QUANTUM TUNNELING IN THE LANDSCAPE OF OPTIMIZATION

ARNAB DAS*

This is a non-technical review, intended for a broad audience from diverse fields of science. Here we discuss searches as solutions to optimization problems, and how quantum mechanics can possibly help in it. To maintain unrestricted search over the whole space of interest, the search dynamics must be ergodic. This is often not the case when the searcher is "classical", i.e., behaving like a classical system moving through a rugged energy landscape. This has remained a major hindrance for versatile heuristics like simulated thermal annealing in finding a good (not to talk about the best) solution in presence of high energy/cost barriers. The stagnated situation was stirred fundamentally by the idea of quantum tunnelling, or, to be more precise, it's dramatic role proposed in the context of spin glass[1], which eventually helped forming the rationale behind one of the most promising form of quantum computation of the present day, namely, quantum annealing. The potential of the idea, however, is not limited to the strict framework of its most popular version, the adiabatic quantum annealing, but can possibly extend more effectively to other, less restrictive forms of quantum search heuristics that could be designed.

I. Optimization by Searching

Optimization is the key to the survival in the real world with limited time and resources. One has to do a complex bargains keeping several factors in mind, in order to get the best out of a situation. For example, a delivery person might have to visit several points in a city within a given time spending minimum fuel, or a circuit has to be completed using minimum wire length connecting millions of elements on a chip. All these are examples of optimization problems. In some cases (mostly of academic interest) the structure of the problem is understood well enough to be used for designing exact and efficient algorithms for solving it. Here, by efficient we loosely mean an algorithm whose time and resource requirement is upper-bounded by a polynomial in the system-size. In most cases of practical importance, however, such algorithms cannot be constructed, and one has to resort to heuristics. These are usually approximate algorithms based on generic strategies designed using knowledge about some general characteristics of the problem in questions. Hope is to obtain a good approximate solution when applied to a specific instance of the problem. Of course, these algorithms cannot guarantee the best solution, but are considered to be most useful because of their satisfactory performance over a vast variety of problems, since the strategies are quite generic.

One such heuristic is simply to search the configuration space. Suppose our optimization problem can be cast as the problem of finding the global minimum of a complicated (real valued) cost/energy function (we call it so)

$$H_c = H_c\left(\sigma_1^z, \sigma_2^z, ..., \sigma_N^z\right), \tag{1}$$

where σ_i^z are classical variables (the idea behind the notations will be clearer later). The task is to find the configuration $\left\{\sigma_i^z\right\}$; i.e., a set of values for the variables for which the value of H_c is the minimum. A search heuristic consists of a strategy to search the entire space of configurations in a systematic manner and come up with a configuration for which the value of H_c is considerably low (if not the least). As is apparent, the size/dimension $\mathcal{D}$ of the configuration space grows exponentially with N : $\mathcal{D}$

* School of Physical Sciences, Indian Association for the Cultivation of Science, 2A & 2B Raja S. C. Mullick Road, Kolkata 700032, India

Email : tpad5@iacs.res.in

$\sim p^N$; where p is the number of values a variable σ_i^z can take. This means, a random search will be futile for any appreciable N; and a search heuristic will be meaningful only if it has a strategy to select from a very small subset of the configurations and search within that. It is crucial to understand that a systematic heuristic, doing any better than a random search, has to exploit the structure of H_c. In a completely unstructured (with respect to H_c) space of configurations, no search heuristic can do better than a random search on an average. In the following section we give an elementary illustration in terms physical examples what we mean by structures helping a search.

