tag{51}$$

where the random external cavity field $h \sim \mathcal{N}(0, q)$. Here, we drop the index 0 from both the new spin x_0 and its cavity field $h^{/0}$, because, under the random mean field connectivity in Eq. (15), there is nothing special about removing and adding back spin 0. We could have done this for any spin x_i, yielding its own cavity field $h^{/i}$ which is also distributed as $\mathcal{N}(0, q)$. Moreover, each cavity field $h^{/i}$ in the absence of x_i is independent of any other cavity field $h^{/j}$ in the absence of x_j. Therefore, the empirical distribution of spins x_i across the index i, defined as $P(x) = (1/N)\sum_{i=1}^{N}\delta(x - x_i)$ can be obtained, in the large-N limit, as

$$P(x) \propto \left\langle \sum\nolimits_{x_* \in \mathrm{Crt}(E_{\mathrm{MF}}[h])} \delta(x - x_*) \right\rangle_h, \tag{52}$$

where $\mathrm{Crt}(E_{\mathrm{MF}}[h])$ denotes the set of critical points of the function $E_{\mathrm{MF}}[h]$ in Eq. (51) and $\langle\cdot\rangle_h$ denotes an average with respect to the Gaussian cavity field $h \sim \mathcal{N}(0, q)$. The normalization factor in Eq. (52) is simply the mean number of critical points in the random ensemble of mean-field energy functions $E_{\mathrm{MF}}[h]$.

Now, with the distribution of spins $P(x)$ in a typical critical point in hand, we can derive self-consistent equations for the order parameters. In particular, it is clear that q in Eq. (49) is simply the second moment of $P(x)$, yielding the self-consistent equation

$$q = \int dx x^2 P(x). \tag{53}$$

Furthermore, t in Eq. (50) is simply the mean of the diagonal elements of the inverse Hessian. The Hessian of the mean-field energy function is given by $H(x) = \partial^2 E_I(x) - t$ and is independent of the cavity field h. Taking the average of its inverse yields the self-consistent equation

$$t = \int dx \frac{P(x)}{\partial^2 E_I(x) - t}. \tag{54}$$

Together, Eqs. (51)–(54) constitute a theoretical prediction for the distribution of spins in a typical critical point [i.e., the special case of $\beta = \mu = 0$ in Eq. (35)]. Interestingly, the cavity result appealingly and intuitively replaces the problem of summing over critical points in a large-N-dimensional system [i.e., Eqs. (2) and (35) with $\mu = \beta = 0$] with the problem of summing over critical points in a random ensemble of one-dimensional systems [i.e., Eqs. (51) and (52)].

B. Equivalence of the naive cavity method with the supersymmetric replica method

We next show that these cavity results are exactly equivalent to those of the replica method in the further special case where SUSY is preserved (i.e., $A = C = 0$). We can demonstrate the equivalence of the cavity result for $P(x)$ in Eq. (52) with the replica result for $P(x)$ in Eq. (36) with $\mu = A = C = 0$ as follows. First, we can apply the Kac-Rice formula in Eq. (32) to Eq. (52) and perform the resulting integral over x_* which simply fixes it to x, yielding

$$P(x) \propto \langle \delta[\partial E_I(x) - tx - h] | \partial^2 E_I(x) - t| \rangle_h. \tag{55}$$

Then, performing the integral over h fixes it to be $h(x) = \partial E_I(x) - tx$, and, recalling that $\langle\cdot\rangle_h$ denotes an average with respect to the Gaussian distribution $\mathcal{N}(0, q)$, we obtain

$$P(x) \propto |\partial^2 E_I(x) - t| \exp\left(-\frac{h(x)^2}{2q}\right), \tag{56}$$

where $h(x) = \partial E_I(x) - tx$ is the external field h required to make x a critical point of the mean-field energy function in Eq. (51). Thus, the distribution of spins $P(x)$ in Eq. (56), and, therefore, in Eq. (52), is entirely equivalent to the replica expression for $P(x)$ when $\mu = A = C = 0$.

Moreover, given this equivalence of $P(x)$, the cavity-derived self-consistent equations for the order parameters q in Eq. (53) and t in Eq. (54) are entirely equivalent to the first two self-consistent equations derived via the replica method in Eq. (39). Thus, overall, the naive cavity method recovers the results of the *supersymmetric* solution but cannot account for supersymmetry breaking.

C. Beyond the naive cavity method: Accounting for supersymmetry breaking

Why does the naive cavity method recover the replica results only in the case of preserved SUSY—i.e., Eq. (36) when $A = C = 0$ and the first two equations in Eq. (39)? Here, we resolve this issue as well as generalize to nonzero μ. The key idea is that the naive cavity method makes an implicit assumption about the nature of the perturbative reaction of the cavity system to the addition of a single new spin in Eq. (46). In particular, this account of the reaction assumes that the only effect of adding a new spin x_0 is to move every critical point of the cavity system a small amount to generate a critical point of the full system. Thus, it is assumed that critical points of the cavity system and the full system are in one-to-one correspondence with each other.

This assumption is likely to be valid if the Hessian matrix $H(\mathbf{x})$ has an eigenvalue distribution $\rho_H(\lambda)$ which vanishes in a finite region about $\lambda = 0$. Because the susceptibility matrix χ is the inverse Hessian, such a gap in the Hessian spectrum yields a nondegenerate, structurally stable critical point that is unlikely to undergo a bifurcation or change its index upon the addition of a new spin. However, if the Hessian spectral density $\rho_H(\lambda)$ extends continuously to $\lambda = 0$, such a critical point is degenerate with extremely soft modes, and the addition of a single spin could cause it to either disappear or bifurcate to create additional critical points. If the landscape has exponentially many critical points whose typical Hessian eigenspectra are gapless, then the addition of a single new spin x_0 could lead to exponentially more or fewer critical points of any given index, depending on the realization of the couplings J_{i0} and the value of x_0 at its own equilibrium. This extreme reactivity of the landscape to the addition of a single spin, marked by an exponential change in the number of critical points, is a fundamental possibility that is not accounted for by the naive cavity method.

We provide a generalized cavity method that can account for this extreme reactivity. We provide a detailed derivation in Appendix A. Readers who are not familiar with spin-glass methods can refer to the step-by-step derivations in Ref. [44], Sec. S-IV, as well. Here, we simply outline the key ideas and intermediate results. Our generalized cavity method starts from the expression for the grand potential:

$$-\beta\Omega = \frac{1}{N}\ln\langle Z\rangle_J, \tag{57}$$

where Z is the partition function given in Eq. (33). Thus, we start from an annealed approximation. Next, because of the critical importance of the presence of soft modes in the energy landscape in the vicinity of critical points, corresponding to eigenvectors of the Hessian with small eigenvalues, we soften the δ functions of the gradient in Eq. (33) and replace them with Gaussians via

$$\delta[\partial_i E(\mathbf{x})] \to \sqrt{\frac{\gamma}{\pi}} e^{-\gamma[\partial_i E(\mathbf{x})]^2}. \tag{58}$$

We work at finite γ throughout the calculation, taking $\gamma \to \infty$ at the end. A finite γ crucially allows the partition function Z in Eq. (33) to receive contributions not only from critical points, but also from the geometry of the landscape in the vicinity of critical points, including the nature of the nonzero gradient in the neighborhood of each critical point.

Next, we split the degrees of freedom $\mathbf{x}$ into that of a cavity system $\mathbf{x}^{/0}$ with components $x_i^{/0}$ for $i = 1, \ldots, N-1$ and a single spin x_0. Mirroring this split, we would like to express the grand potential of the full system in Eq. (57) in terms of the grand potential of the cavity system $\mathbf{x}^{/0}$ (taking into account the effect of the new spin on it) and an effective mean-field grand potential of the new spin x_0 (taking into account the effect of the cavity on it in terms of certain cavity fields and order parameters). Achieving this decomposition *prima facie* poses several challenges, because x_0 and $\mathbf{x}^{/0}$ appear intricately coupled in the expressions for the Hessian determinant $|\det H(\mathbf{x})|$ and the Hessian index $\mathcal{I}(\mathbf{x})$ in Z in Eq. (33). Despite this seemingly intricate coupling, we can show that, upon averaging over the random choice of coupling J_{i0} between the cavity $\mathbf{x}^{/0}$ and the new spin x_0, the interaction between them depends on the cavity system $\mathbf{x}^{/0}$ *only* through the mean cavity susceptibility order parameter t, defined in Eq. (50).

In particular, for the Hessian determinant, we show that, after averaging over J_{i0},

$$|\det H(\mathbf{x})| = |\partial^2 E_I(x_0) - t||\det H(\mathbf{x}^{/0})|. \tag{59}$$

The first term is nothing other than the absolute value of the Hessian of the mean-field energy function E_{MF} in Eq. (51) evaluated at $x = x_0$, while the second term is the same Hessian determinant for the cavity system.

Similarly, for the index of the Hessian, we show that, after averaging over J_{i0},

$$\mathcal{I}(\mathbf{x}) = \bar{\mathcal{I}}(x_0) + \mathcal{I}(\mathbf{x}^{/0}). \tag{60}$$

Here, $\bar{\mathcal{I}}(x_0)$ is defined in Eq. (37) and can be interpreted simply as the index of the mean-field energy function E_{MF} in Eq. (51) evaluated at $x = x_0$. Thus, remarkably, the index of the full system is simply the sum of the index of the mean-field system and the cavity system, on average.

Now assuming formulas (59) and (60) are self-averaging (i.e., they also hold to high accuracy for typical random choices of J_{i0}), we can substitute these formulas into Eq. (33), thereby achieving a partial decomposition of the full partition function Z into that of a cavity system of size $N-1$ and a mean-field system of size 1, coupled so far

only through the cavity susceptibility order parameter t in Eq. (50).

However, to fully complete this decomposition, we must also account for interactions between the cavity system $\mathbf{x}^{/0}$ and the single spin x_0 through the Gaussian softening in Eq. (58) of the δ functions in Eq. (33). We show that these interactions are mediated precisely by two fields:

$$\bar{h} \equiv -\mathbf{J}_0 \cdot \mathbf{x}^{/0},$$
$$\bar{z} \equiv -\gamma \mathbf{J}_0 \cdot \nabla E(\mathbf{x}^{/0}). \tag{61}$$

Here, $\mathbf{J}_0$ is the $N-1$ dimensional coupling vector between x_0 and the cavity system $\mathbf{x}^{/0}$. Note that $\bar{h}$ and $\bar{z}$ are jointly Gaussian distributed with a 2-by-2 covariance matrix that depends on cavity order parameters specified by inner products of $\mathbf{x}^{/0}$ and $\nabla E(\mathbf{x}^{/0})$. In particular, $\bar{h}$ is simply the Gaussian cavity field that already appears in the naive cavity method in Eq. (48) with variance q in Eq. (49).

But, most importantly, $\bar{z}$ is a new cavity field that appears only in our generalized cavity method and plays a fundamental role in accounting for the extreme sensitivity of the landscape to the addition of a new spin x_0. In particular, as detailed in Ref. [44], Sec. S-IV.B.4, the field $\bar{z}$ couples the new spin x_0 to the cavity system through an exponential modification of the partition function Z in Eq. (33) via a multiplicative factor $\exp(x_0\bar{z})$. Given the form of $\bar{z}$ in Eq. (61), this means that if the coupling vector $\mathbf{J}_0$ were aligned to the cavity gradient $\nabla E(\mathbf{x}^{/0})$ in the vicinity of a typical critical point, so that $\bar{z}$ is negative, then the partition function Z would be exponentially enhanced (diminished) if x_0 were to assume larger negative (positive) values. Conversely, if $\mathbf{J}_0$ were antialigned to the cavity gradient $\nabla E(\mathbf{x}^{/0})$ so that $\bar{z}$ were positive, then Z would be exponentially enhanced (diminished) if x_0 were to assume larger positive (negative) values. The end result of the field $\bar{z}$ is then to exponentially reweight the distribution of spins in the mean-field theory of a single spin x_0 according to the exponential weight that different values of x_0 exert on the cavity partition function and, therefore, on the grand potential and the complexity. Thus, while the usual cavity field $\bar{h}$ exerts a force on the new spin x_0 through an energy term $-\bar{h}x_0$ in the mean-field energy function E_{MF} in Eq. (51), we see that the new field $\bar{z}$ yields an entropic force on the new spin x_0 through the proliferation or destruction of exponentially many critical points in the cavity system for different values of x_0.

Now, in order to take the $\gamma \to \infty$ limit, it is useful not to work directly with the fields $\bar{h}$ and $\bar{z}$ but to perform a change of variables (detailed in Appendix A and Ref. [44], Sec. S-IV.B.4) to h and z which remain jointly Gaussian distributed with density $P(h,z)$ given by

$$P(h,z) \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}q & A\\ A & C\end{bmatrix}\right). \tag{62}$$

Here, the covariance parameters at finite γ are given by

$$q = \frac{1}{N-1}|\mathbf{x}^{/0}|^2,$$
$$A = \frac{2}{N-1}\gamma \nabla E(\mathbf{x}^{/0}) \cdot \mathbf{x}^{/0} - t,$$
$$C = \frac{4}{N-1}\gamma^2 |\nabla E(\mathbf{x}^{/0})|^2 - 2\gamma \tag{63}$$

and correspond to cavity order parameters involving inner products of $\mathbf{x}^{/0}$ and $\nabla E(\mathbf{x}^{/0})$ in the vicinity of critical points.

Now, with the definition of the cavity order parameters q, A, and C in Eq. (63) and t in Eq. (50), as well as the Gaussian fields h and z with distribution $P(h,z)$ in Eq. (62), we can achieve a decomposition of the partition function Z in Eq. (33), and, therefore, of the grand potential Ω in Eq. (57), into a cavity system $\mathbf{x}^{/0}$ and a single spin x_0. However, there is one remaining issue: The resultant grand potential of the cavity system has a mismatched variance; the size of the cavity system is $N-1$, while the variance of its connectivity in Eq. (15) for $g=1$ is $1/N$. Given the potentially extreme reactivity of the energy landscape, we cannot ignore this mismatch. Indeed, to obtain self-consistent equations for the order parameters q, A, C, and t of the full system, we must analyze the susceptibility of the grand potential in response to small changes in the variance of its connectivity. We obtain a simple formula for this susceptibility in terms of the cavity order parameters:

$$\left.\frac{d\Omega(g)}{dg}\right|_{g=1} = \frac{1}{2}(qC + A^2) + At. \tag{64}$$

Here, $\Omega(g)$ denotes the grand potential of the full system with a general variance parameter g in Eq. (15). (See Ref. [44], Sec. S-IV.B.6, for details.)

Finally, putting everything together and taking the $\gamma \to \infty$ limit, we find that the grand potential in Eq. (57) or, equivalently, the annealed connectivity average of the grand potential in Eq. (27) (in the special case of $\beta = 0$ relevant to typical critical points and minima) is given by

$$-\beta\Omega(0,\mu) = -\left.\frac{d\Omega(g)}{dg}\right|_{g=1} - \Omega_{\mathrm{MF}}. \tag{65}$$

Here, the first term is a simple function of the order parameters given in Eq. (64), while the second term is the mean-field grand potential Ω_{MF} of a single spin given by

$$-\Omega_{\mathrm{MF}} = \log\langle Z_{\mathrm{MF}}[h,z]\rangle_{h,z}, \tag{66}$$

where $\langle\cdot\rangle_{h,z}$ denotes an average over the Gaussian distribution $P(h,z)$ of cavity fields in Eq. (62) and $Z_{\mathrm{MF}}[h,z]$ denotes the mean-field partition function of a single spin in the presence of cavity fields h and z, given by

$$Z_{\mathrm{MF}}[h,z] = \sum_{x \in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)}. \tag{67}$$

Here, as above, $\mathrm{Crt}(E_{\mathrm{MF}}[h])$ denotes the set of critical points of the mean-field energy function $E_{\mathrm{MF}}[h](x)$ in Eq. (51), and the mean-field index function $\bar{\mathcal{I}}(x)$ defined in Eq. (37) is simply the index of E_{MF} evaluated at x. Finally, our generalized cavity method computes the distribution of spins in a typical critical point, defined in Eq. (35), to be (see Ref. [44], Sec. S-IV.C, for details)

$$P(x) \propto \left\langle \sum\nolimits_{x_* \in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)} \delta(x - x_*) \right\rangle_{h,z}. \tag{68}$$

Appealingly, Eqs. (66)–(68) all correspond to the problem of counting critical points in a random ensemble of one-dimensional systems with mean-field energy functions $E_{\mathrm{MF}}[h]$ subject to a random external field cavity field h, in addition to a random exponential factor e^{xz} that reweights both the partition function $Z_{\mathrm{MF}}[h,z]$ in Eq. (67) and the spin distribution $P(x)$ in Eq. (68). Notably, when the variance C of z and, therefore, the covariance A between h and z is 0, the reweighting factor e^{xz} plays no role, and $P(x)$ in Eq. (68) reduces to the prediction of the naive cavity method in Eq. (52) (when $\mu = 0$). The origin of this reweighting factor for nonzero A and C, as summarized above and described in detail in Ref. [44], Sec. S-IV.B, arises from entropic effects in the cavity system due to exponential changes in the number of critical points, depending on the value x of an added spin and the random alignment z of its coupling vector to the cavity system gradient near critical points. This entropic effect of the cavity on the new spin is encapsulated in the mean-field theory of the new spin simply through the reweighting factor e^{xz}.

Finally, we can *directly* obtain self-consistent equations for the order parameters q, A, C, and t through our generalized cavity method (see Appendix A or Ref. [44], Sec. S-IV.C, for details). The self-consistent equations for q and t are identical in form to those obtained in the naive cavity method in Eqs. (53) and (54), respectively, with the sole difference being that the distribution of spins $P(x)$ obtained in the naive cavity method in Eq. (52) is replaced with the reweighted distribution of spins $P(x)$ obtained in the generalized cavity method in Eq. (68). The generalized cavity method also enables us to find self-consistent equations for the two new order parameters A and C (see Appendix A or Ref. [44], Sec. S-IV.C, for details):

$$\begin{aligned} A &= e^{\Omega_{\mathrm{MF}}} \left\langle \frac{\partial}{\partial h} \left[\sum\nolimits_{x \in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)} x \right] \right\rangle_{h,z} - t, \\ C &= e^{\Omega_{\mathrm{MF}}} \left\langle \frac{\partial^2 Z_{\mathrm{MF}}}{\partial h^2} \right\rangle_{h,z}. \end{aligned} \tag{69}$$

Furthermore, we show in Ref. [44], Sec. S-IV.C, that the cavity-derived self-consistent equations for the order parameters q in Eq. (53), t in Eq. (54), and A and C in Eq. (69) are collectively equivalent to the four equations obtained from extremizing the grand potential Ω in Eq. (65) with respect to q, t, A, and C. This extremization yields the highly compact self-consistent equations

$$\begin{bmatrix} q & A \\ A & C \end{bmatrix} = -\begin{bmatrix} 2\partial_C \Omega_{\mathrm{MF}} & \partial_t \Omega_{\mathrm{MF}} \\ \partial_t \Omega_{\mathrm{MF}} & 2\partial_q \Omega_{\mathrm{MF}} \end{bmatrix}, \qquad t = \partial_A \Omega_{\mathrm{MF}} - A. \tag{70}$$

D. Equivalence between the generalized cavity method and the supersymmetry-broken replica method

We now show the equivalence between the generalized cavity method and SUSY-breaking replica method, identifying the cavity order parameters A and C in the generalized cavity method in Eq. (63) with the SUSY-breaking order parameters A and C in Eqs. (34), (36), and (39).

First, we note that the distribution of spins $P(x)$ in the generalized cavity method in Eq. (68) is entirely equivalent to the distribution of spins derived via the SUSY-broken replica method in Eq. (36) for any values of the order parameters A and C (as well as q and t). This can be seen by applying the Kac-Rice formula to Eq. (68) and directly performing the integrals over h and then z (see Ref. [44], Sec. S-IV.F, for details).

