Efficient Population Transfer via Non-Ergodic Extended States in Quantum Spin Glass

Kostyantyn Kechedzhi[1]
QuAIL and USRA, NASA Ames Research Center, Moffett Field, CA 94035, USA
Google Inc., Venice, CA 90291, USA
kostyantyn.kechedzhi@nasa.gov

Vadim Smelyanskiy
Google Inc., Venice, CA 90291, USA
smelyan@google.com

Jarrod R. McClean
Google Inc., Venice, CA 90291, USA
jmcclean@google.com

Vasil S. Denchev
Google Inc., Venice, CA 90291, USA
denchev@google.com

Masoud Mohseni
Google Inc., Venice, CA 90291, USA
mohseni@google.com

Sergei Isakov
Google Inc., Venice, CA 90291, USA
iserge@google.com

Sergio Boixo
Google Inc., Venice, CA 90291, USA
boixo@google.com

Boris Altshuler
Physics Department, Columbia University,
538 West 120th Street, New York, New York 10027 , USA
bla@phys.columbia.edu

Hartmut Neven
Google Inc., Venice, CA 90291, USA
neven@google.com

Abstract

Quantum tunneling has been proposed as a physical mechanism for solving binary optimization problems on a quantum computer because it provides an alternative to simulated annealing by directly connecting deep local minima of the energy landscape separated by large Hamming distances. However, classical simulations using Quantum Monte Carlo (QMC) were found to

---

efficiently simulate tunneling transitions away from local minima if the tunneling is effectively dominated by a single path. We analyze a new computational role of coherent multi-qubit tunneling that gives rise to bands of non-ergodic extended (NEE) quantum states each formed by a superposition of a large number of deep local minima with similar energies. NEE provide a coherent pathway for population transfer (PT) between computational states with similar energies. In this regime, PT cannot be efficiently simulated by QMC. PT can serve as a new quantum subroutine for quantum search, quantum parallel tempering and reverse annealing optimization algorithms. We study PT resulting from quantum evolution under a transverse field of an n-spin system that encodes the energy function $E(z)$ of an optimization problem over the set of bit configurations z. Transverse field is rapidly switched on in the beginning of algorithm, kept constant for sufficiently long time and switched off at the end. Given an energy function of a binary optimization problem and an initial bit-string with atypically low energy, PT protocol searches for other bitstrings at energies within a narrow window around the initial one. We provide an analytical solution for PT in a simple yet nontrivial model: M randomly chosen marked bit-strings are assigned energies $E(z)$ within a narrow strip $[-n-W/2, n+W/2]$, while the rest of the states are assigned energy 0. The PT starts at a marked state and ends up in a superposition of L marked states inside the narrow energy window whose width is smaller than W. The best known classical algorithm for finding another marked state is the exhaustive search. We find that the scaling of a typical PT runtime with n and L is the same as that in the multi-target Grover's quantum search algorithm, except for a factor that is equal to $\exp(n/(2B^2))$ for finite transverse field $B \gg 1$. Unlike the Hamiltonians used in analog quantum unstructured search algorithms known so far, the model we consider is non-integrable and the transverse field delocalizes the marked states. As a result, our PT protocol is not exponentially sensitive in n to the weight of the driver Hamiltonian and may be initialized with a computational basis state. We develop the microscopic theory of PT by constructing a down-folded dense Hamiltonian acting in the space of marked states of dimension M. It belongs to the class of preferred basis Levy matrices (PBLM) with heavy-tailed distribution of the off-diagonal matrix elements. Under certain conditions, the band of the marked states splits into minibands of non-ergodic delocalized states. We obtain an explicit form of the heavy-tailed distribution of PT times by solving cavity equations for the ensemble of down-folded Hamiltonians. We study numerically the PT subroutine as a part of quantum parallel tempering algorithm for a number of examples of binary optimization problems on fully connected graphs.

1   Introduction

Analog quantum enhanced search and optimization algorithms could, at the very least, provide a stop-gap solution for quantum applications prior to fault-tolerant universal quantum computation [34]. Typically the classical cost function in binary optimization problems is encoded in a n-qubit Hamiltonian $H_{cl} = \sum_z \mathcal{E}_z |z\rangle\langle z|$ diagonal in the computational basis $\{|z\rangle\}$. The energy landscape $\{\mathcal{E}_z\}$ of a typical hard optimization problem is characterized by a large number of spurious local minima. While close in energy, they can be separated

by large Hamming distances. This landscape gives rise to an interesting computational primitive: given an initial bit-string with sufficiently low energy, we are to produce other bit-strings within a certain narrow range of energies $\Delta\mathcal{E}_{\mathrm{cl}}$ in the vicinity of the initial state. In general, this is a hard computational task. For instance, given a solution of a SAT problem finding another solution of a similar quality is generally as hard as finding the initial solution. Inspired by previous work in analogue quantum computing [22, 16, 8], we propose the following quantum population transfer (PT) protocol: given an initial computational state $|z_0\rangle$ with classical energy $\mathcal{E}_{z_0}$, we evolve with the Hamiltonian

$$H = H_{\mathrm{cl}} + H_D, \quad H_D = -B_\perp \sum_{j=0}^{n} \sigma_j^x, \tag{1}$$

where H_D is the driver Hamiltonian, without fine-tuning the evolution time or the strength of the transverse field $B_\perp$. Finally we measure in the computational basis and check the energy $\mathcal{E}_z$ of the outcome $|z\rangle$ if $z \neq z_0$. Practically, analog implementation requires rapid (diabatic) ramp on/off of the transverse field at the beginning/end of the protocol.

The distribution $\rho(E_\gamma)$ of the eigenvalues of H, where $H|\psi_\gamma\rangle = E_\gamma|\psi_\gamma\rangle$, is typically well localized around the mean classical energy, with an exponentially decaying tail reaching towards the low energy states. The initial state $|z_0\rangle$ is located at a deep local minimum of the classical landscape $\{\mathcal{E}_z\}$, at the tail the distribution $\rho(E_\gamma)$. The non-diagonal matrix elements $-B_\perp$ give rise to hopping between states separated by one bit-flip. Matrices with diagonal disorder and hoping between neighbors correspond to the Anderson model introduced in the context of transport and localization in disordered media [5]. The transverse field $B_\perp$ couples the local minima via perturbation theory in a high order given by the Hamming distance between them. In this model (1), as well as in the original Anderson model, there exist bands of localized and extended states separated in energy by a so-called "mobility edge". It was demonstrated in Ref. [4] that localization is responsible for the failure of Quantum Annealing to find a solution of the constraint satisfaction problem (although, the detailed analysis of this effect is still lacking [24, 23]). Nonetheless, extended states could provide a mechanism for population transfer away from the initial state. In spin-systems with transverse field the existence of a so-called "mobility edge" at the tail of the distribution $\rho(E_\gamma)$, separating in the energy spectrum localized and delocalized eigenstates of H, has been recently studied in Refs. [37, 27, 31].

