Energy landscapes of combinatorial optimization in Ising machines

Dmitrii Dobrynin,[1,2,*] Adrien Renaudineau,[1] Mohammad Hizzani,[1,2] Dmitri Strukov,[3] Masoud Mohseni,[4] and John Paul Strachan[1,2,†]

[1]*Peter Grünberg Institut (PGI-14), Forschungszentrum Jülich GmbH, Jülich, Germany*
[2]*RWTH Aachen University, Aachen, Germany*
[3]*University of California Santa Barbara, Santa Barbara, California, USA*
[4]*LSIP, Hewlett Packard Labs, Milpitas, California, USA*

(Received 7 May 2024; accepted 25 September 2024; published 18 October 2024)

Physics-based Ising machines (IM) have been developed as dedicated processors for solving hard combinatorial optimization problems with higher speed and better energy efficiency. Generally, such systems employ local search heuristics to traverse energy landscapes in searching for optimal solutions. Here, we quantify and address some of the major challenges met by IMs by extending energy-landscape geometry visualization tools known as *disconnectivity graphs*. Using efficient sampling methods, we visually capture landscapes of problems having diverse structure and hardness manifesting as energetic and entropic barriers for IMs. We investigate energy barriers, local minima, and configuration space clustering effects caused by locality reduction methods when embedding combinatorial problems to the Ising hardware. To this end, we sample disconnectivity graphs of PUBO energy landscapes and their different QUBO mappings accounting for both local minima and saddle regions. We demonstrate that QUBO energy-landscape properties lead to the subpar performance for quadratic IMs and suggest directions for their improvement.

DOI: 10.1103/PhysRevE.110.045308

I. INTRODUCTION

Recent years have seen an increasing interest in using classical and quantum Ising machines (IM) for solving combinatorial optimization problems relevant for fundamental research and industrial applications [1]. Most of these devices rely on algorithms and physical principles implementing heuristic local search routines, e.g., discrete Monte Carlo (MC) sampling (simulated annealing, parallel tempering) or noisy/chaotic continuous dynamics. Examples of the former are memristive crossbar arrays employed to efficiently perform vector-matrix multiplication [2,3], or digital ASIC annealers [4]. The latter versions of IMs include coherent Ising machines, oscillator networks, quantum annealers, and others [5–8]. The main attraction for the use of IM is the intrinsic compatibility of the algorithm operations with their physical implementations, which offers reducing time-to-solution and/or energy-to-solution metrics polynomially, or by a significant prefactor [9–12].

In this context, there are several outstanding challenges faced by IMs on both algorithmic and hardware levels, resulting in strong compromises being adopted in their practical deployment. One, and possibly the most important, difficulty concerns their application to practically interesting (large) problem sizes. The support of only second-order couplings of "spins", together with connectivity topology constraints (e.g., the chimera graph [13]) results in the introduction of multiple auxiliary variables in order to either avoid higher-order terms, or reach necessary levels of sparsity. The added new variables can scale super-linearly in the number of original variables, not only further challenging the scaling to large problems, but also increasing the search space and modifying the optimization energy landscape. As a result, IMs can be limited to smaller-scale problems, and even these can become harder than their native formulation [14–17] because of the worsened landscape geometry.

A second challenge lies in the algorithmic limitations of IMs. In particular, their reliance on local search heuristics fundamentally puts a bound on the problem classes they are capable of solving [18]. Being inherently local, IMs are prone to suffer from energy barriers rejecting MC moves, and from entropic barriers or degeneracies hampering both sensible exploration and exploitation [19]. However, nonlocal Monte Carlo algorithms have been recently proposed that could significantly accelerate exploration by unmasking certain underlying structures in the configuration space [20]. Clear understanding of geometrical or energy-landscape features of benchmark problems and the corresponding constraints of the Ising hardware is essential to facilitate future advances in the field.

A major challenge for designing discrete optimization/sampling solvers is the lack of understanding or representation of the high-dimensional configuration space. Only a few methods have been developed over the years to visualize high-dimensional cost/energy functions of such problems. One example is disconnectivity graphs (DG, also called barrier trees) [21–23], which aim to simplify the exponentially large configuration space by capturing local minima and their connectivity through energy barriers. It is possible to use DGs to gain quantitative insights into phenomena in a variety of

*Contact author: d.dobrynin@fz-juelich.de
†Contact author: j.strachan@fz-juelich.de

applications ranging from metastable states of protein folding [24] to thermodynamic effects in Lennard-Jones systems [25], biomolecules [26], and spin glasses [27]. However, becaue of the exponential complexity of DG construction and high degeneracy of the solution space, attaining energy-landscape visualization is a computational feat on its own [28].

The contributions of this paper are as follows. Firstly, in Sec. III A we describe an extension for the efficient sampling algorithm of [29] to support DGs of energy landscapes featuring strong degeneracy of the configuration space (millions of states), capturing not only local minima but also saddle regions. We further modify this approach providing means to construct DGs for quadratic optimization problems resulting from locality reduction caused by IM hardware mapping. We achieve this by meaningfully reducing the search space over auxiliary variables and defining "effective" barriers. Secondly, in Secs. III B and III C using 3-SAT as a representative higher-order problem class, we plot DGs for problems of sizes inaccessible to the methods reported previously. With our methods we compare easy to hard instances, and random to industrial (structured) instances. Finally, in Sec. III D we demonstrate suboptimal energy-landscape features of hardware embedding quadratization methods for 3-SAT from the perspective of clustering and entropy of energy minima, which are some of the culprits of algorithmic hardness [30,31].

II. BACKGROUND

The conventional (2-local) Ising Hamiltonian, which IMs natively solve, is

$$H_{\text{Ising}} = \sum_{i<j}^{N} J_{ij} s_i s_j + \sum_{i}^{N} h_i s_i, \tag{1}$$

where $s_i \in \{-1, 1\}$, J_{ij} are spin interaction strengths, and h_i denote local magnetic fields. Finding the ground state of Eq. (1) is an NP-complete problem [32], and therefore approximately solving this Hamiltonian efficiently is of profound interest. Alternatively, the Ising Hamiltonian can be formulated as a quadratic pseudo-boolean function

$$H_{\text{QUBO}} = \sum_{i<j}^{N} Q_{ij} x_i x_j + \sum_{i}^{N} b_i x_i + C, \tag{2}$$

with binary variables $x_i \in \{0, 1\}$. Deciding the ground state of this function among 2^N possible configurations is commonly called quadratic unconstrained binary optimization (QUBO) problem. In this paper, we will use Ising and QUBO terms interchangeably because of their equivalence.

The generalization of QUBO to support higher-order interactions of variables is usually referred to as PUBO ("P" for polynomial),

$$f(\mathbf{x}) = f(x_1, x_2, \ldots, x_N) = \sum_{\{i\}_k \subseteq V} a_{\{i\}} \prod_{\{i\}_k} x_i + C, \tag{3}$$

which is correspondingly equivalent to the k-local Ising (historically called the p-order Ising spin glass [33]). Here, $\{i\}_k \subseteq V$ denote all possible subsets of the set of variables with the order of interaction not larger than the highest $k \geqslant 1$.

The present paper devotes particular attention to the k-SAT problem [see below Eq. (4)], one of the oldest and well-studied NP-complete problems [34–36]. The motivation behind this choice lies in the fact that, apart from being practically important for a variety of applications [37], k-SAT highlights the hardware and algorithmic challenges of IMs [38]. As will be discussed in this paper, it features strong degeneracy of the solution space, an abundance of energy barriers, clustering of solutions, and can only be natively supported by the PUBO formulation, making it a formidable problem class for local search-based quadratic IMs.

A general statement of the k-SAT decision problem is simple: Is there a binary variable assignment $\mathbf{x} \in \mathbb{B}^N$ of the following conjunctive normal form (CNF):

$$\left(l_{i_{1,1}} \vee l_{i_{1,2}} \cdots \vee l_{i_{1,k}}\right) \wedge \cdots \wedge \left(l_{i_{m,1}} \vee l_{i_{m,2}} \cdots \vee l_{i_{m,k}}\right), \tag{4}$$

where $i \in \{1, N\}$, $m \in \{1, M\}$, $l = x$ or $l = \bar{x}$, so that all M clauses are satisfied? With $k \geqslant 3$, it is NP-complete like Ising/QUBO and thus worst case exponentially hard [39]. The k-local PUBO cost function [Eq. (3)] is easily obtained from Eq. (4) as shown below in Methods by Eq. (5).

Many optimization landscape features have been established for hard constraint satisfaction problems [35,40], of which k-SAT is a conventional example. By increasing the number of constraints from the small number, where the problem is easily satisfiable, to larger values up to a point of unsatisfiability, optimization landscapes undergo phase transitions where the dominating "simple" configuration region of connected solutions gets shattered into exponentially many *clusters* of solutions. Each cluster consists of several configurations, which can be easily accessed from each other by local dynamics [41]. Furthermore, some of the variables in such cluster configurations could also be "frozen" [42,43], i.e., remain unchanged regardless of the state of others. In other words, not only can it be difficult to traverse the landscape in search of isolated clusters, but also to transition between such clusters, it is imperative to modify an extensive fraction of variables simultaneously; thus, *nonlocal* moves can be essential [20]. Recently, there has been renewed interest to quantify geometrical aspects of the algorithmic hardness near a computational phase transition by introducing the notion of overlap gap properties (OGP) [18,44,45]. In order to illustrate energy-landscape geometry features as a cause of hardness of combinatorial optimization in IMs, in this paper we focus on illustrating how the landscapes are perceived by *local* search.

Early efforts to visualize energy/fitness landscapes arose in the context of theoretical chemistry and biology [21–23]. Authors of these papers introduced the concept of disconnectivity graphs (DG) implementing a map of exponentially large potential energy configuration spaces to a two-dimensional tree. Figure 1 sketches the idea behind such mapping: every leaf corresponds to a local minimum, while the branches represent the magnitude of energy barriers and connectivity (lowest barrier separation) of local minima with respect to each other.

In principle, arbitrary energy landscapes can be defined by a triplet [46]: X being a set of configurations, neighborhood $\mathcal{N}(x)$ of every state x in X, and energy/fitness function $f(X) \in \mathbb{R}$. We say that a solver explores the energy landscape if a local search move from any given configuration x chooses a state

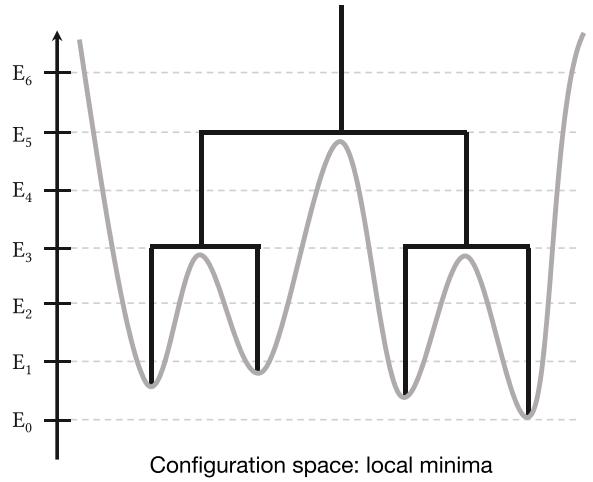


FIG. 1. A simplified view of a disconnectivity graph. Every local minimum corresponds to a leaf; the height of energy barriers is reflected by the energy of branch connections. Horizontal arrangement of minima *does not* represent distance, i.e., by default, has no explicit meaning.

in $\mathcal{N}(x)$. For instance, one may choose a random neighbor (random walk) or the one with the largest energy decrease (steepest descent).

Special attention, however, should be given to the *degeneracy* of such landscapes: Many configurations form neighborhoods, which can be traversed by a local rule at no energy cost. Furthermore, the concept of a local minimum becomes ambiguous and *nonlocal* in degenerate landscapes [47,48]. As Figs. 2(a) and 2(b) demonstrate, it is impossible

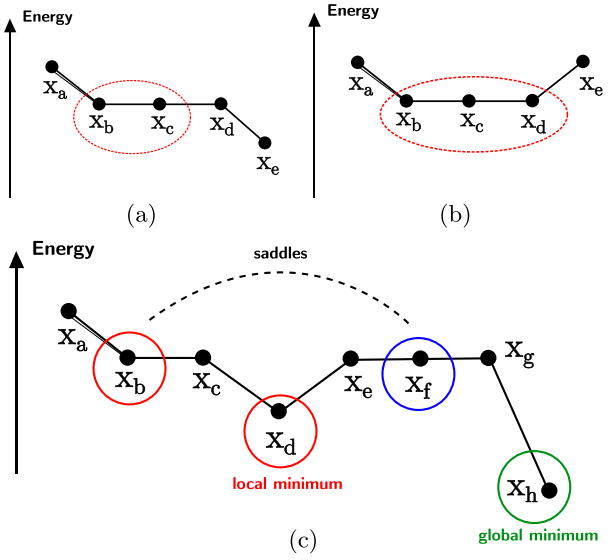


FIG. 2. (a), (b) Two types of degenerate landscapes: a saddle cluster and a local minimum cluster. State $\mathbf{x}_d$ is a zero barrier exit point from the saddle. Outlined are the stable states. (c) The highlighted saddle points can be treated as connected [47] or disconnected [27] based on the adopted definition of disconnectivity graphs. Blue (red) indicates (dis)connectivity to a global minimum (green).

to know if a descending energy path exists from the leftmost stable state $\mathbf{x}_b$ unless exploration finding the rightmost state $\mathbf{x}_d$ is performed. In this paper we will call a stable "plateau" of Fig. 2(a) *a saddle cluster*, while the plateau in Fig. 2(b) will be called *a local/global minimum cluster*. The terminology of saddles/local minima of this paper is chosen to resemble similar terms from continuous optimization. There, multiple studies highlight profound difficulties of navigating high-dimensional landscapes arising from both types of critical points [49,50].

The studies of [28,47,48] have addressed the complexity of constructing DGs of degenerate landscapes with exhaustive enumeration of states. While being computationally infeasible for problems larger than $\approx 30-40$ variables, these studies carried out classifications of saddles or local minima and the ways the states can be connected within a cluster and to other clusters. For example, a difference in possible connectivity of stable points is illustrated in Fig. 2(c). Approach of this paper is closest to that of [27] in which the highlighted saddle points are treated as being disconnected. This choice is motivated by the golf-course-type energy landscapes of 3-XORSAT problems [31], where the paths to good solutions are mostly impeded by the entropic barriers, rather than the energy barriers.