Second, we note that the formula for the grand potential Ω derived by the generalized cavity method in Eq. (65) is entirely equivalent to that obtained by the SUSY-breaking replica method in Eq. (34). However, the generalized cavity method now provides a simple interpretation of each of the terms in Eq. (34). In particular, the first part $-\frac{1}{2}(Cq + A^2) - At$ of Eq. (34) is equivalent to the first term in Eq. (65) and is simply the susceptibility of the grand potential to a change in variance, derived in Eq. (64). Its origin lies in the mismatch between the size of the cavity system $(N-1)$ and its connectivity variance $(1/N)$. The remaining term in Eq. (34) is equivalent to the remaining term $-\Omega_{\mathrm{MF}}$ in Eq. (65) and is simply the grand potential of an ensemble of *single* spins defined in Eqs. (66) and (67). This equivalence can be seen by applying the Kac-Rice formula to the mean-field partition function in Eq. (67) and performing the integrals over h and z in Eq. (66). This calculation yields the final term in Eq. (34) (see Ref. [44], Sec. S-IV.F, for details).

Thus, we conclude that the expressions for Ω in the generalized cavity method in Eq. (65) and the SUSY-breaking replica method in Eq. (34) are equivalent. However, the generalized cavity method provides the important intuition, embodied in Eq. (65), that the grand potential density Ω of the full N-dimensional system is

simply the mean-field grand potential Ω_{MF} of an ensemble of random one-dimensional systems in Eq. (66) plus a correction given in Eq. (64) due to the extreme reactivity of the landscape to changes in connectivity variance.

Finally, given the equivalence of the grand potentials derived via the generalized cavity and SUSY-breaking replica method, as well as the demonstration in the previous subsection that the self-consistent equations for the order parameters in the generalized cavity method can be obtained by extremizing the grand potential, we can conclude that these self-consistent equations are the same in both methods. To further corroborate this conclusion, we provide a direct proof in Ref. [44], Sec. S-IV.F, that the self-consistent equations derived via the generalized cavity method for A and C in Eq. (69) are equivalent to those derived via the SUSY-breaking replica method in Eq. (39). Moreover, the equations for q and t in Eqs. (53) and (54), respectively, are manifestly equivalent to those derived via the SUSY-breaking replica method in Eq. (39) given the equivalence of the distributions $P(x)$ in Eqs. (68) and (36).

E. Supersymmetry-breaking order parameters in terms of landscape susceptibility

The equivalence of the generalized cavity method and the SUSY-breaking replica method at the level of the grand potential Ω, self-consistent equations for order parameters, and distribution of spins $P(x)$, thus, identifies the parameters A and C in the generalized cavity method with the supersymmetry-breaking order parameters of the SUSY-breaking replica method. We now provide a further geometric interpretation of these order parameters in terms of the susceptibility of the grand potential Ω to changes in the energy landscape. In particular, consider adding two extra perturbative terms to the original energy landscape $E_{\mathrm{tot}}(\mathbf{x})$ in Eq. (2) to obtain

$$E'_{\mathrm{tot}}(\mathbf{x}) = E_{\mathrm{tot}}(\mathbf{x}) - \frac{1}{2}a_0|\mathbf{x}|^2 + \sqrt{2s_0}\mathbf{g}\cdot\mathbf{x}. \quad (71)$$

Here, $\mathbf{g}$ is a zero mean random Gaussian vector with identity covariance. Given the structure of the single-site energy function $E_I(x)$ in Eq. (3), the first perturbation in Eq. (71) corresponds in the CIM to changing the laser gain from a to $a + a_0$. The second perturbation corresponds to applying a random Gaussian field on the landscape with a variance on each component of $2s_0$. We can then consider computing the grand potential density $\Omega(a_0, s_0)$ (we suppress the dependence on $\beta = 0$ and μ here) in Eq. (27) by replacing E_{tot} in Eq. (2) with E'_{tot} in Eq. (71) and further averaging over the random field $\mathbf{g}$. We show in Appendix A (see Ref. [44], Sec. S-IV.D, for detailed explanations) that the supersymmetry-breaking order parameters A and C are very simply related to the (connectivity and field averaged) susceptibility of the grand potential Ω with respect to the perturbation strengths a_0 and s_0, respectively:

$$A = \left.\frac{\partial\Omega}{\partial a_0}\right|_{a_0=0,s_0=0}, \qquad C = \left.\frac{\partial\Omega}{\partial s_0}\right|_{a_0=0,s_0=0}. \quad (72)$$

This result *directly* connects the supersymmetry-breaking order parameters to the extreme reactivity of the landscape to two specific small perturbations of the energy function. In particular, A and C are nonzero if and only if the potential Ω and, therefore, the landscape complexity Σ are sensitive to these perturbations.

The expressions for A and C in Eq. (72) also have a counterpart in the mean-field theory of a single spin. Consider adding the same two perturbations to the mean-field energy function $E_{\mathrm{MF}}[h]$ in Eq. (51), obtaining the perturbed energy function

$$E'_{\mathrm{MF}}[h, a_0, s_0](x) = E_{\mathrm{MF}}[h](x) - \frac{1}{2}a_0x^2 + \sqrt{2s_0}gx, \quad (73)$$

where g is now a zero mean unit variance random Gaussian scalar field. We can then consider computing the mean-field grand potential $\Omega_{\mathrm{MF}}(a_0, s_0)$ obtained by replacing $E_{\mathrm{MF}}[h](x)$ with $E'_{\mathrm{MF}}[h, a_0, s_0](x)$ in the defining formula for Ω_{MF} in Eq. (66) and also further averaging over g. Then, we show in Appendix A that the supersymmetry-breaking order parameters A and C are also very simply related to the susceptibility of the mean-field grand potential Ω_{MF} with respect to the perturbation strengths a_0 and s_0, respectively:

$$A = \left.\frac{\partial\Omega_{\mathrm{MF}}}{\partial a_0}\right|_{a_0=0,s_0=0}, \qquad C = \left.\frac{\partial\Omega_{\mathrm{MF}}}{\partial s_0}\right|_{a_0=0,s_0=0}. \quad (74)$$

This result provides an additional way to interpret the SUSY-breaking order parameters A and C within the mean-field theory and exhibits an appealing correspondence to Eq. (72).

F. Structural stability of critical points implies preserved supersymmetry

We further connect the Hessian eigenspectrum to SUSY breaking by showing in Appendix A (or in Ref. [44], Sec. S-IV.E) that if the typical Hessian eigenspectrum of critical points has a gap away from 0, then the SUSY is preserved, and $A = C = 0$. We do this by working at large but finite γ and directly calculating A and C through Eq. (63) and averaging over $\mathbf{x}^{/0}$ (or, equivalently, $\mathbf{x}$) with respect to a distribution with partition function given by Z in Eq. (31), with δ functions softened to Gaussians via Eq. (58). The key idea is that this distribution concentrates in the vicinity of critical points, and if the Hessian has a gap at typical critical points, one can perform a change of variables from $\mathbf{x}$ to $\nabla E(\mathbf{x})$, since there is a one-to-one map between these quantities in the neighborhood of any critical point with a gapped Hessian eigenspectrum. Direct calculation of the integral over gradients in the vicinity of a critical point then reveals that $A = C = 0$.

Thus, if typical critical points are structurally stable (i.e., with gapped Hessian eigenspectra), SUSY is preserved. The contrapositive of this statement then tells us that SUSY breaking implies a vanishing gap in the Hessian eigenspectrum and, therefore, structural instability in typical critical points. If exponentially many critical points have such structural instability, then SUSY will be broken. This analytic calculation provides further justification for why the generalized cavity method (and not just the naive cavity method) is necessary when exponentially many critical points are structurally unstable.

G. Convexity of the mean-field energy landscape implies preserved supersymmetry

Suppose the mean-field energy function $E_{\mathrm{MF}}[h]$ in Eq. (51) is strictly convex. This happens only if $a+t<0$. We provide a proof in Appendix A (or in Ref. [44], Sec. S-IV.E) that, under this assumption of convexity, the self-consistent equations for the order parameters in Eq. (69) admit SUSY-preserving solutions with $A=C=0$. We find below in Sec. VI that these SUSY-preserving solutions correctly predict the distribution of spins and Hessian eigenspectra when $a+t<0$. In contrast, we also see that when $a+t>0$, and, therefore, the mean-field energy function $E_{\mathrm{MF}}[h]$ is nonconvex, we must use SUSY-breaking solutions to correctly predict the distribution of spins and Hessian eigenspectra. Thus, just as SUSY breaking implies a highly reactive landscape in the full N-dimensional system, as evidenced by the susceptibility formulas in Eq. (72), at the level of the mean-field one-dimensional system, SUSY breaking is closely related to nonconvexity in the mean-field energy landscape.

We further show in Appendix A that, when the mean-field energy function is convex, SUSY-preserving solutions also exhibit vanishing complexity Σ. Thus, convexity of the mean-field landscape implies simplicity of the full landscape under SUSY. Thus, in summary, the predictions of our replica and generalized cavity theories are that, for $a+t<0$, SUSY is preserved and landscape complexity is 0, and, when $a+t>0$, SUSY is broken.

VI. NUMERICAL TESTS OF SUPERSYMMETRY BREAKING FOR TYPICAL CRITICAL POINTS AND MINIMA

A. A supersymmetry-breaking phase transition in the properties of typical critical points

Here, we test our theoretical predictions for the structure of typical critical points derived from the grand potential in Eq. (34) or, equivalently, Eq. (65) with $\mu=0$. We directly sample critical points of all indices in many finite-size SK models. Across this ensemble of critical points, the distribution of intensive index r and energy E peak sharply at their respective most likely values. Focusing on these typical critical points (see Appendix B 1 for details of the numerical sampling of critical points), we can measure the distribution of spins $P(x)$, the distribution diagonal elements $H^I(\mathbf{x})_{ii}$ in Eq. (6), and the distribution of Hessian eigenvalues $\rho_H(\lambda)$, shown as blue histograms in Figs. 3(a)–3(c), respectively. We can further compare these observables to the theoretical predictions for $P(x)$ in Eq. (36) or (52), the distribution of $H^I(\mathbf{x})_{ii}$ derived from $P(x)$, and the Hessian eigenspectrum $\rho_H(\lambda)$ derived from Eqs. (21) and (22). We obtain an excellent match between theory and experiment for a range of laser gain a [compare orange curves and blue histograms in Figs. 3(a)–3(c)].

A key feature of these results is that, as the laser gain a is increased, the distribution of OPO amplitudes in Fig. 3(a) bifurcates into a bimodal then trimodal distribution with an increasing density of uncommitted spins with values near the origin. Correspondingly, the distribution of diagonal Hessian eigenvalues in Fig. 3(b) exhibits an increasing density of negative values originating from these uncommitted spins, which then corresponds to an increasing density of negative Hessian eigenvalues in Fig. 3(c) via Dyson's Brownian motion in Eq. (17), starting from the initial condition in Fig. 3(b).

We further compute the order parameters q, t, A, and C arising from solutions of Eq. (39) or, equivalently, Eq. (69) or (70) (see Appendix B 2 for numerical details of solving these self-consistent equations). Figure 3(d) shows the evolution of A and C with increasing laser gain a, indicating a supersymmetry-breaking phase transition at $a=a_t\sim-0.93$, when A and C first acquire nonzero values. Figure 3(e) shows the evolution of $a+t$, which, in both theory and experiment, transitions from negative to positive also at $a=a_t\sim-0.93$. Recall that the mean-field energy function $E_{\mathrm{MF}}[h](x)$ in Eqs. (51) and (3) is convex if and only if $a+t<0$. Thus, together, Figs. 3(d) and 3(e) confirm our theoretical prediction that SUSY is broken precisely when the mean-field energy function becomes nonconvex.

Finally, Figs. 3(f)–3(i) demonstrate an excellent match between theory (orange curves) and experiments (blue dots) for the order parameter q, the complexity Σ [derived from Eq. (28)], the intensive index r [derived from Eq. (42)], and the intensive energy E [derived from Eq. (41)], respectively. In particular, the complexity in Fig. 3(g) becomes nonzero at the same transition $a=a_t\sim-0.93$ when SUSY is broken and the mean-field energy function becomes nonconvex.

Quite remarkably, the intensive index r and complexity Σ are exactly zero for $a<a_t$ [Figs. 3(g) and 3(h)]. This means that most critical points have a vanishing intensive index, and the number of critical points is subexponential. Therefore, we can expect that the energy landscape is relatively flat and not so rugged at such low laser gain a. In contrast, at very large a, the complexity approaches $\log 3$ and the intensive index r approaches $1/3$, as expected from the discussion in Sec. II B 2, which suggests the existence of 3^N critical points at large a located near the points $\{-\sqrt{a},0,\sqrt{a}\}^N$, in which a typical critical point has $1/3$ of

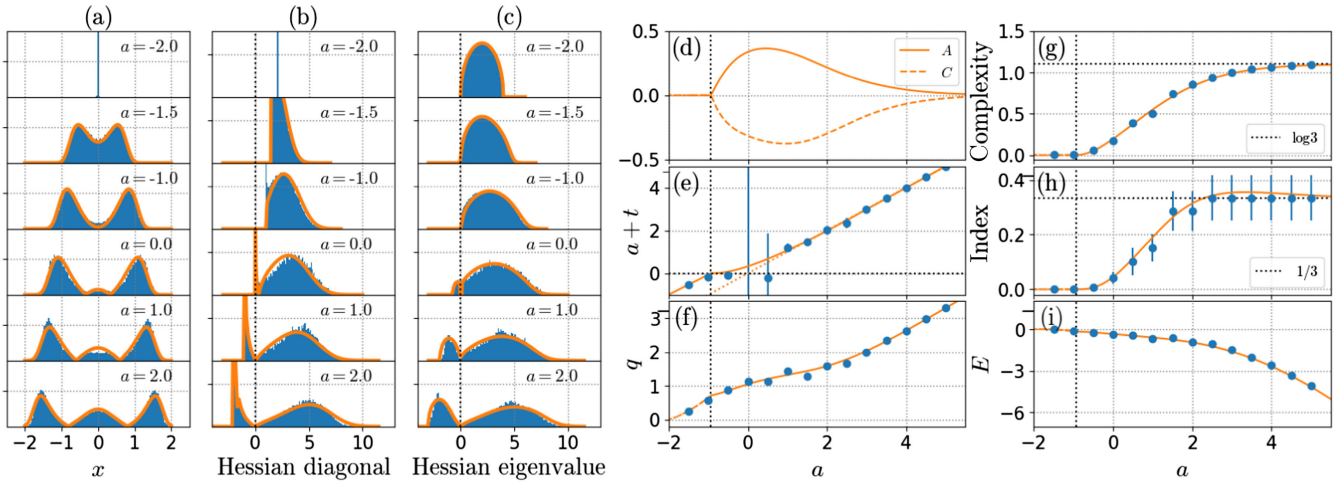

FIG. 3. A supersymmetry-breaking phase transition in the properties of typical critical points. All panels indicate theoretical results (orange curves) and experimental results (blue histograms and markers). Details of experimental results are in Appendix B 1, and details of theoretical solutions are in Appendix B 2. (a) The distribution of spins or OPO amplitudes as laser gain a increases. (b) The distribution of Hessian diagonal elements $H^I(\mathbf{x})_{ii}$. (c) The Hessian eigenspectrum $\rho_H(\lambda)$. The Hessian eigenspectra in (c) can be understood intuitively as the outcome of Dyson's Brownian motion starting from the initial condition of diagonal elements in (b). (d)–(f) The order parameters A and C, $a+t$, and q, respectively. A supersymmetry-breaking phase transition is observed at the theoretically predicted point $a = a_t \sim -0.93$ (vertical dotted line). (g)–(i) The complexity Σ, the intensive index r, and the intensive energy E, respectively, of typical critical points.

its spins uncommitted near 0, contributing to a typical index of $r = 1/3$.

In summary, our combined theory and experiment uncovers a phase transition between a supersymmetric phase (when $a < a_t \sim -0.93$), where the intensive index and complexity of typical critical points is 0, the number of critical points is subexponential in N, and the mean-field energy function is convex, and a supersymmetry-broken phase (when $a > a_t \sim -0.93$), where the intensive index and complexity of critical points is finite, there are exponentially many structurally unstable critical points, and the mean-field energy function is nonconvex.

B. A supersymmetry-breaking phase transition in the properties of typical minima

We next test our theoretical predictions for the structure of typical minima, derived from the grand potential in Eq. (34) or, equivalently, Eq. (65) with $\mu \to -\infty$. The theoretical calculations are entirely parallel to those of the previous subsection, with the sole replacement of $\mu = 0$ with $\mu \to -\infty$, and the experimental results are also parallel with the sampling restricted to minima as opposed to saddle points of any index. We further compute the binned energy of all minima found and focus on the typical minima with the most likely binned value of energy (see Appendix B 1 for details). As we see below, the intensive energy of typical minima can be strictly higher than the energy of the global minimum, especially at large laser gain.

Figures 4(a)–4(c) demonstrate an excellent match between theory and experiment for the distribution of spins $P(x)$, the distribution diagonal elements $H^I(\mathbf{x})_{ii}$ in Eq. (6), and the distribution of Hessian eigenvalues $\rho_H(\lambda)$, respectively. Interestingly, the distribution of spins $P(x)$ at large laser gain exhibits exactly 0 density for a range of x values around $x = 0$ [see, e.g., the cases of $a = 0, 1, 2$ in Fig. 4(a)]. This vanishing density can be understood through the cavity method as a simple consequence of the structure of the mean-field energy function $E_{\mathrm{MF}}[h](x)$ in Eqs. (51) and (3) and its associated mean-field index function $\bar{\mathcal{I}}(x)$ defined in Eq. (37), which is simply the index of E_{MF} evaluated at x. $\bar{\mathcal{I}}(x)$ plays a role in determining $P(x)$ through Eq. (68), and, when $\mu \to -\infty$, this equation indicates that $P(x)$ must vanish whenever $\bar{\mathcal{I}}(x) > 0$, or, equivalently, $P(x)$ must vanish over any range of x where $E_{\mathrm{MF}}[h](x)$ has a negative Hessian. For the particular double-well form of $E_{\mathrm{MF}}[h](x)$ in Eqs. (51) and (3) with $a + t > 0$, the theory implies $P(x)$ must vanish exactly when $|x| \le \sqrt{(a+t)/3}$. Remarkably, this striking prediction of vanishing density in $P(x)$ for typical minima at large a when $E_{\mathrm{MF}}[h](x)$ is nonconvex is verified in experiments [vanishing of blue histograms in Fig. 4(a)].

Furthermore, Figs. 4(d) and 4(e) show a match between theory and experiment for the order parameters $a + t$ and q, respectively, while Fig. 4(f) shows the evolution of A and C with a. Figure 4(f) indicates a supersymmetry-breaking phase transition [50] at $a = a_t \sim -0.93$, when A and C first become nonzero as laser gain a increases. This is exactly the same transition value at which supersymmetry breaking occurs for typical critical points. Indeed, this phase transition shares several similar properties with that of typical critical points. At this transition, $a + t$ first becomes positive as a increases [Fig. 4(d)], which means the

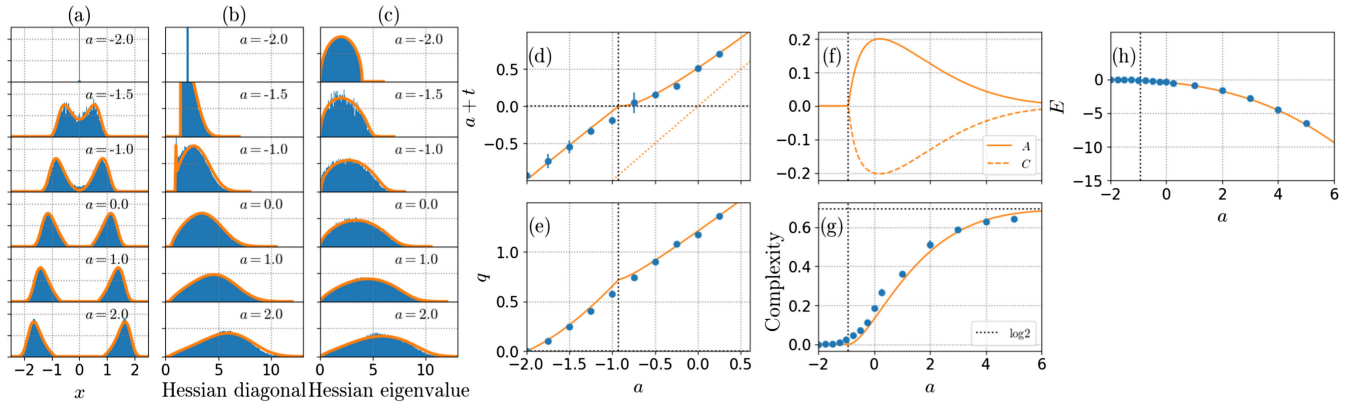

FIG. 4. A supersymmetry-breaking phase transition in the properties of typical minima. All panels indicate theoretical results (orange curves) and experimental results (blue histograms and markers). Details of experimental results are in Appendix B 1, and details of theoretical solutions are in Appendix B 2. (a) The distribution of spins or OPO amplitudes as laser gain a increases. (b) The distribution of Hessian diagonal elements $H^I(\mathbf{x})_{ii}$. (c) The Hessian eigenspectrum $\rho_H(\lambda)$. The Hessian eigenspectra in (c) can be understood intuitively as the outcome of Dyson's Brownian motion starting from the initial condition of diagonal elements in (b). (d)–(f) The order parameters $a+t$, q, and A and C, respectively. A supersymmetry-breaking phase transition is observed at the theoretically predicted point $a = a_t \sim -0.93$ (vertical dotted line). (g),(h) The complexity Σ and the intensive energy E, respectively, of typical minima.

mean-field energy function $E_{\mathrm{MF}}[h](x)$ first becomes nonconvex. This nonconvexity of $E_{\mathrm{MF}}[h](x)$ then generates an increasingly large region of vanishing density in $P(x)$ for typical minima around $x = 0$, as a increases beyond a_t [Fig. 4(a)], as discussed above.

