Quantum annealing (QA) is a means for solving optimization problems with the help of quantum mechanics. Recently, QA processors with up to 512 qubits have been developed by D-Wave Systems Inc. and tested by many independent research groups. Such quantum annealing processors implement a Hamiltonian that can be approximated as

\[ H(s) = A(s)\mathcal{H}_D + B(s)\mathcal{H}_P, \]

\[ \mathcal{H}_D = -\sum_i \sigma_i^x, \]

\[ \mathcal{H}_P = \sum_i h_i \sigma_i^z + \sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z, \]

where \( \sigma_i^x, \sigma_i^z \) are Pauli matrices acting on qubit \( i \), \( s = t/t_a \), with \( t \) being time and \( t_a \) the annealing time, \( h_i \) and \( J_{ij} \) are tuneable dimensionless parameters, and \( A(s) \) and \( B(s) \) are monotonic functions of \( s \) such that \( A(0) \gg B(0) \approx 0 \) and \( B(1) \gg A(1) \approx 0 \) (see Fig. 3). If the computation is successful, the state of the system at the end of the evolution, when all qubits are measured, is in a global minimum or an acceptable low energy local minimum of \( \mathcal{H}_P \).

By now, the presence of quantum mechanics, including entanglement, has been established in D-Wave processors. It has also been shown that multiqubit tunneling can provide computational benefit compared to some classical annealing algorithms. However, whether or not quantum mechanics can lead to any scaling advantage (quantum speedup) over available classical solvers remains an open question. Recently, there have been many attempts to detect signatures of quantum speedup in D-Wave Two quantum annealing processors. In these studies the performance of the quantum annealer is compared with some classical algorithms such as simulated annealing (SA), quantum Monte Carlo simulation of quantum annealing (SQA), or some algorithms that are tailored to the type of problems that the processor solves. The computation time is typically expressed in terms of the time to solution (TTS) with 99% certainty:

\[ TTS = \frac{\log(0.01)}{\log(1-P_0)} t_a \]

where \( P_0 \) is the final ground state probability and \( t_a \) is the annealing time. Notice that for \( P_0 \ll 1 \), Eq. (3) becomes \( TTS \approx 4.6(t_a/P_0) \). To make our analysis simpler, here we define:

\[ TTS = t_a/P_0. \]

It is clear that (3) and (4) give the same scaling for TTS in the asymptotic (small \( P_0 \)) limit. For a classical solver such as SA or SQA, \( t_a \) is taken to be proportional to the number sweeps (number of iterations in which all spins are updated in a single anneal).

One hopes to determine how TTS scales with the problem size \( N \) and whether the quantum annealer provides a better scaling compared to the available classical solvers. For problems with essentially 2 dimensional connectivity, such as the Chimera graph (the native graph of D-Wave Two), the TTS is expected to be an exponential function of \( \sqrt{N} \) (i.e., the tree-width of the graph). The slope of \( \log(TTS) \) versus \( \sqrt{N} \), therefore, provides the coefficient in the exponent. Quantum speedup then means that the quantum annealer provides a smaller slope than other classical solvers. More precisely, quantum speedup is defined in Ref. as

\[ S(N) = \frac{TTS_{\text{Classical}}(N)}{TTS_{\text{Quantum}}(N)}, \]

A positive slope in \( S(N) \) means a quantum speedup and a negative slope means an unfavorable scaling for QA. In practice, complications arise due to the dependence of the scaling curves on the annealing time \( t_a \). Figure shows schematically TTS curves commonly obtained from some annealing algorithms such as SA or SQA for three different annealing times \( t_{a1} < t_{a2} < t_{a3} \). The slope of the curves are smaller at smaller \( N \) because \( t_a \) is typically too long for such easy problems. The slope

\[ S(N) = \frac{TTS_{\text{Classical}}(N)}{TTS_{\text{Quantum}}(N)}. \]
increases at large $N$ reflecting the fact that the chosen $t_a$ is not enough to find a solution for large size problems. It is clear that the curves depend on $t_a$ and therefore no nuclear calculation can be made about the asymptotic scaling of the algorithm. To resolve this problem, Ref. [14] suggests to optimize all of the annealers. If for each $N$ one finds the optimal $t_a$ that minimizes $TTS$, then the resulting “optimal” curve will be independent of a particular choice of $t_a$ (the blue line in Fig. 1). Assuming that the same arguments hold for QA, one needs to also optimize the quantum annealer in order to be able to make a fair comparison between the solvers.