The population transfer corresponds to the formation of a "conduction band" that could be understood from the following arguments. We express the probability of a transition from $|z_0\rangle$ to $|z\rangle$,

$$P(t, z|z_0) = \left| \sum_\gamma \langle z|\psi_\gamma\rangle \langle \psi_\gamma|z_0\rangle e^{-iE_\gamma t} \right|^2, \tag{2}$$

in terms of the eigenstates $\{|\psi_\gamma\rangle\}$ and eigenvalues $\{E_\gamma\}$ of the system Hamiltonian H. In the delocalized phase the state $|z_0\rangle$ has a sizable overlap with a large set of eigenstates of size Ω with energies within some range $|E_\gamma - E_{\gamma'}| \sim \Delta E$. These are the eigenstates that dominate the sum in Eq. (2). The eigenstates in this set posses an important property extensively studied in the theory of transport in disordered systems [26, 3, 2]. They have largely overlapping supports over a support set $\mathscr{S}$ of bit-strings. This implies, from Eq. (2), that after a typical population transfer time $t_{\mathrm{PT}} \propto 1/\Delta E$ and for any initial state $z_0 \in \mathscr{S}$ the population is spread over the entire set $\mathscr{S}$, that is $P(t_{\mathrm{PT}}, z_1|z_0) \sim 1/|\mathscr{S}|$ for all $z_1 \in \mathscr{S}$. The conduction band is formed by the eigenstates within a spectrum width ΔE associated with

the bit-strings in $\mathscr{S}$. From the point of view of condensed matter physics, the eigenstates that overlap with $|z_0\rangle$ are non-ergodic. Nevertheless, they form a conduction miniband with energies below the quantum spin glass transition at the tail of $\rho(E_\gamma)$ [27, 2].

The formation of the conduction band explained above is the physical mechanism that we intend to exploit in the PT protocol to solve the computational primitive defined above. It is well established that simulating unitary time quantum dynamics in the delocalized phase to approximate $P(t_{\mathrm{PT}}, z_1|z_0)$ can not be done efficiently by known numerical techniques, such as quantum Monte-Carlo or tensor network methods, due to the coherent many-body nature of this transport phenomena. In addition to tunneling the transverse field $B_\perp$ gives rise to shifts in the classical energies $\mathcal{E}_{\mathrm{cl}}$ distributed over the width $\Delta\mathcal{E}_{\mathrm{cl}}$. This limits how narrow the target window of classical energies in the primitive can be.

More generally, the PT protocol can provide a useful primitive to explore energy landscapes on the way to lower energy states for optimization, reverse annealing [33] and quantum machine learning [10, 7]. The output of PT z can be used as an input of a classical optimization heuristic such as simulated annealing or parallel tempering in a "hybrid" optimization algorithm [32] where quantum and classical steps can be used sequentially to gain the complementary advantages of both [11]. A quantum advantage of the population transfer primitive would imply an advantage of such quantum parallel tempering over similar classical algorithms.

We propose a theoretical approach to analyze this problem with detailed analysis presented in [36]. Here we provide the results. Our approach exploits the existence of two relevant energy scales. The first scale is the width of the non-ergodic conduction miniband ΔE. The second scale is the typical change in classical energy $\mathcal{E}_{\mathrm{cl}}$ corresponding to one spin flip. Because of the large Hamming distances separating states in the support of the conduction band $\mathscr{S}$, the effective coupling elements that couple them correspond to high order in perturbation theory in $B_\perp$, and therefore $|\mathcal{E}_{\mathrm{cl}}| \gg \Delta E$. The dynamics within the miniband is described by the effective downfolded Hamiltonian

$$\mathscr{H} = \sum_{j=1}^{M} \varepsilon_j |j\rangle \langle j| + \sum_{j,k=1}^{M} V_{jk} |j\rangle \langle k| \ . \tag{3}$$

The sum is over a size M of the subset of computational basis states $|z_j\rangle$ sufficiently wide such that it contains the support set $\mathscr{S}$. The ε_j's are appropriately renormalized energies of the Hamiltonian H_{cl}. The non-diagonal matrix elements V_{jk} correspond to the sum over all elementary spin-flip processes that begin in state $|j\rangle$, proceed through virtual states separated by energies at least $\mathcal{E}$ from the miniband, and return back to the miniband only at the last step, at the state $|k\rangle$. We emphasize that in general V_{jk} takes into account all loops where the process returns back to the same virtual state without visiting the miniband.

In this paper we apply this analytical framework to "impurity band" model which demonstrates a quantum spin glass behavior yet allows analytical description of the quantum dynamics in the course of the PT protocol. In the second part of the paper we present numerical analysis of a set of more practical models defined by 2-local Hamiltonians. The impurity band model is defined by the Hamiltonian,

$$H = H_{\mathrm{cl}} - B_\perp \sum_{j=0}^{n} \sigma_j^x, \quad H_{\mathrm{cl}} = \sum_{j=1}^{M} \mathcal{E}(z_j) |z_j\rangle \langle z_j| \tag{4}$$

where the n-bit-strings $\{z_j\}_{j=1}^{M}$ are chosen uniformly at random from all bit-strings of length n, there are $M \gg 1$ marked states $|z_i\rangle$, with energies $\mathcal{E}_{z_j} = -n + \varepsilon_j$. The ε_j's are

independently distributed around 0 with a narrow width $W \ll 1$ to be discussed below. All other states have energy 0 and are separated by a large gap $\sim n$ from the very narrow band of marked states. The typical distance between marked states is $n/2$. If M is exponentially large in n the typical distance $d_{\min}$ to the nearest marked state is much smaller than $n/2$ but remains extensive $d_{\min} = \mathcal{O}(n)$. As such, each marked state $|z_j\rangle$ is a deep local minimum of $\mathcal{E}(z)$ coupled to other marked states via transverse field induced multiqubit tunneling with amplitude decreasing exponentially with Hamming distance d.

We obtain an explicit analytical form for the statistical properties of the PT dynamics in the above model (4) by deriving in the form of Eq. (3) an effective down-folded Hamiltonian in the energy strip associated with the PT [36]

$$\mathscr{H}_{ij} = \delta_{ij}\epsilon_j + (1 - \delta_{ij})\mathcal{V}_{ij}\sqrt{2}\sin\phi(d_{ij}) . \tag{5}$$

Here the diagonal elements ϵ_j are given by the marked state energies counted off from the center of the impurity band shifted due to the effect of the transverse field $\sim B_\perp^2$. Their PDF is assumed to be exponentially bounded with some width W.