Finally, the complexity Σ first becomes nonzero just above $a = a_t$ [Fig. 4(g)]. In the supersymmetric phase with $a \le a_t$ and $A = C = 0$, the complexity Σ of typical minima is 0, indicating the number of minima in the energy landscape is subexponential in N. On the other hand, there are exponentially many minima in the supersymmetry-broken phase. As a becomes large, we expect 2^N minima, and indeed the complexity converges to $\log 2$ in the large-a limit. Finally, Fig. 4(h) depicts the evolution of the intensive energy of typical minima with increasing a, again indicating an excellent match between theory and experiment.

VII. FULL REPLICA SYMMETRY BREAKING AND RIGIDITY PHASE TRANSITIONS IN GLOBAL MINIMA

We next move on from typical critical points and typical minima to the properties of global energy minima. We define global minima as those with the lowest intensive energy. For a single sample, different global minima with the same intensive energy in the large-N limit could have different *extensive* energies with subleading $o(N)$ differences. Note that this definition allows the landscape to have multiple global minima. We find that, at large laser gain a, global minima have lower intensive energies than local minima, and, to describe such low-energy global minima, we must break replica symmetry, just like in the SK model. In contrast for local minima, as described above, replica symmetric solutions, albeit with broken SUSY, sufficed to match numerical experiments.

A. A full replica symmetry-breaking theory of global minima

We performed a full replica symmetry-breaking calculation (see Ref. [44], Secs. S-II.G and S-III, for details) for global minima, which yields the following formula for the grand potential in the low-temperature $\beta \to \infty$ limit:

$$\Omega(\infty, -\infty) = \lim_{\beta\to\infty} \frac{1}{2} q_d t + \frac{\beta}{4}\left(q_d^2 - \int_0^1 dy q^2(y)\right) - f(0,0), \tag{75}$$

where the function $f(y,h)$ obeys the Parisi differential equation

$$\frac{\partial}{\partial y} f(y,h) = -\frac{1}{2}\frac{dq}{dy}\left[\frac{\partial^2 f}{\partial h^2} + \beta y\left(\frac{\partial f}{\partial h}\right)^2\right], \tag{76}$$

with the boundary condition

$$f(1,h) = \beta^{-1} \log\left[\sum_{x\in \mathrm{Crt}_0[E_{\mathrm{MF}}(\cdot,h)]} e^{-\beta E_{\mathrm{MF}}(x,h)}\right]. \tag{77}$$

This expression has order parameters t, q_d, and $q(y)$, where $q(y)$ is a nondecreasing non-negative function defined in $y \in [0,1]$. The values of these order parameters are chosen to extremize the grand potential in Eq. (75). The order parameter t reflects, as above in Eq. (50), the trace of the susceptibility matrix of the system to a small external field, but this time while the system is in a global minimum. q_d reflects the self-overlap (i.e., the average of $(1/N)\sum_i x_i^2$). The function $q(y)$ is called the overlap function, whose functional inverse represents the cumulative probability

density of the overlap $(1/N)\sum_i x_i^1 x_i^2$ of two different randomly sampled global minima $\mathbf{x}^1$ and $\mathbf{x}^2$.

Furthermore, the distribution $P(x)$ can be obtained from a solution to the following differential equation for a propagator $P(y,h)$ [85]:

$$\frac{\partial P}{\partial y} = \frac{1}{2}\frac{dq}{dy}\left(\frac{\partial^2 P}{\partial h^2} - 2\beta y \frac{\partial f}{\partial h}\frac{\partial P}{\partial h}\right), \tag{78}$$

with a boundary condition at $y = 0$ given by

$$P(0,h) = [2\pi q(0)]^{-1} \exp\left(-\frac{h^2}{2q(0)}\right). \tag{79}$$

The distribution of spins $P(x)$ in global minima can then be written in terms of the propagator evaluated at $y = 1$ and the value of $h = E_I(x) - tx$ which solves the extremization condition for the mean-field energy function $E_{\mathrm{MF}}[h]$ (see Ref. [44], Sec. S-II.G, details):

$$P(x) = |\partial^2 E_I(x) - t| P[1, \partial E_I(x) - tx]. \tag{80}$$

Finally, with $P(x)$ in hand, we can calculate the distribution of Hessian eigenvalues $\rho_H(\lambda)$ as above, using Eqs. (21) and (22).

B. Numerical tests of full replica symmetry breaking for global minima

To test our theory for global minima, we numerically solve the extremization conditions for the grand potential in Eq. (75) by approximating the overlap function $q(y)$ to be a sum of 37 step functions. This corresponds to a 37-step replica symmetry-breaking solution, approximating full replica symmetry breaking. We exploit numerical techniques addressed in Ref. [83] to find the order parameters satisfying the extremization conditions.

In Fig. 5, we compare our theoretical predictions with numerical experiments on finite-size systems. For the numerical experiments, we find the lowest energy minimum among many sampled minima for each sample of J (see Appendix B 1 for details of the sampling). Figures 5(a)–5(c) demonstrate an excellent match between theory and experiment for the distribution of spins $P(x)$, the distribution of diagonal elements $H^I(\mathbf{x})_{ii}$ in Eq. (6), and the distribution of Hessian eigenvalues $\rho_H(\lambda)$, respectively. Figures 5(d)–5(g) show a match between theory and experiment for the order parameters $a + t$, q_d, and the intensive energy, respectively. Finally, Fig. 5(h) shows the overlap function $q(y)$.

Figure 5 implies the existence of two phase transitions as a increases. First, a phase transition occurs at $a = \lambda_{\min}(J) = -2$. This is the point where the origin $\mathbf{x} = 0$ first bifurcates and the landscape starts to be nonconvex. Indeed, Fig. 5(e) shows that the self-overlap q_d starts to have a finite value at $a = -2$, which implies that the global minimum is no longer at the origin. Moreover, the overlap function $q(y)$ function undergoes a transition from a vanishing flat function for $a < -2$, indicating replica symmetry, to a continuously increasing function for $a > -2$, indicating full replica symmetry breaking (FRSB) [Fig. 5(h)].

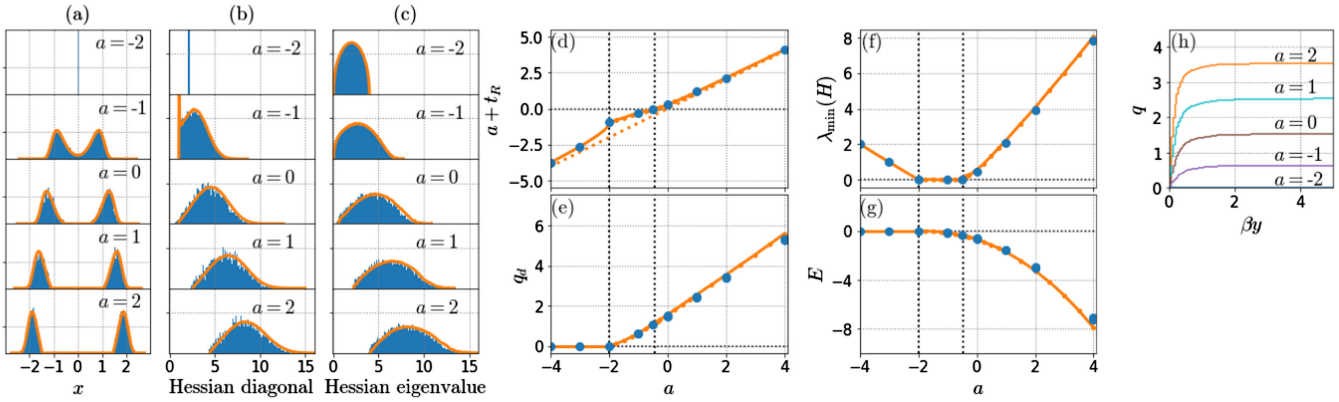

FIG. 5. Replica symmetry breaking and rigidity phase transitions in the properties of global minima. All panels indicate theoretical results (orange curves) and experimental results (blue histograms and markers). Details of experimental results are in Appendix B 1, and details of theoretical solutions are in Appendix B 2. (a) The distribution of spins or OPO amplitudes as laser gain a increases. (b) The distribution of Hessian diagonal elements $H^I(\mathbf{x})_{ii}$. (c) The Hessian eigenspectrum $\rho_H(\lambda)$. The Hessian eigenspectra in (c) can be understood intuitively as the outcome of Dyson's Brownian motion starting from the initial condition of diagonal elements in (b). (d)–(f) The order parameters $a + t$, q_d, and $\lambda_{\min}(H)$, respectively. At $a = a_r = -2$ there is a replica symmetry-breaking phase transition where q_d first acquires a nonzero value as a increases [see (e)]. At $a = a_g \sim -0.45$, there is a rigidity phase transition in the global minimum when $a + t$ first becomes positive (d) and the minimum Hessian eigenvalue transitions from 0 to positive (f). (g) The intensive energy E of global minima. (h) The overlap function $q(y)$ transitions from flat for $a < -2$ to continuously increasing for $a > -2$, indicating a replica symmetric to full replica symmetry-breaking transition at $a = a_r = -2$. The two phase transitions in replica symmetry breaking at $a = a_r = -2$ and rigidity at $a = a_g \sim -0.45$ are shown as dotted vertical lines in (d)–(g).

Within this FRSB regime, we find another phase transition similar to the case of the typical minima, which is again characterized by the sign of the $a + t$, which now goes from negative to positive at the critical point $a = a_g \sim -0.45$ [Fig. 5(d)]. The global minimum is supersymmetric (see Ref. [44], Sec. S-II.G, for details), so this transition is not characterized by spontaneous SUSY breaking as in the case of typical minima. Instead, it is characterized by the minimum eigenvalue of the Hessian. Figures 5(c) and 5(f) clearly show that this minimum eigenvalue is close to vanishing for $a_r < a < a_g$. This indicates that even the global minima, like all typical minima, are marginally stable with soft or flat directions corresponding to vanishingly small Hessian eigenvalues. On the other hand, for $a > a_g$, the global minimum undergoes a rigidity phase transition in which the soft, flat directions disappear, because the Hessian eigenspectrum is gapped away from 0 [Figs. 5(c) and 5(f)]. Thus, the global minimum is rigid or stable to small perturbations. In contrast, typical local minima remain soft and flat for all values of a considered, since the Hessian eigenspectrum always reaches to 0 [Fig. 4(c)].

Note that, while the global minima are marginally stable when $-2 < a < a_g$, they can still be described without SUSY-breaking order parameters A and C, in contrast to the case of the typical local minima. This phenomenon can happen when the complexity of global minima is zero. In this case, even if states are highly reactive to perturbations of the energy landscape, the change in the number of the global minima, in response to small changes in the energy landscape, is still subexponential. Thus, the susceptibility of the grand potential and complexity to such perturbations is 0, which implies via Eq. (72) that the SUSY-breaking order parameters obey $A = C = 0$.

VIII. THE PHASE DIAGRAM OF GEOMETRIC LANDSCAPE ANNEALING

We can now put together a global view of the geometry of the evolving energy landscape as the laser gain a is annealed. Overall, the energy landscape experiences three important phase transitions: (i) the replica symmetry-breaking transition for global minima at $a = a_r = -2$; (ii) the SUSY-breaking transition for typical minima and typical critical points at $a = a_t \sim -0.93$; and (iii) the rigidity phase transition for global minima at $a = a_g \sim -0.45$. The entire phase diagram is shown in Fig. 6.

When $a < a_r = -2$, the landscape is convex, and the single global minimum occurs at the origin $\mathbf{x} = 0$. Then, at the first phase transition at $a = a_r$, the origin bifurcates, and just above $a = a_r$ many minima start to appear. Figure 6 shows that the energy of global minima, typical minima, and typical critical points are all essentially equivalent for $a_r < a < a_t$. This means that the majority of critical points are minima, along with associated saddles

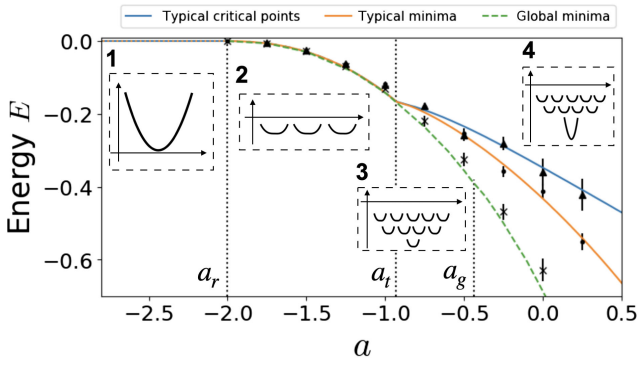

FIG. 6. The phase diagram of the energy landscape. The curves are theoretically predicted energy of the typical critical points (top blue line), typical minima (middle orange line), and global minima (green dashed line). The markers are the numerically obtained energies for typical critical points (triangles), typical minima (circles), and global minima (crosses), respectively. We observe three phase transitions at $a = a_r$, a_t, and a_g. The insets are sketches of the energy landscape in the four different phases.

of finite or at most subleading $o(N)$ index and energy barrier heights. Moreover, typical minima are also almost global minima. Finally, due to zero complexity of typical critical points and minima in the range $a < a_t$ [Figs. 3(g) and 4(g), respectively], the total number of critical points is subexponential within this phase, and, hence, so is the number of minima. Thus, SUSY is preserved due to zero complexity in the range $a_r < a < a_t$, despite the fact that typical critical points, minima, and global minima have a Hessian eigenspectrum that extends to 0 in this range [Figs. 3(c), 4(c), and 5(c)]. Thus, overall, in the SUSY phase $a_r < a < a_t$, the nonconvex energy landscape is relatively flat, with all subexponentially many critical points having essentially the same intensive energies and all having soft or flat directions with near zero Hessian eigenvalues.

At $a = a_t$, both typical critical points and minima experience SUSY breaking, due to the proliferation of exponentially many critical points and minima with nonzero complexity [Figs. 3(g) and 4(g), respectively] in conjunction with their marginal stability [Figs. 3(c) and 4(c)]. Moreover, an intensive energy gap starts to appear between typical minima and global minima. Hence, we expect that finding the lowest CIM energy state starts to get difficult at $a > a_t$ due to the exponential number of higher-energy typical minima.

Finally, while global minima are marginally stable until $a = a_g$, with many soft or flat modes, they become fully rigid for $a > a_g$ due to a Hessian eigenspectrum gapped away from 0 [Fig. 5(c)].

In addition to a global view of how typical critical points, minima, and global minima evolve as a function of laser gain a, as depicted in Fig. 6, we can also obtain a global view of the energies and locations of critical points of *all* indices at a fixed laser gain a. Figure 7 depicts this global view for both theory and experiment at large a (in this case, $a = 4$), which is the important case when the CIM energy

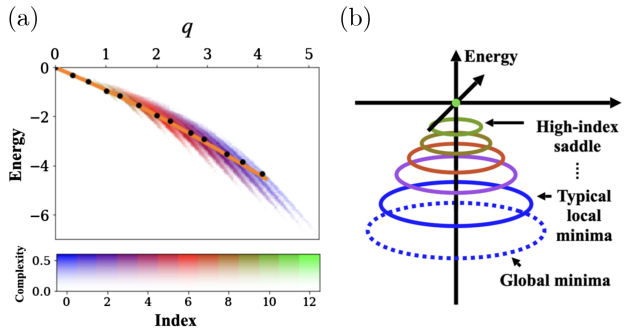

FIG. 7. A schematic view of the energy landscape at large laser gain. (a) A two-dimensional heat map of the experimentally derived distribution of intensive energy and squared radius q of critical points with $a = 4$, $N = 12$, derived from the sampled points depicted in Fig. 3. The color gradient denotes the index of critical points, while the opacity illustrates the complexity. The black points are the experimentally derived most likely location of critical points for each index $I = 0, \ldots, 12$. The orange curve is a theoretically predicted relationship between the energy and squared radius q of critical points obtained by solving Eq. (39) for varying μ in Eq. (34) and plotting $q(\mu)$ versus $E(\mu)$ [given in Eq. (41)]. (b) A schematic depiction of the energy landscape at large laser gain, consisting of concentric shells of increasing radius and decreasing index and energy.

function approximates well the Ising energy function of interest [see Eq. (13)].

In particular, Fig. 7(a) shows a heat map of the numerically estimated complexity of critical points in a finite-size ($N = 12$) system, as a joint function of their energy and squared radius $q = (1/N)\sum_i x_i^2$ and colored by their index. This complexity heat map shows a clear correlation between index, radius, and energy, with lower index critical points occurring at lower energy and larger radius. Moreover, for each index we plot the most likely location in the energy-squared radius plane (i.e., the location where the complexity of critical points of that index is maximized) as black dots in Fig. 7(a).

We then compare the locations of these black dots with theoretical predictions derived from our SUSY-breaking theory of critical points. In particular, we continuously vary the chemical potential μ in the expression for the grand potential in Eq. (34) [or, equivalently, Eq. (65)] and solve the self-consistent equations for the order parameters in Eq. (39) [or, equivalently, Eqs. (69) and (70)] as a function of μ, as well as computed the intensive energy as a function of μ via Eq. (41). Altogether, this yields the average squared radius $q(\mu)$ and intensive energy $E(\mu)$ over a weighted averaged of critical points controlled by μ, as in the partition function in Eq. (27) with $\beta = 0$. As $\mu \to -\infty$, this weighted averaged is dominated by index 0 critical points, or minima. As μ increases, the weighted average is dominated by higher index critical points. Thus, a theoretical prediction is that the curve $E(\mu)$ versus $q(\mu)$ as μ varies from $-\infty$ to $+\infty$ should provide information about the most likely location of saddle points of increasing index in the E-q plane, thereby going through all the black points in Fig. 7(a). Remarkably, this prediction is confirmed in Fig. 7(a): The orange curve is a plot of the theoretically derived curve $E(\mu)$ versus $q(\mu)$, and it does indeed go through all the black points, which indicate the experimentally derived most likely locations in the E-q plane for critical points of each index.

A schematic view of the energy landscape which is justified by Fig. 7(a) is shown in Fig. 7(b). Schematically, at large a, the CIM energy landscape exhibits a highly rough structure with concentric shells of critical points of increasingly lower index occurring at increasing lower energy and increasingly larger radius. In particular, the global minimum occurs at the largest radius and lowest energy. But, just above this in energy and at a smaller radius, there is a wall of exponentially many typical local minima that stand as a potential barrier. Thus, despite the fact that at large a the CIM energy landscape has the nice property that it mimics the Ising energy landscape of interest [see Eq. (13)], direct optimization at large a starting from the origin poses a difficult problem, as energy minimization must traverse successively lower index saddles and minima at lower energy and larger radius that may prevent reaching the deepest global minima at the largest radius.