In Fig. 2(c) there is no barrier between $\mathbf{x}_b$ and the global minimum $\mathbf{x}_h$, but the path to it lies through a local minimum $\mathbf{x}_d$. As a result, joining the states separated by a "hole" would result in a deceiving visualization hiding landscape features important for local search routines. With the methods of this paper (see Secs. III A 2 and III A 3), we will address such diversity of scenarios by distinguishing the states with connections to global minimum (blue color) from those separated from it by either barriers or "holes" (red color). This will provide a clear explanation of why *second*-order IMs can be greatly challenged by higher-order combinatorial optimization problems (Sec. III D).

III. RESULTS

A. Methods

1. Locality reduction of k-SAT

The k-SAT problem of maximizing the number of satisfied clauses of Eq. (4) is reformulated as PUBO (3) minimization as follows (inverting the expression and using De-Morgan law):

$$\left(l_{i_{1,1}} \vee l_{i_{1,2}} \cdots \vee l_{i_{1,k}}\right) \wedge \cdots \to \bar{l}_{i_{1,1}} \bar{l}_{i_{1,2}} \ldots \bar{l}_{i_{1,k}} + \ldots, \qquad (5)$$

where each literal $l_i = x_i$ or $l_i = \bar{x}_i$. The issue of its locality reduction to support the formulation of Eq. (2) has been heavily investigated over the recent years [51], with the efforts aimed at introducing quadratizations that have the smallest possible number of auxiliary variables, minimize bit-precision requirements on the weights, or have algorithmically favorable properties, e.g., submodularity [52].

Perhaps the simplest method to meet the first requirement is to use quadratization by substitution, i.e., to introduce *auxiliary* variables y for each pair of variables $x_p x_q$ in the original PUBO function of Eq. (3) until the problem of required order is obtained, i.e., second order for QUBO (2). The constraints

are then enforced by either explicitly considering the equalities $x_p x_q = y$, or by the addition of quadratic penalty terms in the cost function for each substitution,

$$f(\mathbf{x}) = \pm x_1 x_2 \dots x_k \rightarrow$$
$$\rightarrow g(\mathbf{x}, \mathbf{y}) = \pm y x_3 \dots x_k + P_{\pm}(x_1, x_2, y). \quad (6)$$

The choice of the P function is not unique; for instance, one can make sure that

$$f(\mathbf{x}) = \min_y g(\mathbf{x}, \mathbf{y}) \quad (7)$$

is satisfied, thereby preserving global minima of the original problem. Additionally, the choice of $x_i x_j$ admits some freedom and can be optimized for the minimum number of auxiliary variables by solving a vertex cover problem [53]. For simplicity, we use an efficient greedy routine to perform such optimization (for more details on quadratization methods outlined below cf. Appendix B).

For example, a commonly used quadratization penalty choice for locality reduction was suggested by Rosenberg (thirdrd-order example) [54],

$$\pm x_p x_q x_k = \min_y [\pm y x_k + (3y - 2x_p y - 2x_q y + x_p x_q)], \quad (8)$$

where $x_p x_q$ was replaced by y, and the remaining terms penalize the mismatch of $x_p x_q$ and y. Thus, every appearance of $x_p x_q$ in the $k \geqslant 3$ terms of the PUBO function is substituted by the same y, and for each such substitution $3y - 2x_p y - 2x_q y + x_p x_q$ penalty is added. This mapping is also implicitly used when the approach of reversible logic of [55,56] is employed.

Performing standard simulated annealing optimization of 3-SAT problems we found a different mapping to be computationally superior to the Rosenberg version. The new mapping extends the quadratization ideas [57,58] and [53] by approaching the monomials with positive and negative coefficients differently,

$$-x_1 \dots x_k = \min_y \left[(k-1)y - \sum_{i=1}^{k} x_i y \right],$$
$$x_1 \dots x_k = x_2 \dots x_k - \bar{x}_1 \dots x_k$$
$$= x_2 \dots x_k + \min_y \left[(k-1)y - \bar{x}_1 y - \sum_{i=2}^{k} x_i y \right], \quad (9)$$

and thus we call it KZFD-BG after the authors. However, instead of applying these penalties individually for each term in the PUBO function [59], the variable substitution in this paper is done as in the Rosenberg case, i.e., sharing substituted pairs across multiple monomials (see Appendix B). As a result, this yields the same number of native and auxiliary variables regardless of the mapping used. We address simulated annealing performance difference of the mappings in the context of PUBO and QUBO comparison in Sec. III D.

Any locality reduction method modifies the "native" optimization landscape in nontrivial ways and can make its exploration algorithmically more challenging. In particular, Eq. (7) guarantees that for every stable state $\mathbf{x}^*$ of $f(\mathbf{x})$ with respect to a single bit-flip, $f(\dots, \bar{x}_i^*, \dots) - f(\dots, x_i^*, \dots) \geqslant 0$, $\forall i$, the quadratization $g(\mathbf{x}^*, \mathbf{y}^*)$ is also in a stable state, which is given by $\min_{\mathbf{y}} g(\mathbf{x}^*, \mathbf{y}) \equiv g(\mathbf{x}^*, \mathbf{y}^*)$.

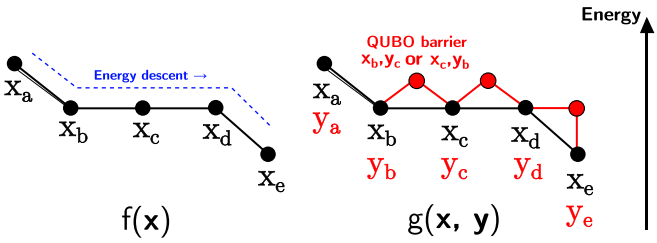


FIG. 3. (Left) PUBO landscape sketch, neighboring states are connected by a single flip. (Right) QUBO mapping landscape; the auxiliary $\mathbf{y}$ adaptation may introduce new energy barriers preventing the otherwise possible descent in energy.

Indeed, the bit-flip energy changes with respect to the auxiliary variables are non-negative because of the definition of g: $g(\mathbf{x}^*, \dots, \bar{y}_i^*, \dots) \geqslant g(\mathbf{x}^*, \dots, y_i^*, \dots) = \min_y g(\mathbf{x}^*, \mathbf{y})$. In turn, the energy change of flipping x is also non-negative because of the following chain:

$$g(\dots, \bar{x}_i^*, \dots, \mathbf{y}^*) \geqslant \min_{\mathbf{y}} g(\dots, \bar{x}_i^*, \dots, \mathbf{y}) \quad (10)$$
$$= f(\dots, \bar{x}_i^*, \dots) \geqslant f(\mathbf{x}^*) = g(\mathbf{x}^*, \mathbf{y}^*). \quad (11)$$

However, such correspondence does not hold in the opposite direction, i.e., a stable state of $g(\mathbf{x}, \mathbf{y})$ is not guaranteed to be a stable state of $f(\mathbf{x})$.

For illustration, in Fig. 3 a "linear" landscape represents states connected by a bitflip local move in the N-dimensional hypercube. The left sketch depicts a degenerate case with states $\mathbf{x}_b$, $\mathbf{x}_c$ being stable, but $\mathbf{x}_a$ and $\mathbf{x}_d$ unstable. In turn, the right sketch shows how the quadratization mapping induces a rugged structure on top of the original manifold because of the auxiliary variables and the penalty terms. For every state $\mathbf{x}$ there is a corresponding minimizing auxiliary state $\mathbf{y}$ (possibly nonunique) according to Eq. (7). The low-energy state $\mathbf{x}_e$ that was easily accessible by a greedy local search descend can now be separated by energy barriers because of the necessity to adapt $\mathbf{y}$ for every $\mathbf{x}$.

If $\mathbf{x}$ and $\mathbf{y}$ are treated on equal footing, then one is forced to explore a configuration space $2^{|\{y\}|}$ times bigger than the native problem. The problem that already had highly nontrivial landscape structure caused by frustrations and long-distance correlations of variables, after quadratization, will have these features hidden or worsened by the mismatch of "gradients" and energy barriers, ultimately causing significant deterioration of the IMs' ability to find solutions [14,16,17,59].

The effect of penalty-based locality-reduction methods may be different depending on a combinatorial problem class that is being quadratized. For example, a popular benchmarking 3-regular 3-XORSAT problems [12] feature variables that appear in only three third-order clauses. Thus, the QUBO formulation has only three native-auxiliary interactions per native variable, which are responsible for the QUBO energy barriers (Fig. 3). In comparison, the phase transition random 3-SAT problems [36] have on average $\approx 3 \times 4.267$ appearances of variables in different clauses.

Finally, we note that the sparsifying approaches that aim to reduce degrees of interaction between variables can introduce even more energy barriers into the problem because of auxiliary variables and penalties akin to the locality reduction

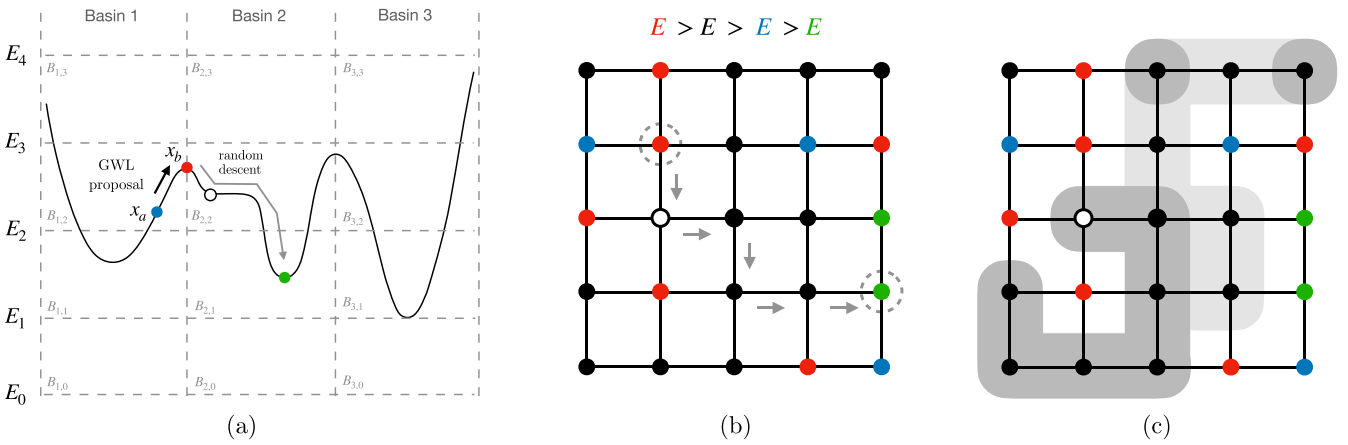


FIG. 4. (a) Generalized Wang-Landau (GWL) for sampling and barrier estimation. GWL proposal $x_a \to x_b$ is sampled with probability of Eq. (12). The basin of attraction is identified by an algorithm of choice. (b) Random descent illustration. The highlighted red circle state belongs to the basin of the green highlighted local minimum state. (c) If necessary, breadth-first search accurately calculates cluster sizes at the end of sampling. The light grey states are unstable exits, the dark grey states are stable saddles.

methods. We do not focus on sparsification in this paper; nonetheless, one example is given in Appendix A.

2. *Sampling algorithm outline*

In order to study and visualize with DGs the energy landscapes of degenerate optimization problems and their QUBO mapping modifications, we extend the generalized Wang-Landau (GWL) [60,61] sampling approach of the Refs. [29,62]. The GWL non-Markov chain Monte Carlo algorithm carries out random walks in the configuration space aiming to achieve approximately uniform attendance of all predefined energy levels $l \in [1, L]$ and all recorded basins of attraction $k \in [1, K]$.

During preprocessing steps [see Fig. 4(a)] one defines the landscape partition into sectors in energy $[E^1, E^2, \ldots, E^L]$ and affinity to a basin of attraction of a local minimum $\mathbf{x}^k$: $\mathbf{x} \in B_{k,l}$, if $E(\mathbf{x}) \in [E^l, E^{l+1})$ and $\textbf{descent}(\mathbf{x}) = \mathbf{x}^k$. We note that the **descent** routine can be defined differently, and it makes sense to choose its definition similar to the the actual solver algorithm that would be used for solving studied problems in practice (see Sec. III A 3 below). Next, the sampling of states is performed with the following acceptance probability:

$$p_{a\to b} = \min\left[1, \exp\{\beta(E_a - E_b)\}\frac{\gamma_{k_a,l_a}}{\gamma_{k_b,l_b}}\right], \tag{12}$$

where $\gamma_{k,l}$ is a current estimation of the statistical weight of a sector

$$\frac{1}{Z}\sum_{\mathbf{x}\in B_{l,k}} \exp(-\beta E(\mathbf{x})) \approx \gamma_{k,l}. \tag{13}$$

This estimation is constantly updated, when the sector $B_{l,k}$ is visited, by

$$\gamma_{l,k}^{t+1} = \gamma_{l,k}^{t} e^{f}, \tag{14}$$

where f can follow a decreasing schedule usually starting from the value $f = 1$. It is numerically convenient to also define a histogram

$$\theta_{l,k}^{t+1} \equiv \ln \gamma_{l,k}^{t+1} = \theta_{l,k}^{t} + f, \tag{15}$$

which is initialized at 0 for all l, k at the beginning of the algorithm. If the exploration of as many minima as possible is preferred, then f is not decreased over time [62], but in this case the estimation of γ would not be accurate [63]. We do not decrease f, because our goal is construction of DGs with rapid discovery of distinct local minima/saddles.

When a step $\mathbf{x}_a \to \mathbf{x}_b$ is tried (accepted or not), one saves the energy $\max[E_a, E_b]$ {$\max[E_a, E_a + \Delta E_{\text{QUBO}}(\mathbf{x}_a \to \mathbf{x}_b)]$ for QUBO, see Sec. III A 4} as a current energy barrier estimation between basins k_a and k_b. This "educated guess" can then potentially be improved with the ridge descent algorithm [29]. If the zero-energy barrier is found for a state perceived as local minimum [e.g., $\mathbf{x}_f \to \mathbf{x}_g \to \mathbf{x}_h$ in Fig. 2(c)], the status of such minimum is changed to a saddle, and its histogram is joined (max values of each row) with the corresponding lower basin of attraction (e.g., $\mathbf{x}_h$). If a saddle is connected to several lower basins, then the visits are distributed uniformly at random among them.