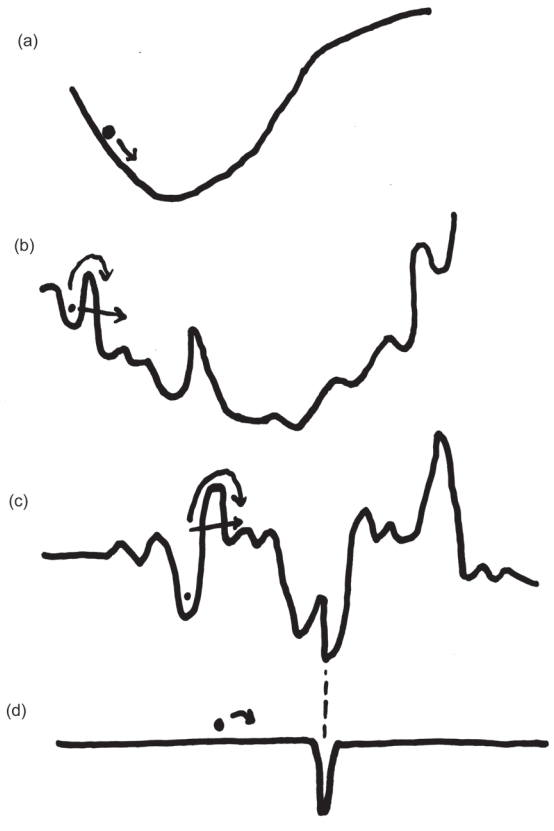

Fig. 1. Cartoons of different kinds of search landscapes. (a) shows the simplest possible structured scenario - a particle rolling along the smooth concave side of a bowl to reach the optimal point (the bottom/minimum) following the local potential gradient from anywhere. (b) illustrates the scenario where the local gradients might not be sufficient to reach the ground state, as it might lead to local minima. However, there is an overall gradient towards the global minimum, and if the occasional local barriers are allowed to be scaled/tunnelled through stochastically, then the overall structure of the landscape may lead to the global minimum. (a-b) constitute examples of \structured" search spaces. (c) illustrates a case where there is no global gradient or hierarchy, and there is basically no useful correlation between the depths of the minima and their locations. In such cases, no search algorithm can do any better than random searches. In fact, physically motivated search dynamics can do even worse (in finding the true global minimum), since they would spend a lot of extra time in narrow local minimum compared to random searches. (d) is the example of the extreme case of unstructured search space - a completely at landscape with a single potential minimum - entries are absolutely uncorrelated, and the global minimum can occur just anywhere on the landscape. (c-d) are examples of unstructured search space. Actually any problem of the form (c) can be reduced to (d).

II. Searching the Structured and Unstructured Landscape

As mentioned in the previous section, an effective search heuristics must exploit the structure of the cost/ energy function itself. To illustrate this point, let us consider a potential profile (Fig. 1a) with single minimum (like a smooth bowl in presence of gravity). In order to find the global minimum of such a problem, all one needs to do is to follow the local gradient, starting from any point on the landscape (just roll down a marble from any point along the bowl surface). The structure of the potential will guide it to the lowest point without any distraction.

The situation can however be a bit worse than this. The surface of the "bowl" might not be smooth locally, and can be punctuated with occasional local minima, guarded by potential barriers (Fig. 1b). In that case, following the local gradient would not be sufficient - one needs to jump/tunnel out of these shallow local minima in order to get into deeper minima. One can resort to (artificial) thermal/quantum dynamics to get out of these minima.

The scenario can be even worse. The landscape can be globally almost flat in the sense that the potential has no global structure that leads to the global minimum (or preferentially to the deeper minima - the position and depth/ heights of the local minima/potential barriers are mostly uncorrelated, or, at best have correlations which does leading to the global minimum) is observed (Fig. 1c). Consider an overall flat road, scarred with deep potholes and lofty speed breakers. In such a case, any local search heuristic is as poor as a random search. This is not to say that quantum tunneling based search does not have any advantage over the classical ones - advantage is indeed there, but that is upper bounded by the quantum advantage one gets for searching a completely unstructured database[2].

Finally there can be the completely unstructured configuration space, which means the position and the depth of the global minimum has absolutely no effect on the potential even at its closest neighbourhood - an example is a completely flat potential energy landscape with a single potential minimum (so called golf-course potential; Fig. 1d). In order to avoid additional complexities and delays involved in searching a rugged yet unstructured landscape, it is easy to see that one can actually cast the problem of this sort to the golf- course form. In that case, the optimal quantum advantage is the square-root speed up, i.e., in the

quantum case, resource/time ~ $\sqrt{\mathcal{D}}$, as opposed to the classical case, where resource/time ~ $\mathcal{D}$. This quantum speedup is optimal in the sense that no unitary quantum algorithm can provide any better scaling with N.