IX. THE RELATIONSHIP BETWEEN ANNEALING PERFORMANCE AND ENERGY LANDSCAPE GEOMETRY

We next discuss the relationship between the phase transitions in the energy landscape geometry discussed above and the performance of geometric landscape annealing. This analysis also reveals an optimal annealing schedule to arrive at a low value of the Ising energy. Indeed, it is this annealing schedule that we use to attain good performance in Fig. 1 in Sec. III.

We simulate the annealing processes with various annealing schedules with a system size of $N = 10^4$. The schedules $a(t)$ are chosen as

$$a(t) = \min\left(\frac{t}{\tau} + a(0), a_{\max}\right); \tag{81}$$

i.e., $a(t)$ linearly increases from $a(0)$ with slope of τ^{-1} until it saturates at $a = a_{\max}$ (see top panel in Fig. 8). We set $a(0)$ to be the smallest eigenvalue of J, i.e., approximately $a_r = -2$, because the state $\mathbf{x}$ is always trivially at the origin for a smaller than the eigenvalue. The initial state is chosen as a random Gaussian vector with independent components each drawn from a Gaussian distribution with zero mean and standard deviation 0.1. We verify that the annealing performance is not influenced by the choice of the standard deviation unless it is *much* smaller than $O(1)$. In that case, the initial state is very close to the origin and takes a long time to escape the saddle point at the origin. When it does,

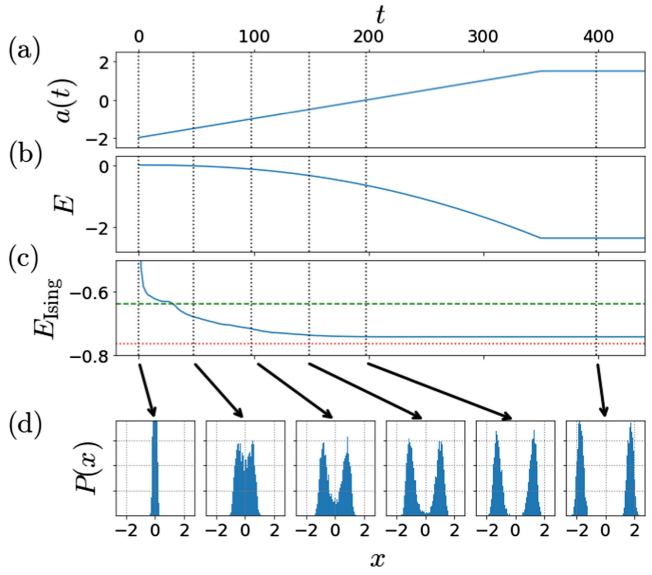

FIG. 8. An example of annealing trajectories (a) The annealing schedule given by Eq. (81) with $a_{\max} = 1.5$ and $\tau = 10^2$. (b) the trajectory of soft-spin network's energy E. (c) The trajectory of Ising energy E_{Ising} of the corresponding spin configuration. The green dashed line represents the Ising energy obtained by the spectral method $E_{\mathrm{Ising}} = -(2/\pi)$, and the red dotted line the ground-state energy in the large-N limit (approximately -0.76). (d) The local variables' distributions $P(x)$. The plots represent the snapshots of the distribution at time $t = 0$, 50, 100, 150, 200, 400, which are shown as vertical dotted lines above.

it aligns with the principal eigenvector of J, which is the most negative curvature direction around the saddle at the origin. Therefore, initializing very close to the origin is almost equivalent to the case of initializing the state along the principal eigenvector of J, a possibility which we discuss further below.

In Fig. 8, we show the trajectory of an annealing process with $a_{\max} = 1.5$ and $\tau = 100$ for the CIM energy (second panel), Ising energy (third panel), and several snapshots of the distribution of OPO amplitudes $P(x)$ (fourth panel). Since the initial state is a random vector around the origin, both the CIM energy E and the Ising energy E_{Ising} are close to zero [86]. As the laser gain $a(t)$ increases, both energies decrease monotonically, and the distribution $P(x)$ gradually transforms from an unimodal shape around the origin to a bimodal shape. During this transformation, small OPO amplitudes x_i around the origin are driven to either large positive or negative values, causing sign flips of x_i that lower E_{Ising}. However, once the distribution $P(x)$ gets completely separated into positive and negative parts at $t \sim 200$, fewer amplitudes x_i can flip their signs, and the Ising energy freezes.

To understand the dependency of the final CIM and Ising energies on the annealing schedule, we simulate the annealing processes with various $a_{\max}$ and τ. Figure 9(a) shows the CIM and Ising energy trajectories for three different $a_{\max}$ and ten different τ. Each row and color corresponds to a certain value of $a_{\max}$ and τ, respectively. Figures 9(b) and 9(c) show the *final* achieved CIM and Ising energies, respectively at $t = 6 \times 10^2$, averaged over five different realizations of the random initial state and the connectivity J. In particular, Fig. 9(b) shows for each $a_{\max}$ the difference ΔE between the final CIM energy achieved by annealing to $a_{\max}$ and the corresponding global minimum energy at $a_{\max}$, while Fig. 9(c) shows the corresponding Ising energy E_{Ising} of the Ising sign pattern of the CIM state found by annealing. The colored solid lines represent the final achieved energy for different $a_{\max}$ on the x axis and different colors for different τ. The dotted black curve with triangle markers above these colored solid lines corresponds to $\tau = 0$, i.e., a rapid quench or gradient descent from a random initial state at fixed $a_{\max}$. On the other hand, the solid black curve with inverted triangle markers below all the colored solid lines corresponds to $\tau = \infty$, or the slowest possible annealing process obtained by integrating the following adiabatic evolution:

$$\frac{d\mathbf{x}}{da} = H^{-1}\mathbf{x}, \tag{82}$$

with Hessian H given in Eq. (5). Note that this equation can be obtained by differentiating the stationary condition $[dE_{\mathrm{tot}}(\mathbf{x})/dx_i] = 0$ with respect to a. For comparison, Figs. 9(b) and 9(c) also show the results of gradient descent starting from the principal eigenvector (the black dashed line with rectangle markers) [87]. This corresponds to the limit of an extremely small standard deviation of the initial random Gaussian state. In Fig. 9(b), we also show for reference the CIM energy of typical minima (the highest black dotted line with "x" markers).

In the following subsections, we discuss the major features observed in Fig. 9 when the annealing process terminates in different phases of the energy landscape geometry revealed in previous sections. In particular, we discuss in succession (i) the small-gain supersymmetric phase where $a_r < a_{\max} < a_t$ [the left column in Fig. 9(a)]; (ii) the intermediate-gain supersymmetry-breaking phase where $a_t < a_{\max} < a_g$ [the middle column in Fig. 9(a)]; and (iii) large-gain rigid global minimum phase of $a_{\max} > a_g$ [the right column in Fig. 9(a)]. Note that when the gain is smaller than $a_r = -2$, the energy landscape is convex and the CIM state is confined to the origin.

A. The small-gain supersymmetric phase $a_r < a_{\max} < a_t$

As is shown in Fig. 9(b), all the final achieved CIM energies are very close to the $\Delta E = 0$ horizontal line, regardless of annealing time constant τ. This is because the majority of minima in the supersymmetric phase are almost global minima (Fig. 6). The final achieved Ising energy is also almost independent of the annealing schedule, as is shown in Fig. 9(c). However, unlike the final achieved

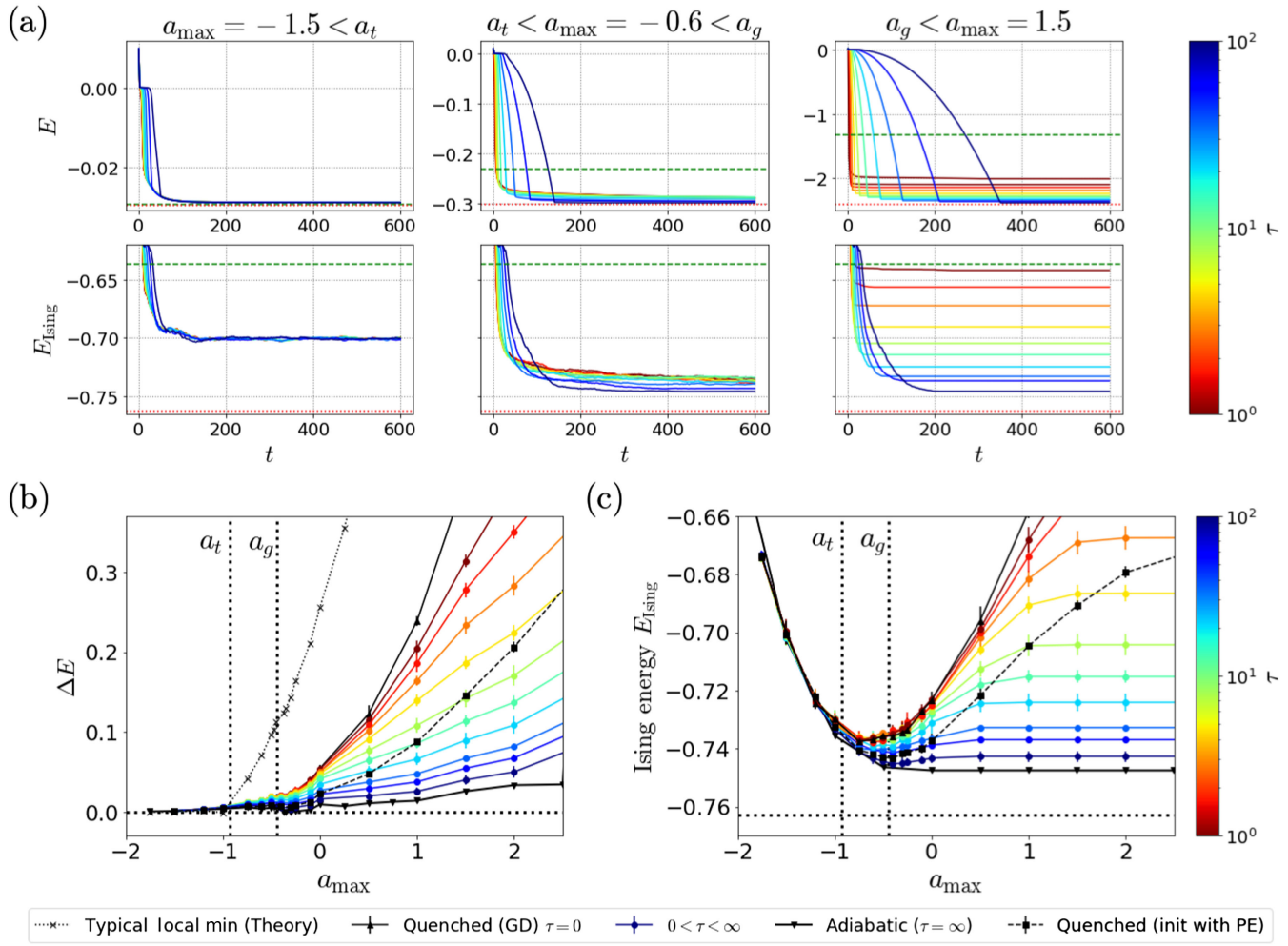

FIG. 9. The performance of geometric landscape annealing and its dependence on annealing schedules. We simulate geometric landscape annealing with a two-parameter family of annealing schedules in Eq. (81) parametrized by the final laser gain $a_{\max}$ and time constant τ. (a) Trajectories of the CIM energy E (top) and corresponding Ising energy E_{Ising} (bottom) in the three different phases for $a_{\max}$. The color bar indicates the annealing time constant τ. In the three top plots of CIM energy, the green dashed and red dotted lines represent the energy of typical local minima and the global minima, respectively. In the bottom plots of Ising energy, the green dashed and red dotted horizontal lines represent $E_{\text{sp}} = -2/\pi$ and $E_{\text{SK}} \sim -0.763$, respectively. (b) The solid colored lines (with color indicating annealing time constant according to the color bar) indicate the final achieved CIM energy E for each $a_{\max}$ at $t = 6 \times 10^2$ *minus* the theoretically calculated CIM ground-state energy for the same value of $a_{\max}$. The dotted horizontal line of $\Delta E = 0$ represents the baseline ground-state energy. The very top dotted black line indicates the theoretically calculated CIM energy of typical local minima (again, minus the energy of the corresponding global minima). (c) The final achieved Ising energy at $t = 6 \times 10^2$ as a function of annealing time constant τ indicated by color and final gain $a_{\max}$ on the horizontal axis. The horizontal dotted line is the ground-state Ising energy $E_{\text{SK}} \approx -0.763$. In (b) and (c), the solid colored lines are the annealing processes with mean and standard deviation computed across five different initializations and connectivities J. The black line above these colored lines is the case of $\tau = 0$, which corresponds to rapid quench from $a = a_r$ to $a = a_{\max}$. The black line below all the colored lines is the trajectory of the energy obtained by integrating the adiabatic differential equation (82), which essentially corresponds to $\tau = \infty$. The black dashed line in the midst of the colored lines represents the energy trajectory of gradient descent dynamics initialized along the principal eigenvector of J when a is fixed at $a_{\max}$.

CIM energy in Fig. 9(b), the final achieved Ising energy in Fig. 9(c) decreases rapidly with increasing $a_{\max}$ in the range $a_r < a_{\max} < a_t$. This decrease occurs because the distribution of OPO amplitudes $P(x)$ has a finite density at the origin, and so ramping up the laser gain allows some of these small amplitude spins to flip their signs, thereby lowering the achieved Ising energy.

B. The intermediate-gain SUSY-breaking phase $a_t < a_{\max} < a_g$

Once a exceeds a_t, the complexity of minima becomes positive [Fig. 4(g)], and the energy of typical minima becomes strictly larger than that of global minima (Fig. 6). This means there are exponentially many local minima above the global minima in the energy landscape. Thus, if

we rapidly increase the laser gain a and relax the system from a high-energy state near the origin, the resultant trajectory is likely to be trapped by a high-energy local minimum. This effect makes the final achieved CIM energy of rapid annealing higher than that of slow annealing [Fig. 9(a), middle top, and Fig. 9(b)]. Moreover, since the CIM energy is correlated with the Ising energy, rapid annealing also yields higher Ising energy than slow annealing [Fig. 9(a), middle bottom, and Fig. 9(c)]. However, since the complexity of minima is still relatively low in the range $a_t < a_{\max} < a_g$, the increment of final achieved CIM and Ising energies with annealing speed is relatively small.

On the other hand, another effect decreases the Ising energy with increasing $a_{\max}$ at all annealing speeds in the range $a_t < a_{\max} < a_g$ [Fig. 9(c)]. As seen in the top middle panel in Figs. 9(a) and 9(b), the final CIM energy achieved by annealing is still lower than that of typical minima and is rather closer to that of global minima. Therefore, the states obtained by annealing are likely to have features of the global minima rather than the typical minima. Indeed, the slowest annealing process has energy very close to that of global minima. This observation is bolstered by the numerical results shown in Fig. 10, which indicates that the trajectories of the slowest possible annealing processes have the features of the global minima's rigidity phase

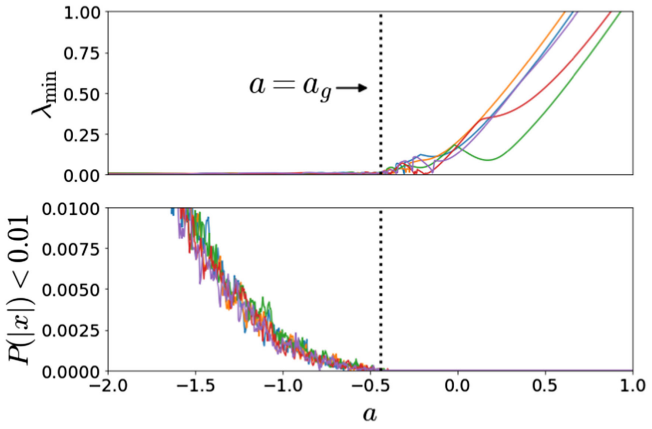

FIG. 10. Slow annealing trajectories exhibit the same phase transition as that of CIM global minima. We simulate the geometric landscape annealing process by integrating Eq. (4) with five different samples of J with system size $N = 10^4$. The annealing schedule is given by Eq. (81) with $\tau = 10^2$ and $a_{\max} = 4$. Each color represents a trajectory of a single instance. Top: the minimum eigenvalue of the Hessian along the trajectory. We observe a similar rigidity phase transition as that of global minima in Fig. 5(f) wherein the minimal eigenvalue of the Hessian transitions from 0 to nonzero values at $a = a_g$. Bottom: The fraction of the number of OPOs with small amplitudes $|x_i| < 0.01$ starts to vanish at $a = a_g$, similar to how the distribution $P(x = 0)$ for global minima vanishes at $a = a_g$. These two observations provide evidence that the slow annealing process can find the near-global CIM energy minima at least around $a = a_g$.

transition; i.e., the trajectory experiences the transition from marginal stability to full stability and the localization of $P(x)$ exactly at $a = a_g$. Since $P(x)$ has a finite density around the origin below a_g, the signs of some OPO amplitudes can still be flipped by ramping up the gain up to $a = a_g$. This effect allows E_{Ising} to decrease further in the range $a_t < a_{\max} < a_g$, especially for slow annealing, as seen in Fig. 9(c).

C. The large-gain rigid global minima phase $a_{\max} > a_g$

First, we focus on the black bottom lines in Figs. 9(b) and 9(c) representing the trajectory of the slowest possible annealing process. As we discuss above, this trajectory experiences a rigidity phase transition similar to that of global minima, and the Hessian of the energy landscape along this trajectory becomes gapped away from 0 for $a > a_g$, just as it does for global minima [Fig. 5(f)]. This implies that the adiabatic evolution (82) is nonsingular, and the time derivative of x_i cannot be large. Because the distribution $P(x)$ is localized and separated into two sets of x_i with positive and negative signs, only a few numbers of x_i can flip their signs by this bounded state evolution. Hence, it is unlikely that many x_i flip their signs, and, therefore, it is also unlikely that the Ising energy is lowered for $a > a_g$. Indeed, the final achieved Ising energy for the slowest annealing process [lower solid black curve in Fig. 9(c)] is flat for $a_{\max} > a_g$. Thus, interestingly, as we ramp up the laser gain beyond a_g, even though the CIM energy landscape becomes more like the Ising landscape, the final achieved Ising energy via annealing cannot be lowered. In other words, the geometric landscape annealing process at slow annealing speeds is effectively terminated by the rigidity phase transition in global minima at $a = a_g$, well before the CIM energy landscape looks like the Ising energy landscape at large a as in Eq. (13).

When the annealing speed is faster, the trajectory is more likely to be trapped by a higher-energy local minimum, leading to both higher final CIM and Ising energies. This increase of final energies with increased annealing speeds becomes stronger as the final gain $a_{\max}$ increases because of both the complexity growth of typical minima with a [Fig. 4(g)] and the growing energy gap between typical and global minima (Fig. 6). Likely because of both of these landscape properties, the final achieved CIM and Ising energies are significantly larger under faster annealing compared to slower process at very large $a_{\max} > a_g$ [Fig. 9(a), right, and Figs. 9(b) and 9(c)].

D. The optimal annealing schedule terminates at the rigidity phase transition for global minima

In summary, out of the general space of annealing schedules in Eq. (81), the optimal schedule with lowest achievable Ising energy is given by $a_{\max} = a_g$ and large τ. If the annealing speed is slow enough, a further increase of

$a_{\max}$ does not further lower the Ising energy, because the annealing trajectory tracks the evolution of a CIM global minimum found at $a = a_g$, which remains rigid for $a > a_g$. Also, more rapid annealing (smaller τ) yields much less optimal, much higher Ising energies at larger $a_{\max} > a_g$. However, at $a = a_g$, the dependence of the final achieved Ising energy on the speed of the annealing process is remarkably weak, ranging from approximately -0.75 to approximately -0.73 as τ ranges from 1 to 100. Thus, geometric landscape annealing in this case is surprisingly robust to annealing speed, provided annealing is optimally terminated at the rigidity phase transition for global minima.