Explicit analytical form of the off-diagonal elements is obtained using WKB approach. In Eq. (5) $\phi(d) \equiv \phi(E^{(0)}, d)$ is a WKB phase that describes the oscillation of the matrix elements with the Hamming distance. The tunneling amplitude $\mathcal{V}_{ij}$ equals

$$\mathcal{V}_{ij} \equiv V(d_{ij}), \quad V(d) = \sqrt{A(d/n, B_\perp)}\,\frac{n^{5/4}\,e^{-n\theta(B_\perp)}}{\sqrt{\binom{n}{d}}} , \tag{6}$$

where $i \neq j$ and the coefficient $A(\rho, B_\perp)$ is a smooth function of its arguments. The function $\theta(B_\perp)$ is given in [36]. Below we use its asymptotical form in the limit $B_\perp \gg 1$,

$$\theta \simeq \frac{1}{4B_\perp^2} + \frac{1}{24B_\perp^4} + \frac{1}{60B_\perp^6} + \cdots . \tag{7}$$

In this limit $\theta \ll 1$. We shall refer to $\mathscr{H}$ in (5) as the Impurity Band (IB) Hamiltonian.

The typical matrix element corresponds to tunneling at distance $n/2$ given by,

$$V_{\text{typ}} \sim n^2 2^{-n/2} e^{-n/(4B_\perp^2)}. \tag{8}$$

At the same time the typical distance from a marked state to its nearest neighbor is extensive $d_{\min} = \mathcal{O}(n)$ which corresponds to a matrix element, see Eq. (6), exponentially larger than the typical value V_{typ}. Therefore there is a hierarchy of off-diagonal matrix elements of $\mathscr{H}_{ij}$. The off diagonal matrix elements of random realizations of $\mathscr{H}$ are described by a heavy-tailed probability density function [14, 30]. Such random matrices are called Levi matrices.

The PDF of the rescaled squared amplitudes $w_{ij} = V^2(d_{ij})/V_{\text{typ}}^2$ can be obtained in the explicit form [36],

$$\text{PDF}(w) = \frac{1}{w^2\sqrt{\pi \log w}}, \quad w \in [1, \infty). \tag{9}$$

The particular form of scaling is the direct consequence of the fact that our problem has no "structure": the tunneling matrix elements depend only on Hamming distance and marked states are chosen at random.

The key difference of the ensemble of matrices $\mathscr{H}_{ij}$ from Levi matrices studied in the literature [38, 30, 29, 14] is that the dispersion, W, of the diagonal matrix elements is much larger than the typical magnitude of the off-diagonal elements V_{typ}. Therefore $\mathscr{H}_{ij}$ can be called preferred basis Levi matrices (PBLM).

We note that the existence of heavy tails in the PDF of the off-diagonal matrix elements of the down-folded Hamiltonian $\mathscr{H}$ is due to the infinite dimension of the Hilbert space of the original problem (1) for $n \to \infty$. This happens because the exponential decay of the matrix elements with the Hamming distance d is compensated by the exponential growth of the number of states at the distance d from a given state. We expect that this PBLM structure is a generic feature of the effective Hamiltonians for PT at the tail of the density of states in quantum spin glass problems.

The competition between the exponential decrease of the matrix element and the increase of the number of neighbors at distance d can result in eigenstates $|\psi_\beta\rangle$ of H associated with the impurity band becoming delocalized over a large subset of marked states $\mathscr{S}_\beta$ with size $1 \ll |\mathscr{S}_\beta| \propto M^\alpha$ and $0 < \alpha \leq 1$. For $\alpha = 0$ the eigenstate $|\psi_\beta\rangle$ is localized, for $\alpha = 1$ the eigenstate is delocalized in the entire space of marked states. For $0 < \alpha < 1$ the eigenstate can be considered "non-ergodic" and its support set $\mathscr{S}_\beta$ is sparse in the space of the marked states. The PBLM matrices support non-ergodic delocalized states when the width W is much bigger than the largest off-diagonal matrix element in a typical row of $\mathscr{H}_{ij}$ and much smaller than the largest off-diagonal element in a matrix

$$V_{\text{typ}}M^{1/2} \ll W \ll V_{\text{typ}}M . \tag{10}$$

For smaller dispersion $W \lesssim V_{\text{typ}}M^{1/2}$ the matrix eigenstates are ergodic while for $W \gtrsim V_{\text{typ}}M$ the eigenstates are localized. This non-ergodic regime is a distinct feature of the PBLM and is absent in Levi matrices. Such phase diagram resembles the one in the Rosenzweig-Porter (RP) model [26, 15]. The difference of RP from PBLM is that the statistics of the off-diagonal matrix elements in the RP ensemble are Gaussian [35] rather than polynomial (9). In this paper we focus on PT transfer within the non-ergodic delocalized phase, which is more likely to generalize to other models. We note that the localized phase does not support population transfer.

In the delocalized phase eigenstates with largely overlapping supports $\cap_\beta \mathscr{S}_\beta \approx \mathscr{S}(z_j)$ form narrow mini-bands. The mini-band width Γ may be interpreted as the inverse scrambling time and determines the width of the plateau in the Fourier-transform of the typical transition probability $\tilde{P}(\omega, z|z_j)$ [26].[2] In other words, the significant PT of $P(t, z|z_j)$ from the initial marked state $|z_j\rangle \in \mathscr{S}$ into other states of the same miniband $\mathscr{S}$ occurs over time $t_{\text{PT}} \sim 1/\Gamma$.

Because of the PBLM structure of the Hamiltonian $\mathscr{H}$ one can expect that the runtime of the PT protocol t_{PT} will have a heavy-tailed PDF whose form is of practical interest. It is closely related to the PDF of the miniband widths $\Gamma \sim 1/t_{\text{PT}}$. We obtained the PDF($\Gamma$) using the cavity method for random symmetric matrices [1, 14, 9, 38].

Previously cavity equations were solved only in their linearized form, i.e., near the localization transition. We were able to solve the fully nonlinear cavity equations in the delocalized non-ergodic phase [36]. We obtained boundaries of the non-ergodic phase analytically in terms of the ratio of W/V_{typ} and the form of the PDF $\mathscr{P}(\Gamma)$ inside the phase. It is given by the alpha-stable Levi distribution [19, 14] with the tail index 1, see Fig. 1

$$\mathscr{P}(\Sigma'') = \frac{1}{C} L_1^{1,1} \left(\frac{\Sigma'' - \Sigma''_{\text{typ}}}{C} \right) , \tag{11}$$

$$\Sigma''_{\text{typ}} = \mu_\Omega \Sigma''_*, \quad C = \sigma_\Omega \Sigma''_* . \tag{12}$$

---

[2] The same plateau width characterizes the frequency dependence of the eigenfunction overlap correlation coefficient $K(\omega) = M \sum_{j=1}^{M} \sum_{\beta,\beta'} |\langle j|\psi_\beta\rangle|^2 |\langle j|\psi_{\beta'}\rangle|^2 \delta(\omega - E_\beta + E_{\beta'})$ [26].