We keep track of maximum K number of lowest in energy local minimum/saddle clusters adaptively uniting them by discovered connectivity and thus allowing space for additional clusters to be taken into account. If K is too small, then only a few energy levels will be available for the DG construction. In addition, we define a special $K + 1$ column of the histogram for all of the states that do not fit into the first K clusters [62].

The implementation of the sampling method of this paper is publicly avaiable for reproducibility of the results and is described in Appendix D 1. Additional details, including uniformity of sampled histograms, accuracy of DG construction, computational cost, and hyperparameters for all disconnectivity graphs of this paper are presented in Appendices D 2–D 4.

3. *Extension for degeneracy*

By design, GWL uniformly samples states across basins of attraction of local minima and energy levels. Its main purpose in this paper is to discover as many regions of the landscape as possible without being stuck in a particular place, thereby not biasing the DG estimation. What is crucial in the definition

of the algorithm is the **descent** routine, which identifies local minima and saddle points. When a problem has no degeneracies, e.g., S-K spin glass with Gaussian weights, one can define the descend (hill climbing) as the steepest descent, i.e., spins with the highest energy reduction are flipped. However, in this paper we are interested in highly degenerate integer valued optimization problems, where such definition is not possible.

As briefly discussed above in Fig. 2, local minima and saddles are perceived differently depending on the algorithm employed for solving such problems. In this case, constructing exact DGs, apart from being infeasible for large problems, may result in misleading conclusions. For instance, a very large saddle point may have only one zero barrier exit from itself, which may never be found by a local search routine, effectively being a local minimum, but it would still be depicted as a saddle on a DG, or even worse, not shown at all if saddles are not considered.

Here, we aim to balance between efficient exploration of the landscape and visualization of relevant landscape features. For this purpose, we use *random descent* [see Fig. 4(b)], in which a greedy local move is performed in the first-seen random direction decreasing the energy. This descent routine corresponds to the MC sampling approach we use in simulated annealing benchmarking but at $T = 0$ (see Appendix C for more details on SA).

Once a stable state is encountered [white circle in Fig. 4(b)], a limited exploration of the "plateau" region is performed until either the budget of allowed moves is exhausted, or an exit from the saddle is found ([green circle in Fig. 4(b)]. We defined a hyperparameter, which determines for how long an algorithm can explore a stable cluster before registering it as a local minimum/saddle in the histogram. If a cluster is easily escapable, then there is no reason to keep track of it.

The states encountered during such exploration of saddles/local minima are stored in a single cluster (including the unstable exit states, i.e., four states are stored in Fig. 4(b). If some of the stored states are encountered again during GWL sampling, all of the states that belong to a single cluster are joined, with their histograms united by their max values.

Previously, the Refs. [64,65] addressed the difficulty of clustering in the context of improving the uniform sampling of the *ground states*. Once a ground state $\mathbf{x}$ was found, a ballistic search (BS) routine was carried out: starting from some global minimum state, a chain of zero-energy states was constructed by flipping every variable maximum once. With the use of such chains, the cluster sizes and thus connectivity of states were estimated more reliably. We experimented with this method for clusters at every energy level and found it useful for clustering remote configurations when the number of states becomes infeasibly large.

Finally, in Fig. 4 we illustrate breadth-first search (BFS) that we use to exactly evaluate sizes of clusters at the end of sampling, when such statistics are of interest. Both stable and unstable states (exits) participate in BFS, and we confine their number by a predefined bound of states per energy level (usually 10^7 in this paper). While only the stable states are later shown on the DGs, the ratio of stable to unstable number of configurations can potentially be used to estimate the probability of escaping saddle regions of the landscape.

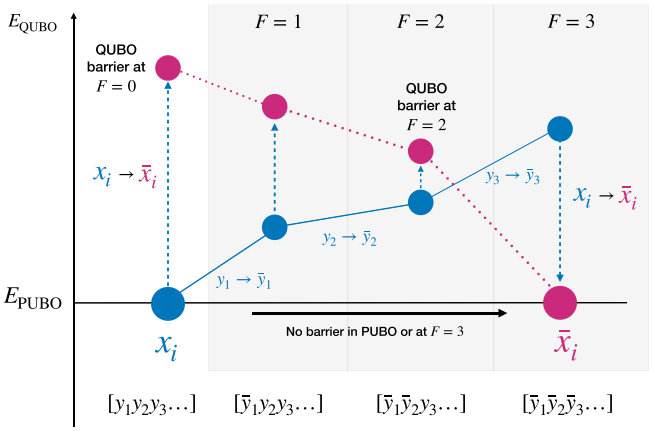


FIG. 5. QUBO factor $F = |\{y_a\}|$ motivation example. By perturbing $F = 3$ auxiliary variables one is able to restore the PUBO zero barrier between x_i and $\bar{x}_i$ states. For $F \in (0, 3)$ the effective barrier is defined, taking intermediate values between QUBO and PUBO.

4. Extension for QUBO mapping

Two neighboring configurations are considered to be a part of a single degenerate cluster in the native [PUBO Eq. (3)] landscape if a local move separating them is of zero-energy cost, as shown for states $\mathbf{x}_b$ and $\mathbf{x}_c$ in Fig. 3. The state $\mathbf{x}_d$ does not belong to a cluster since it has a move of negative energy to the state $\mathbf{x}_e$. However, in the special case of the QUBO mapping landscape, the same state $\mathbf{x}_d$ would now be considered a saddle point since the decrease in energy is only achieved through an intermediate $\mathbf{y}$ adjustment.

The presence of a barrier in QUBO for a transition $\mathbf{x}_b \to \mathbf{x}_c$ (when originally there could be no barrier at all) puts the local search at a disadvantage because of the higher rejection rate of local moves. Raising the temperature of sampling, e.g., of simulated annealing, would not fully solve the problem since it would harm the necessary exploitation of the low-energy manifold. Additionally, once local search is complete, a solver discards $\mathbf{y}$ values using the states of $\mathbf{x}$ as a solution. The search over the subspace of $\mathbf{y}$, thus, does not look for new solutions, but rather varies the induced QUBO barriers between the neighbors in the $\mathbf{x}$ space.

To highlight the significance of landscape ruggedness of quadratization compared to the native space and facilitate fairer comparison, here we provide QUBO with additional capabilities by assuming that a local search solver can "look beyond" the QUBO landscape barriers to a certain adjustable degree. Figure 5 depicts the case where the penalty terms of a QUBO mapping introduce interactions that favor auxiliary variable states different by Hamming distance 3 for two native configurations separated by a single bit-flip, i.e., x_i and $\bar{x}_i$.

If the problem is approached head-on, one would need to either climb a steep barrier of $x_i \to \bar{x}_i$ and then adapt three auxiliary variables, or sequentially flip each of $y_a \to \bar{y}_a$, i.e., climb a long barrier. Such scenario of long barriers is argued to be difficult for tunneling in quantum annealers [66],

considering that the mapping quadratization is essential because of strict hardware limitations. We note, however, that with every sequential flip of y_a, the barrier of the $x_i \to \bar{x}_i$ move is reduced, i.e., allowing more y_a to be explored raises the chance to overcome the QUBO barriers introduced by the mapping in the first place.

As a result, we augment disconnectivity graph analysis by introducing a QUBO factor F, which stands for the maximum number of allowed auxiliary variable flips of nonzero energy for every native move $x_i \to \bar{x}_i$. With large enough F the original (PUBO) landscape is recovered, while for small F values "effective" energy barriers are still present, and thus the landscape connectivity is worsened by the mapping. In addition, F serves as means to compare different QUBO mappings head-to-head, with mappings allowing small F being arguably better for the local search of IMs. We perform such comparison supported by the simulated annealing results in Sec. III D.

The algorithm to compute the effective barriers is as follows. First, at a fixed position in the space of $\mathbf{x}$ we set the auxiliary variables $\mathbf{y}$ in a valid state required by Eq. (7) [67]. Next, the bit-flip energy change $\Delta E_{\mathrm{QUBO}}(\ldots, x_i \to \bar{x}_i, \ldots \mathbf{y})$ is computed, which corresponds to the "vanilla" QUBO barrier at $F = 0$. Second, in order to calculate the effective QUBO barrier of $x_i \to \bar{x}_i$, we list all auxiliary variables $\{y_a\}$ that interact with x_i, i.e., $Q_{ia} \neq 0$. Out of all listed y_a, we choose F variables with the minimum values of $\Delta E(\ldots, \bar{x}_i, \ldots, y_a \to \bar{y}_a, \ldots) < 0$. Finally, the effective barrier (see Fig. 5) $\Delta E_{\mathrm{QUBO,F}}$ is obtained by (y variables do not interact with each other in 3-SAT mappings),

$$\Delta E_{\mathrm{PUBO}}(x_i \to \bar{x}_i) \leqslant \Delta E_{\mathrm{QUBO,F}}(x_i \to \bar{x}_i) \equiv$$
$$\Delta E_{\mathrm{QUBO}}(x_i \to \bar{x}_i) + \sum_{a=1}^{F} \Delta E(\bar{x}_i, y_a \to \bar{y}_a) <$$
$$\Delta E_{\mathrm{QUBO}}(x_i \to \bar{x}_i). \tag{16}$$

5. *Disconnectivity graphs notation*

In the following sections we adopt the below convention when plotting DGs (e.g., see Fig. 6). The y axis stands for the PUBO/QUBO energy. Every circle represents a separate local minimum/saddle cluster. The diameter of such circle corresponds to the square root of the cluster degeneracy, i.e., the area of a circle is proportional the number of connected stable configurations within a cluster. There is no explicit meaning behind the x axis distance between the DG leaves and branches. If a circle is shown to have a zero-energy connection to lower clusters, then it represents a saddle cluster. If two or more saddles appear connected, then the situation depicted in Fig. 2(c) between $\mathbf{x}_b$ and $\mathbf{x}_f$ is in place. Red clusters in Figs. 8(a) and 8(b) have no direct connection [not found during sampling) to the global minimum denoted by green, i.e., all local minima are red, as well as some saddles (e.g., the state $\mathbf{x}_b$ in Fig. 2(c)]. Blue saddle clusters in Figs. 8(a) and 8(b) were found to be connected to the global minimum by a descent algorithm of choice without energy barriers [e.g., $\mathbf{x}_f$ in Fig. 2(c)].

Every DG is accompanied by a histogram of the number of states obtained with BFS at each energy level. The degeneracy

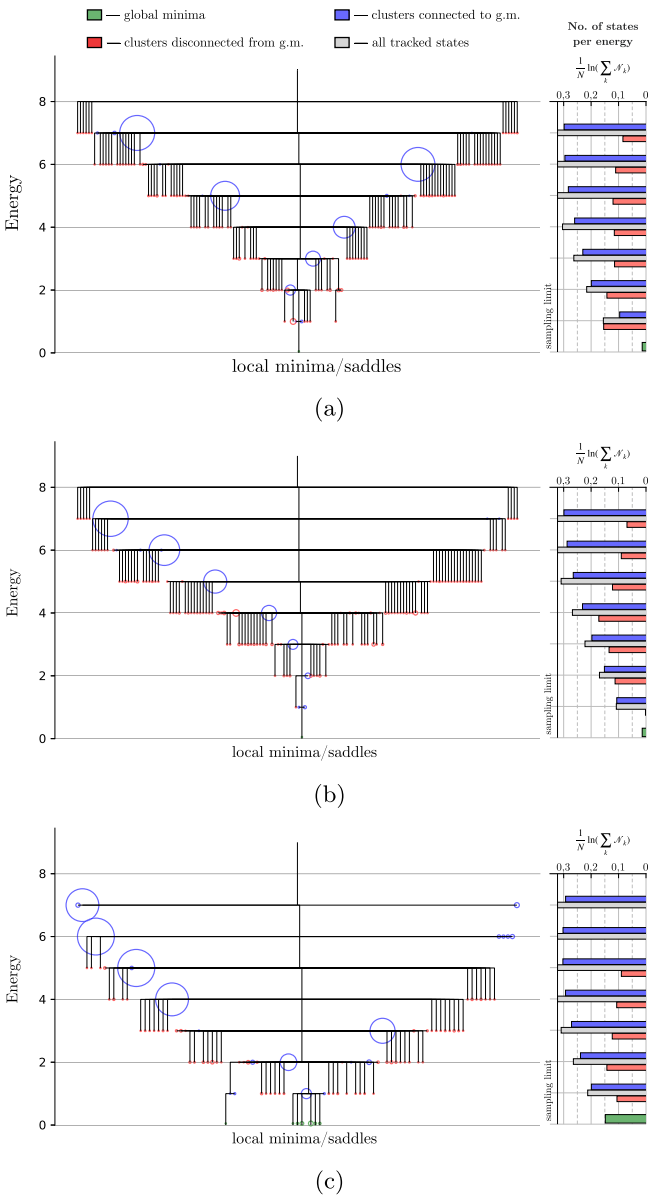


FIG. 6. Disconnectivity graphs (PUBO) of "very hard" (a) (instance *uf50-981*), "hard" (b) (instance *uf50-920*), and "easy" (c) (instance *uf50-933*) 3-SAT problems. States truncated at $E \leqslant 7$. Sampling limit per energy level: 10^7. (a) 156 clusters, two global minimum states. (b) 148 clusters, two global minimum states. (c) 99 clusters, 1654 global minimum states.

of every separate cluster is denoted by $\mathcal{N}_k$, while the total number of states per energy is plotted as a normalized by N (number of native variables) natural logarithm of $\sum_k \mathcal{N}_k$. The grey histogram shows the total number of BFS aggregated states (including unstable saddle exits). The blue and red histograms count the corresponding stable states shown by circles on the DG.