Note that we also discover that the suboptimal higher Ising energies found by fast annealing can be mitigated by initializing the state along the principal eigenvector of the connectivity J. The performance of gradient descent at fixed $a_{\max}$ with this initialization is shown as a dashed back line with square markers in Figs. 9(b) and 9(c). The final Ising energy of this process also achieves the lowest value around $a = a_g$, and the Ising energy difference from the slowest annealing process is < 0.01. Since the slow annealing process can reach a global minimum of the CIM energy at $a = a_g$, the best achievable Ising energy can be simply characterized as the Ising energy of the CIM global minimum *specifically* when $a = a_g$. Note, however, that as $a_{\max}$ increases beyond a_g, even at extremely slow annealing, the CIM energy found by annealing is *no longer* equivalent to the CIM energy of the global minimum, as reflected by the detachment of the lower black solid line from the lowest horizontal dotted line of $\Delta E = 0$ in Fig. 9(b) for $a \gg a_g$. This means geometric landscape annealing cannot find a CIM global energy minimum for $a \gg a_g$. Since the CIM energy global minimum at very large a is also an Ising energy global minimum, according to Eq. (13), this means the annealing process cannot find the exact Ising ground state either, as reflected by the gap in Fig. 9(c), between the solid black bottom curve and the dotted horizontal line of $E_{\mathrm{SK}} \sim -0.763$. As discussed above, the adiabatic evolution for $a > a_g$ is continuous due to the nondegeneracy of Hessian along the trajectory. This implies that the near-global CIM energy minimum at $a = a_g$ that originates from the bifurcation at the origin is itself not continuously connected to global CIM energy minima at very large a. This type of discontinuity around phase transitions has been known as a major challenge for annealing processes such as simulated annealing and quantum annealing.

X. DISCUSSION

In an effort to develop a theoretical understanding of how a physical computing device, the coherent Ising machine, solves discrete combinatorial optimization problems by embedding them in annealed nonlinear analog dynamics, we engaged in an extensive study of the geometry of the energy landscape of this system and how it evolves as the laser gain is annealed, when the system is attempting to find the ground state of the SK spin glass. We were able to quantitatively describe the geometry of the landscape at all laser gains in terms of the number of critical points, and their locations (distance from the origin), energies, indices, and Hessian eigenspectra. We found at large laser gain, when the CIM energy function mimics the Ising energy function, the CIM energy landscape exhibits a complex hierarchical concentric shell structure in which saddle points of successively lower index and energy are located at successively larger radii (Fig. 7). This complex landscape presents a challenge to dissipative gradient descent dynamics, which at a fixed large laser gain cannot come close to either the CIM or Ising energy global minimum [top solid quenched $\tau = 0$ black line with triangle markers in Figs. 9(b) and 9(c)].

However, annealing the laser gain takes the CIM landscape through a sequence of phase transitions, each one introducing successive complexity. For $a < a_r$, the landscape is convex with a single global minimum at the origin. Then, for $a_r < a < a_t$, there are many (though subexponential in N) critical points. The intensive energies of all critical points are close to those of both typical and global minima. Hessian eigenspectra of all critical points extend to zero, indicating extensively many soft modes. This represents a highly flat landscape with many minima with similar energies tightly concentrated around a specific value, separated by saddle points whose energy barrier heights relative to minima and whose indices both scale sublinearly in N. The existence of such soft modes implies that the CIM undergoes subsequent bifurcations, called retarded bifurcations [29], flipping the signs of frustrated soft spins. This situation is described by supersymmetric solutions. Then, for $a_t < a < a_g$, supersymmetry for typical critical points and minima is broken; there are exponentially many of them with Hessian eigenspectra extending to 0. Furthermore, the intensive energies of typical critical points, typical minima, and global minima start to separate, indicating the beginnings of a rugged landscape (Fig. 6). The global minimum still has extensively many soft modes. Finally, for $a > a_g$, the global minimum undergoes a rigidity phase transition and all its soft modes disappear. Moreover, our cavity method for deriving these results yields conceptual insight into the meaning of SUSY breaking and the resultant order parameters, in terms of the extreme reactivity of the landscape to specific external perturbations, originating from exponentially many critical points with extensively many soft modes.

This detailed analysis of the landscape not only provides conceptual insights into why geometric landscape annealing works, through annealing the laser gain of the CIM, but also suggests an optimal annealing schedule. Basically, the Ising energy along a slow CIM annealing trajectory continuously decreases as a increases until a hits a_g. At this point, the CIM

annealing trajectory, whose energy has been following that of the CIM global minimum [bottom solid adiabatic $\tau = \infty$ black line with inverted triangle markers in Fig. 9(b)], becomes trapped in a rigid minimum that it cannot escape with further annealing. Thus, no further sign flips can occur and the Ising energy is fixed; there is no advantage to terminating the annealing process at any $a_{\max} > a_g$. In fact, there is a disadvantage: If one terminates annealing at some $a_{\max} > a_g$, the results can depend strongly on the annealing speed τ [i.e., substantial height variation of the colored lines for $a_{\max} \gg a_g$ in Figs. 9(b) and 9(c)]. However, if one terminates at $a_{\max} = a_g$, the final achieved CIM or Ising energies do not depend strongly on annealing speed [i.e., very little height variation of the colored lines for $a_{\max} = a_g$ in Figs. 9(b) and 9(c)]. This robustness to annealing speed is a consequence of the landscape geometry: For a up to a_g the intensive energy gap between typical local minima and global minima is not so large (Fig. 6), so if faster annealing results in trapping by local minima, such trapping cannot lead to substantially higher CIM energy.

All of this landscape analysis together points to an optimal and robust annealing schedule: Simply anneal a to $a_{\max} = a_g$ when the global minima of the CIM energy landscape become rigid. The slower the annealing, the better, but excessive slowness is not required due to the robustness of the final energies to annealing speed. Indeed, this landscape-derived annealing schedule allowed us to find SK spin configurations with energies in the large-N limit close to within about 1% of the true ground-state energy (Fig. 2). This final mismatch between the Ising energy found by CIM annealing and the actual Ising energy means that the global minimum of the CIM energy landscape at $a = a_g$ (or at least the state, with energy close to that of the global minimum, found by annealing) is not continuously connected to the global minimum of the CIM energy landscape at $a \gg a_g$, when the CIM energy landscape approximates well the Ising energy landscape. One possible scenario is an energy level crossing between two far apart local minima between a_g and large a, which leads to a different state becoming the global minimum at large a than the state that is a global minimum at $a = a_g$.

Overall, this extensive analysis of energy landscape geometry and its relation to annealing dynamics opens the door to several interesting directions. Most importantly, while we have focused on the SK spin-glass problem due to its combined mathematical simplicity and high degree of frustration, it would be interesting to extend our proposed theoretical framework to more realistic problem instances, which typically exhibit more structure than the SK instances. We believe our theoretical framework can be adapted to such structured instances by introducing additional order parameters, thereby providing a pathway toward tackling more realistic problem instances. A logical first step would be to generalize our results for previously studied structured instances, such as the Wishart planted ensemble [88] or the spiked Wigner model [89]. These problem instances incorporate low-rank term(s) in their Ising coupling matrices of the form ss^T, where s is an N-dimensional vector. In such cases, it would be natural to introduce an order parameter quantifying the inner product between s and the position of typical critical points of CIM. Elucidating the phase diagram of the energy landscape with this new order parameter presents a challenging yet promising avenue for future research.

Moreover, there exist many other ensembles of random optimization problems that can be efficiently mapped to Ising energy minimization, including, for example, partitioning, covering, packing, matching, clique finding, graph coloring, minimum spanning trees, and the traveling salesman problem [17]. Each of these ensembles of random problems could exhibit different geometric properties under landscape annealing, and analyzing the relationship between the evolution of landscape geometry, optimal annealing schedules, and annealing performance in these different ensembles could shed light on different universality classes of possible scenarios. To address these varied random problem instances, we can no longer rely on Pastur's formula (22), which stems from the properties of the Gaussian orthogonal ensemble. Consequently, we must develop and employ alternative theoretical methods tailored to each ensemble. Furthermore, while our current work assumes that the eigenvalue density of the Hessian vanishes at the origin (an assumption validated by our numerical results), this may not hold true for other cases. Investigating the conditions under which this assumption breaks down represents another intriguing area for future exploration.

Second, we have considered gradient descent dynamics on an evolving energy landscape. One could also analyze nongradient descent dynamics. For instance, a few previous works discuss the addition of nonconservative feedback such as the error-correcting scheme [28] or manifold reduction method [30], which introduce better control of soft-spin amplitudes. Another possibility is the addition of asymmetric parts to the connectivity matrix [90]. Such additional nongradient dynamics can typically induce chaos and destabilize the least stable minima. Just as the Kac-Rice formula can be used to count critical points of an energy landscape, as we have done here, it can also be used to count fixed points in nongradient dynamical systems, for example, in neural network dynamics [91] or ecological dynamics [92–94]. Such a Kac-Rice analysis of the CIM dynamics with an asymmetric connectivity component, which can be implemented physically in the CIM hardware, may provide an intriguing window into whether and how chaos might aid optimization [95,96].

Third, our analysis methods may also be useful for exploring the potential utility of nondegenerate OPO dynamics [42] for evading obstacles in the CIM energy

landscape. Such nondegenerate OPO dynamics can be modeled as a set of coupled oscillatory phase variables, akin to a network of Kuramoto oscillators [97], which have also been employed in physical computing devices to solve Ising energy minimization problems [98–100]. The geometric landscape annealing considered here can be thought of as gradually interpolating between soft-spin variables to strongly bistable binary variables while keeping the Ising connectivity fixed. On the other hand, the flexibility of physical OPO devices also opens the door to more general *dynamics* annealing strategies that interpolate between nondegenerate oscillatory phaselike dynamics and degenerate soft-spin dynamics or strongly bistable binary dynamics [42]. Exploring and analyzing the utility of this broader class of annealing strategies in solving diverse optimization problems constitutes an interesting direction for future research.

Fourth, and perhaps most interestingly, OPO networks can be constructed in ways that interpolate between classical and quantum operating regimes, as a function, e.g., of linear decoherence rates relative to coherent nonlinear dynamical rates [18]. Our work in the classical setting here provides a foundation for exploring how novel emergent information dynamics in the classical-quantum crossover [19] may impact optimization performance. Indeed, a key open question is, how does open dissipative quantum dynamics negotiate high-dimensional spaces riddled with saddle points and local minima as in Fig. 7? Is there some balance between coherent quantum evolution and environment-induced dissipation that can aid in optimization through energy minimization? Perhaps an interesting place to start is small systems of $N = 4$ coupled OPOs whose open dissipative quantum dynamics can both be tractably simulated on classical computers [101], as well as physically implemented in circuit QED [102] or nanophotonic [18] devices. An interesting question is to map out the computational phase diagram of such problems, parametrized by 4-by-4 connectivity matrices, and determine the boundaries between two computational phases in which the classical CIM either succeeds or fails. Then, one could explore how the quantum CIM behaves differently in each of these phases.

Along these quantum lines, recent work has examined how quantum or other physical effects in open dissipative physical systems modify their classical dissipative dynamics, yielding optimization benefits. For example, in a multimode cavity QED system whose classical dynamics mimics a Hopfield associative memory [103], the natural cavity dynamics yield the steepest energy descent dynamics that enhance both the capacity and robustness of memories relative to that of the classical Hopfield model [104]. Also, when the same cavity QED system implements an SK spin glass, simulations of the system reveal that the open dissipative quantum dynamics drive the coupled spins to enter highly entangled quantum states, which, in turn, allow the system to evade semiclassical energy barriers, thereby arriving at lower-energy states more quickly relative to the more semiclassical dissipative dynamics [105]. It would be interesting to explore whether analogous effects related to optimization benefits arise in quantum versions of the CIM.

In summary, the solution of combinatorial optimization algorithms using novel physical computing hardware is a rich and emerging field. Our initial theoretical analysis of the coherent Ising machine in the classical limit reveals a rich theory with diverse connections across physics and mathematics, spanning spin glasses, the replica method, the cavity method, supersymmetry breaking, Dyson's Brownian motion, random matrix theory, and the statistical mechanics of random landscapes. Moreover, analysis combining these topics yields geometric insights into the nature of optimal annealing schedules and the computational power of geometric landscape annealing in optimization. Future directions of theory suggest the potential for usefully connecting to even more diverse topics, including Kuramoto networks, chaos, and open dissipative quantum dynamics. Given the recent emerging interest in diverse physical computing devices, spanning spintronic [106], memristor [107], photonic and optical [2,108], and CMOS substrates [109], for solving diverse NP-hard combinatorial optimization problems, we hope our theoretical analysis may inspire much future work aimed at understanding general approaches for how annealed nonlinear analog dynamical systems can aid in solving discrete optimization problems, thereby merging the primarily analog worlds of physics with the primarily discrete worlds of computer science.

ACKNOWLEDGMENTS

This work is supported by the National Science Foundation (NSF) under Grant No. CCF-1918549 and by NTT Research. A. Y. is supported by the Masason Foundation. S. G. acknowledges funding from NSF CAREER Grant No. 1845166 and the Schmidt Foundation through its Science Polymath Award. The authors thank Timothée Leleu, Evan Laksono, Daniel Wennberg, Niharika Gunturu, Edwin Ng, and Ryotatsu Yanagimoto for discussions.

APPENDIX A: DERIVATION OF GRAND POTENTIAL

In this appendix, we present two detailed derivations of the grand potential. Our first derivation employs the replica method, while our second derivation involves a novel generalized cavity method. For readers less acquainted with these methodologies, we recommend consulting Ref. [44], where step-by-step derivations are provided for pedagogical purposes.

1. Replica-based calculation

As is shown in Eq. (30), the grand potential can be calculated by the ensemble average of Z^n where the partition function Z is given by Eq. (31). By introducing new auxiliary variables u^a for $a = 1, 2, \ldots, n$, Z^n can be written as

$$Z^n = \int \prod_{i,a} dx_i^a \int_{-i\infty}^{i\infty} \prod_{i,a} \frac{du_i^a}{2\pi i} |\det[H(x^a)]| e^{\mu \mathcal{I}(\mathbf{x}^a)} \exp\left[\sum_i u_i^a \partial_i E(\mathbf{x}^a) - \beta E(\mathbf{x}^a)\right].$$

Then, inserting this expression for Z^n, performing a change of variables in J_{ij} [45], and introducing new auxiliary variables via Hubbard-Stratonovich transformations [73], we obtain

$$-\beta\Omega(\beta,\mu) = \lim_{n\to 0} \frac{1}{n} \mathrm{Ext}_\Theta \left(\Omega_0'(\Theta) + \frac{1}{N} \log\left[\int D\mathbf{x} \int D\mathbf{u} e^{\sum_i S'(\Theta, \mathbf{x}_i, \mathbf{u}_i)} \left\langle \prod_a |\det H(\mathbf{x}^a)| e^{\mu \mathcal{I}(\mathbf{x}^a)} \right\rangle_J \right] \right),$$

where $\int D\mathbf{x} = \int \prod_{i,a} dx_i^a$ and $\int D\mathbf{u} = \int_{-i\infty}^{i\infty} \prod_{i,a} (du_i^a / 2\pi i)$. $\Theta := (\{q^{ab}\}, \{w^{ab}\}, \{\lambda^{ab}\})$ is a tuple of several auxiliary variables, Ext_Θ means extremization with respect all variables Θ, and $\Omega_0'(\Theta)$ and $S'(\Theta, x, u)$ are given, respectively, by

$$\Omega_0'(\Theta) = \sum_{a,b} \left[-\frac{1}{2}(w^{ab})^2 - \beta w^{ab} q^{ab} - \frac{\beta^2}{4}(q^{ab})^2 - \lambda^{ab} q^{ab} \right],$$

$$S'(\Theta, x, u) = \sum_a [-\beta E_I(x^a) + u^a \partial E_I(x^a)] + \sum_{ab} \left[w^{ab} x^a u^b + \frac{1}{2} q^{ab} u^a u^b + \lambda^{ab} x^a x^b \right].$$

We assume that the average of the product of the determinant and the chemical potential factor factorizes into the product of the averages:

$$\left\langle \prod_a |\det H(\mathbf{x}^a)| e^{\mu \mathcal{I}(\mathbf{x}^a)} \right\rangle_J \approx \prod_a \langle |\det H(\mathbf{x}^a)| \rangle_J \left\langle e^{\mu \mathcal{I}(\mathbf{x}^a)} \right\rangle_J.$$

To evaluate the chemical potential term, we exploit the formula [110]

$$\mathcal{I}(\mathbf{x}) = \lim_{\epsilon \to 0} \frac{1}{2\pi i} (\log\det[H(\mathbf{x}) - i\epsilon] - \log\det[H(\mathbf{x}) + i\epsilon]).$$

Further assuming that the Hessian's eigenspectrum density at the origin vanishes, the average of the chemical potential term is calculated as

$$\log \left\langle e^{\mu \mathcal{I}(\mathbf{x})} \right\rangle_J = \mu \sum_i \bar{\mathcal{I}}(x_i),$$

where $\bar{\mathcal{I}}(x)$ is given in Eq. (37). Similarly, we obtain

$$\log \langle |\det[H(\mathbf{x})]| \rangle_J = \sum_i \log[\partial^2 E_I(x_i) - t].$$

We substitute these equations and introduce new variables A^{ab} and $C^{ab} (= C^{ba})$, as is done for supersymmetry-breaking complexity calculations [54]. Specifically, we perform the change of variables

$$w^{ab} = -t^a \delta_{ab} - \beta q_{ab} - A^{ab},$$

$$\lambda^{ab} = \frac{\beta}{2} t^a \delta_{ab} + \frac{\beta^2}{2} q^{ab} + \frac{\beta}{2}(A^{ab} + A^{ba}) + \frac{1}{2} C^{ab}.$$

While w^{ab} and λ^{ab} are difficult to interpret, the new variables A^{ab} and C^{ab} have clear physical interpretations as the susceptibility of the complexity to certain perturbations in the energy landscape, which we discuss later. In terms of the new variables A^{ab} and C^{ab}, the grand potential is expressed as

$$-\beta\Omega(\beta,\mu) = \lim_{n\to 0} \frac{1}{n} \mathrm{Ext} \left(\Omega_0 + \log \int \prod_a [dx^a du^a w(x^a)] e^{S + \mu \sum_a \bar{\mathcal{I}}(x^a)} \right)$$

with $w(x^a) = \partial^2 E_I(x^a) - t^a$ and

$$\Omega_0 = \sum_{a,b}\left[-\frac{1}{2}(A^{ab})^2 - A^{aa}t_R^a\delta_{ab} - \frac{\beta}{2}q^{aa}t^a\delta_{ab} - \frac{\beta^2}{4}(q^{ab})^2 - \beta A^{ab}q^{ab} - \frac{1}{2}C^{ab}q^{ab}\right],$$
$$S = \sum_{ab}\left[\frac{1}{2}q^{ab}u^a u^b + A^{ab}u^a x^b + \frac{1}{2}C^{ab}x^a x^b\right] + \sum_a\left[-u^a h^a + \beta x^a h^a - \beta(E_I(x^a) - \frac{t^a}{2}(x^a)^2)\right], \tag{A1}$$

where $h^a = \partial E_I(x^a) - t^a x^a$.

We next derive the grand potential for $\beta = 0$ with the replica symmetric ansatz. Under this ansatz, q^{ab} is parametrized as $q_{ab} = q\delta_{ab} + \bar{q}$, with two parameters q and $\bar{q}$. Similarly, A and C are parametrized as $A^{ab} = A\delta_{ab} + \bar{A}$ and $C^{ab} = C\delta_{ab} + \bar{C}$, respectively, and t^a is now independent of the replica index a, which we denote t. It is easy to see that the terms including $\bar{q}$, $\bar{A}$, and $\bar{C}$ are proportional to n^2, where n is the number of replicas, and, hence, we ignore such terms, meaning that all the replicas are decoupled. Then, under this ansatz, we obtain

$$-\beta\Omega(0,\mu) = Ext\left[-\frac{1}{2}(Cq + A^2) - At + \log\int_{-i\infty}^{i\infty}\frac{du}{2\pi i}\int dx w(x)\exp\left(\frac{1}{2}\begin{pmatrix}u\\x\end{pmatrix}^T\begin{pmatrix}q & A\\A & C\end{pmatrix}\begin{pmatrix}u\\x\end{pmatrix} - uh + \mu\bar{\mathcal{I}}(x)\right)\right].$$

By integrating out u, we get Eq. (34). The self-consistent equations [Eq. (39)] are obtained by the stationary condition of the extermination.