Here Σ_{typ}'' is a shift of the distribution and C its scale parameter (characteristic width) and we introduced a notation $\Sigma_*'' = \pi V_{\text{typ}}^2/(W/M)$.

$$\mu_\Omega \simeq \frac{1}{\sigma_\Omega} + \frac{2\sigma_\Omega(1 - \gamma_{\text{Euler}})}{\pi} \, . \tag{13}$$

$$\sigma_\Omega = \sqrt{\frac{\pi}{4\log\Omega}} \, . \tag{14}$$

Here Ω is the number of states in the miniband. This number $\Omega = (\pi M V_{\text{typ}}/W)^2$ is a square function of the ratio of the typical tunneling matrix element V_{typ} to the level separation W/M.

We introduce the scaling of the width of the distribution of ϵ_m with the matrix size M,

$$W = \lambda M^{\gamma/2} V_{\text{typ}} \, , \tag{15}$$

where γ is a real non-negative parameter that controls the scaling of the ratio of the typical diagonal to off-diagonal matrix element V_{typ} given in Eq. (8), and λ is an auxiliary constant of order one. With this scaling ansatz we get [36],

$$\Omega = \left(\frac{\pi}{\lambda}\right)^2 M^{2-\gamma} \, . \tag{16}$$

Using the above scaling ansatz (15) and the expressions for σ_Ω (14) and μ_Ω (13) we obtain,

$$\Sigma_{\text{typ}}'' \simeq \frac{2\pi^{1/2}}{\lambda} V_{\text{typ}} M^{1-\gamma/2}(\log\Omega)^{1/2} \, , \tag{17}$$

$$C \simeq \frac{\pi^{3/2}}{2\lambda} V_{\text{typ}} M^{1-\gamma/2}(\log\Omega)^{-1/2} \, . \tag{18}$$

The most probable value of the miniband width is $\Gamma_{\text{typ}} = V_{\text{typ}}(\pi\Omega\log\Omega/4)^{1/2}$, and its characteristic dispersion $\pi\Gamma_{\text{typ}}/(4\log\Omega)$. In a non-ergodic delocalized phase $M \gg \Omega \gg 1$ and the typical PT time $t_{\text{PT}} \sim 1/\Gamma_{\text{typ}}$ obeys the condition
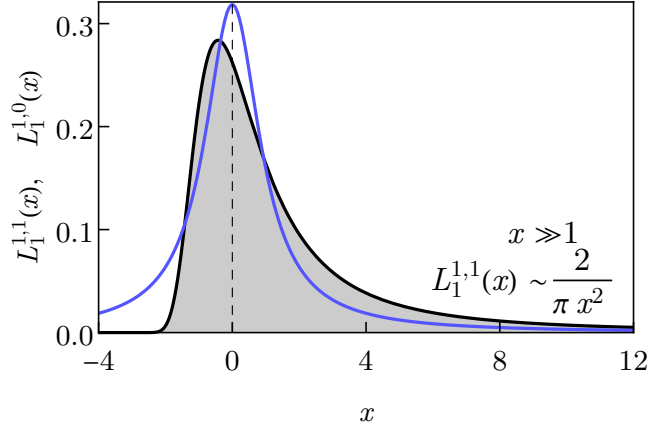
$$(M\log M)^{-1/2} \ll t_{\text{PT}}V_{\text{typ}} \sim (\Omega\log\Omega)^{-1/2} \ll 1 \, . \tag{19}$$

2 Complexity of the Population Transfer protocol

Starting at $t = 0$ from a marked state $|z_j\rangle$ the probability for the population to be transferred to other marked states is $1 - \psi^2(z_j, t)$. At the initial stage the survival probability $\psi^2(z_j, t)$ decays exponentially with the mean decay time $1/\Gamma_j = 1/(2\Sigma_j'')$.

The initial marked state $|z_j\rangle$ decays into the eigenstates $|\psi_\beta\rangle$ of the IB Hamiltonian $\mathscr{H}$ with typical energies E_β inside the narrow interval corresponding to the miniband associated with $|z_j\rangle$. It has width Σ_j'' and is centered around $\mathscr{H}_{jj} = \epsilon_j$. Typical classical energies ϵ of the bit-strings measured at the end of PT protocol will obey the probability distribution $\mathcal{P}(\epsilon - \epsilon_j - \Sigma_j')$ where Σ_j' is the self-energy shift of the marked state ϵ_j and $\mathcal{P}$ is the rescaled Cauchy distribution shown in Fig. 1 which reads [36],

$$\mathcal{P}(\Sigma') = \frac{1}{\pi} \frac{\Sigma_{\text{typ}}'}{\left(\Sigma_{\text{typ}}'\right)^2 + (\Sigma')^2}, \quad \Sigma_{\text{typ}}' = \Sigma_*''\sqrt{\frac{4\log M}{\pi}}. \tag{20}$$

Figure 1 Black solid line shows the plot of the Levi alpha-stable distribution $L_\alpha^{C,\beta}(x)$ [14] with tail index $\alpha = 1$, asymmetry parameter $\beta = 1$ and unit scale parameter $C = 1$. Inset shows asymptotic behavior of the distribution at large positive x. At $-x \gg 1$ the function decays steeply as a double exponential, $\log L_1^{1,1}(x) \propto -\exp(-\frac{\pi}{2}x)$. Blue line shows the Cauchy distribution $L_1^{1,0}(x) = \frac{1}{\pi(1+x^2)}$. We follow here the definition introduced in [14] and used in subsequent papers on Levi matrices in physics literature. In mathematical literature [39] a different definition is usually used, corresponding to $f(x;\alpha,\beta,C^{1/\alpha},0) = L_\alpha^{C,\beta}(x)$.

The success of PT protocol is to find a bit-string distinct from z_j at a time t with energy inside the window $\Delta\mathcal{E}_{cl}$ around ϵ_j. The PT success time therefore equals

$$t_{PT}^j = \frac{1}{2\Sigma_j'' p_{\Delta\mathcal{E}}}, \quad p_{\Delta\mathcal{E}} = \int_0^{\Delta\mathcal{E}_{cl}} \mathcal{P}\left(\epsilon - \Sigma_j' - \frac{\Delta\mathcal{E}_{cl}}{2}\right) d\epsilon .$$

Here $p_{\Delta\mathcal{E}}$ is the probability of detecting a bit-string inside the target window $\Delta\mathcal{E}_{cl}$ under the condition that initial state has decayed. Assume that the PT window is as wide as the typical miniband width, $\Delta\mathcal{E}_{cl} = \Sigma_{typ}''$. In this case $p_{\Delta\mathcal{E}}$ differs from 1 only by a constant factor that does not depend on M. Therefore after a sufficiently long time a solution, a bit-string inside the PT window, is detected with finite probability. Because the initial state $|z_j\rangle$ is picked at random the typical time to success of PT $t_{PT} \sim 1/\Sigma_{typ}''$ corresponds to the inverse typical width of the miniband. All of the states in a miniband are populated at (roughly) the same time t_{PT} because the transition rate to a subset of states on a distance d away from $|z_j\rangle$ depends on d very weakly. This is a result of a cancellation between the combinatorial number $\binom{n}{d}$ of states (and hence decay channels) at distance d from a given marked state and the dependence of the matrix element squared on d, see (6).