B. Easy and hard problems

The finite-size fluctuations of relatively small random 3-SAT problems usually employed for IM benchmarking in practice results in a strong spread of their hardness [16]. In this

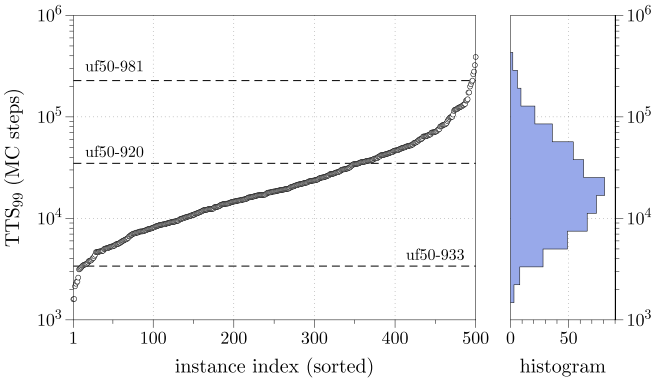


FIG. 7. The distribution of the TTS_{99} of the PUBO (native) SA for 500 instances (*uf50 500-1000*). Dashed lines for the TTS_{99} of instances used for DG construction in Fig. 6(b). For SA implementation details and code availability cf. Appendix C.

section, with the help of DGs and using the open benchmarking library SATLIB [68], we aim to highlight the landscape features exhibited by such instances of different hardness. As a means of benchmarking we employ a simulated annealing (SA) solver described in detail in Appendix C and available at [69]. For every optimization problem instance it outputs time-to-solution 99% (TTS_{99}) value (in Monte Carlo steps), which is the time needed for a stochastic solver to reach a solution at least once with probability $p = 0.99$.

In Fig. 7 we show the SA hardness distribution of 500 instances from SATLIB of size $N = 50$ and $M = 218$ clauses ($\alpha = M/N$ near the phase transition). The instance *uf50-920* visualized by the DG in Fig. 6(b) was found to be relatively "hard" with $(34.9 \pm 1.0) \times 10^3$ TTS_{99} algorithmic steps, the instance *uf50-933* in Fig. 6(c) was "easy" with 3390 ± 40 steps, and the instance *uf50-981* in Fig. 6(a) was "very hard" with $(229 \pm 16) \times 10^3$ steps (with respect to the observed range of TTS).

A clear distinction is seen in both the number of global minimum configurations, as well as the number of distinct global minimum clusters between the easy and hard instances (7 vs 1). It is, in general, unclear what is the property of optimization landscapes that would measure hardness best for a particular solver. In [20] authors use the number of global minimum clusters as a proxy for predicting survey propagation's ability to find global minima in 4-SAT random instances. However, instead of trying to predict the hardness of instances by DGs, we demonstrate how DGs can be used to gain insights into experimentally observed algorithm behaviors by a diverse set of sampled landscape properties.

In addition to much larger cardinality of the set of global minima, the easy problem in Fig. 6(c) features many saddle point states at $E = 1$ (20877 states), which are only connected to global minima and act as a basin of attraction for solvers. In comparison, the "hard" instance contains similar saddles with only 206 states. There were no local minima found at $E > 5$ in the easy instance, while the "hard" and "very hard" instances feature local minima even at $E = 8$. We also highlight the higher ratio of local minima/disconnected saddles in the "hard" instance compared to the "easy" one (the number of "red" states).

At every energy level we observe massive saddle points, which are connected to the global minimum. We note, however, that it becomes very important for local search not to descend into a wrong local minimum cluster, even though in principle it is possible to descend to a solution without overcoming any barriers. This is particularly highlighted by the difference between "hard" and "very hard" instances. In the histogram of Fig. 6(a) we observe a large number of local minima at $E = 1$, as well as a distinct basin of attraction separated from the global minimum by the barrier $\Delta E = 2$, compared to Fig. 6(b).

The majority of energy barriers in the tested 3-SAT problems is the minimum possible one, $\Delta E = 1$ (as also previously observed in [70]). In other words, the constructed DGs illustrate the significance of entropic barriers that are determined by probabilities of descending into better areas of the landscape, which resulted in the observed more than two orders of magnitude spread of the time-to-solution metric in Fig. 7.

C. Random and industrial problems

The uniform random 3-SAT problems are a common benchmark for testing the performance of heuristic solvers. In the thermodynamic limit of $N \to \infty$ and $M \to \infty$ their static properties are understood within the framework of the replica symmetry breaking (or cavity) methods of statistics [40]. In general, the lack of structure of random CSP causes state-of-the-art exact solvers to struggle near the phase transition ratio $\alpha = M/N$ and ultimately take exponential time to find solutions because of the difficulty of truncating the search space based on exponentially growing deep decision trees [34].

On the other hand, it is not a difficult task to engineer a structured problem to challenge a heuristic solver. A very small basin of attraction of a global minimum with overall rugged landscape would make a local search heuristic relying on stochastic exploration get lost. As a result, stochastic by design, IMs can have a hard time outperforming exact routines exploiting inherent structures of problems. In order to draw conclusions about the capabilities that IMs would need to tackle both combinatorial optimization classes, we employ DGs to visualize the distinction in landscape properties between fully random and structured "industrial" instances.

To represent the structured industrial class, we generated a 3-SAT formulation of the factoring problem of the number 55 using the method from [71]. This resulted in a 3-SAT instance having 68 boolean variables and 248 clauses with only one global minimum. For comparison, a random uniform instance near the phase transition ratio α was generated of the same 68 variable size, but with 295 clauses. We obtained an instance with a single global minimum cluster having four configurations. The DG of the uniform random instance is shown in Fig. 8(a), while the DG of factoring—in Fig. 8(b).

With the chosen value of $K = 500$, the DG of the random problem was truncated at the energy levels $E = 5$ or lower, resulting in 474 distinct local minimum/saddle clusters after postprocessing. In comparison, the semiprime factoring DG has managed to fit only clusters at the energies $E \leqslant 4$ with $K = 1000$. This indicates much more pronounced ruggedness

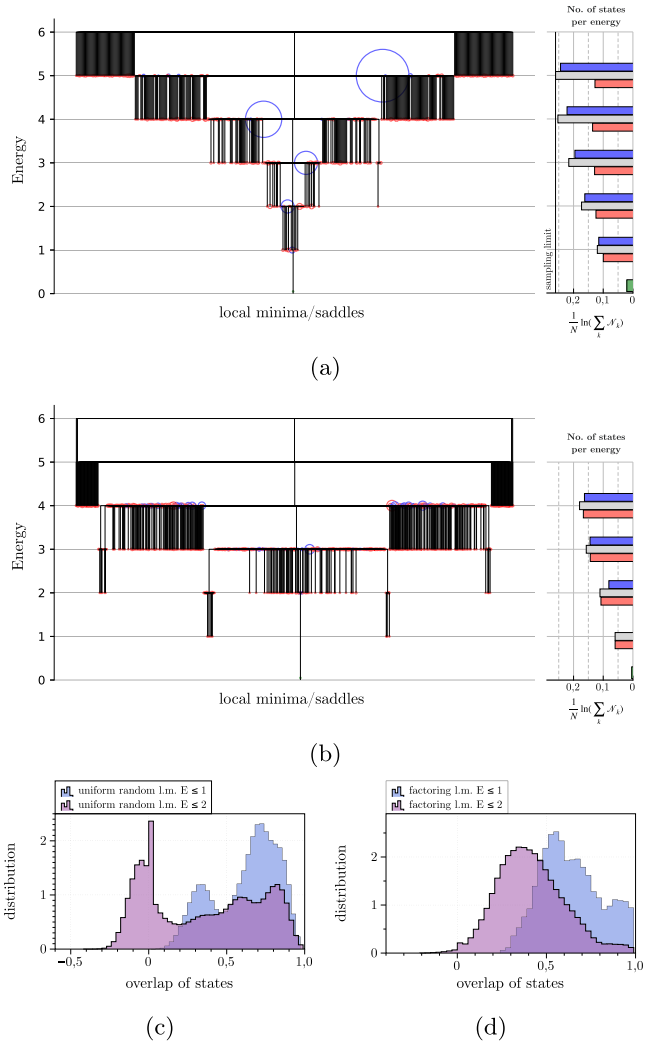


FIG. 8. Disconnectivity graph examples of 3-SAT problems with local minima truncated at $E \leqslant 4$. (a) Uniform random of 68 variables and 295 clauses. 5×10^7 sampling limit of states per energy. 474 clusters, four global minima. (b) Semi-prime factoring of 55 mapped to the 3-SAT problem of 68 variables and 248 clauses [71]. 998 clusters, one global minimum. (c), (d) Overlap distributions of local minimum states.

of the factoring 3-SAT problem with weak connectivity of saddle points.

As in the previous section, we observed exponentially large connected saddle clusters at every energy level the of random 3-SAT instance: It is possible to traverse huge distances in the optimization landscape without the need to overcome any energy barriers. It means that the hardness of this problem class arises mainly from the entropic barriers, leaving gradient-based solvers oblivious about meaningful exploration directions.

In comparison, the number of saddle clusters disconnected from the global minimum in the factoring problem constitute a much bigger fraction of the overall number of captured clusters. Moreover, the number of stable states of the factoring problem at the given energies is much smaller than in the random case (we did not need to impose sampling limits), indicating that the energy barriers are the main contributor to the hardness. These results support the conclusion of [72], where authors argue that there is no evidence for an advantage of employing SAT reductions for factoring problems, both using classical SOTA SAT solvers, and their hypothetical classical or quantum physics-inspired counterparts.

The limitation of DGs is that they compress combinatorial landscape information to local minima and barriers between them, while the distances in solution space are left aside. Since we are able to store all of the discovered by GWL + BFS states, one approach to probing such distance information is by calculating the mutual overlaps of local minima. The mutual overlap of states for Ising formulated problems is defined by $q_{ab} = \frac{1}{N}\sum_{i=1}^{N} s_i^a s_i^b$ [36], where $s_i = 2x_i - 1$. We show computed histograms of overlaps of local minima for the given random and industrial instances in Figs. 8(c) and 8(d).

The overlap distributions of random and structured problems exhibit distinct behavior with the random instance having the majority of states at zero overlap values. This property is not explicitly shown by the DG visualization. It is implied, however, by the very large saddle clusters. Compared to the random instance, the local minima of the factoring problem at $E \leqslant 2$ are closer to each other without showing evidence of a gap in the overlaps [18]. Thus, IMs can be challenged by different landscape features depending on the problem class, suggesting a strong algorithmic need for specialization.

D. QUBO mappings of 3-SAT

In this section we study energy landscapes of QUBO mappings of 3-SAT using the DGs construction extension introduced in Sec. III A 4. We explicitly state the F factor when mappings are compared with each other. We say that $F = \infty$ when all auxiliary variables are probed, essentially recovering the native (PUBO) landscape. At a given F the "effective" barrier definition is illustrated in Fig. 5.

In Fig. 9 we plot a KZFD-BG QUBO mapping landscape of the instance from Fig. 6(b) truncated to the subspace of energies $E \leqslant 5$. The QUBO barrier factor was chosen to be $F = 1$, meaning that at every step $\mathbf{x}_a \to \mathbf{x}_b$ only one QUBO auxiliary variable with $\Delta E > 0$ is allowed to be flipped in order to overcome the QUBO barriers between the native states $\mathbf{x}$. One can observe the following features of the QUBO landscape of 3-SAT:

(i) The connectivity of states is drastically reduced with large saddle point clusters of the PUBO landscape shattered into multiple disconnected saddles or local minima in the QUBO landscape (in total 1377 clusters). This has a direct negative effect on the ability to find global minima for the local search heuristics at low temperatures/noise.

(ii) The global minimum cluster (which consists of two states for this instance) is preserved, but only 18 compared to the original 206 configurations were found to be connected from energy $E = 1$ towards the solution. In other words, blue clusters have become red clusters. The same behavior is observed higher in energies, i.e., local search faces new energy barriers *in addition* to entropic barriers.

In order to highlight the necessity to carefully approach mapping into Ising hardware, we would like to directly compare the QUBO mappings introduced in Sec. III A 1 and

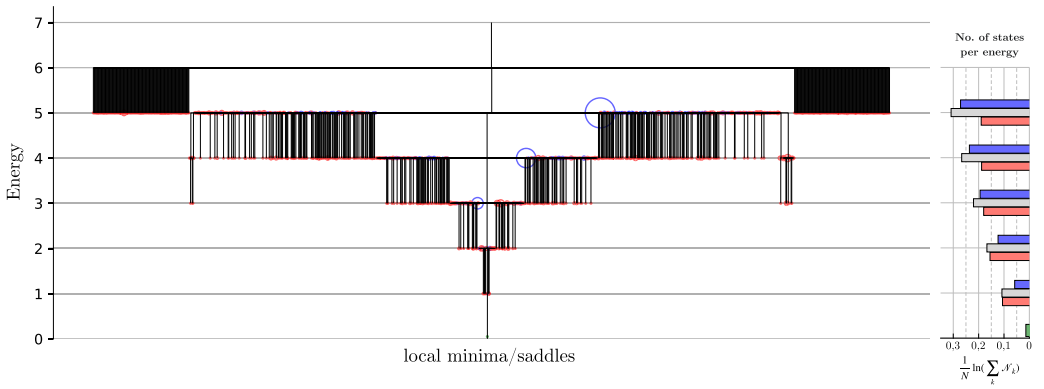


FIG. 9. QUBO mapping landscape (KZFD-BG) truncated to the stable states with $E \leqslant 5$ of a 3-SAT instance in Fig. 6(b). Problem size: 50 (native) + 136 (auxiliary) variables. QUBO factor $F = 1$. Total number of local minimum/saddle clusters is 1377; two global minimum states.

described in detail in Appendix B to each other from the perspective of connectivity of states (clustering) at different values of the QUBO factor F. We use 500 SATLIB instances of size 50 with 218 clauses (*uf50 501-1000*) to accumulate statistics from sampled DGs. On average, our QUBO mapping scheme of variable substitution and penalty terms introduced 138 ± 4 auxiliary variables.

In Fig. 10(a) we show the histogram of the number of local minimum/disconnected saddle clusters $\mathcal{N}(s) \equiv \exp(N\Sigma(s))$ of size $S \equiv \exp(Ns)$ for KZFD-BG and Rosenberg mappings, and for the native space. The parameter $\Sigma(s)$ is usually referred to as cluster *complexity*, while s is the cluster *entropy* [35]. We consider the states sampled at energies $e = E/N \leqslant 0.04$. $\Sigma(s)$ is computed as the logarithm of the number of clusters of entropy s averaged over 500 used instances. As mentioned in Sec. III A 3, the distinct clusters are sampled with the GWL algorithm, while the cluster entropy estimations are improved further by BFS. At the given energy levels we never reached the limit of BFS (10^7 states), which means that the size of every discovered cluster was exactly refined with BFS. As discussed in Appendix D 3, we also made sure that the GWL sampling histogram was uniform for every mapping, and that on average every local minimum had approximately the same number of visits.