As a small detour, we note that it is possible to map any problem of type (c) to a problem of type (d). This can be done by disregarding the landscape of the problem Hamiltonian altogether, and instead, marking the target configuration by a property which we want to optimize (e.g., we can ask if the configuration has a certain energy, i.e., value for H_c), and assign a negative potential energy to a state only if it meets the search criteria, or else assign zero potential energy to it. With this assignment, and a ergodic kinetic energy term, the problem is like a search on a glof-course potential. It is easy to see that this can be done for all kinds of problem for which the value of H_c can be calculated efficiently given its argument (a configuration). This breaks the original question to a set of subquestions, e.g., in each go we can ask if there is a configuration whose energy is below a given value. All such configurations can be assigned a negative potential energy, while the rests with zero potential energy. This will be just the analog version of Grover's search algorithm, possibly with multiple solutions. Clearly, this iteration can be used to solve the problem efficiently to any finite accuracy in the $N \to \infty$ limit, if each question is answered efficiently.

III. Physical Cost Functions: Glasses - Structured or Unstructured?

Physical example of cost functions are Hamiltonians, where σ_i^z are interacting physical degrees of freedom. One simply stated example is an Ising Hamiltonian

$$H_{Ising} = \sum_{i>j} J_{ij}\sigma_i^z\sigma_j^z \ , \qquad (2)$$

where J_{ij} are couplings varying randomly in magnitude and sign, drawn from some distribution and σ_i^z are Ising spins taking values ±1 each. Physical systems corresponding to these Hamiltonians are called spin-glasses in general. The ground state(s) of such a simple looking Hamiltonian can be extremely complex, and so can be its energy landscape. In particular, the presence of disorder (random variation of the size of the couplings) and frustration (-ve signs of the couplings and the graph connectivity making it impossible to satisfy all bonds simultaneously) makes is rugged, plagued with several local minima and high energy barriers surrounding those[3]. Frustration often leads to exponential number of local minima which are quite stable. This leads to the scenario closer to that caricatured in (c). It is possible that glass like problems do have some kind of a structure, which could help laying out a general principle for tailoring efficient algorithms for solving them, but those are still not known to us. On the contrary, some of them are considered NP-complete, and even NP-hard (see, e.g.[4] for a discussion). Hence they are usually difficult to solve with local search algorithms.

IV. Searching with Quantum Tunnelling: From Quantum Spin Glass to Quantum Annealing

In order to carry out the search, the Hamiltonian H_c (in this case, H_{Ising}) is to be endowed with a physical dynamics. The dynamics is to be steered in such a manner, that at the end of it, the system is highly likely to end up at the global minimum of H_c (or, at the least, a satisfactorily deep minimum).

A. Classical Annealing : One very successful and largely used strategy is the thermal simulated annealing[5]. The essential idea is to introduce an artificial thermal probability of getting from a configuration C_i to a configuration C_f : The probability of the transition is proportional to the Boltzmann factor

$$e^{-\beta\left(H_c(C_f)-H_c(C_i)\right)}$$

for $H_c(C_f) > H_c(C_i)$; and unity otherwise. Initially the inverse temperature β is kept very low, so that the system just indiscriminately jumps from one configuration to the other without paying much attention to the landscape structure. Then the temperature is reduced systematically and slowly enough so that the system starts getting to see the landscape, and prefers lower energy configurations. If the cooling is slow enough, then the system gets enough time to sample a substantial section of the configuration space, and can eventually settle into a deep enough minimum.

However, there are two major problems with this approach - first is the lack of ergodicity, and second is the lack of typicality. We elaborate on these a bit in the following.