In practice, quantum annealing processors come with their own limitations in terms of the control parameters or the available annealing schedules. For example, D-Wave Two annealing curves, $A(s)$ and $B(s)$ are fixed, and the minimum available annealing time is 20 $\mu$s. The latter is typically too long for the type of problems that have been studied and therefore the computation is always suboptimal. This makes the task of determining quantum speedup difficult for the reasons outlined above. It may still be possible to put a bound on the optimal scaling from a suboptimal one based on a reasoning originally provided in Refs. [8, 9]. Suppose Fig. 1 is the $TTS$ obtained from a quantum annealer with the minimum allowed annealing time $t_{a3}$ and maximum 512 qubits ($\sqrt{N} < 23$). The red curve in Fig. 1 therefore, represents the experimentally accessible region. It is clear that the slope of the red curve in Fig. 1 is not a representative of the asymptotic performance. But if we were allowed to decrease the annealing time to an extent that optimization could become feasible, then it would have been possible to obtain the blue (optimized) curve. It appears from the figure that the blue curve has more slope than the red one. Therefore, the slope of the measured (red) curve provides a lower bound for that of the optimal (blue) curve.

The above statement becomes more intuitive when we think in terms of the final ground state probability $P_0$ as a function of $N$ for the situation depicted in Fig. 1. We have assumed a nonzero temperature to correspond better with a realistic quantum annealer. At zero temperature, the asymptotic probability is flat at $P_0 = 1$. The optimal curve in (b) is plotted by connecting the corresponding tangency points in (a).

![FIG. 1: A sketch of time to solution $TTS$, for a given annealer, as a function of the problem size $N$ for different values of annealing time $t_{a1} < t_{a2} < t_{a3}$. Such behavior has been observed in SA and SQA and conjectured for QA [8,9]. The blue line shows the scaling of the optimized solver. The red line represents the region accessible by a quantum annealer with $t_a \geq t_{a3}$ and $N \leq 512$.](image1)

![FIG. 2: (a) The inverse of $TTS$ and (b) the ground state probability $P_0$ as a function of $N$ for the situation depicted in Fig. 1. We have assumed a nonzero temperature to correspond better with a realistic quantum annealer. At zero temperature, the asymptotic probability is flat at $P_0 = 1$. The optimal curve in (b) is plotted by connecting the corresponding tangency points in (a).](image2)
by dividing $P_0$ by $t_a$ its absolute slope will always increase from Fig. 2b to Fig. 2h (assuming that the annealing times increase from left to right, which seems reasonable according to the figure). This again justifies the above statement that the absolute slope of the suboptimal curve provides a lower bound for that of the optimal curve. Before discussing the validity of this argument, let us try to understand the origin of the asymptotic behavior.