We characterize the PT by the relation between the typical success time of PT t_{PT} and the number of states Ω over which the population is spread during PT

$$t_{PT} \sim \frac{1}{V_{typ}\sqrt{\Omega\log\Omega}} \sim \left(\frac{2^n}{n\Omega\log\Omega}\right)^{1/2} e^{2\theta n}, \tag{21}$$

where we set $\Delta\mathcal{E}_{cl} \sim \Sigma_*''$ (see discussion above). The time t_G for the Grover algorithm for unstructured quantum search to find Ω items in a database of the size 2^n is $t_G \sim (2^n/\Omega)^{1/2}$. PT time t_{PT} scales worse than Grover time t_G by an additional exponential factor $e^{2\theta n} \simeq e^{\frac{n}{2B_\perp^2}}$ (7). At large transverse fields $1 \ll B_\perp = \mathcal{O}(n^0)$ the scaling exponent is small $2\theta \ll 1$.

3 Comparison with the analogue Grover search

Inspired by the Hamiltonian version of Grover algorithm proposed in [18] we consider the PT protocol in the IB model H_{cl} starting from the ground state of H_D which is a fully symmetric state $|S\rangle = 2^{-n/2} \sum_{j=1}^{n} |z\rangle$ in a computational basis. This protocol can be implemented by adjusting the value of transverse field $B_\perp \approx 1$ so that the ground state energy of the driver is set near the center of the IB. Then we can replace the full driver with the projector on its ground state, $H_D \to -nB_\perp |S\rangle \langle S|$. The quantum evolution is described by the Hamiltonian

$$H_{\text{G}} = -nB_\perp |S\rangle \langle S| + \sum_{j=1}^{M} \mathcal{E}(z_j) |z_j\rangle \langle z_j| \ , \tag{22}$$

with the initial condition $|\psi(0)\rangle = |S\rangle$. In the case where all impurity energies are equal to each other, $\{\mathcal{E}(z_j) = -n\}_{j=1}^{M}$, and $B_\perp = 1$ the Hamiltonian H_{G} is a generalization of the analog version of Grover search [18] for the case of M target states. The system performs Rabi oscillations between the initial state $|S\rangle$ and the state which is an equal superposition of all marked (solution) states. Time to solution is the half-period of the oscillations, the "Grover time" t_{G}

$$t_{\text{G}} = \frac{\pi}{2nB_\perp} \sqrt{\frac{2^n}{M}} \ . \tag{23}$$

Hamiltonian versions of Grover search with transverse field driver whose ground state were tuned at resonance with that of the solution state were considered in [17, 12].

We assume as before that marked state energies take distinct values $\mathcal{E}(z_j) = -n + \epsilon_j$ randomly distributed over some narrow range W. We investigate the effect of systematic error in the Grover diffusion operator [20]. In the Hamiltonian formulation [18] this corresponds to the deviation from unity of the parameter $B_\perp$ that controls the weight of the driver in (23). We will define the driver error ϵ_0 by,

$$B_\perp = 1 - \frac{\epsilon_0}{n} \ . \tag{24}$$

Assuming that $N \gg M$ one can instead of the state $|S\rangle$ consider the decay of the state $|0\rangle \equiv \frac{1}{\sqrt{N-M}} \sum_{j=M+1}^{N} |z_j\rangle$, where $j \in [1, M]$ corresponds to marked states. We use (24) and omit constant terms and small corrections $\mathcal{O}(M/N)$ in H_G. The non-zero matrix elements $H_G^{ij} = \langle i| H_G |j\rangle$ in this subspace $\mathscr{S}$ have the form

$$H_G^{jj} = \epsilon_j, \quad H_G^{j0} = -(1 - \delta_{j0})V, \quad V = n2^{-n/2} \ , \tag{25}$$

where $j \in [0, M]$ and $H_G^{j0} = H_G^{0j}$. On a time scale $t \ll 1/\delta\epsilon = M/W$ much smaller than the inverse spacing of the energies ϵ_j the quantum evolution with initial condition $|\psi(0)\rangle = |0\rangle$ corresponds to the decay of the discrete state with energy ϵ_0 into the continuum [28] with the finite spectral width W [25]. We assume that $\epsilon_0 \gg W$ while the spread of the marked state energies $W \lesssim V\sqrt{M}$, so that absent driver errors, PT time would follow a Grover-like scaling law $t \sim 1/(V\sqrt{M})$.

The state $|0\rangle$ is coupled non-resonantly to a continuum with narrow bandwidth. The expression for the population transfer to the marked states can be obtained from the time-dependent perturbation theory in the parameter ϵ_0/W

$$\sum_{m=1}^{M} |\psi_m(t)|^2 = \frac{2MV^2}{\epsilon_0^2} \left(1 - \cos(\epsilon_0 t) \frac{\sin(Wt/2)}{Wt/2} \right) \ .$$

Maximum transfer occurs at the time $t_0 = \pi/\epsilon_0$ with the total transferred probability $p_0 = 4MV^2/\epsilon_0^2$. Typical time $t_{\mathrm{PT}} \simeq t_0/p_0$ to achieve the successful population transfer to marked states involves repeating the experiment $1/p_0$ times

$$t_{\mathrm{PT}} = \frac{1}{\Gamma_0} \frac{\pi^2 \epsilon_0}{W} \; , \tag{26}$$

where $\Gamma_0 = 2\pi V^2/(W/M)$ and the first factor in r.h.s gives the typical transfer time in absence of driver errors. Errors increase the transfer time by a large factor ϵ_0/W.

For the maximum possible bandwidth W when nearly all states are populated, $W \sim \Gamma_0 \sim V\sqrt{M}$, the time of population transfer (26) is

$$t_{\mathrm{PT}} \sim t_{\mathrm{G}} \left(t_{\mathrm{G}} \epsilon_0 \right) \quad \left(\epsilon_0 \gg t_{\mathrm{G}}^{-1} \sim V\sqrt{M} \right) . \tag{27}$$

As expected, when the driver error exceeds inverse Grover time $1/t_{\mathrm{G}}$ the performance of analogue Grover algorithms (22) degrades relative to t_{G}. This is a direct consequence of the fact that the quantum evolution begins from fully symmetric state which is a ground state of the driver Hamiltonian whose energy is tuned at resonance with the marked states. In this case the transverse field Hamiltonian driver effectively corresponds to the projector (22). Because the ground state is not degenerate, the resonance region is exponentially narrow ($\sim 2^{-n/2}\sqrt{M}$). This results in the exponential sensitivity of the Grover algorithm performance to the value of driver weight. This critical behavior was studied in the work on quantum spatial search [13] for the case of one marked state.