We explicitly plot the global minima ($E = 0$) distribution as a sanity check: The RSB theory predicts its maximum value in the thermodynamic limit being at the entropy $s \approx 0.06$, while the curve itself should be below 0 complexity when the clause-to-variable ratio of 3-SAT is above the phase transition value 4.267 (we have 4.36) [40]. Both features are present for our sampled data.

We observe shattering of the native landscape clusters by the Rosenberg mapping to be stronger than that of the KZFD-BG mapping. This result can be interpreted as follows. On average, in order to transition (overcome the barrier) from state $\mathbf{x}_a$ to state $\mathbf{x}_b$ having the same energy, the Rosenberg mapping needs to overcome barriers for at least F_{Ros} auxiliary variables, introduced by quadratization, while KZFD-BG would safely transition after passing only $F_{\text{KZFD}} < F_{\text{Ros}}$. As a result, F can be seen as a measure of the ruggedness or shattering of the quadratized optimization landscape induced by the mapping.

As displayed by the histogram in Fig. 10(a), both QUBO mappings feature large clusters that are not accounted for in the PUBO case. These are the native space saddle clusters connected to the global minimum (thus not shown on the PUBO histogram), which were transformed by quadratization to either local minima or disconnected saddles. In Sec. III B we discussed the value of connected saddle points

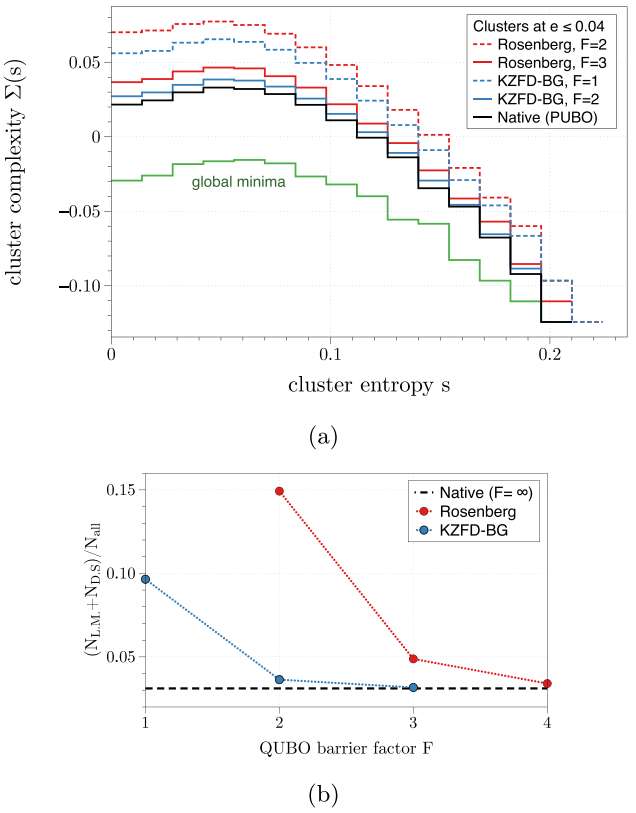


FIG. 10. (a) Local minimum and disconnected saddle cluster complexities vs cluster entropies for the native (PUBO) and QUBO landscapes at energies $E \leqslant 2$ for 500 sampled SATLIB 3-SAT problems of size $N = 50$. (b) The ratio of the number of local minima and saddles disconnected from the global minimum to all sampled stable states.

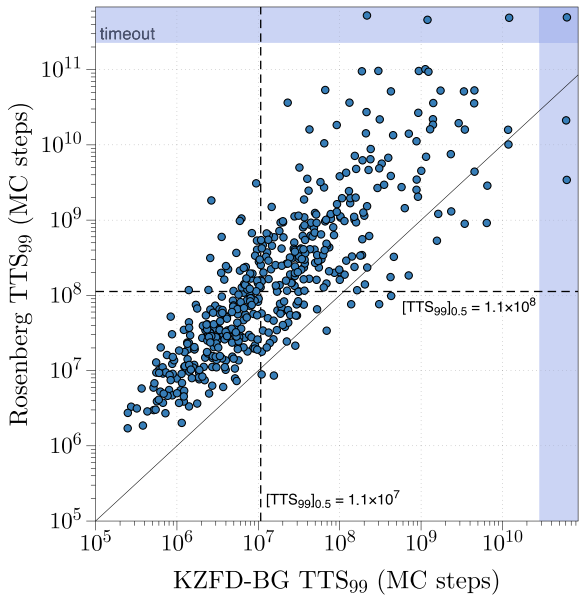


FIG. 11. TTS_{99} of simulated annealing for the QUBO mappings: non-termwise Rosenberg and KZFD-BG. Instances SATLIB *uf50 500-1000*. The timeouts are different because of distinct optimal number of sweeps (see Appendix C3); the solid line denotes the equality of TTS_{99}; the dashed lines indicate the medians of TTS_{99}.

at low temperatures/noise for finding global minima using IMs. QUBO mappings, thus, can transform saddle points into local minima effectively impeding the descend in energy. The ratio of local minima/disconnected saddles to all stable states is shown in Fig. 10(b). With increasing F we approach the native landscape faster for the KZFD-BG mapping than for the Rosenberg mapping. This suggests a potential algorithm for IMs that are forced by hardware to use quadratization methods. With sufficient exploration of the auxiliary space, it is possible to recover the native (PUBO) landscape geometry and thereby benefit from the reduced number of local minima/disconnected saddles, provided that the costs of specific hardware implementations do not outweigh such benefits in terms of time-to-solution/energy-to-solution metrics.

With our analysis we would also like to highlight the effect of choosing quadratization methods on the performance of solvers. While all such methods preserve global minima, the geometry of the configuration space changes, thereby drastically decreasing the local search efficiency in terms of the time-to-solution prefactor and empirical scaling with the problem size [14,16,17,59]. To support the observed energy landscape advantage of KZFD-BG mapping over Rosenberg, we performed simulated annealing for a collection of SATLIB 3-SAT problem instances in Fig. 11. The advantage of the KZFD-BG mapping clearly exhibits itself in the solver performance giving smaller TTS_{99} for the majority of instances ($\approx$96%). For this problem size of $N = 50$ we observe an order of magnitude improvement of the median time-to-solution. As a result, we would like to distinguish the energy-landscape geometry features of different individual problem instances from the features of quadratization methods. In the former case, the details of the energy landscape result in the spread of computational hardness as we showed in the Fig. 7. In the latter case, the QUBO mappings can lead to orders of magnitude penalties on performance for any problem instance (comparing Figs. 7 and 11). We refer to Appendix C for SA implementation details, including hyperparameter optimization, error estimation, and timeout definition. In Appendix C 3 we also test a larger problem size demonstrating increasing advantage of PUBO over QUBO, as well as of one QUBO mapping over the other, even though their corresponding QUBO embedding size is the same.

IV. CONCLUSIONS

In this paper we have suggested methods to sample disconnectivity graphs of degenerate combinatorial optimization problems, while also introducing extensions for quadratic embeddings motivated by hardware constraints of Ising machines. DGs have proven to be able to visually capture energy-landscape properties of instances with different structure (industrial and random), hardness, and order (quadratic and higher order). To characterize clustering/ruggedness of the configuration space arising from locality reduction, we have introduced a method, QUBO factor F. From this perspective we have discussed the reasons behind observed experimental performance gap between different QUBO mappings, as well as between QUBO and PUBO.

The directions for future work include investigating other definitions of neighborhoods beyond the simple bit-flip in order to visualize and gain intuition into how optimization landscapes are perceived by different local (or nonlocal) search routines. For example, isoenergetic cluster moves [73] allow solvers to make large Hamming distance steps, defining a different neighborhood for each configuration, thus a new DG with distinct connectivity of states. Moreover, understanding of the energy-landscape geometry is of great importance in a variety of fields ranging from inference and learning in energy-based models [74] to attractor dynamics and storage capacity in associative memories [75,76]. One application example is nonequilibrium inhomogeneous sampling methods [20,77], which essentially modify energy barriers reducing the hardness of sampling of high-quality and diverse solutions.

Other embedding methods motivated by the available connectivity topology or the bit-precision requirements of the Ising hardware constitute a complementary problem, which can also be studied with the methods of this paper. Finally, the distinct properties of auxiliary variables imply the possibility to introduce adaptive algorithms leveraging the specific native-auxiliary interactions within the constrains of Ising machines.

ACKNOWLEDGMENTS

This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under the Air Force Research Laboratory (AFRL) Agreement No. FA8650-23-3-7313. The views, opinions, and/or findings expressed are those of the authors and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. The authors gratefully acknowledge computing time on the supercomputer JURECA [78] Forschungszentrum Jülich under Grant No. "optimization". We also gratefully acknowledge

TABLE I. Sparsification truth table.

x	y	$g(\mathbf{x}, y)$	$f(\mathbf{x})$
0	0	0	0
0	1	$B + \|B\|$	
1	0	$A + \|B\|$	$A + B$
1	1	$A + B$	

the generous funding of this work under NEUROTEC II (Verbundkoordinator/Förderkennzeichen: Forschungszentrum Jülich/16ME0398K) by the Bundesministerium für Bildung und Forschung.

APPENDIX A: SPARSIFICATION BY AUXILIARY VARIABLES

Let us assume that because of the hardware limitations we are unable to support full interaction connectivity of a variable x_1 of the PUBO function in Eq. (3). Because of its multilinear form, we can split the interactions of x_1, i.e.,

$$f(\mathbf{x}) = A\big(x_{i_a}, \ldots, x_{k_a}\big)x_1 + B\big(x_{i_b}, \ldots, x_{k_b}\big)x_1, \tag{A1}$$

where A and B are independent functions. x_1 in the second term can be substituted by an auxiliary variable y with the introduction of a penalty as follows:

$$g(\mathbf{x}, y) = Ax_1 + By + P(x_1 + y - 2x_1 y), \tag{A2}$$

which obeys $f(\mathbf{x}) = \min_y g(\mathbf{x}, y)$ as in locality reduction methods if $P \geqslant |B|$. As a result, the local search move $(x_1, y) = (0, 0) \to (\bar{x}_1, \bar{y}) = (1, 1)$ can be made with single flips through a higher-energy barrier $A + |B|$ than in the denser original formulation (see Table I).

APPENDIX B: 3-SAT TO QUBO MAPPINGS

In this appendix we describe in detail the mappings of 3-SAT problems formulated as conjugate normal forms (CNF) to quadratic pseudo-boolean functions (QUBO).

(a) Notation. x are boolean variables, l are literals that stand for either x or its negation $\bar{x} \equiv 1 - x$, y are boolean auxiliary variables.

The problem of maximizing the number of satisfied clauses of size $k = 3$ is reformulated as a minimization problem of a third-order pseudo-boolean polynomial of literals as follows (inverting the expression and using De-Morgan law):

$$\begin{aligned}&\big(\bar{l}_{1_1} \vee \bar{l}_{2_1} \vee \bar{l}_{3_1}\big) \wedge \big(\bar{l}_{1_2} \vee \bar{l}_{2_2} \vee \bar{l}_{3_2}\big) \wedge \ldots \\ &\quad \to l_{1_1} l_{2_1} l_{3_1} + l_{1_2} l_{2_2} l_{3_2} + \ldots,\end{aligned} \tag{B1}$$

where each $l_{a_i} = x_{a_i}$ or $l_{a_i} = \bar{x}_{a_i}$, $i \in [1, 3M]$, $a \in [1, N]$. A straightforward mapping of this expression to QUBO (quadratization) would be to introduce the Rosenberg penalties for every term in the sum,

$$\begin{aligned}l_{1_1} l_{2_1} l_{3_1} + l_{1_2} l_{2_2} l_{3_2} + \ldots &= \min_{y \in \mathbb{B}} g(l(x), y), \\ g(l(x), y) &= \sum_{i \in [1,M]} y_i l_{3_i} + \big(3y_i - 2y_i l_{1_i} \\ &\quad - 2y_i l_{2_i} + l_{1_i} l_{2_i}\big).\end{aligned} \tag{B2}$$

The validity of such quadratizaton [Eq. (7)] directly follows from the fact that auxiliary variables y_i are introduced independently for each term of Eq. (B1) and $l_1 l_2 l_3 = \min_y [y l_3 + (3y - 2y l_1 - 2y l_2 + l_1 l_2)]$.

1. Non-termwise Rosenberg

In order to get the "classic" Rosenberg [54] quadratizaton, we write the PUBO of Eq. (B1) for variables x,

$$\begin{aligned}&l_{1_1} l_{2_1} l_{3_1} + l_{1_2} l_{2_2} l_{3_2} + \ldots \\ &\quad = \sum_{i<j<k} S_{ijk} x_i x_j x_k + \sum_{i<j} W_{ij} x_i x_j + \sum_i B_i x_i + C.\end{aligned} \tag{B3}$$

The pairs $x_m x_n$ in a set covering all terms of order three are substituted by auxiliary variables $y_{(mn)}$ with the addition of a penalties as in Eq. (B2),

$$\begin{aligned}\sum_{i<j<k} S_{ijk} x_i x_j x_k + \ldots = \min_y &\sum_{(mn),k} S_{(mn)k} y_{(mn)} x_k \\ &+ \sum_{(mn)} P^R_{(mn)} (3y_{(mn)} \\ &- 2y_{(mn)} x_n - 2y_{(mn)} x_m + x_m x_n) \\ &+ \ldots [\leqslant \text{2nd order terms}],\end{aligned} \tag{B4}$$

where the lower-bound penalty coefficients are now index dependent [79],

$$\begin{aligned}&P^R_{(mn)} \geqslant \max\left[\sum_k S^+_{(mn)k}, -\sum_k S^-_{(mn)k}\right] \\ &S^+_{(mn)k} > 0, \; S^-_{(mn)k} < 0.\end{aligned} \tag{B5}$$

2. Non-termwise KZFD-BG

Here we modify the Rosenberg mapping of Appendix B 1 applying quadratization ideas of [53,57,58], where the positive and negative monomials get different penalty terms of Eq. (9) (here third order),

$$\begin{aligned}-x_1 x_2 x_3 &= \min_y \left[2y - \sum_{i=1}^{3} y x_i\right], \\ x_1 x_2 x_3 &= x_2 x_3 - \bar{x}_1 x_2 x_3 \\ &= x_2 x_3 + \min_y \left[2y - y\bar{x}_1 - \sum_{i=2}^{3} y x_i\right].\end{aligned} \tag{B6}$$

Rearranging the summands in these equations, we get for the positive monomial

$$x_m x_n x_k \to y x_k + (y - y x_m - y x_n + x_m x_n), \tag{B7}$$

and for the negative monomial

$$-x_m x_n x_{k'} \to -y x_{k'} + y - x_m x_n + (y - y x_m - y x_n + x_m x_n). \tag{B8}$$

TABLE II. KZFD-BG QUBO mapping truth table for every substituted pair $x_n x_m$ and an auxiliary $y_{(mn)}$.