Breaking of Ergodicity: As mentioned in the previous section, the landscape can have several local minima surrounded by energy/cost barriers that scales with system-size N. In that case, a thermal jump to come out of one such local minimum can take exponentially long time in N. Hence in the thermodynamic limit, the entire configuration will be broken down into numerous (typically

exponential in N for frustrated systems) disjoint sectors/ basins, such that dynamics initiated in one of those cannot reach the other in any finite time. This is called breaking of ergodicity. Clearly, if this is indeed the case, the only way to go around it is to try out exponentially many initial configurations to find the right basin, where the global minimum lies. The scaling of the algorithm is then no better than a random search.

Lack of typicality: The other problem is, a classical system can remain in any one of the exponentially many possible configurations at a time. Hence within any finite time, the fraction of space searched by the system has to be exponentially small in N. This is also the case while simulating a simple system (e.g. a uniform ferromagnet) using, say, Monte Carlo. What comes to our save is the "typicality" – for simple systems, we need not actually sample the whole configuration space, but only a handful of configurations do the job for the statistical purpose. This is because if the energy is given, the configuration to configuration fluctuation of local observables are negligible for large enough N. This is true also for the quantum case - for a generic interacting Hamiltonian, the eigenstate to eigenstate fluctuation of local observable is a smooth function of the energy difference between the eigenstates. This is called the Eigenstate Thermalization Hypothesis[6]. For systems with disorder, this breaks down. Two energetically close configurations can have very different configurations, and the reverse. Since the simulations of the annealing are based on energetics, the above implies, it is not sufficient to sample a small part of the configuration space and consider that to be a good representative of the whole space.

B. Quantum Tunnelling : Before the era of quantum algorithms, in 1989, Chakrabarti and his collaborators[1], proposed a scenario which implied that quantum tunnelling in many-body systems, even in presence of disorders, can potentially combat the above two problems. The system studied was the Sherrington-Kirkpatrick model in the transverse field given by the Hamiltonian

$$H = -\sum_{i,j} J_{ij}\sigma_i^z\sigma_j^z - \Gamma\sum_i \sigma_i^x, \qquad (3)$$

where J_{ij} are drawn randomly from a Gaussian distribution of zero mean. The system is a paradigmatic example of a spin glass. Based on numerical simulations of the model within a finite size, it was claimed that the macroscopic quantum tunnelling induced by the transverse field can make the system ergodic for a finite range of parameters, where the system would be strongly non-ergodic in the absence of the quantum fluctuations. While the quantitative claims made in the said paper is still a matter of debate, the physical picture it put forth had all the ingredients to inspire and boost the idea of using quantum fluctuations in annealing algorithms - either in addition to or in lieu of thermal fluctuations. The proposal appeared rather daring at that time, yet the possibility it indicated to, continues to bear promises even these days. First, quantum tunnelling was proposed, to restore ergodicity in a classically glassy system under consideration. The rationale given was, in quantum mechanics, a system can tunnel through even a lofty barrier if it is narrow enough, and the long-range spin-glasses have many such barriers. The tunneling would also allow (as implied by the proposal) an extended wave-function that can "feel" the entire landscape *simultaneously*, apparently evading the problem of limited sampling of the cofiguration space (which might lack typicality in the sense described above) within finite sampling time.

However, it was no earlier than in 1998, that the above physical ideas got their application in a well defined formulation of quantum optimization. In the said year, Kadowaki and Nishimori[7] proposed the idea of simulated quantum annealing (this was the work where the term "quantum annealing" was first coined with its present day meaning). They demonstrated that annealing by tuning quantum fluctuations can be quantitatively much faster than purely thermal annealing. They performed simulated annealing of a few variants of Ising model $H_c(\sigma_i^z)$ (including the Sherrington-Kirkpatrick model studied in[1]) by tuning a transverse field $\left(H_x(t) = -\Gamma(t)\sum_i \sigma_i^x\right)$ from a very high initial value $\left(\left|H_x(0)\right| \gg \left|H_c\right|\right)$ to zero slowly. In almost all the cases, a substantial quantitative improvement in the probability of reaching the ground state was observed compared to purely thermal annealing for the same annealing time. Soon after, a concrete formulation of quantum annealing at zero temperature, based on the adiabatic theorem of quantum mechanics was proposed by Farhi et. al.[8]. According to their scheme, to anneal a classical Hamiltonian H_c (e.g., a classical Ising Hamailtonian) by tuning a quantum Hamiltonian H_Q (e.g., the transverse field term), one might stick to the following annealing schedule.