In contrast, in the PT protocol considered in this paper there was no need to fine-tune the value of $B_\perp$ other than making it large, $B_\perp \gg 1$. This happened because the effective coupling between the marked states described by the down-folded Hamiltonian $\mathscr{H}$ (5) was not due to any one particular eigenstate of the driver (such as the state $|S\rangle$ for the Grover case). Instead this coupling was formed due to an exponentially large (in n) number of non-resonant, virtual transitions between the marked states and highly exited states of the transverse field Hamiltonian H_D. This resulted in a significant improvement in robustness for the proposed PT relative to the analogue Grover algorithm.

4   Numerical simulations of population transfer

We illustrate by numerical simulations the potential of the PT subroutine for hybrid quantum-classical search and optimization algorithms. We consider a model defined by the following 2-local Hamiltonian,

$$\mathcal{H} = \mathcal{H}_{\mathrm{cl}} + \Gamma \sum_{i,j}^{n} \left(\delta_{ij}(|h_i|+1)\sigma_i^x + (|J_{ij}|+1)\sigma_i^x \sigma_j^x \right), \;\; \mathcal{H}_{\mathrm{cl}} = \sum_{i=1}^{n} h_i \sigma_i^z + \sum_{i,j}^{n} J_{ij}\sigma_i^z \sigma_j^z, \;\; (28)$$

where δ_{ij} is the Kronecker symbol, $J_{ij} \in [-1,1]$ and $h_i \in [-1,1]$ are uniformly distributed random numbers with finite 6-bit precision. A subset of $n/2$ bonds (i,j) are chosen to be dimers, non-overlapping pairs of spins with strong ferromagnetic interactions $J_{ij} = -4$. Note that we are using "matched" driver whose strength scales with the total longitudinal field acting on a given qubit.

We simulate of $n = 25$ qubit system. A single bit flip steepest descent (SD) algorithm starting from all possible bistrings identifies all local minima of a given realization of the model $\mathcal{H}_{\mathrm{cl}}$ in Eq. (28). Optimization of simulated annealing parameters for this type of
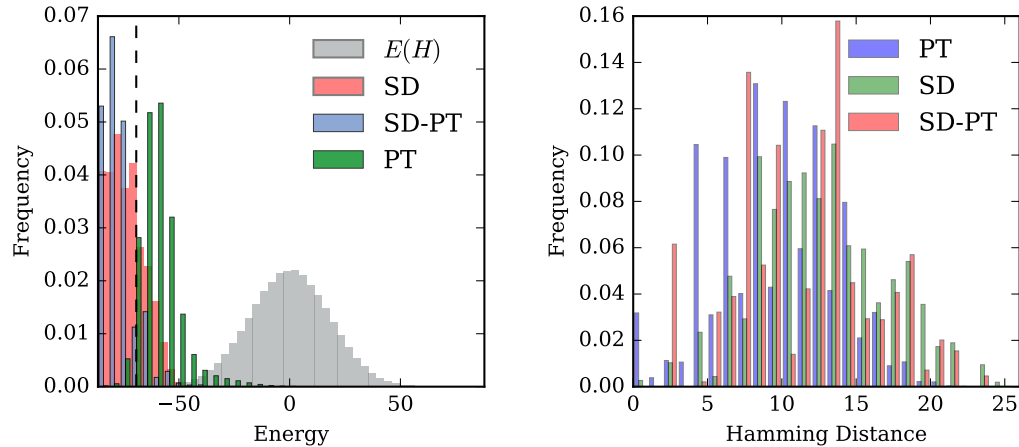
Figure 2 *Left panel* shows a histogram of normalized weights of classical energies in $H_{cl}^{(2)}$. $E(H)$ is the density of states in the original Hamiltonian spectrum. The classical steepest descent (SD) distribution shows the probability of ending up in a local minimum following a steepest descent run performed greedily by single spin flips. PT histogram shows the weight of classical energies in the output wave function following a population transfer (PT) run relative to the initial state energy (dotted black line). The SD-PT run is the distribution of local minimum starting from a state measured from the PT state. We note that the SD and SD-PT distributions are plotted as adjacent histograms to increase visibility of the data, but depicted bins are actually overlapping and not alternating. *Right panel* shows the distribution of Hamming distances from the initial state for states that fall within 1 standard deviation of the energy from the peak of the population transfer (PT) distribution for several methods.

instances suggested that SD (low temperature limit) is near optimal and therefore can serve as a proxy for hardness of finding a given bitstring, see histogram in the left panel of Fig. 2. Starting from a low-energy local minimum we perform a population transfer, Hamiltonian evolution with a fixed strength of the driver strength $\Gamma = 0.2$ for sufficiently long time. The quantum evolution is simulated using Trotter decomposition with 300 steps. The evolution time is chosen sufficiently long for the population transfer to approximately saturate. The histogram of weights of classical energies in the output wave function, Fig. 2, shows significant weight remaining in the low energy region of the spectrum in the vicinity of the initial bitstring energy. The PT output wave function has support on bitstrings separated from the initial state by large Hamming distances, see right panel in Fig. 2. Moreover, repeated sampling from the PT output produces bitstrings separated by large Hamming distances from each other, see left panel in Fig. 3. Bitstrings sampled from the PT output wave function can be used as starting states for SD which finds some of the low energy minima with higher probability than SD initialized with uniformly random bitstrings, see right panel in Fig. 3. Therefore PT protocol provides a quantum coherent pathway between low energy states that is complementary to SD and simulated annealing and therefore PT could potentially serve as useful subroutine in hybrid algorithms such as quantum parallel tempering.
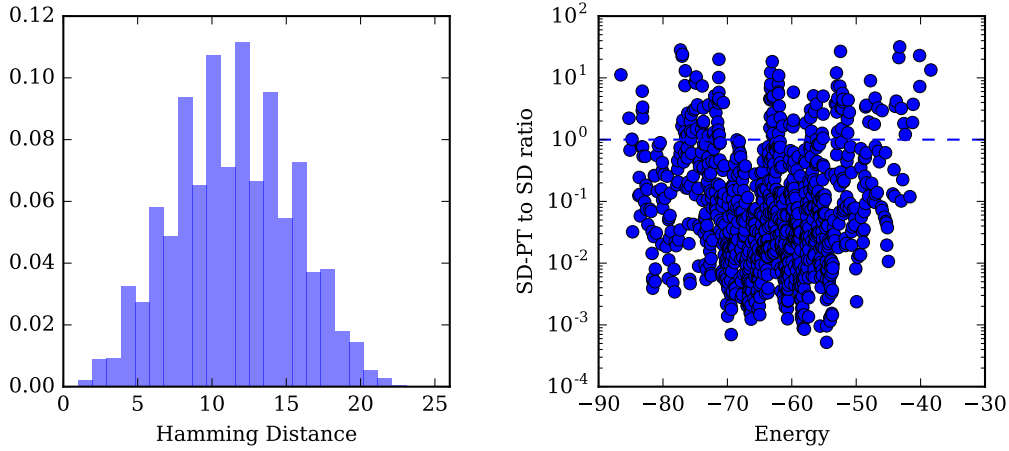
■ **Figure 3** *Left panel* shows the histogram of Hamming distance between all pairs of states within 1 standard deviation of the energy from the peak of the population transfer (PT) distribution. Hamming distances are weighted by their joint probability within the PT distribution. The 0 hamming distance self contributions are excluded. The even-odd pattern that is observed results from the dimer structure of the Hamiltonian and the corresponding match driver. *Right panel* plots every local minimum with respect to single spin flips in the spectrum, which is concentrated to the low energy sector of the total Hamiltonian. For each local minimum we depict the ratio of the probability of ending up in that local minimum when starting from the PT distribution against a uniformly random initial state. We see that a significant fraction of states are enriched in the PT case, including the global minimum.