$y_{(mn)}$	x_m	x_n	$g(x, y_{(mn)})$	$f(x)$
0	0	0	0	0
1	0	0	$N^+ + N^- + \lvert N^-\rvert + P^K_{(mn)}$	
0	1	0	0	0
1	1	0	$N^+ + N^- + \lvert N^-\rvert$	
0	1	1	$-\lvert N^-\rvert + P^K_{(mn)}$	$N^+ + N^-$
1	1	1	$N^+ + N^-$	

As a result, an arbitrary third-order pseudo-boolean function is quadratizatized as

$$\begin{aligned}
&\sum_{i<j<k} S_{ijk} x_i x_j x_k + \dots \\
&\quad = \min_y \sum_{(mn),k} S^+_{(mn)k} y_{(mn)} x_k \\
&\qquad + \sum_{(mn),k'} S^-_{(mn)k'} (y_{(mn)} x_{k'} - y_{(mn)} + x_m x_n) \\
&\qquad + \sum_{(mn)} P^K_{(mn)} (y_{(mn)} - y_{(mn)} x_n - y_{(mn)} x_m + x_m x_n) \\
&\qquad + \dots [\leqslant \text{2nd order terms}],
\end{aligned} \tag{B9}$$

where $S^+_{(mn)k'}$, $S^-_{(mn)k'}$ denote coefficients of positive and negative monomials. The penalty parameters $P^K_{(mn)}$ are chosen as

$$\begin{aligned}
&P^K_{(mn)} \geqslant \sum_k S^+_{(mn)k} - \sum_k S^-_{(mn)k}, \\
&S^+_{(mn)k} > 0, \; S^-_{(mn)k} < 0.
\end{aligned} \tag{B10}$$

Indeed, for every auxiliary variable index (mn) we have

$$\begin{aligned}
g(x, y_{(mn)}) = (N^+ + N^-) y_{(mn)} - \lvert N^-\rvert (x_m x_n - y_{(mn)}) \\
+ P^K_{(mn)} (y_{(mn)} - y_{(mn)} x_n - y_{(mn)} x_m + x_m x_n),
\end{aligned} \tag{B11}$$

where we defined

$$\begin{aligned}
&N^+ \equiv \sum_k S^+_{(mn)k} x_k, \; N^- \equiv \sum_{k'} S^-_{(mn)k'} x_k \\
&\lvert N^+\rvert = \sum_k S^+_{(mn)k}, \; -\lvert N^-\rvert = \sum_k S^-_{(mn)k}.
\end{aligned} \tag{B12}$$

As a result, $f(x) = \min_y g(x, y)$ because Eq. (B10) is guaranteed, as shown in Table II. Compared to the Rosenberg mapping, non-termwise KZFD-BG has smaller dynamic range second-order interactions, since that $2P^R > P^K$.

The same native variable pairs $x_i x_j$ are chosen for substitution for both mappings in this paper for fair comparison. Their choice is a result of a greedy (i.e., efficient) optimization algorithm choosing the most frequent variable pairs, which achieves significant reduction (possibly not the optimal [80]) of the QUBO configuration space compared to the term-wise methods of Eq. (B2).

APPENDIX C: BENCHMARKING METHODS

1. Simulated annealing

Simulated annealing (SA) [81] is one of the simplest yet often powerful physics-inspired heuristic algorithms, which performs a MCMC (Markov chain Monte Carlo) sampling following a predefined decreasing temperature schedule. There are two common MCMC transition probability rules [82]: the heat-bath, $p(\mathbf{x} \to \mathbf{x}') = [1 + \exp(\beta \Delta E(\mathbf{x} \to \mathbf{x}'))]^{-1}$, and the Metropolis-Hastings (used in this paper),

$$p(\mathbf{x} \to \mathbf{x}') = \min[1, \exp(-\beta \Delta E(\mathbf{x} \to \mathbf{x}'))],$$

where $\beta = 1/T$.

The SA implementation we used to generate data for Figs. 7 and 11 follows an exponential temperature schedule $T(k) = T_{\text{init}} \exp(-\tau k/(N_{\text{sweeps}} - 1))$, where $\tau = \log(T_{\text{init}}/T_{\text{final}})$ and $k \in [0, N_{\text{sweeps}} - 1]$. At each k, we carry out one "sweep" over a permutation of N (N^{QUBO} for the QUBO mapping) variables of the problem applying the $p(x_i \to \bar{x}_i)$ rule. This results in a total $N \times N_{\text{sweeps}}$ ($N^{\text{QUBO}} \times N_{\text{sweeps}}$ for QUBO maps) MC steps for one SA run. The implementation of the SA used in this paper is publicly available with the DG sampler code in [69].

2. Error estimation

Simulated annealing, being a heuristic probabilistic solver without guarantees, outputs a problem solution with a certain probability of success (POS) θ. POS is defined as the number of successful runs s out of all independent SA number of repetitions N_{reps}. In this paper, every such SA run gets its own random seed, which results in an independent starting state and a sampling "trajectory" followed. POS θ can exhibit strong instance-to-instance variation within a single problem class because of the distribution of problem hardness. Moreover, θ also depends on the algorithm hyperparameters and the problem size N [83].

The total effort R_p of finding the ground state (or some predefined approximate solution) by a heuristic solver is commonly defined as the number of times the algorithm needs to be independently repeated in order to find a solution with probability p (%),

$$R_p(\theta) = \frac{\log(1 - p)}{\log(1 - \theta)}. \tag{C1}$$

R_p is then multiplied by a single SA run length to get the time-to-solution metric $\text{TTS}_p = N \times N_{\text{sweeps}} \times R_p$ (in MC steps). As a result, the wall-clock time can be readily estimated using one SA step cycle physical time of the CPU/GPU or an Ising-machine/dedicated hardware implementation. Because of the focus on the energy-landscape geometry and the corresponding algorithmic penalties of the QUBO mappings, in this paper we report all results in MC steps. Additionally, in Fig. 11 we define an artificial "timeout" value equal to $\text{TTS}^{\text{timeout}}_{99} = N \times N_{\text{sweeps}} \times R_{99}(0.5/N_{\text{reps}})$. This threshold value indicates instances with $\text{TTS}_{\text{p}} > \text{TTS}^{\text{timeout}}_{\text{p}}$ having zero observed successful trials s.

We followed the studies [83,84] for the error estimation of the SA benchmarking data. Using the recorded number

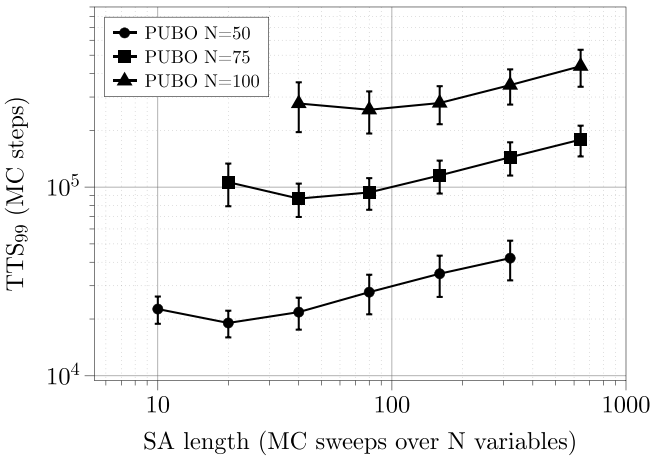


FIG. 12. Hyperparameter optimization of the number of simulated annealing sweeps for PUBO. $N_{\text{reps}} = 5120,\ 7680,\ 10240$ of 50 instances each for $N = 50,\ 75,\ 100$, respectively. Mean and the standard deviation of the median TTS_{99} estimated with bootstrapping.

of successful trials s from a number N_{reps} of independent SA repetitions, the probability distribution of the POS θ is modelled using the beta distribution

$$\beta[1/2 + s,\ 1/2 + (N_{\text{reps}} - s)]. \tag{C2}$$

In order to generate the error bars for a given value of interest $\mathcal{F}$ and a given set of instances $\mathcal{S}$, we use a simple bootstrapping method. A new set of instances $\mathcal{S}_i$ of the same cardinality as $\mathcal{S}$ is resampled with replacement from $\mathcal{S}$ 10000 times. For each such instance j in $\mathcal{S}_i$ the POS θ is sampled from the beta distribution of Eq. (C2). Finally, the statistics of $\mathcal{F}$ is obtained using the set $\mathcal{F}_i = \mathcal{F}(\{\theta_j\}_i)$. For example, in Figs. 12–14 below we report the mean and the standard deviation of the median TTS_{99} using this bootstrapping method. The same rule applies when we report the median of the ratios of TTS_{99}.

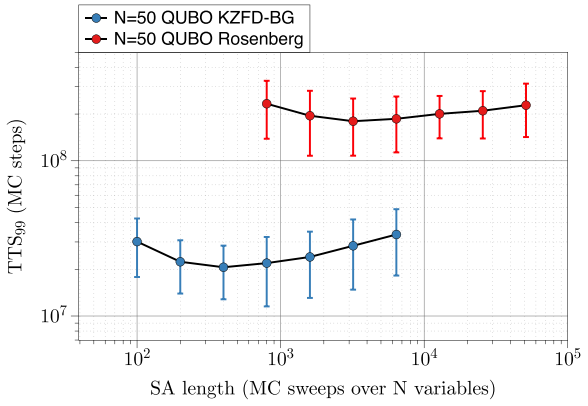


FIG. 13. Hyperparameter optimization of the number of simulated annealing sweeps for the QUBO mappings. Using 50 instances of size $N = 50$, with mappings giving $N^{\text{QUBO}} = 188 \pm 5$. $N_{\text{reps}} = 20480$ for each instance. Mean and the standard deviation of the median TTS_{99} estimated with bootstrapping.

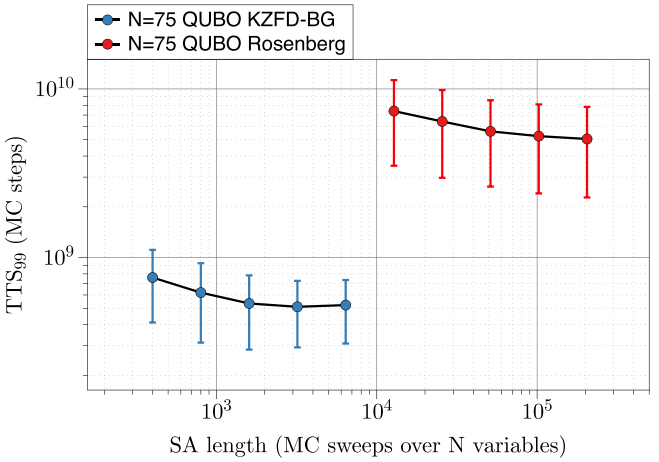


FIG. 14. Hyperparameter optimization of the number of simulated annealing sweeps for the QUBO mappings. Using 100 instances of size $N = 75$, with mappings giving $N^{\text{QUBO}} = 303 \pm 5$. $N_{\text{reps}} = 81920$ for each instance. Mean and the standard deviation of the median TTS_{99} estimated with bootstrapping.

3. Annealing hyperparameter optimization

The initial and final temperatures of SA set for the benchmarking of PUBO and QUBO mappings were chosen as $T_i = 1.5,\ T_f = 0.1$. As a reminder, the minimum nonzero $|\Delta E|$ for the problems tested in this paper equals to 1. The chosen values of T resulted on average in the initial $0.615 \pm 0.005,\ 0.634 \pm 0.010,\ 0.523 \pm 0.012$ and the final $0.10 \pm 0.03,\ 0.20 \pm 0.03,\ 0.14 \pm 0.03$ MCMC sampling acceptance rates for PUBO, QUBO KZFD-BG, and QUBO Rosenberg, respectively.

PUBO and each QUBO mapping with the chosen temperature schedule favor different N_{sweeps} for optimal performance. In order to facilitate fairer benchmarking, in Figs. 12–14 we optimized this hyperparameter of SA. First, the PUBO performance at problem sizes $N = 50,\ 75,\ 100$ is optimized in Fig. 12 using first 50 satisfiable SATLIB 3-SAT instances at each size [*uf(N) 1-50*]. Second, the QUBO performance optimization at problem size $N = 50$ for the same 50 instances and at $N = 75$ for all available 100 instances is shown in Figs. 13 and 14.

The mean and the standard deviation of the median TTS_{99} were estimated using success probabilities obtained from N_{reps} experiment repetitions for each instance and each value of N_{sweeps}. As a result, we found the optimum N_{sweeps} and used these established values to generate results for 500 instances *uf50 501-1000* in Figs. 7 and 11 with increased number of the repetitions for even better statistics: 10240 and 40960, respectively.

Finally, we note the scaling differences between PUBO and QUBO with the problem size. The change of median TTS_{99} from $N = 50$ to $N = 75$ in PUBO is: 18400 ± 900 for instances *uf50 500-1000* in Fig. 7 to 83000 ± 12000 for instances *uf75 1-100*, i.e., ≈5 times increase. In QUBO it equals to $(1.08 \pm 0.14) \times 10^7$ and $(1.07 \pm 0.17) \times 10^8$ at $N = 50$ (Fig. 11) and $(5.1 \pm 2.2) \times 10^8$ and $(5.0 \pm 2.8) \times 10^9$ at $N = 75$ for KZFD-BG and Rosenberg mappings respectively

(Fig. 14). As a result, the increase of the median TTS_{99} with increasing problem size for both mappings is ≈47 times.