$$H_{Tot}(t) = \frac{t}{\tau}H_c + \left(1 - \frac{t}{\tau}\right)H_Q. \qquad (4)$$

One must start with the ground state of H_Q, which should be easy to prepare, and which must have non-zero (preferably of same order) overlap with all classical configurations (i.e., eigenstates of H_c). Then according to the schedule, at $t = 0$ the total Hamiltonian is H_Q, and the

system is in its ground state. Then as one reduces the strength of H_Q and tube up the strength of H_c following the protocol, the system evolves. Now the adiabatic theorem of quantum mechanics guarantees that the system starting at the ground state of $H_{Tot}(t = 0)$; will follow the ground state of the instantaneous Hamiltonian $H_{Tot}(t)$ throughout the evolution, if the annealing is slow enough. To be precise, according to the adiabatic theorem, for a non-degenerate spectrum with a finite energy gap between the ground state and first excited state, the adiabatic evolution is assured if the evolution time τ satisfies the following condition -

$$\tau >> \frac{\left|\left\langle H_{tot}\right\rangle\right|_{\max}}{\Delta^2_{\min}}, \tag{5}$$

where

$$\left|\left\langle \mathcal{H}_{tot}\right\rangle\right|_{\max} = \max_{0\le t\le\tau}\left[\left|\left\langle\phi_0(t)\right|\frac{d\mathcal{H}_{tot}}{ds}\left|\phi_1(t)\right\rangle\right|\right],$$

$$\Delta^2_{\min} = \min_{0\le t\le\tau}\left[\Delta^2(t)\right];\ s = t/\tau;\ 0\le s\le 1, \tag{6}$$

$\left|\phi_0(t)\right\rangle$ and $\left|\phi_1(t)\right\rangle$ being respectively the instantaneous ground state and the first excited state of the total Hamiltonian $\mathcal{H}_{tot}$, and $\Delta(t)$ the instantaneous gap between the ground state and the first excited state energies.

C. Localization and Beyond : As is apparent from the above discussion, in the adiabatic formulation of quantum annealing, the time required for successful annealing strictly depends on the minimum energy gap between the ground state and the first excited state along the annealing path. This exposes adiabatic versions of quantum annealing to a very generic, apparently insurmountable barrier, namely, localization. Most of the useful problems have intrinsic disorder in it, and that leads to localization (either single particle Anderson type[9], or many-body localization[10]). Localization implies occurrence of gaps ($\Delta_{\min}$) which are exponentially small in system-sizes (even for trivial problems - say for a disordered ferromagnet), implying an annealing time that scales exponentially with system-size for finite success rate[11].

However, the adiabatic quantum annealing, though well formulated with a clear speed bound, is too restrictive. After all, one does not really intend to find the exact global minimum, but in general is more interested in good approximate solutions. This leaves space for improvements, departing from the unitary dynamics (and hence the strict condition of adiabaticity). Several success stories kept on appearing, including astronomical speed-up over purely thermal annealing in the context of short-ranged glassy systems[12], successful annealing of moderately sized samples of NP-complete spin glasses[13], quantum speed-up in annealing of kinetically constrained models[14] and several more[4,15,16].

Recently, a Canada based company D-Wave produced a quantum annealer with thousands of units, each of which is a six-qubit coherent quantum device[17]. Several stories of success (i.e., quantum speed-up) as well of failure have been reported from the runs of different versions of the said machine, from NASA, University of Southern California, Harvard University, ETH Zurich and several other places[17].