Lastly, we discuss how we obtain the expression for distribution $P(x)$ [Eq. (36)] and the Hessian eigenvalue spectrum [Eq. (22)]. For an arbitrary well-behaved function $O(x_i)$, we define the following expectation value:

$$\langle O\rangle := \left\langle \frac{1}{Z}\sum_{\alpha\in \mathrm{Crt}(E)} e^{-\beta E(\mathbf{x}^\alpha)+\mu\mathcal{I}(\mathbf{x}^\alpha)}\left(\frac{1}{N}\sum_i O(x_i)\right)\right\rangle_J.$$

This quantity is equivalent to the following derivative:

$$\frac{1}{N}\frac{d}{ds}\bigg|_{s=0}\left\langle \log\sum_{\alpha\in \mathrm{Crt}(E)} e^{-\beta E(\mathbf{x}^\alpha)+\mu\mathcal{I}(\mathbf{x}^\alpha)+s\sum_i O(x_i)}\right\rangle_J.$$

Following the derivation of grand potential above, we can calculate this quantity, resulting in $\langle O\rangle = \int dx P(x)O(x)$, where $P(x)$ is given by Eq. (36).

The Hessian eigenspectrum can be obtained from the resolvent $R(z)$ via the inverse Stieltjes transform. The resolvent can be written as a derivative as follows:

$$R(z) = -\frac{1}{N}\frac{d}{d\hat{z}}\bigg|_{\hat{z}=z}\frac{\det[H(\mathbf{x}) - \hat{z}I]}{\det[H(\mathbf{x}) - zI]}. \tag{A2}$$

We calculate the average of $\log\det[H(\mathbf{x}^\alpha) - \hat{z}I] - \log\det[H(\mathbf{x}^\alpha) - zI]$ and then obtain an ensemble-averaged resolvent through Eq. (A2). Following this strategy, the average of $\log\det[H(\mathbf{x}^\alpha) - \hat{z}I] - \log\det[H(\mathbf{x}^\alpha) - zI]$ is given by

$$\langle\log\det[H(\mathbf{x}^\alpha) - \hat{z}I] - \log\det[H(\mathbf{x}^\alpha) - zI]\rangle_{\beta,\mu} = N^{-1}\frac{d}{ds}\bigg|_{s=0}\log\sum_{\alpha\in \mathrm{Crt}(E)}\exp\Big(-\beta E_I(\mathbf{x}^\alpha) + \mu\mathcal{I}(\mathbf{x}^\alpha) + s\{\log\det[H(\mathbf{x}^\alpha) - \hat{z}I] - \log\det[H(\mathbf{x}^\alpha) - zI]\}\Big).$$

Here, $\langle\cdot\rangle_{\beta,\mu}$ represents the weighted average over all critical points $\mathbf{x}^\alpha$ with the Boltzmann weights $\exp(-\beta E_I(\mathbf{x}^\alpha) + \mu\mathcal{I}(\mathbf{x}^\alpha))$. The modified free energy parametrized by s can be calculated using the replica method in a manner parallel to the previous calculation and also by exploiting the formula for the average of the determinant. In the end, we obtain

$$\left\langle\log\sum_{\alpha\in \mathrm{Crt}(E)}\exp\Big(-\beta E_I(\mathbf{x}^\alpha) + \mu\mathcal{I}(\mathbf{x}^\alpha) + s\{\log\det[H(\mathbf{x}^\alpha) - \hat{z}I] - \log\det[H(\mathbf{x}^\alpha) - zI]\}\Big)\right\rangle_J = \lim_{n\to 0}\frac{1}{n}\mathrm{Ext}\left\{\Omega_0 + \log\int\prod_a[dx^a du^a w(x^a)]e^{S+\mu\sum_a\bar{\mathcal{I}}(x^a)}\exp\left(s\left[\frac{N}{2}\{[t^a(\hat{z})]^2 - [t^a(z)]^2\} + \log\frac{\partial E_I(x^a) - \hat{z} - t^a(\hat{z})}{\partial E_I(x^a) - z - t^a(z)}\right]\right)\right\},$$

where the stationary condition gives $t^a(z) = \langle[\partial^2 E(x^a) - z - t^a(z)]^{-1}\rangle_S$, with Ω_0 and S defined in Eq. (A1). By taking the derivative with respect to s, we obtain

$$\langle \log\det[H(\mathbf{x}^\alpha)-\hat{z}I]-\log\det[H(\mathbf{x}^\alpha)-zI]\rangle_{\beta,\mu,J}=\frac{N}{2}[t(\hat{z})]^2-\frac{N}{2}[t(z)]^2+N\left\langle \log\frac{\partial E_I(x^a)-\hat{z}-t(\hat{z})}{\partial E_I(x^a)-z-t(z)}\right\rangle_S.$$

Here, we exploit the fact that t^a is independent of the replica index a and write it as t without its replica index. Assuming that the quantity above is self-averaging,

$$\begin{aligned}R(z)&=-N^{-1}\frac{d}{d\hat{z}}\Big|_{\hat{z}=z}\exp\left(\frac{N}{2}[t(\hat{z})]^2-\frac{N}{2}[t(z)]^2+N\left\langle\log\frac{\partial E_I(x^a)-\hat{z}-t(\hat{z})}{\partial E_I(x^a)-z-t(z)}\right\rangle_S\right)\\&=t(z)t'(z)-\langle[\partial E_I(x^a)-z-t(z)]^{-1}[1+t'(z)]\rangle_S\\&=t(z).\end{aligned}$$

Therefore, the function t defined in Eq. (39) is nothing other than the resolvent of the Hessian at the origin $z=0$. Moreover, through the inverse Stieltjes transform, the eigenvalue density of the Hessian is simply proportional to the imaginary part of $t(z)$ near the real axis.

2. The generalized cavity method

In this section, we derive the grand potential at the annealed level via a generalized cavity method with more detailed mathematical expressions. We start from the Kac-Rice formula for the grand potential, but we relax the delta-function constraint on the gradient and replace it with a soft Gaussian function with effective inverse temperature β. The relaxed grand potential is given as follows:

$$\exp(N\Omega_\beta)=\mathbb{E}_J\left[\int\prod_{i=0}^{N-1}dx_i\left(\frac{\beta}{\pi}\right)^{N/2}|\det H(\mathbf{x})|e^{\mu I(\mathbf{x})}\exp\left(-\beta\sum_{i=0}^{N-1}\left(\partial E_I(x_i)+\sum_{j=0}^{N-1}J_{ij}x_j\right)^2\right)\right].$$

We split the system of N soft spins into a single soft spin x_0 and the rest of the spins, i.e., the cavity system $x^{/0}=(x^1,\ldots,x^N)$. Note that the joint distribution of $x=x_0$ and gradient $y=\nabla_0 E(\mathbf{x})$ is given by

$$P_\beta(x,y)\propto\mathbb{E}_J\left[\int\prod_{i=0}^{N-1}dx_i\exp\left(-\beta\sum_{i=0}^{N-1}\left(\partial E_I(x_i)+\sum_{j=0}^{N-1}J_{ij}x_j\right)^2\right)|\det H(\mathbf{x})|e^{\mu I(\mathbf{x})}\delta(x_0-x)\delta[\nabla_0 E(\mathbf{x})-y]\right].\quad\text{(A3)}$$

The grand potential can be factorized as follows:

$$\exp(N\Omega_\beta)=\mathbb{E}_{J^{/0}}\left[\int\prod_{i=0}^{N-1}dx_i\mathbb{E}_{\mathbf{J}_0}\left[\omega(x_0\mathbf{J}_0,\mathbf{x}^{/0})\sqrt{\frac{\beta}{\pi}}\exp\left(-\beta\left(\sum_{i=1}^{N-1}J_{0i}x_i+\partial E_I(x_0)\right)^2\right)\frac{|\det H(\mathbf{x})|}{|\det H(\mathbf{x}^{/0})|}e^{\mu(\mathcal{I}(\mathbf{x})-\mathcal{I}(\mathbf{x}^{/0}))}\right]\right].\quad\text{(A4)}$$

Here, $H(\mathbf{x}^{/0})$ is the $N-1\times N-1$ submatrix of $H(\mathbf{x})$ and is simply the Hessian of the cavity system in the absence of x_0. $\mathbb{E}_{\mathbf{J}_0}$ represents an average over the vector $\mathbf{J}_0:=(J_{01},J_{02},\ldots,J_{0(N-1)})$ which couples x_0 to the cavity system $\mathbf{x}^{/0}$. Also, $\mathbb{E}_{J^{/0}}$ represents an average over $J^{/0}$ which is the $N-1\times N-1$ submatrix of J corresponding to the connectivity matrix of the cavity. Finally, $\omega(\mathbf{s},\mathbf{x}^{/0})$ is the grand potential density of the cavity system $\mathbf{x}^{/0}$ in the absence of x^0 but in the presence of an external field $\mathbf{s}$ that tilts the gradient:

$$\omega(\mathbf{s},\mathbf{x}^{/0})=\left(\frac{\beta}{\pi}\right)^{(N-1)/2}|\det H(\mathbf{x}^{/0})|e^{\mu\mathcal{I}(\mathbf{x}^{/0})}\exp\left(-\beta\sum_{i=1}^{N-1}\left(\partial E_I(x_i)+\sum_{j=1}^{N-1}J_{ij}x_j+s_i\right)^2\right).$$

Assuming that $t:=N^{-1}\mathrm{Tr}H^{-1}(\mathbf{x})$ is self-averaging, the determinant ratio in Eq. (A4) is calculated as $|\det H(\mathbf{x})/\det H(\mathbf{x}^{/0})|=|\partial^2 E_I(x_0)-t|$. Furthermore, the difference of indices $\mathcal{I}(\mathbf{x})-\mathcal{I}(\mathbf{x}^{/0})$ in Eq. (A4) is calculated as follows:

$$\begin{aligned}&\mathcal{I}(\mathbf{x})-\mathcal{I}(\mathbf{x}^{/0})\\&=\lim_{\epsilon\to 0}\frac{1}{2\pi i}\left(\log\frac{\det[H(\mathbf{x})-i\epsilon]}{\det[H(\mathbf{x}^{/0})-i\epsilon]}-\log\frac{\det[H(\mathbf{x})+i\epsilon]}{\det[H(\mathbf{x}^{/0})+i\epsilon]}\right)\\&=\bar{\mathcal{I}}(x_0),\end{aligned}$$

where $\bar{\mathcal{I}}(x)$ is given by Eq. (37). The grand potential density $\omega(\mathbf{s},\mathbf{x}^{/0})$ can be factored into the product of $\omega(0,\mathbf{x}^{/0})$ and a function of s: $\omega(x_0\mathbf{J}_0,\mathbf{x}^{/0})=\omega(0,\mathbf{x}^{/0})\exp(-2x_0\beta\sum_{i=1}^{N-1}\nabla_i E(\mathbf{x}^{/0})J_{0i}-\beta x_0^2)$, where $\nabla_i E(\mathbf{x}^{/0})=\partial E_I(x_i)+\sum_{j=1}^{N}J_{ij}x_j$. Substituting these expressions into Eq. (A4), we notice that the expression depends on the coupling vector $\mathbf{J}_0$ only through two variables $\bar{h}:=-\sum_{i=1}^{N-1}J_{0i}x_i$ and $\bar{z}:=-2\beta\sum_{i=1}^{N-1}\times\nabla_i E(\mathbf{x}^{/0})J_{0i}$. These two scalars are jointly Gaussian distributed with zero mean, and the integration over $\mathbf{J}_0$ can be replaced with the two-dimensional integration over $(\bar{h},\bar{z})$. Then, the integration over $\mathbf{J}_0$ can be replaced with the expectation value $\langle\cdot\rangle_{\bar{h},\bar{z}}$ with respect to random Gaussian variables $\bar{h}$ and $\bar{z}$:

$$\exp(N\Omega_\beta)=\mathbb{E}_{J^{/0}}\left[\int\prod_{i=0}^{N-1}dx_i\omega(0,\mathbf{x}^{/0})\left(\frac{\beta}{\pi}\right)^{1/2}\left\langle\exp\left(x_0\bar{z}-\beta(\partial E_I(x_0)-\bar{h})^2-\beta x_0^2\right)\right\rangle_{\bar{h},\bar{z}}|\partial^2E_I(x_0)-t_R|e^{\mu\bar{\mathcal{I}}(x_0)}\right].$$

Furthermore, the expectation value can be rewritten by a change of Gaussian variables $\bar{h}\to h$ and $\bar{z}\to z$ as follows:

$$\begin{aligned}&\left\langle\exp\left(x_0\bar{z}-\beta(\partial E_I(x_0)-\bar{h})^2-\beta x_0^2\right)\right\rangle_{\bar{h},\bar{z}}\\&\quad=\left\langle\exp\left(x_0z-\beta(\partial E_I(x_0)-t_Rx_0-h)^2\right)\right\rangle_{h,z},\end{aligned}$$

where the covariance of the new centered Gaussian random variables z and h is now given by

$$\begin{aligned}\langle h^2\rangle&=N^{-1}|\mathbf{x}^{/0}|^2,\\\langle hz\rangle&=2N^{-1}\beta\nabla E(\mathbf{x}^{/0})\cdot\mathbf{x}^{/0}-t,\\\langle z^2\rangle&=4N^{-1}\beta^2|\nabla E(\mathbf{x}^{/0})|^2-2\beta.\end{aligned}$$

After this change of variables, it is not difficult to see that the expression is equivalent to the following:

$$\begin{aligned}\exp(N\Omega_\beta)=\mathbb{E}_{J^{/0}}\Bigg[&\int\prod_{i=1}^{N-1}dx_i\omega(0,\mathbf{x}^{/0})\\&\times\left\langle\int dy\sqrt{\frac{\beta}{\pi}}e^{-\beta y^2}\sum_{x\in\mathrm{Crt}(E_{\mathrm{MF}}[y+h])}e^{xz+\mu\bar{\mathcal{I}}(x)}\right\rangle_{h,z}\Bigg],\end{aligned}$$

where $E_{\mathrm{MF}}[h+y](x):=E_I(x)-(t/2)x^2-(h+y)x$.

In this equation, the Gaussian average $\langle\cdot\rangle_{h,z}$ has a covariance matrix that depends on the cavity system $\mathbf{x}^{/0}$ only through the following three quantities, which we define to be $q(\mathbf{x}^{/0})$, $A(\mathbf{x}^{/0})$, and $C(\mathbf{x}^{/0})$:

$$\begin{aligned}q(\mathbf{x}^{/0})&=(N-1)^{-1}|\mathbf{x}^{/0}|^2\approx\langle h^2\rangle,\\A(\mathbf{x}^{/0})&=2(N-1)^{-1}\beta\nabla E(\mathbf{x}^{/0})\cdot\mathbf{x}^{/0}-t\approx\langle hz\rangle,\\C(\mathbf{x}^{/0})&=4(N-1)^{-1}\beta^2|\nabla E(\mathbf{x}^{/0})|^2-2\beta\approx\langle z^2\rangle.\end{aligned}$$

We assume that $q(\mathbf{x}^{/0})$, $A(\mathbf{x}^{/0})$, and $C(\mathbf{x}^{/0})$ concentrate around their expectation values q, A, and C under the cavity grand potential density $\omega(0,\mathbf{x}^{/0})$. Under this assumption, the average $\langle\cdot\rangle_{h,z}$ can be taken outside of the integral over $\mathbf{x}^{/0}$, obtaining

$$\begin{aligned}\Omega_\beta=&(N-1)(\tilde{\Omega}_\beta-\Omega_\beta)\\&+\log\left\langle\int dy\sqrt{\frac{\beta}{\pi}}e^{-\beta y^2}\sum_{x\in\mathrm{Crt}(E_{\mathrm{MF}}[y+h])}e^{xz+\mu\bar{\mathcal{I}}(x)}\right\rangle_{h,z},\end{aligned}$$

where the covariance of h and z is given by

$$\begin{pmatrix}q&A\\A&C\end{pmatrix}$$

and $\tilde{\Omega}_\beta$ is defined as

$$\exp[(N-1)\tilde{\Omega}_\beta]:=\mathbb{E}_{J^{/0}}\left[\int\prod_{i=1}^{N-1}dx_i\omega(0,\mathbf{x}^{/0})\right].$$

In the limit of large N, the first term of the equation above converges to a derivative, which can be written in terms of order parameters:

$$\begin{aligned}\lim_{N\to\infty}(N-1)(\tilde{\Omega}_\beta-\Omega_\beta)&=-\left.\frac{d\Omega_\beta(\sigma)}{d\sigma}\right|_{\sigma=1}\\&=\frac{1}{2}qC+\frac{1}{2}A^2+At.\end{aligned}$$

Substituting this expression and taking the limit of large β, we recover Eq. (65).

Lastly, we discuss the derivation of the self-consistent equations via the cavity method. We assume that the joint distribution of $\mathbf{x}$ and $\nabla E(\mathbf{x})$ can be factorized into a product of N copies of the independent distribution

$P_\beta(x,y)$, defined in Eq. (A3). Under this assumption, any self-averaging function of $\mathbf{x}$ and $\nabla E(\mathbf{x})$ that tightly concentrates about its mean can be well approximated as an average over the mean-field distribution $P_\beta(x,y)$. Under this assumption, the functions $q(\mathbf{x})$, $A(\mathbf{x})$, and $C(\mathbf{x})$ concentrate about their means as follows:

$$\begin{aligned} q(\mathbf{x}) &= \int dxdy P_\beta(x,y)x^2, \\ A(\mathbf{x}) &= 2\beta \int dxdy P_\beta(x,y)xy - t, \\ C(\mathbf{x}) &= 4\beta^2 \int dxdy P_\beta(x,y)y^2 - 2\beta. \end{aligned}$$

The distribution $P_\beta(x,y)$ can be obtained similarly to the grand potential, resulting in

$$P_\beta(x,y) = Z_\beta^{-1}\left\langle \int dy_* \delta(x_*-x)\delta(y_*-y)\sqrt{\frac{\beta}{\pi}}e^{-\beta y_*^2} \times \sum_{x_* \in \mathrm{Crt}(E_{\mathrm{MF}}[y_*+h])} e^{x_* z+\mu\bar{\mathcal{I}}(x_*)} \right\rangle_{h,z}, \tag{A5}$$

where

$$Z_\beta = \left\langle \int dy \left(\frac{\beta}{\pi}\right)^{1/2} e^{-\beta y^2} \sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[y+h])} e^{xz+\mu\bar{\mathcal{I}}(x)} \right\rangle_{h,z}.$$

Substituting this expression into Eq. (A5) and taking the low-temperature limit $\beta\to\infty$, we obtain self-consistent equations for the order parameters associated with critical points:

$$\begin{aligned} q &= Z^{-1}\left\langle \sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)}x^2 \right\rangle_{h,z}, \\ A &= Z^{-1}\left\langle \frac{d}{dh}\sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)}x \right\rangle_{h,z} - t, \\ C &= Z^{-1}\left\langle \frac{d^2}{dh^2}\sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)} \right\rangle_{h,z}, \end{aligned}$$

with $Z = \langle \sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)}\rangle_{h,z}$. Note that, while the derivatives with respect to h can be undefined with some values of h, we can always define the expectation value of the derivatives through integration by parts. Similarly, the self-consistent equation for t can be obtained under the limit $\beta\to\infty$:

$$\begin{aligned} t &= \lim_{\beta\to\infty}\int dxdy P_\beta(x,y)\frac{1}{\partial^2 E_I(x)-t} \\ &= Z^{-1}\left\langle \sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)}\frac{dx}{dh} \right\rangle_{h,z}. \end{aligned}$$

These self-consistent equations can also be given a variational characterization. Indeed, they are equivalent to the equations obtained by extremizing the right-hand side of Eq. (65) with respect to q, A, C, and t, i.e.,

$$\Omega = \operatorname*{ext}_{(q,A,C,t)}\left[-\frac{1}{2}(Cq+A^2)-At + \log\left\langle \sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)} \right\rangle_{h,z}\right]. \tag{A6}$$

This equivalence can be seen by explicitly calculating the stationary condition of the right-hand side.