We also estimated the TTS_{99} ratio of the QUBO mappings, namely the median of $TTS_{99}^{Ros}/TTS_{99}^{KZFD}$. The resulting medians of ratios are 9.3 ± 0.7 at $N = 50$ and 15.8 ± 2.4 at $N = 75$. As a result, the following conclusions can be made:

(i) the advantage of PUBO vs QUBO grows with N, i.e., the scaling of PUBO is exponentially better than both considered QUBO mappings; and

(ii) the scaling advantage with growing N when comparing two QUBO mappings is also observed; however, reliable functional fitting of scaling and extrapolation to larger problem sizes requires extensive testing of the mappings at $N > 75$ and is left for future work.

APPENDIX D: DISCONNECTIVITY GRAPHS SAMPLER

1. Code availability

The original code developed for this paper uses GWL sampling described in Sec. III A and has the following output: GWL histogram of visits to basins of attraction and energy levels, sampled clusters degeneracies, symmetric matrix of energy barriers between clusters, local minimum states. The information about the connectivity of clusters and their type (local minimum/saddle) we then derive from the barrier matrix during post-processing and DG construction. As input the program takes the conjugate normal form of a 3-SAT problem.

The examples of DG sampling hyperparameters that can be tuned are: number of parallel threads of sampling, total GWL steps per thread, limit on the cluster exploration and breadth-first search limits. In principle, it is possible to tune all hyperparameters to optimize the sampling for a particular problem class. The 3-SAT (PUBO/QUBO) GWL sampling code with the hyperparameters used in this paper as well as the simulated annealing implementation following Appendix C is available in a public repository at Ref. [69]. Extended disconnectivity graphs construction from the aforementioned data sampled with the GWL algorithm is based on the functions from **pele** library [85] and can be made available upon reasonable request.

2. GWL sampling uniformity

Generalized Wang-Landau (GWL) algorithm discussed in Sec. III A 2 aims to sample the configuration space as uniformly as possible. The uniformity of sampling is being tracked by the histogram $\theta_{l,k}$ (15) of visits to a particular energy level l and local minimum/saddle basin of attraction k. Figure 15 showcases one histogram example that we observed while sampling the energy landscape for the DG construction of Fig. 6(b). We note that some of the independent clusters reported by the histogram can in fact be the same cluster with connections between them not yet discovered during sampling. An extensive BFS search described in Sec. III A 3 is employed in this paper at postprocessing to join such clusters together.

By tracking the distribution properties of the histogram, one is able tune the hyperparameters of GWL sampling and/or observe its convergence. To verify our choices of hyperparameters, in Fig. 16 we plot the relative standard deviation (RSD)

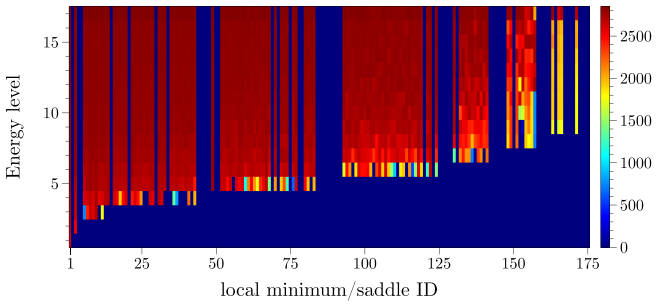


FIG. 15. Sampling histogram of the 3-SAT instance in Fig. 6(b). Total number of GWL steps is 4×10^6. The visits to saddle clusters are counted towards their corresponding local minima lower in energy, i.e., saddles show zero visits in the histogram (cf. Sec. III A 2).

of $\theta_{l,k}$ as a function of the number of GWL MC steps. The maximum histogram energy E^L was chosen to be 16 for *uf50* instances in PUBO and QUBO, 18 for the random $N = 68$ instance, and 17 for the semiprime factoring instance. Since the set of local minima/saddles during sampling is not fixed, in general the histogram may show temporary increases in its deviation because of the discovery of new clusters. We have chosen RSD of ≈10–15% for DG construction, which resulted in 4×10^6 GWL steps for Fig. 6 and 1.6×10^7 for Figs. 8(a), 8(b), and 9.

3. Disconnectivity graphs convergence

Uniformity of the GWL histogram, while indicative of high quality of sampling, is not a guarantee of DG construction accuracy. We have also tested the convergence (saturation) of the discovered local minimum/saddle clusters and of the energy barriers between them. In Fig. 17(a) we plot occurrence percentage of the cluster sets in an independent sampling run at different numbers of GWL sampling steps to the sets in the sampling runs used for DGs in this paper. These distinct runs differ only by a unique random seed choice. In Fig. 17(a) we observe that at $4-8 \times 10^6$ all local minimum clusters are

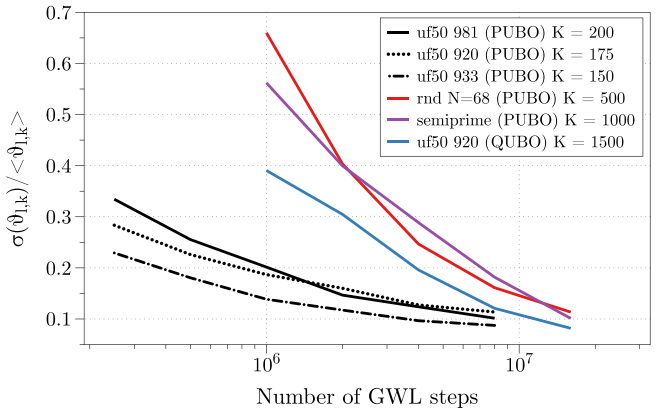
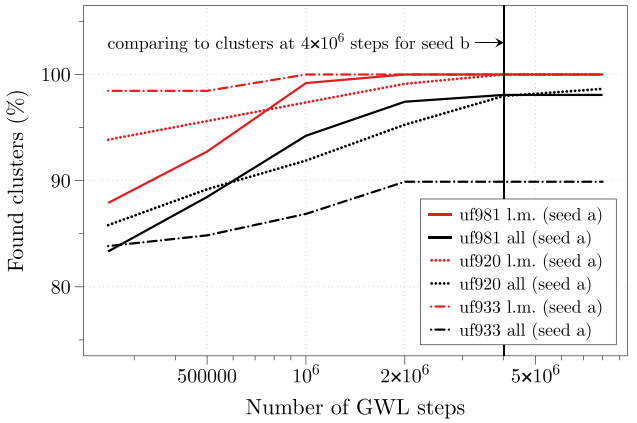
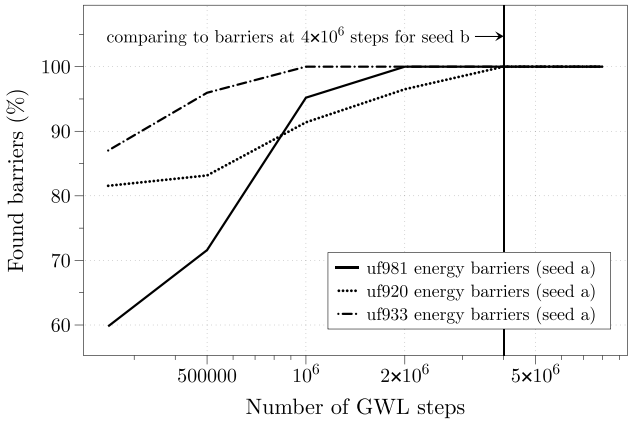


FIG. 16. Relative standard deviation of the nonzero GWL histogram elements $\theta_{l,k}$ as a function of the number of GWL sampling steps for every instance visualized with DGs (see histogram example in Fig. 15).



(a)



(b)

FIG. 17. The percentage of found number of local minimum clusters and all clusters (local minima + saddles) (a) and energy barriers (b) at different values of GWL sampling steps and for an independent run (seed "a") with respect to the result used in this paper (4×10^6 steps, seed "b") in Fig. 6.

discovered with respect to the clusters obtained at 4×10^6 in the GWL run used in this paper, i.e., sampling saturated.

The saddle clusters did not fully saturate because of the 10^7 sampling limit, which restricted our ability to discover all stable states at very high-energy levels and exactly match the clusters. However, this did not affect the accuracy of energy barrier construction, as shown in Fig. 17(b). Here, we test how many of the $K'(K'-1)/2$ barriers between K' local minima of the "seed b" runs have been reconstructed in the independent "seed a" runs. For every instance, we observe that the barriers between local minima saturate, indicating the reproducibility and accuracy of DGs construction.

In Figs. 8(a), 8(b), and 9 we found all discovered local minima at $E \leqslant 4$, $E \leqslant 3$, and $E \leqslant 5$, respectively, to coincide with an independent sampling run using 1.6×10^7 GWL steps. At higher energy levels our limits on the number of distinct clusters K (given in Fig. 16) truncated different sets of local minima in independent runs; therefore, we do not compare the local minima found with seeds "a" and "b". The barriers between the matched local minima were observed to 100% match in Figs. 8(a) and 8(b) and 99.85% match in Fig. 9.

4. Sampling complexity

Sections III A 2–III A 4 describe a variety of primitives that were implemented in [69] in order to construct DGs. Above we reported the numbers of GWL MC steps that were used to obtain the data about local minima/saddles, as well as about the energy barriers between then. Each GWL step consists of the following routines, each having its corresponding complexity.

When a new state is proposed, random descend is performed to establish the affinity to a particular basin of attraction. Each step of the random energy descend requires the computation of ΔE of bit-flip neighbors. If no negative ΔE is found, the worst case number of computations is $O(N)$, assuming a sparse problem without scaling of the number of interactions for each variable. When a plateau region is encountered during the descend, we perform a fixed predefined number of exploration steps before terminating the descend. It is a hyperparameter and chosen to be 20 in this paper. We do not scale this number with the problem size, thus the complexity is also $O(N)$. Since the total number of descend steps scales as $O(N)$, the resulting complexity is $O(N^2)$.

When a saddle or a local minimum is identified, we need to either find an existing cluster it is connected to, or insert it as a new cluster to the set of all clusters. At each energy level we store a sorted set container of all so far discovered stable states. Let us assume that there are exponentially many already found states, i.e., the container size is worst case $O(\exp cN)$. The search and insertion into such a sorted container has complexity $O(\log(\text{size}))$, i.e., $O(N)$. Since we need to identify and insert $O(N)$ states, because of $O(N)$ possible energy levels and a fixed number of newly discovered states at each level, the overall complexity of search and insertion is $O(N^2)$.

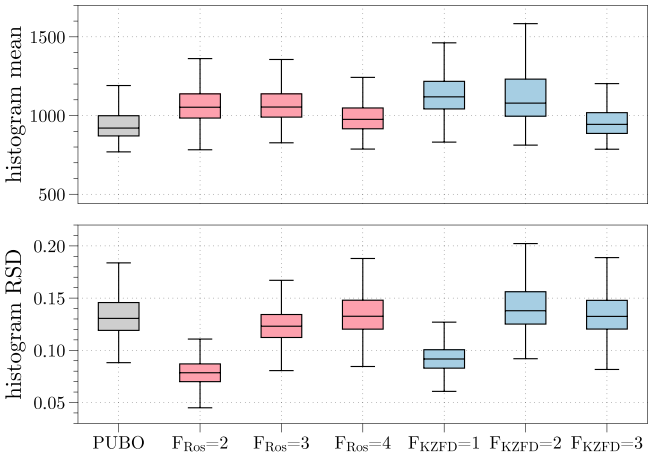


FIG. 18. Statistics of the means and of relative standard deviations (RSD) of the nonzero GWL histogram elements $\theta_{l,k}$ for instances and mappings in Fig. 10.

Finally, the breadth-first search can be executed in our implementation when a new cluster is discovered, or at the end of GWL sampling to exactly calculate the the size of each cluster. In the former case, the limit is a hyperparameter, which in this paper was chosen to be 500. In the latter case, for each energy level we set the total limit on states to 10^7 for the DGs of problems of native size $N = 50$ (PUBO and QUBO) and 5×10^7 for DGs of the problems of size $N = 68$. The number of required iterations depends on the degeneracy of the problem of interest.

In this paper, our main focus was the accuracy of the introduced DGs sampling and construction method demonstrated for the chosen hyperparameter in the sections above. Our machine (single thread of a CPU) took ≈20–30 minutes for GWL sampling of DGs in Figs. 6 (4×10^6 steps), ≈2–3 hours in Figs. 8 and 9 (1.6×10^7 steps), each having tracked tens of of millions of local minimum/saddle states. The program [69] supports multitheading speedup because of parallel independent sampling of a single-energy landscape. We leave the exhaustive hyperparameter optimization and benchmarking of our implementation of the method for future work.

5. QUBO sampling details

For every QUBO mapping and every value of F in Fig. 10, we have chosen the histogram size limits K so that we are able to fit all distinct clusters at the energy levels $E \leqslant 2$. Next, the numbers of GWL sampling steps were chosen so that in each case we obtain good levels of histogram relative standard deviation ($\lessapprox 15\%$) and approximately equal number of absolute visits to each basin of attraction (≈1000). For most instances, $K = 70$ was sufficient for PUBO, $K = 500,\ 200,\ 100$ for the Rosenberg mapping with $F = 2,\ 3,\ 4$, respectively, and $K = 400,\ 150,\ 80$ for the KZFD-BG mapping with $F = 1,\ 2,\ 3$, respectively. We found the value of $N_{\text{steps}} = K \times 10^4$ to satisfy the aforementioned requirements. The resulting statistics of $\theta_{l,k}$ of instances in Fig. 10 are shown in Fig. 18.

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

[2] M. R. Mahmoodi, M. Prezioso, and D. B. Strukov, Versatile stochastic dot product circuits based on nonvolatile memories for high performance neurocomputing and neurooptimization, Nat. Commun. **10**, 5113 (2019).

[3] F. Cai, S. Kumar, T. Van Vaerenbergh, X. Sheng, R. Liu, C. Li, Z. Liu, M. Foltin, S. Yu, Q. Xia, J. J. Yang *et al.*, Power-efficient combinatorial optimization using intrinsic noise in memristor hopfield neural networks, Nat. Electron. **3**, 409 (2020).

[4] M. Aramon, G. Rosenberg, E. Valiante, T. Miyazawa, H. Tamura, and H. G. Katzgraber, Physics-inspired optimization for quadratic unconstrained problems using a digital annealer, Front. Phys. **7**, 48 (2019).