All these clearly indicate that disorder, though certainly rules out quantum annealing from being a panacea to all ailments, does not essentially rule out advantages over classical algorithms due to quantum tunnelling in its extremely unpredictable forms in a complex many-body systems, especially outside the rigid condition of adiabaticity. Especially, a many-body localized phase can be intrinsically metastable[18], or can be fragile in presence of external baths[19], and periodic perturbations[20]. Hence the role of quantum tunnelling in optimizing disordered many-body systems in the presence of auxilliary degrees of freedom and/or external perturbations is far from clear. In other words, the help one might get from the structure of the landscape in presence of quantum fluctuations in conjunction with other types of favourable noises, is a deep open issue, which can possibly be settled over time only by trials. The quest continues. ❐

References

1. P. Ray, B. K. Chakrabarti, and A. Chakrabarti, *Phys. Rev.* **B 39**, 11828 (Jun 1989), https: //link.aps.org/doi/10.1103/*Phys. Rev.* B.39.11828.
2. L. K. Grover, *Phys. Rev. Lett.* 79, 325 (Jul 1997), https:// link.aps.org/doi/10.1103/ *Phys. Rev. Lett.* 79. 325.
3. K. Binder and A. P. Young, *Rev. Mod. Phys.* **58**, 801 (Oct 1986), https://link.aps.org/doi/10.1103/RevModPhys.58.801.
4. A. Das and B. K. Chakrabarti, *Rev. Mod. Phys.* **80**, 1061 (Sep 2008), https://link.aps.org/doi/10.1103/RevModPhys.80.1061.
5. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983), ISSN 0036-8075, http://science.sciencemag.org/ content/220/4598/671.full.pdf, http://science.sciencemag. org/ content/220/4598/671.
6. M. Rigol, V. Dunjko, and M. Olshanii, *Nature* **452**, 854 (2016), http://dx.doi.org/10.1038/nature06838.
7. T. Kadowaki and H. Nishimori, *Phys. Rev.* **E 58**, 5355 (Nov 1998), https://link.aps.org/doi/10.1103/PhysRevE.58.5355.

8. E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, *Science* **292**, 472 (2001).

9. P. W. Anderson, *Phys. Rev.* **109,** 1492 (Mar 1958), https://link.aps.org/doi/10.1103/*Phys. Rev.* 109, 1492.

10. D. M. Basko, I. L. Aleiner, and B. L. Altshuler, *Annals of Physics* **321**, 1126 (2006), ISSN 0003-4916, http://www.sciencedirect.com/science/article/pii/S0003491605002630.

11. B. Altshuler, H. Krovi, and J. Roland, *Proceedings of the National Academy of Sciences* **107**, 12446 (2010), ISSN 0027-8424, https://www.pnas.org/content/107/28/12446.full.pdf, https://www.pnas.org/content/107/28/12446.

12. G. E. Santoro, R. Martoñák, E. Tosatti, and R. Car, *Science* **295**, 2427 (2002), ISSN 0036-8075, http://science.sciencemag.org/content/295/5564/2427.full.pdf, http://science.sciencemag.org/content/295/5564/2427.

13. A. Das and B. K. Chakrabarti, *Phys. Rev.* **E 78**, 061121 (Dec 2008), https://link.aps.org/doi/10.1103/PhysRevE.78.061121.

14. A. Das, B. K. Chakrabarti, and R. B. Stinchcombe, *Phys. Rev.* **E 72**, 026701 (Aug 2005), https://link.aps.org/doi/10.1103/PhysRevE.72.026701.

15. A. Ghosh and S. Mukherjee, *Science and Culture* **79**, 485 (2013).

16. T. Albash and D. A. Lidar, *Rev. Mod. Phys.* **90**, 015002 (Jan 2018), https://link.aps.org/doi/10.1103/RevModPhys.90.015002.

17. S. Suzuki and A. E. Das, *Eur. Phys. J. Spec. Top.* 224 (2015), doi:nbibinfofdoigfhttps://doi.org/10.1140/epjst/e2015-02337-1g.

18. D. A. Huse, *Physics* **9**, 76 (Mar 2016).

19. S. Banerjee and E. Altman, *Phys. Rev. Lett.* **116**, 116601 (Mar 2016), https://link.aps.org/doi/10.1103/PhysRevLett.116.116601.

20. A. Lazarides, A. Das, and R. Moessner, *Phys. Rev. Lett.* **115**, 030402 (Jul 2015), https://link.aps.org/doi/10.1103/PhysRevLett.115.030402.