It is not difficult to see that Eq. (A6) is equivalent to an extremizer of Eq. (34). This means that the cavity method and the aforementioned replica-based calculation give the equivalent result.

3. A geometric interpretation of the supersymmetry-breaking order parameters

Here, we show that the supersymmetry-breaking order parameters A and C can be interpreted as susceptibilities of the grand potential Ω to certain perturbations. Specifically, we consider the perturbations of the quadratic term and a random magnetic field:

$$E'(\mathbf{x}) = E(\mathbf{x}) - \frac{a}{2}|\mathbf{x}|^2 + \sqrt{2s}\mathbf{g}\cdot\mathbf{x},$$

where a is a constant for the quadratic term and the other term $\mathbf{g}\cdot\mathbf{x}$ represents the coupling with the external random field $\mathbf{g}$, a centered Gaussian vector with unit variance. s is the scalar coupling constant. The corresponding mean-field energy is defined as

$$E'_{\mathrm{MF}}[h,a,\sqrt{2s}g](x) = E_{\mathrm{MF}}[h](x) - \frac{a}{2}x^2 + \sqrt{2s}gx.$$

We first consider the case where we have only the quadratic perturbation ($a>0$ and $s=0$). The grand potential $\Omega(a)$ of the perturbed energy function has the form of Eq. (A6) with different mean-field energy function $E'_{\mathrm{MF}}[h,a,\sqrt{2s}g](x)$. The order parameters $q(a)$, $A(a)$, $C(a)$, and $t(a)$ all depend on a and satisfy the stationary condition for $\Omega(a)$. Hence, the derivative of $\Omega(a)$ with respect to a is given by

$$\left.\frac{d\Omega}{da}\right|_{a=0} = \left.\frac{\partial\Omega}{\partial a}\right|_{a=0} = \frac{1}{Z}\frac{\partial}{\partial t}\left\langle \sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)} \right\rangle_{h,z} = A. \tag{A7}$$

This equality shows that A can be interpreted as a susceptibility of the grand potential with respect to the strength of the quadratic perturbation.

Similarly, we consider the case only with the coupling with the random external field (i.e., $a = 0$, $s > 0$). We define the averaged grand potential as follows:

$$\exp[N\Omega(s)] = \mathbb{E}_g \mathbb{E}_J \left[\int \prod_{i=0}^{N-1} dx_i |\det H(\mathbf{x})| e^{\mu I(\mathbf{x})} \times \prod_{i=0}^{N-1} \delta\left(\partial E_I(x_i) + \sum_{j=0}^{N-1} J_{ij} x_j + \sqrt{2s} g_i \right) \right],$$

where $\mathbb{E}_g$ represents the average over the random external field $\mathbf{g}$. By calculating the average over $\mathbf{g}$, we get $\exp[N\Omega(s)] = \exp(N\Omega_{\frac{1}{4s}})$. Therefore, the derivative with respect to s is given by

$$\left.\frac{d\Omega}{ds}\right|_{s=0} = 2 \left.\frac{d\Omega_\beta}{d\frac{1}{2}\beta^{-1}}\right|_{\beta=\infty} = \frac{1}{Z} \left\langle \frac{d^2}{dh^2} \sum\nolimits_{x \in \mathrm{Crt}(E'_{\mathrm{MF}}[h])} e^{xz + \mu \bar{\mathcal{I}}(x)} \right\rangle_{h,z} = C.$$

Thus, C can be interpreted as the susceptibility of the grand potential with respect to the coupling with the random external field.

It is not difficult to see that A and C can be understood also as the derivatives of the *mean-field* grand potential $\Omega_{\mathrm{MF}}(a,s) \coloneqq \log \langle \sum_{x \in \mathrm{Crt}(E'_{\mathrm{MF}}[h,a,\sqrt{2s}g])} e^{xz+\mu\bar{\mathcal{I}}(x)} \rangle_{h,z,g}$. Indeed, for the order parameter A, it is clear that Eq. (A7) implies $A = (\partial/\partial a)|_{a=0}\Omega_{\mathrm{MF}}$. Similarly, the following argument shows $C = (\partial/\partial s)|_{s=0}\Omega_{\mathrm{MF}}$. Notice that the quantity inside the bracket $\sum_{x \in \mathrm{Crt}(E'_{\mathrm{MF}}[h,0,\sqrt{2s}g])} e^{xz+\mu\bar{\mathcal{I}}(x)}$ is a function of $h + \sqrt{2s}g$, which is a centered random Gaussian variable. Its variance and covariance with z are $\langle (h+\sqrt{2s}g)^2 \rangle = q + 2s$ and $\langle (h+\sqrt{2s}g) z \rangle = A$, respectively. Hence, the Gaussian average over h, z, and g is equivalent to the average over $h + \sqrt{2s}g$ and z with the covariance matrix

$$\begin{pmatrix} q+2s & A \\ A & C \end{pmatrix}.$$

Therefore, the derivative by s is equivalent to the derivative by $q/2$. Thus,

$$\left.\frac{\partial \Omega_{\mathrm{MF}}}{\partial s}\right|_{s=0} = 2\frac{d}{dq} \left\langle \sum\nolimits_{x \in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz+\mu\bar{\mathcal{I}}(x)} \right\rangle_{h,z} = C.$$

4. Nondegeneracy of critical points implies supersymmetry and structural stability

Next, we show that if typical critical points have eigenvalues bounded away from 0, then its supersymmetry-breaking order parameters are vanishing, which implies the structural stability of the typical critical points (since A and C are susceptibilities of the grand potential to certain perturbations of the energy landscape.)

Recall that, with a finite inverse temperature, A and C are defined as $A(\mathbf{x}) = 2N^{-1}\beta \nabla E(\mathbf{x}) \cdot \mathbf{x} - t$ and $C(\mathbf{x}) = 4N^{-1}\beta^2 |\nabla E(\mathbf{x})|^2 - 2\beta$, respectively. If β is large enough, we anticipate that the order parameters of a typical critical point can be obtained by averaging $A(\mathbf{x})$ and $C(\mathbf{x})$ over a small neighborhood U around the critical point, i.e.,

$$A + t_R = \left[\int_U \prod_{i=0}^{N-1} dx_i e^{-\beta|\nabla E(\mathbf{x})|^2} |\det H(\mathbf{x})| \right]^{-1} \times \int_U \prod_{i=0}^{N-1} dx_i 2\beta \frac{\nabla E(\mathbf{x}) \cdot \mathbf{x}}{N} e^{-\beta|\nabla E(\mathbf{x})|^2} |\det H(\mathbf{x})|.$$

Here, we exploit the fact that $\mathcal{I}(\mathbf{x})$ is constant in the small neighborhood because the eigenvalues are bounded away from zero. Similarly,

$$C + 2\beta = \left[\int_U \prod_{i=0}^{N-1} dx_i e^{-\beta|\nabla E(\mathbf{x})|^2} |\det H(\mathbf{x})| \right]^{-1} \times \int_U \prod_{i=0}^{N-1} dx_i 4\beta^2 \frac{|\nabla E(\mathbf{x})|^2}{N} e^{-\beta|\nabla E(\mathbf{x})|^2} |\det H(\mathbf{x})|.$$

By the assumption, we can apply the inverse function theorem to $y_i \coloneqq \nabla_i E(\mathbf{x})$ around the neighborhood of the critical point. The Jacobian is given by the Hessian $(dy_i/dx_j) = H_{ij}$, and we approximately see that $x_i = H^{-1}_{ij} y_j$. Therefore,

$$A \approx \frac{2\beta}{N} \int_{\mathbb{R}^N} d\mathbf{y} \mathbf{y}^T H^{-1} \mathbf{y} e^{-\beta|\mathbf{y}|^2} \Big/ \int_{\mathbb{R}^N} d\mathbf{y} e^{-\beta|\mathbf{y}|^2} - t = N^{-1}\mathrm{Tr}H^{-1} - t = 0,$$

$$C \approx \frac{2\beta^2}{N} \int_{\mathbb{R}^N} d\mathbf{y} |\mathbf{y}|^2 e^{-\beta|\mathbf{y}|^2} \Big/ \int_{\mathbb{R}^N} d\mathbf{y} e^{-\beta|\mathbf{y}|^2} - 2\beta = 0,$$

where $d\mathbf{y} \coloneqq \prod_{i=0}^{N-1} dy_i$.

The contrapositive of this result then immediately tells us that if supersymmetry-breaking order parameters are nonzero, then typical critical points are marginally unstable (with no gap in the eigenvalue density away from 0). Earlier results additionally imply such marginally unstable critical points are also structurally unstable to small perturbations in the energy landscape.

a. Convexity of the mean-field energy landscape implies vanishing complexity and supersymmetry

In the following discussion, we address two statements on the convexity of the mean-field energy landscape and supersymmetry of the typical critical points. First, we show that the strict convexity of the mean-field energy landscape $E_{\mathrm{MF}}[0](x)$ is a sufficient condition for the self-consistent equations to have a supersymmetric fixed point with $A = C = 0$, resulting in vanishing complexity. Suppose that the supersymmetry-breaking order parameters vanish $A = C = 0$. In this setting, the Gaussian random variable z is always zero, and the average $\langle\cdot\rangle_{h,z}$ is only over h with variance of q, which we denote $\langle\cdot\rangle_h$. Therefore, the self-consistent equations for A and C are reduced to $A = Z^{-1}\langle (d/dh)\sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} x\rangle_h - t$ and $C = Z^{-1}\langle (d^2/dh^2)\sum_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} 1\rangle_h$, respectively. It is easy to see that these self-consistent equations hold when the mean-field energy function without external field $E_{\mathrm{MF}}[0](x)$ is strictly convex and has a monotonically increasing gradient. Indeed, $\mathrm{Crt}(E_{\mathrm{MF}}[h])$ contains only a single element for any $h\in\mathbb{R}$, and, therefore,

$$A = Z^{-1}\left\langle \frac{d}{dh}\sum\nolimits_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} x\right\rangle_h - t,$$
$$C = Z^{-1}\left\langle \frac{d^2}{dh^2}\sum\nolimits_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} 1\right\rangle_h. \tag{A8}$$

Hence, the convexity of the mean-field energy implies supersymmetry and, therefore, also implies structural stability of typical critical points. In this case, the grand potential (and, therefore, the complexity) of typical critical points vanishes:

$$\Omega = -\frac{1}{2}(Cq + A^2) - At + \log\left\langle \sum\nolimits_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} e^{xz}\right\rangle_{h,z}$$
$$= \log\left\langle \sum\nolimits_{x\in \mathrm{Crt}(E_{\mathrm{MF}}[h])} 1\right\rangle_h = 0.$$

Next, we prove that the convexity is a necessary condition when the mean-field energy is given by the quartic function $E_{\mathrm{MF}}[0](x) = x^4/4 - a_{\mathrm{eff}}x^2/2$; i.e., $a_{\mathrm{eff}} \le 0$ is a necessary condition for the existence of supersymmetric solutions—i.e., if $a_{\mathrm{eff}} > 0$, the self-consistent equations do not have any supersymmetric solution. Suppose $a_{\mathrm{eff}} > 0$ and a supersymmetric solution exists, with which Eq. (A8) holds. The second equation of Eq. (A8) can be calculated as $C = (2\tilde{h}/\sqrt{2\pi q^3})\exp[-(\tilde{h}^2/2q)]$, where $\tilde{h} = (2a_{\mathrm{eff}}^{2/3}/3\sqrt{3})$. This is strictly positive and contradicts our assumption of supersymmetry $A = C = 0$. Hence, the SUSY solution does not exist.

APPENDIX B: DETAILS OF COMPARISON OF THEORY AND NUMERICAL EXPERIMENTS IN FIGS. 3–5

1. Numerical exploration of critical points

Here, we explain how we numerically sample typical critical points in Fig. 3, typical minima in Fig. 4, and global minima in Fig. 5 for different realizations of J.

In Fig. 3, we sample critical points via Newton's method with many different initializations. Note that Newton's method converges not only to local minima, but also to critical points of any index [111] and, hence, works as an efficient sampler of all critical points. The detailed experimental parameters are shown in Table I, including the system size N, chosen for each value of a. The chosen system size N decreases as a grows. This is because the number of critical points increases exponentially, and, hence, it is difficult to sample all the critical points with large N in the large-a regime. The initial states are sampled from a centered Gaussian distribution for $a \le 1$, while for $a > 1$ we initialize at every point of $\{-\sqrt{a}, 0, \sqrt{a}\}^N$. After the deduplication of the sampled critical points, we estimate the most frequent values of the energy and the index. In order to select specifically the most typical critical points, we focus only on sampled critical points with the most frequent values of both the energy and index. The most frequent index is plotted in the figure with the error bar of length $1/N$ reflecting the minimum discretization of the fractional index r at finite N. To find the most likely energy, we discretize energy into N_{bins} bins and include only

TABLE I. Parameters for sampling typical critical points.

Param	Description	Values													
a	Gain parameter	−1.5	−1.0	−0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
N	System size	400	400	120	48	20	20	14	14	12	12	12	12	12	12
N_{sample}	Number of sampled minima for each instance	1×10^3	2×10^3	1×10^5	3×10^5	4×10^5	1.6×10^6	3^{14}	3^{14}	3^{12}	3^{12}	3^{12}	3^{12}	3^{12}	3^{12}
N_J	Number of sampled instances	20	20	20	10	10	5	5	5	5	5	5	5	5	5
N_{bins}	Number of bins for energy histogram	100	100	100	100	100	100	100	100	100	100	100	100	100	100

TABLE II. Parameters for sampling typical minima.

Param	Description	Values														
a	Gain parameter	-2.0	-1.75	-1.5	-1.25	-1.0	-0.75	-0.5	-0.25	0.0	0.25	1.0	2.0	3.0	4.0	5.0
N	System size	1×10^3	1×10^3	1×10^3	1×10^3	400	200	150	100	50	30	30	20	20	15	20
N_J	Number of sampled instances	20	20	20	20	20	20	20	20	20	20	20	20	20	20	20
N_{sample}	Number of sampled minima for each instance	7×10^1	7×10^3	7×10^3	7×10^3	7×10^3	7×10^5	1.4×10^6	2.8×10^6	2.8×10^6	2.8×10^6	2.8×10^6	7×10^6	7×10^6	7×10^6	7×10^6
N_{bins}	Number of bins for energy histogram	10	10	10	20	25	25	50	100	100	50	25	500	500	500	500

critical points whose energy is in the most likely bin. From this restricted set of the most typical critical points, we can directly compute order parameters q in Eq. (49) and t in Eq. (50) for each typical critical point. Then, we compute the average values of q and t over all the typical critical points for each instance. For each sampled instance J, we compute the means of order parameters across all the typical critical points. In the plots of q and t in Fig. 3, the error bars represent the standard deviation of the mean over the different instances of J. We sample more realizations of J at smaller a to compensate for the fact that there are fewer critical points at smaller a. To compute the distribution of spins $P(x)$, we compute the empirical histogram of individual spin values across the ensemble of typical critical points with the most frequent energy and index. The Hessian eigenspectrum is computed similarly as ensembles of the eigenvalues of those sampled critical points.

In Fig. 4, we minimize the energy function by the Newton-Conjugate-Gradient algorithm from many randomly initialized points to sample minima. The detailed experimental parameters are shown in Table II, such as the number of samples N_{sample} and the system size N chosen for each value of a. After the deduplication of the sampled minima, we focus on the typical minima, defined as the minima with the most likely energy in a bin among N_{bins} energy bins (just as we do for critical points above). We compute the order parameters q and t from this set of minima. The error bars in the figure represent the standard deviation across instances of J of the values over all typical minima for each instance. The complexity is calculated for $a > -1.5$ as $N^{-1}\log(\mathcal{N}_{\max}/\delta E)$, where $\mathcal{N}_{\max}$ is the number of minima in the most likely energy bin and δE is the bin width. For $a \le -1.5$, the numbers of minima we obtain are not enough to estimate the density $\mathcal{N}_{\max}/\delta E$, and, hence, we instead estimate the complexity as the logarithm of the total number of minima, divided by N. We compute the ensemble of $P(x)$ [Fig. 4(a)] and the Hessian spectrum [Fig. 4(b)] across all minima in the most likely energy bin. The system size N is selected according to the value of a, from $N = 15$ to $N = 10^3$. Since the complexity Σ is smaller for smaller a, we use a larger N for smaller a to be able to more accurately estimate the smaller complexity.

In Fig. 5, we sample minima in the same manner above using the Newton-CG method with N_{sample} different initialized points and choose the lowest energy state from each of N_J instances. The values of N, N_{sample}, and N_J for each value of a are displayed in Table III. The order parameters and the distributions are computed as the ensemble of those sampled lowest-energy states.

2. Solving self-consistent equations for order parameters

In the following, we describe our approach to solving the self-consistent equations (39), for each value of the gain

TABLE III. Parameters for sampling global minima.

Param	Description	Values									
a	Gain parameter	−4.0	−3.0	−2.0	−1.25	−1.0	−0.5	0.0	1.0	2.0	4.0
N	System size	1000	1000	1000	1000	800	200	100	40	40	20
N_J	Number of sampled instances	20	20	20	20	20	20	20	200	200	200
N_{sample}	Number of sampled minima for each instance	7×10^1	7×10^1	7×10^1	7×10^3	2.8×10^6	1.4×10^6	2.8×10^6	7×10^6	7×10^6	7×10^6

parameter a. Our solution is obtained by iteratively updating the order parameters until they satisfy the self-consistent relations. To simplify the iterative process, we fix a value of $a_{\text{eff}} = a + t$ rather than fixing the gain parameter a.

With a fixed value of a_{eff}, we proceed to iteratively update q, A, and Cq with the following equations starting from $q = 1.0$, $A = 0.5$, and $Cq = 2.0$:

$$q \leftarrow \langle x^2 \rangle,$$
$$A \leftarrow \frac{\langle xh(x) \rangle}{2q} - \frac{1}{2}\left\langle \frac{1}{3x^2 - a_{\text{eff}}} \right\rangle,$$
$$Cq \leftarrow -1 + q^{-1}\langle h^2(x) \rangle - 2q^{-1}A\langle xh(x) \rangle + A^2.$$

We perform these updates for a total of 300 iterations, after which the variables converge within an error margin of 10^{-6}. Subsequently, we compute t using the following equation: $t = \langle 1/3x^2 - a_{\text{eff}} \rangle$. Finally, we determine the value of a by calculating $a = a_{\text{eff}} - t$.

[1] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity* (Courier Corporation, New York, 1998).

[2] N. Mohseni, P. L. McMahon, and T. Byrnes, *Ising machines as hardware solvers of combinatorial optimization problems*, Nat. Rev. Phys. **4**, 363 (2022).

[3] K. P. Kalinin and N. G. Berloff, *Computational complexity continuum within Ising formulation of NP problems*, Commun. Phys. **5**, 20 (2022).

[4] M. Syed and N. G. Berloff, *Physics-enhanced bifurcation optimisers: All you need is a canonical complex network*, IEEE J. Sel. Top. Quantum Electron. **29**, 1 (2023).

[5] F. Molnár, S. R. Kharel, X. S. Hu, and Z. Toroczkai, *Accelerating a continuous-time analog sat solver using GPUs*, Comput. Phys. Commun. **256**, 107469 (2020).

[6] S. Dutta, A. Khanna, A. S. Assoa, H. Paik, D. Schlom, Z. Toroczkai, A. Raychowdhury, and S. Datta, *An Ising Hamiltonian solver using stochastic phase-transition nano-oscillators*, Nat. Electron. Rev. **4**, 502 (2020).

[7] T. Honjo, T. Sonobe, K. Inaba, T. Inagaki, T. Ikuta, Y. Yamada, T. Kazama, K. Enbutsu, T. Umeki, R. Kasahara, K. ichi Kawarabayashi, and H. Takesue, *100,000-spin coherent Ising machine*, Sci. Adv. **7**, eabh0952 (2021).