[5] X. Peng, Z. Liao, N. Xu, G. Qin, X. Zhou, D. Suter, and J. Du, Quantum adiabatic algorithm for factorization and its experimental implementation, Phys. Rev. Lett. **101**, 220405 (2008).

[6] W. A. Borders, A. Z. Pervaiz, S. Fukami, K. Y. Camsari, H. Ohno, and S. Datta, Integer factorization using stochastic magnetic tunnel junctions, Nature (London) **573**, 390 (2019).

[7] A. Mallick, M. K. Bashar, D. S. Truesdell, B. H. Calhoun, S. Joshi, and N. Shukla, Using synchronized oscillators to compute the maximum independent set, Nat. Commun. **11**, 4689 (2020).

[8] H. Goto, K. Endo, M. Suzuki, Y. Sakai, T. Kanao, Y. Hamakawa, R. Hidaka, M. Yamasaki, and K. Tatsumura, High-performance combinatorial optimization based on classical mechanics, Sci. Adv. **7**, eabe7953 (2021).

[9] R. Hamerly, T. Inagaki, P. L. McMahon, D. Venturelli, A. Marandi, T. Onodera, E. Ng, C. Langrock, K. Inaba, T. Honjo *et al.*, Experimental investigation of performance differences between coherent Ising machines and a quantum annealer, Sci. Adv. **5**, eaau0823 (2019).

[10] S. Patel, L. Chen, P. Canoza, and S. Salahuddin, Ising model optimization problems on a FPGA accelerated restricted Boltzmann machine, arXiv:2008.04436.

[11] T. Leleu, F. Khoyratee, T. Levi, R. Hamerly, T. Kohno, and K. Aihara, Scaling advantage of nonrelaxational dynamics for high-performance combinatorial optimization, arXiv:2009.04084.

[12] M. Kowalsky, T. Albash, I. Hen, and D. A. Lidar, 3-regular three-XORSAT planted solutions benchmark of classical and quantum heuristic optimizers, Quantum Sci. Technol. **7**, 025008 (2022).

[13] M. S. Könz, W. Lechner, H. G. Katzgraber, and M. Troyer, Embedding overhead scaling of optimization problems in quantum annealing, PRX Quantum **2**, 040322 (2021).

[14] A. Perdomo-Ortiz, A. Feldman, A. Ozaeta, S. V. Isakov, Z. Zhu, B. O'Gorman, H. G. Katzgraber, A. Diedrich, H. Neven, J. de Kleer, B. Lackey, and R. Biswas, Readiness of quantum optimization machines for industrial applications, Phys. Rev. Appl. **12**, 014004 (2019).

[15] E. Valiante, M. Hernandez, A. Barzegar, and H. G. Katzgraber, Computational overhead of locality reduction in binary optimization problems, Comput. Phys. Commun. **269**, 108102 (2021).

[16] M. Hizzani, A. Heittmann, G. Hutchinson, D. Dobrynin, T. V. Vaerenbergh, T. Bhattacharya, A. Renaudineau, D. Strukov, and J. P. Strachan, Memristor-based hardware and algorithms for higher-order hopfield optimization solver outperforming quadratic Ising machines, arXiv:2311.01171.

[17] C. Bybee, D. Kleyko, D. E. Nikonov, A. Khosrowshahi, B. A. Olshausen, and F. T. Sommer, Efficient optimization with higher-order Ising machines, Nat. Commun. **14**, 6033 (2023).

[18] D. Gamarnik, The overlap gap property: A topological barrier to optimizing over random structures, Proc. Natl. Acad. Sci. USA **118**, e2108492118 (2021).

[19] M. Bernaschi, M. Bisson, M. Fatica, E. Marinari, V. Martin-Mayor, G. Parisi, and F. Ricci-Tersenghi, How we are leading a 3-XORSAT challenge: From the energy landscape to the algorithm and its efficient implementation on GPUs, Europhys. Lett. **133**, 60005 (2021).

[20] M. Mohseni, D. Eppens, J. Strumpfer, R. Marino, V. Denchev, A. K. Ho, S. V. Isakov, S. Boixo, F. Ricci-Tersenghi, and H.

Neven, Nonequilibrium Monte Carlo for unfreezing variables in hard combinatorial optimization, arXiv:2111.13628.
[21] O. M. Becker and M. Karplus, The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics, J. Chem. Phys. **106**, 1495 (1997).
[22] D. J. Wales, M. A. Miller, and T. R. Walsh, Archetypal energy landscapes, Nature (London) **394**, 758 (1998).
[23] J. P. K. Doye, M. A. Miller, and D. J. Wales, The double-funnel energy landscape of the 38-atom Lennard-Jones cluster, J. Chem. Phys. **110**, 6896 (1999).
[24] D. Chakraborty, R. Collepardo-Guevara, and D. J. Wales, Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins, J. Am. Chem. Soc. **136**, 18052 (2014).
[25] F. Calvo, T. V. Bogdan, V. K. de Souza, and D. J. Wales, Equilibrium density of states and thermodynamic properties of a model glass former, J. Chem. Phys. **127**, 044508 (2007).
[26] D. J. Wales, Energy landscapes and properties of biomolecules, Phys. Biol. **2**, S86 (2005).
[27] P. Garstecki, T. X. Hoang, and M. Cieplak, Energy landscapes, supergraphs, and "folding funnels" in spin systems, Phys. Rev. E **60**, 3219 (1999).
[28] K. Biswas and H. G. Katzgraber, Adding color: Visualization of energy landscapes in spin glasses, arXiv:2004.12431.
[29] Q. Zhou, Random walk over basins of attraction to construct Ising energy landscapes, Phys. Rev. Lett. **106**, 180602 (2011).
[30] A. Coja-Oghlan, A. Haqshenas, and S. Hetterich, Walksat stalls well below satisfiability, SIAM J. Discrete Math. **31**, 1160 (2017).
[31] M. Bellitti, F. Ricci-Tersenghi, and A. Scardicchio, Entropic barriers as a reason for hardness in both classical and quantum algorithms, Phys. Rev. Res. **3**, 043015 (2021).
[32] F. Barahona, On the computational complexity of Ising spin glass models, J. Phys. A: Math. Gen. **15**, 3241 (1982).
[33] E. Gardner, Spin glasses with p-spin interactions, Nucl. Phys. B **257**, 747 (1985).
[34] B. Selman and S. Kirkpatrick, Critical behavior in the computational cost of satisfiability testing, Artif. Intell. **81**, 273 (1996).
[35] M. Mézard, T. Mora, and R. Zecchina, Clustering of solutions in the random satisfiability problem, Phys. Rev. Lett. **94**, 197205 (2005).
[36] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, New York, 2009).
[37] J. Marques-Silva, Practical applications of boolean satisfiability, in *Proceedings of the 2008 9th International Workshop on Discrete Event Systems, Gothenburg, Sweden* (IEEE, 2008), pp. 74–80.
[38] T. Bhattacharya, G. H. Hutchinson, G. Pedretti, X. Sheng, J. Ignowski, T. V. Vaerenbergh, R. Beausoleil, J. P. Strachan, and D. B. Strukov, Computing high-degree polynomial gradients in memory, Nat. Commun. **15**, 8211 (2024).
[39] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness* (W. H. Freeman & Co., New York, 1990).
[40] A. Montanari, F. Ricci-Tersenghi, and G. Semerjian, Clusters of solutions and replica symmetry breaking in random k-satisfiability, J. Stat. Mech.: Theory Exp. (2008) P04004.
[41] L. Zdeborová and F. Krząkała, Phase transitions in the coloring of random graphs, Phys. Rev. E **76**, 031131 (2007).
[42] J. Ardelius and L. Zdeborová, Exhaustive enumeration unveils clustering and freezing in the random 3-satisfiability problem, Phys. Rev. E **78**, 040101 (2008).
[43] K. Li, H. Ma, and H. Zhou, From one solution of a 3-satisfiability formula to a solution cluster: Frozen variables and entropy, Phys. Rev. E **79**, 031102 (2009).
[44] G. Bresler and B. Huang, The algorithmic phase transition of random k-sat for low degree polynomials, arXiv:2106.02129.
[45] D. Gamarnik, C. Moore, and L. Zdeborová, Disordered systems insights on computational hardness, J. Stat. Mech.: Theory Exp. (2022) 114015.
[46] C. M. Reidys and P. F. Stadler, Combinatorial landscapes, SIAM Rev. **44**, 3 (2002).
[47] C. Flamm, I. L. Hofacker, P. F. Stadler, and M. T. Wolfinger, Barrier trees of degenerate landscapes, Z. Phys. Chem. **216**, 155 (2002).
[48] J. Hallam and A. Prugel-Bennett, Large barrier trees for studying search, IEEE Trans. Evol. Comput. **9**, 385 (2005).
[49] Y. 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, arXiv:1406.2572.
[50] J. D. Lee, I. Panageas, G. Piliouras, M. Simchowitz, M. I. Jordan, and B. Recht, First-order methods almost always avoid strict saddle points, Math. Program. **176**, 311 (2019).
[51] N. Dattani, Quadratization in discrete optimization and quantum mechanics, arXiv:1901.04405.
[52] A. Schrijver, A combinatorial algorithm minimizing submodular functions in strongly polynomial time, J. Comb. Theory, Ser. B **80**, 346 (2000).
[53] E. Boros and A. Gruber, On quadratization of pseudo-boolean functions, arXiv:1404.6538.
[54] G. Rosenberg, Reduction of bivalent maximization to the quadratic case, Cahiers du Centre d'Etudes de Recherche Operationnelle **17**, 71 (1975).
[55] J. D. Biamonte, Nonperturbative k-body to two-body commuting conversion hamiltonians and embedding problem instances into Ising spins, Phys. Rev. A **77**, 052331 (2008).
[56] N. A. Aadit, A. Grimaldi, M. Carpentieri, L. Theogarajan, J. M. Martinis, G. Finocchio, and K. Y. Camsari, Massively parallel probabilistic computing with sparse Ising machines, Nat. Electron. **5**, 460 (2022).
[57] V. Kolmogorov and R. Zabin, What energy functions can be minimized via graph cuts? IEEE Trans. Pattern Anal. Mach. Intell. **26**, 147 (2004).
[58] D. Freedman and P. Drineas, Energy minimization via graph cuts: Settling what is possible, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (2005), Vol. 2, pp. 939–946.
[59] A. Sharma, M. Burns, A. Hahn, and M. Huang, Augmenting an electronic Ising machine to effectively solve boolean satisfiability, Sci. Rep. **13**, 22858 (2023).
[60] D. P. Landau, S.-H. Tsai, and M. Exler, A new approach to Monte Carlo simulations in statistical physics: Wang-Landau sampling, Am. J. Phys. **72**, 1294 (2004).
[61] F. Liang, A generalized Wang-Landau algorithm for Monte Carlo computation, J. Am. Stat. Assoc. **100**, 1311 (2005).
[62] W. Tang and Q. Zhou, Finding multiple minimum-energy conformations of the hydrophobic-polar protein model via multidomain sampling, Phys. Rev. E **86**, 031909 (2012).

[63] A. Barbu and S.-C. Zhu, *Monte Carlo Methods* (Springer, Singapore, 2020).

[64] A. K. Hartmann, A new method for analysing ground-state landscapes: Ballistic search, J. Phys. A: Math. Gen. **33**, 657 (2000).

[65] A. Mann and A. K. Hartmann, Numerical solution-space analysis of satisfiability problems, Phys. Rev. E **82**, 056702 (2010).

[66] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, What is the computational value of finite-range tunneling? Phys. Rev. X **6**, 031015 (2016).

[67] This valid state may not be unique, i.e. there could be degeneracy in the auxiliary variable space **y** for each fixed **x**. In this case we consider all such configurations and take the minimum of the computed effective QUBO barrier.

[68] H. H. Hoos, T. Stützle, I. P. Gent, H. van Maaren, and T. Walsh, *SATLIB: An Online Resource for Research on SAT* (Kluwer Academic Publishers, The Netherlands, 2000).

[69] Disconnectivity graph sampler for pubo/qubo mappings of 3-sat problems, available at https://github.com/dumdob/sat-qubo-dg-sampler (2024).

[70] J. Frank, P. Cheeseman, and J. Stutz, When gravity fails: Local search topology, arXiv:cs/9712101.

[71] P. Purdom and A. Sabry, CNF generator for factoring problems, https://cgi.luddy.indiana.edu/~sabry/cnf.html.

[72] M. Mosca and S. R. Verschoor, Factoring semi-primes with (quantum) sat-solvers, Sci. Rep. **12**, 7982 (2022).

[73] J. Houdayer, A cluster Monte Carlo algorithm for 2-dimensional spin glasses, Eur. Phys. J. B **22**, 479 (2001).

[74] Y. Lecun, S. Chopra, R. Hadsell, M. A. Ranzato, and F. J. Huang, A tutorial on energy-based learning, *Predicting Structured Data*, edited by G. Bakir, T. Hofman, B. Scholkopt, A. Smola, and B. Taskar (MIT Press, 2006).

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

[76] D. Krotov and J. J. Hopfield, Dense associative memory for pattern recognition, arXiv:1606.01164.

[77] M. Mohseni, M. M. Rams, S. V. Isakov, D. Eppens, S. Pielawa, J. Strumpfer, S. Boixo, and H. Neven, Sampling diverse near-optimal solutions via algorithmic quantum annealing, Phys. Rev. E **108**, 065303 (2023).

[78] Jülich Supercomputing Centre, JURECA: Data centric and booster modules implementing the modular supercomputing architecture at Jülich supercomputing centre, J. Large-Scale Res. Facilities **7** (2021).

[79] R. Babbush, B. O'Gorman, and A. Aspuru-Guzik, Resource efficient gadgets for compiling adiabatic quantum optimization problems, Ann. Phys. (NY) **525**, 877 (2013).

[80] Finding a minimal set of pairs of variables to be substituted is in general an NP-Hard problem.

[81] S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, Optimization by simulated annealing, Science **220**, 671 (1983).

[82] H. G. Katzgraber, Introduction to Monte Carlo methods, arXiv:0905.1629.

[83] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Defining and detecting quantum speedup, Science **345**, 420 (2014).

[84] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, and D. A. Lidar, Probing for quantum speedup in spin-glass problems with planted solutions, Phys. Rev. A **92**, 042325 (2015).

[85] pele: Python energy landscape explorer, https://pele-python.github.io/pele/.