5   Conclusion

We analyze the computational role of coherent multiqubit tunneling that gives rise to bands of nonergodic delocalized quantum states as a coherent pathway for population transfer (PT) between computational states with close energies. In this regime PT cannot be efficiently simulated by QMC.

We solve this problem using the following quantum population transfer (PT) protocol: prepare the system in a computational state $|z_j\rangle$ with classical energy $\mathcal{E}(z_j)$, then evolve it with the transverse-field quantum spin Hamiltonian. Classical energies $\mathcal{E}(z)$ are encoded in the problem Hamiltonian diagonal in the basis of states $|z\rangle$. The key difference between PT protocol and QA [22, 16, 8] or analogue quantum search Hamiltonians [18, 13] is that the transverse field is kept constant throughout the algorithm and is not fine-tuned to any particular value. At the final moment of PT we projectively measure in the computational basis and check if the outcome z is a "solution", i.e., $z \neq z_j$, and the energy $\mathcal{E}(z)$ is inside the window $\Delta\mathcal{E}_{\mathrm{cl}}$.

In this paper we analyzed PT dynamics in Impurity Band (IB) model with a "bimodal" energy function: $\mathcal{E}(z) = 0$ for all states except for M "marked" states $|z_j\rangle$ picked at random with energies forming a narrow band of the width W separated by a large gap $\mathcal{O}(n)$ from the rest of the states. This landscape is similar to that in analogue Grover search [18, 17] with multiple target states and a distribution of oracle values for the targets. The best known classical algorithm for finding another marked state has cost $O(2^n/M)$.
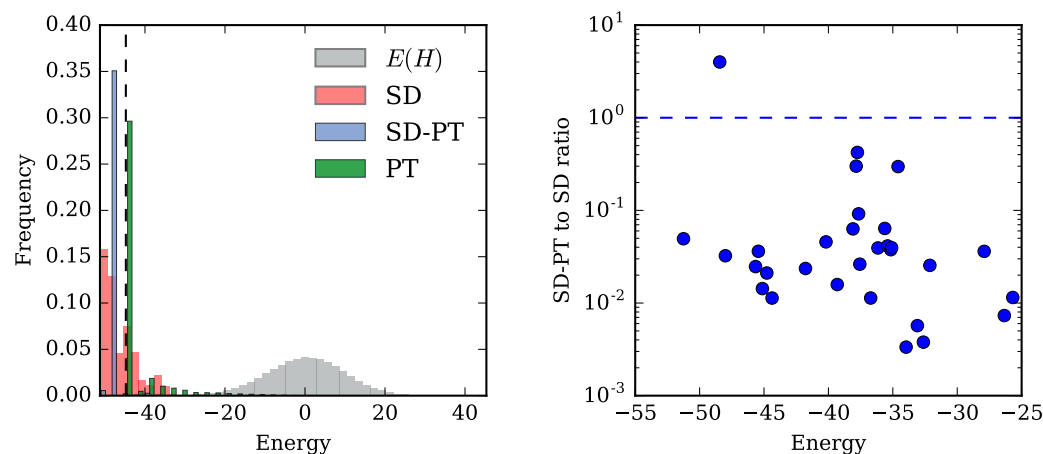
The PT dynamics is described by the down-folded $M \times M$ Hamiltonian $\mathcal{H}$ that is dense in the space of the marked states $|z_j\rangle$. The distribution of matrix elements $\mathcal{H}_{ij}$ has a heavy tail decaying as a cubic power for $V(d) \gg V_{\text{typ}}$. This is a remarkable result of the competition between the very steep decay of the off-diagonal tunneling matrix element with the Hamming distance d, and the steep increase in the number of marked states $M_d \propto \binom{n}{d}$ at distance d. We emphasize that such polynomial tail in the distribution of matrix elements is only possible either in infinite dimension or in presence of long-range interactions (e.g, dipolar glass).

The dispersion of the diagonal elements $\mathcal{H}_{jj} = \mathcal{E}(z_j)$ is expected to be large, $W \sim V_{\text{typ}} M^{\gamma/2} \gg V_{\text{typ}}$ with $\gamma \in [1,2]$. In the range $1 < \gamma < 2$ there exist minibands of non-ergodic delocalized eigenstates of $\mathcal{H}$. Their width is proportional to $1/t_{\text{PT}} \ll W$. Each miniband associated with a support set $\mathscr{S}$ over the marked states.

The distribution of miniband widths Γ obeys alpha-stable Levi law with tail index 1. The typical value of Γ and its characteristic variance exceeds the typical matrix element of $\mathcal{H}$ by a factor $\Omega^{1/2}$ where $\Omega = (M V_{\text{typ}}/W)^2$ is a size of the support set in a typical miniband.

We demonstrate that quantum PT finds another state within a target window of energies Ω in time $t_{\text{PT}} \propto 2^{n/2} \Omega^{-1/2} \exp(n/(2B_\perp^2))$. The scaling exponent of t_{PT} with n differs from that in Grover's algorithm by a factor $\propto B_\perp^{-2}$, which can be made small with large transverse fields $n \gg B_\perp^2 \gg 1$.

Crucial distinctions between this case and the Hamiltonian in the analogue version of Grover's algorithm [18] for the case of multiple target states are the non-integrability of our model, and the delocalized nature of the eigenstates within the energy band W. Furthermore, analogue Grover's algorithm for multiple targets is exponentially sensitive in n to the weight of the driver Hamiltonian, and cannot be initialized with a computational basis state.

The quantum spin Hamiltonian in (1) can be described using n-body, infinite-range interactions. However, it shares key properties with the infinite range spin-glass models involving only p-body interactions at finite values of p >2 that could be implemented on a quantum computer with polynomial in n resources. The evidence of non-ergodic extended states were recently uncover numerically in the low-energy part of the spectrum of the quantum transverse field p-spin model [6].

Similar to the model (1) the low-energy part of the spectrum of transverse field p-spin model is characterized by the proliferation of statistically independent deep local minima separated by large, O(n), number of spin flips. Model of this type are characterized by a RSB-1 type of spin glass transition and were studied in the context of quantum annealing [21]. They represent perhaps the next step to study PT algorithms.