[8] T. Inagaki, Y. Haribara, K. Igarashi, T. Sonobe, S. Tamate, T. Honjo, A. Marandi, P. L. McMahon, T. Umeki, K. Enbutsu, O. Tadanaga, H. Takenouchi, K. Aihara, K.-i. Kawarabayashi, K. Inoue, S. Utsunomiya, and H. Takesue, *A coherent Ising machine for 2000-node optimization problems*, Science **354**, 603 (2016).

[9] P. L. McMahon, A. Marandi, Y. Haribara, R. Hamerly, C. Langrock, S. Tamate, T. Inagaki, H. Takesue, S. Utsunomiya, K. Aihara, R. L. Byer, M. M. Fejer, H. Mabuchi, and Y. Yamamoto, *A fully programmable 100-spin coherent Ising machine with all-to-all connections*, Science **354**, 614 (2016).

[10] Y. Yamamoto, T. Leleu, S. Ganguli, and H. Mabuchi, *Coherent Ising machines—Quantum optics and neural network Perspectives*, Appl. Phys. Lett. **117**, 160501 (2020).

[11] Y. Yamamoto, K. Aihara, T. Leleu, K.-i. Kawarabayashi, S. Kako, M. Fejer, K. Inoue, and H. Takesue, *Coherent Ising machines—Optical neural networks operating at the quantum limit*, npj Quantum Inf. **3**, 49 (2017).

[12] Z. Wang, A. Marandi, K. Wen, R. L. Byer, and Y. Yamamoto, *Coherent Ising machine based on degenerate optical parametric oscillators*, Phys. Rev. A **88**, 063853 (2013).

[13] A. Yamamura, K. Aihara, and Y. Yamamoto, *Quantum model for coherent Ising machines: Discrete-time measurement feedback formulation*, Phys. Rev. A **96**, 053834 (2017).

[14] M. Calvanese Strinati, D. Pierangeli, and C. Conti, *All-optical scalable spatial coherent Ising machine*, Phys. Rev. Appl. **16**, 054022 (2021).

[15] D. Sherrington and S. Kirkpatrick, *Solvable model of a spin-glass*, Phys. Rev. Lett. **35**, 1792 (1975).

[16] F. Barahona, *On the computational complexity of Ising spin glass models*, J. Phys. A **15**, 3241 (1982).

[17] A. Lucas, *Ising formulations of many NP problems*, Front. Phys. **2**, 00005 (2014).

[18] M. Jankowski, R. Yanagimoto, E. Ng, R. Hamerly, T. P. McKenna, H. Mabuchi, and M. M. Fejer, *Ultrafast nonlinear photonics—From classical physics to non-Gaussian quantum dynamics*, Adv. Opt. Photonics **16**, 347 (2024).

[19] R. Yanagimoto, R. Nehra, R. Hamerly, E. Ng, A. Marandi, and H. Mabuchi, *Quantum nondemolition measurements with optical parametric amplifiers for ultrafast universal quantum information processing*, PRX Quantum **4**, 010333 (2023).

[20] G. Bilbro, R. Mann, T. K. Miller, W. E. Snyder, D. E. van den Bout, and M. White, *Optimization by mean field annealing*, in *Advances in Neural Information Processing*

Systems 1, edited by D. S. Touretzky (Morgan-Kaufmann, Burlington, MA, 1989), pp. 91–98.
[21] P. Chaudhari and S. Soatto, *On the energy landscape of deep networks*, arXiv:1511.06485.
[22] Y. V. Fyodorov and P. Le Doussal, *Topology trivialization and large deviations for the minimum in the simplest random optimization*, J. Stat. Phys. **154**, 466 (2014).
[23] Y. Haribara, H. Ishikawa, S. Utsunomiya, K. Aihara, and Y. Yamamoto, *Performance evaluation of coherent Ising machines against classical neural networks*, Quantum Sci. Technol. **2**, 044002 (2017).
[24] R. Hamerly *et al.*, *Experimental investigation of performance differences between coherent Ising machines and a quantum annealer*, Sci. Adv. **5**, aau0823 (2019).
[25] G. Vincent, K. Mirko, and R. Jean-Paul, *Simulated annealing: A proof of convergence*, IEEE Trans. Pattern Anal. Mach. Intell. **16**, 652 (1994).
[26] T. Kadowaki and H. Nishimori, *Quantum annealing in the transverse Ising model*, Phys. Rev. E **58**, 5355 (1998).
[27] T. Leleu, Y. Yamamoto, S. Utsunomiya, and K. Aihara, *Combinatorial optimization using dynamical phase transitions in driven-dissipative systems*, Phys. Rev. E **95**, 022118 (2017).
[28] T. Leleu, Y. Yamamoto, P. L. McMahon, and K. Aihara, *Destabilization of local minima in analog spin systems by correction of amplitude heterogeneity*, Phys. Rev. Lett. **122**, 040607 (2019).
[29] J. Wang, D. Ebler, K. M. Wong, D. S. W. Hui, and J. Sun, *Bifurcation behaviors shape how continuous physical dynamics solves discrete Ising optimization*, Nat. Commun. **14**, 2510 (2023).
[30] J. S. Cummins, H. Salman, and N. G. Berloff, *Classical vs quantum annealing and manifold reduction in soft-spin minimizers of Ising Hamiltonians*, arXiv:2311.17359.
[31] M. Calvanese Strinati, L. Bello, E. G. Dalla Torre, and A. Pe'er, *Can nonlinear parametric oscillators solve random Ising models?*, Phys. Rev. Lett. **126**, 143901 (2021).
[32] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, World Scientific Lecture Notes in Physics Vol. 9 (World Scientific, Singapore, 1986).
[33] A. J. Bray and M. A. Moore, *Chaotic nature of the spin-glass phase*, Phys. Rev. Lett. **58**, 57 (1987).
[34] F. Krzakala and O. Martin, *Chaotic temperature dependence in a model of spin glasses*, Eur. Phys. J. B **28**, 199 (2002).
[35] T. Rizzo and A. Crisanti, *Chaos in temperature in the Sherrington-Kirkpatrick model*, Phys. Rev. Lett. **90**, 137201 (2003).
[36] L. Fernandez, V. Martin-Mayor, G. Parisi, and B. Seoane, *Temperature chaos in 3D Ising spin glasses is driven by rare events*, Europhys. Lett. **103**, 67003 (2013).
[37] A. Montanari, *Optimization of the Sherrington–Kirkpatrick Hamiltonian*, SIAM J. Comput. FOCS19-1 (2021), 10.1137/20M132016X.
[38] C. R. Laumann, A. Pal, and A. Scardicchio, *Many-body mobility edge in a mean-field quantum spin glass*, Phys. Rev. Lett. **113**, 200405 (2014).
[39] L. Foini, G. Semerjian, and F. Zamponi, *Solvable model of quantum random optimization problems*, Phys. Rev. Lett. **105**, 167204 (2010).
[40] B. Altshuler, H. Krovi, and J. Roland, *Anderson localization makes adiabatic quantum optimization fail*, Proc. Natl. Acad. Sci. U.S.A. **107**, 12446 (2010).
[41] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi, *The quantum adiabatic algorithm applied to random optimization problems: The quantum spin glass perspective*, Phys. Rep. **523**, 127 (2013).
[42] A. Roy, S. Jahani, C. Langrock, M. Fejer, and A. Marandi, *Spectral phase transitions in optical parametric oscillators*, Nat. Commun. **12**, 835 (2021).
[43] Note that there exist more efficient algorithms to optimize the SK instances than the CIM. For instance, Montanari recently proposed an algorithm specialized for the SK instances which can achieve an intensive energy within a constant ϵ above the ground energy with high probability. This algorithm's complexity is $O(N^2)$ for any fixed $\epsilon > 0$ independent of N.
[44] A. Yamamura, H. Mabuchi, and S. Ganguli, *Geometric landscape annealing as an optimization principle underlying the coherent Ising machine*, arXiv:2309.08119.
[45] A. J. Bray and M. A. Moore, *Metastable states in spin glasses*, J. Phys. C **13**, L469 (1980).
[46] A. J. Bray and M. A. Moore, *Metastable states in the solvable spin glass model*, J. Phys. A **14**, L377 (1981).
[47] T. Aspelmeier, A. J. Bray, and M. A. Moore, *Complexity of Ising spin glasses*, Phys. Rev. Lett. **92**, 087203 (2004).
[48] A. Crisanti, L. Leuzzi, G. Parisi, and T. Rizzo, *Spin-glass complexity*, Phys. Rev. Lett. **92**, 127203 (2004).
[49] G. Parisi and T. Rizzo, *On supersymmetry breaking in the computation of the complexity*, J. Phys. A **37**, 7979 (2004).
[50] A. Crisanti, L. Leuzzi, and T. Rizzo, *Complexity in mean-field spin-glass models: Ising p-spin*, Phys. Rev. B **71**, 094202 (2005).
[51] T. Rizzo, *TAP complexity, the cavity method and supersymmetry*, J. Phys. A **38**, 3287 (2005).
[52] T. Aspelmeier, R. A. Blythe, A. J. Bray, and M. A. Moore, *Free-energy landscapes, dynamics, and the edge of chaos in mean-field models of spin glasses*, Phys. Rev. B **74**, 184411 (2006).
[53] Z. Fan, S. Mei, and A. Montanari, *TAP free energy, spin glasses, and variational inference*, arXiv:1808.07890.
[54] M. Müller, L. Leuzzi, and A. Crisanti, *Marginal states in mean-field glasses*, Phys. Rev. B **74**, 134431 (2006).
[55] Y. V. Fyodorov, *Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices*, Phys. Rev. Lett. **92**, 240601 (2004).
[56] Y. V. Fyodorov, *Counting stationary points of random landscapes as a random matrix problem*, arXiv:cond-mat/0507059.
[57] A. J. Bray and D. S. Dean, *Statistics of critical points of Gaussian fields on large-dimensional spaces*, Phys. Rev. Lett. **98**, 150201 (2007).
[58] Y. V. Fyodorov and I. Williams, *Replica symmetry breaking condition exposed by random matrix calculation of landscape complexity*, J. Stat. Phys. **129**, 1081 (2007).

[59] Y. V. Fyodorov, *On statistical mechanics of a single particle in high-dimensional random landscapes*, arXiv:0801.0732.

[60] A. Cavagna, J. P. Garrahan, and I. Giardina, *Quenched complexity of the mean-field p-spin spherical model with external magnetic field*, J. Phys. A **32**, 711 (1999).

[61] A. Cavagna, I. Giardina, and G. Parisi, *Stationary points of the Thouless-Anderson-Palmer free energy*, Phys. Rev. B **57**, 11251 (1998).

[62] A. Crisanti, L. Leuzzi, G. Parisi, and T. Rizzo, *Complexity in the Sherrington-Kirkpatrick model in the annealed approximation*, Phys. Rev. B **68**, 174401 (2003).

[63] A. Crisanti, L. Leuzzi, and T. Rizzo, *The complexity of the spherical $\{p\}$-spin spin glass model, revisited*, Eur. Phys. J. B **36**, 129 (2003).

[64] Y. V. Fyodorov, *High-dimensional random fields and random matrix theory*, arXiv:1307.2379.

[65] A. Auffinger and G. B. Arous, *Complexity of random smooth functions on the high-dimensional sphere*, Ann. Probab. **41**, 4214 (2013).

[66] A. Auffinger, G. B. Arous, and J. Černý, *Random matrices and complexity of spin glasses*, Commun. Pure Appl. Math. **66**, 165 (2013).

[67] V. Ros, G. Ben Arous, G. Biroli, and C. Cammarota, *Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima, and phase transitions*, Phys. Rev. X **9**, 011003 (2019).

[68] V. Ros, G. Biroli, and C. Cammarota, *Complexity of energy barriers in mean-field glassy systems*, Europhys. Lett. **126**, 20003 (2019).

[69] S. Becker, Y. Zhang, and A. A. Lee, *Geometry of energy landscapes and the optimizability of deep neural networks*, Phys. Rev. Lett. **124**, 108301 (2020).

[70] R. J. Adler, J. E. Taylor *et al.*, *Random Fields and Geometry* (Springer, New York, 2007), Vol. 80.

[71] M. Potters and J.-P. Bouchaud, *A First Course in Random Matrix Theory: for Physicists, Engineers and Data Scientists* (Cambridge University Press, Cambridge, England, 2020).

[72] A. Cavagna, I. Giardina, G. Parisi, and M. Mezard, *On the formal equivalence of the TAP and thermodynamic methods in the SK model*, J. Phys. A **36**, 1175 (2003).

[73] A. Annibale, A. Cavagna, I. Giardina, G. Parisi, and E. Trevigne, *The role of the Becchi–Rouet–Stora–Tyutin supersymmetry in the calculation of the complexity for the Sherrington–Kirkpatrick model*, J. Phys. A **36**, 10937 (2003).

[74] A. Annibale, A. Cavagna, I. Giardina, and G. Parisi, *Supersymmetric complexity in the Sherrington-Kirkpatrick model*, Phys. Rev. E **68**, 061103 (2003).

[75] A. Annibale, G. Gualdi, and A. Cavagna, *Coexistence of supersymmetric and supersymmetry-breaking states in spherical spin-glasses*, J. Phys. A **37**, 11311 (2004).

[76] A. Cavagna, I. Giardina, and G. Parisi, *Cavity method for supersymmetry-breaking spin glasses*, Phys. Rev. B **71**, 024422 (2005).

[77] F. J. Dyson, *A Brownian-Motion model for the eigenvalues of a random matrix*, J. Math. Phys. (N.Y.) **3**, 1191 (1962).

[78] M. Madan Lal, *Random matrices*, in *Pure and Applied Mathematics* (Elsevier, New York, 2004), Vol. 142, pp. 1–32.

[79] L. A. Pastur, *On the spectrum of random matrices*, Theor. Math. Phys. **10**, 67 (1972).

[80] T. Aspelmeier, A. Billoire, E. Marinari, and M. A. Moore, *Finite-size corrections in the Sherrington–Kirkpatrick model*, J. Phys. A **41**, 324008 (2008).

[81] S. Boettcher, *Extremal optimization for Sherrington-Kirkpatrick spin glasses*, Eur. Phys. J. B **46**, 501 (2005).

[82] A. Crisanti and T. Rizzo, *Analysis of the $\{\infty\}$-replica symmetry breaking solution of the Sherrington-Kirkpatrick model*, Phys. Rev. E **65**, 046137 (2002).

[83] M. J. Schmidt, *Replica symmetry breaking at low temperature*, Ph.D. thesis, Universität Würzburg, 2008.

[84] F. Morone, F. Caltagirone, E. Harrison, and G. Parisi, *Replica theory and spin glasses*, arXiv:1409.2722.

[85] H.-J. Sommers and W. Dupont, *Distribution of frozen fields in the mean-field theory of spin glasses*, J. Phys. C **17**, 5785 (1984).

[86] The Ising energy of random spin states is almost zero for the following reason. Flipping some of the Ising spins, we can make all spins $+1$. By flipping the sign of J_{ij} properly at the same time, we can keep the Ising energy unchanged. This new J matrix follows Gaussian orthogonal ensemble as well. The Ising energy is given by the sum of all the elements of the new J matrix. By the law of large numbers, this quantity converges to its mean 0.

[87] We set the norm of the initialized principal eigenvector to match that of a Gaussian random initialization.

[88] F. Hamze, J. Raymond, C. A. Pattison, K. Biswas, and H. G. Katzgraber, *Wishart planted ensemble: A tunably rugged pairwise Ising model with a first-order phase transition*, Phys. Rev. E **101**, 052102 (2020).

[89] A. El Alaoui and F. Krzakala, *Estimation in the spiked Wigner model: A short proof of the replica formula*, in *Proceedings of the 2018 IEEE International Symposium on Information Theory (ISIT)* (IEEE, New York, 2018), pp. 1874–1878.

[90] M. Stern, H. Sompolinsky, and L. F. Abbott, *Dynamics of random neural networks with bistable units*, Phys. Rev. E **90**, 062710 (2014).

[91] G. Wainrib and J. Touboul, *Topological and dynamical complexity of random neural networks*, Phys. Rev. Lett. **110**, 118101 (2013).

[92] G. Biroli, G. Bunin, and C. Cammarota, *Marginally stable equilibria in critical ecosystems*, New J. Phys. **20**, 083051 (2018).

[93] G. Bunin, *Ecological communities with Lotka-Volterra dynamics*, Phys. Rev. E **95**, 042414 (2017).

[94] J. R. Ipsen and P. J. Forrester, *Kac–Rice fixed point analysis for single- and multi-layered complex systems*, J. Phys. A **51**, 474003 (2018).

[95] L. Chen and K. Aihara, *Chaotic simulated annealing by a neural network model with transient chaos*, Neural Netw. **8**, 915 (1995).

[96] C.-s. Zhou and T.-l. Chen, *Chaotic annealing for optimization*, Phys. Rev. E **55**, 2580 (1997).

[97] J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, *The Kuramoto model: A simple paradigm for synchronization phenomena*, Rev. Mod. Phys. **77**, 137 (2005).

[98] J. Wu, L. Jiao, R. Li, and W. Chen, *Clustering dynamics of nonlinear oscillator network: Application to graph coloring problem*, Physica (Amsterdam) **240D**, 1972 (2011).

[99] D. I. Albertsson, M. Zahedinejad, A. Houshang, R. Khymyn, J. Åkerman, and A. Rusu, *Ultrafast Ising machines using spin torque nano-oscillators*, Appl. Phys. Lett. **118**, 112404 (2021).

[100] T. Wang and J. Roychowdhury, *OIM: Oscillator-based Ising machines for solving combinatorial optimisation problems*, in *Unconventional Computation and Natural Computation: Proceedings of the 18th International Conference, UCNC 2019, Tokyo, Japan, 2019, Proceedings 18* (Springer, New York, 2019), pp. 232–256.

[101] Y. Inui and Y. Yamamoto, *Entanglement and quantum discord in optically coupled coherent Ising machines*, Phys. Rev. A **102**, 062419 (2020).

[102] A. Blais, A. L. Grimsmo, S. M. Girvin, and A. Wallraff, *Circuit quantum electrodynamics*, Rev. Mod. Phys. **93**, 025005 (2021).

[103] J. J. Hopfield, *Neural networks and physical systems with emergent collective computational abilities*, Proc. Natl. Acad. Sci. U.S.A. **79**, 2554 (1982).

[104] B. P. Marsh, Y. Guo, R. M. Kroeze, S. Gopalakrishnan, S. Ganguli, J. Keeling, and B. L. Lev, *Enhancing associative memory recall and storage capacity using confocal cavity QED*, Phys. Rev. X **11**, 021048 (2021).

[105] B. P. Marsh, R. M. Kroeze, S. Ganguli, S. Gopalakrishnan, J. Keeling, and B. L. Lev, *Entanglement and replica symmetry breaking in a driven-dissipative quantum spin glass*, Phys. Rev. X **14**, 011026 (2024).

[106] A. Grimaldi, L. Mazza, E. Raimondo, P. Tullo, D. Rodrigues, K. Y. Camsari, V. Crupi, M. Carpentieri, V. Puliafito, and G. Finocchio, *Evaluating spintronics-compatible implementations of Ising machines*, arXiv:2304.04177.

[107] L. English, A. Zampetaki, K. Kalinin, N. Berloff, and P. G. Kevrekidis, *An ising machine based on networks of subharmonic electrical resonators*, Commun. Phys. **5**, 333 (2022).

[108] D. Pierangeli, G. Marcucci, and C. Conti, *Adiabatic evolution on a spatial-photonic Ising machine*, Optica **7**, 1535 (2020).

[109] T. Wang, L. Wu, and J. Roychowdhury, *New computational results and hardware prototypes for oscillator-based Ising machines*, in *Proceedings of the 56th Annual Design Automation Conference 2019* (Association for Computing Machinery, New York, 2019), pp. 1–2, 10.1145/3316781.3322473.

[110] A. Cavagna, J. P. Garrahan, and I. Giardina, *Index distribution of random matrices with an application to disordered systems*, Phys. Rev. B **61**, 3960 (2000).

[111] Y. N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, *Identifying and attacking the saddle point problem in high-dimensional non-convex optimization*, Adv. Neural Inf. Process. Syst. **27**, 2933 (2014).