Finally, we analyzed numerically the PT initiated at a low energy local minimum of a 2-local spin glass model and observed that sampling PT output together with subsequent application of classical steepest descent allows exploring the energy landscape in a way complimentary to steepest descent and simulated annealing. This suggest possible use of PT as a subroutine for hybrid quantum-classical algorithms for search and optimization such as quantum paralleled tempering.

References

1   Ragi Abou-Chacra, DJ Thouless, and PW Anderson. A selfconsistent theory of localization. *Journal of Physics C: Solid State Physics*, 6(10):1734, 1973.

2   BL Altshuler, E Cuevas, LB Ioffe, and VE Kravtsov. Nonergodic phases in strongly disordered random regular graphs. *Physical Review Letters*, 117(15):156601, 2016.

3   BL Altshuler, LB Ioffe, and VE Kravtsov. Multifractal states in self-consistent theory of localization: analytical solution. *arXiv preprint arXiv:1610.00758*, 2016.

4   Boris Altshuler, Hari Krovi, and Jérémie Roland. Anderson localization makes adiabatic quantum optimization fail. *Proceedings of the National Academy of Sciences*, 107(28):12446–12450, 2010.

5   Philip W Anderson. Absence of diffusion in certain random lattices. *Physical review*, 109(5):1492, 1958.

6   C. L. Baldwin, C. R. Laumann, A. Pal, and A. Scardicchio. Clustering of nonergodic eigenstates in quantum spin glasses. *Phys. Rev. Lett.*, 118:127201, Mar 2017. `doi:10.1103/PhysRevLett.118.127201`.

7   Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195, 2017.

8   J Brooke, D Bitko, G Aeppli, et al. Quantum annealing of a disordered magnet. *Science*, 284(5415):779–781, 1999.

9   Zdzisław Burda, Jerzy Jurkiewicz, Maciej A Nowak, Gabor Papp, and Ismail Zahed. Free random lévy and wigner-lévy matrices. *Physical Review E*, 75(5):051126, 2007.

10   Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602, 2017.

11   Nicholas Chancellor. Modernizing quantum annealing using local searches. *New Journal of Physics*, 19(2):023024, 2017.

12   Andrew M Childs, Enrico Deotto, Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Andrew J Landahl. Quantum search by measurement. *Physical Review A*, 66(3):032314, 2002.

13   Andrew M Childs and Jeffrey Goldstone. Spatial search by quantum walk. *Physical Review A*, 70(2):022314, 2004.

14   Pierre Cizeau and Jean-Philippe Bouchaud. Theory of lévy matrices. *Physical Review E*, 50(3):1810, 1994.

15   Davide Facoetti, Pierpaolo Vivo, and Giulio Biroli. From non-ergodic eigenvectors to local resolvent statistics and back: A random matrix perspective. *EPL (Europhysics Letters)*, 115(4):47003, 2016.

16   Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A quantum adiabatic evolution algorithm applied to random instances of an np-complete problem. *Science*, 292(5516):472–475, 2001.

17   Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum computation by adiabatic evolution. *arXiv preprint quant-ph/0001106*, 2000.

18   Edward Farhi and Sam Gutmann. Analog analogue of a digital quantum computation. *Physical Review A*, 57(4):2403, 1998.

19   BV Gnedenko and AN Kolmogorov. Limit distributions for sums of independent. *Am. J. Math.*, 105, 1954.

20   Lov K Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79(2):325, 1997.

21   Thomas Jörg, Florent Krzakala, Jorge Kurchan, and A. C. Maggs. Simple glass models and their quantum annealing. *Phys. Rev. Lett.*, 101:147204, Oct 2008. `doi:10.1103/PhysRevLett.101.147204`.

22   Tadashi Kadowaki and Hidetoshi Nishimori. Quantum annealing in the transverse ising model. *Physical Review E*, 58(5):5355, 1998.

23   Sergey Knysh. Zero-temperature quantum annealing bottlenecks in the spin-glass phase. *Nature communications*, 7:12370, 2016.

24   Sergey Knysh and Vadim Smelyanskiy. On the relevance of avoided crossings away from quantum critical point to the complexity of quantum adiabatic algorithm. *arXiv preprint arXiv:1005.3011*, 2010.

25   Eugene Kogan. On the analytic structure of green's function for the fano-anderson model. *arXiv preprint quant-ph/0611043*, 2006.

26   V E Kravtsov, I M Khaymovich, E Cuevas, and M Amini. A random matrix model with localization and ergodic transitions. *New Journal of Physics*, 17(12):122002, 2015. URL: `http://stacks.iop.org/1367-2630/17/i=12/a=122002`.

27   Christopher R Laumann, A Pal, and A Scardicchio. Many-body mobility edge in a mean-field quantum spin glass. *Physical Review Letters*, 113(20):200405, 2014.

28   Gerald D Mahan. *Many-particle physics*. Springer Science & Business Media, 2013.

29   Fernando Lucas Metz, Izaak Neri, and Désiré Bollé. Localization transition in symmetric random matrices. *Physical Review E*, 82(3):031135, 2010.

30   Cécile Monthus. Localization transition in random lévy matrices: multifractality of eigenvectors in the localized phase and at criticality. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(9):093304, 2016.

31   Gianni Mossi and Antonello Scardicchio. Many body localization transition in quantum spin glasses on the bethe lattice. *arXiv preprint arXiv:1703.03678*, 2017.

32   Hartmut Neven. Enhancing simulated annealing with quantum annealing, December 2015. US Patent Application Publication, PCT/US2016/068400. URL: `https://patents.google.com/patent/WO2017117016A1`.

33   M. Ohkuwa, H. Nishimori, and D. A. Lidar. Reverse annealing for the fully connected *p*-spin model. *ArXiv e-prints*, 2018. `arXiv:1806.02542`.

34   J. Preskill. Quantum Computing in the NISQ era and beyond. *ArXiv e-prints*, 2018. `arXiv:1801.00862`.

35   Norbert Rosenzweig and Charles E Porter. " repulsion of energy levels" in complex atomic spectra. *Physical Review*, 120(5):1698, 1960.

36   V. N. Smelyanskiy, K. Kechedzhi, S. Boixo, S. V. Isakov, H. Neven, and B. Altshuler. Non-ergodic delocalized states for efficient population transfer within a narrow band of the energy landscape. *arXiv preprint arXiv:1802.09542*, 2018. `arXiv:1802.09542`.

37   Vadim N Smelyanskiy, Udo v Toussaint, and Dogan A Timucin. Dynamics of quantum adiabatic evolution algorithm for number partitioning. *arXiv preprint quant-ph/0202155*, 2002.

38   Elena Tarquini, Giulio Biroli, and Marco Tarzia. Level statistics and localization transitions of levy matrices. *Physical Review Letters*, 116(1):010601, 2016.

39   Johannes Voit. *The statistical mechanics of financial markets.* Springer Science & Business Media, 2013.