The commonly observed weak dependence of $P_0$ on $t_a$ in the small $N$ and/or long $t_a$ limit is a signature of quasistatic behavior. In thermodynamics, a system is called quasistatic when its time-dependence is very slow compared to its relaxation time so that it stays near the equilibrium state at all times. Typically, in QA the system will eventually deviate from equilibrium and the probability freezing is short, the final probabilities will be close to the equilibrium probability distribution at a single (freeze-out) point within the freezing region. To demonstrate this behavior numerically, we consider a 16 qubit problem with $h_i$ selected uniform randomly from the set $\{-1/3, 1/3\}$ and $J_{ij}$ values to be either -1, or uniformly randomly from $\{-1/3, 1/3\}$. Figure 3, shows the 12 lowest energy eigenvalues of $H(s)$ obtained with realistic functions $A(s)$ and $B(s)$ 21. The dashed black lines in the figure represent the corresponding classical energies, i.e., eigenvalues of $B(s)\mathcal{H}_P$. We calculate the occupation probabilities using the (Redfield) open quantum master equation discussed in Refs. 14 20. Similar master equations have proven to provide good qualitative and quantitative descriptions of superconducting quantum annealing processors 6 7 11 14 14. We assume an environment at equilibrium temperature $T = 40$ mK having an ohmic noise spectral density with dimensionless coupling constant $\eta = 0.24$. This value is extracted from independent experiments similar to that in Ref. 22. The same model with the same parameters was successfully used to explain the experimental data in Ref. 14. Here, we have chosen a larger than normal ($\approx 15$ mK) temperature to deliberately populate the excited states. Otherwise, only the ground state would have been populated and the equilibration and freezing effects would have been invisible. Circles in Fig. 3: represent the occupation probabilities of the lowest 12 eigenstates during the evolution. On the same figure, we have also plotted the equilibrium probabilities (solid lines) using Boltzmann distribution:

$$P_n^B(s) = e^{-E_n(s)/k_BT}\left[\sum_k e^{-E_k(s)/k_BT}\right]^{-1}, \quad (6)$$

where $E_n(s)$ is an instantaneous eigenvalue of $H(s)$ and $k_B$ is the Boltzmann constant. As can be seen in the figure, the probabilities closely follow the equilibrium Boltzmann distribution up to almost $2/3$ of the evolution (green region in Fig. 3). As $s \to 1$, the function $A(s)$ becomes smaller making the thermal relaxation slower. When the relaxation becomes too slow, the system cannot follow equilibrium any more and the probabilities start deviating from Boltzmann until they all saturate (freeze). The saturations happen within a freezing region (yellow in Fig. 3). If the freezing region is narrow enough, then the final probabilities will be close to the Boltzmann probability distribution at a single freeze-out point $s = s^*$, marked by the red vertical dotted line in

FIG. 3: (a) Envelope functions, $A(s)$ and $B(s)$ in Eq. (4), for a D-Wave Two quantum annealer. (b) The lowest 12 energy levels of a randomly generated 16 qubit problem as a function of the annealing parameter $s$. Dashed black lines are the classical energies. (c) Occupation probabilities of the same system as in (a) during the annealing calculated using Redfield formalism (circles) and Boltzmann distribution (solid lines), assuming $T = 40$ mK and $t_a = 20 \mu s$. All probabilities follow the Boltzmann distribution in the green (quasistatic) region until they start freezing in the yellow (freezing) region. The probabilities do not change in the blue (frozen) region. The final probabilities are very close to the Boltzmann probabilities at the freeze-out point $s^*$, marked by the vertical (red) dashed line. Notice that in (a) the quantum energy eigenvalues are close to the classical ones at the freeze-out point.
Fig. 3. The exact position of $s^*$ depends weakly on the annealing time. To understand this, let $\gamma(s)$ represent the dominant relaxation rate at $s$. Typically, near the end of annealing $\gamma(s)$ vanishes exponentially due to the exponential decay of $A(s)$. The freeze-out happens when $\gamma(s)$ becomes too small compared to the inverse annealing time. Assuming exponential relation ($\gamma(s) \sim \gamma_0 e^{-\alpha s}$), $s^*$ is the solution of $\gamma_0 e^{-\alpha s^*} = t_a^{-1}$, therefore,

$$s^* \approx \ln(\gamma_0 t_a)/\alpha,$$

where $\gamma_0$ and $\alpha$ are problem dependent constants. The logarithmic dependence of $s^*$ on $t_a$ results in a weak dependence of $P$ on $t_a$. If the Boltzmann probability is approximately linear close to $s^*$, then $P_0$ will be linearly dependent on $\ln(t_a)$

$$P_0(t_a) \sim (\kappa/\alpha) \ln(\gamma_0 t_a).$$

This weak dependence is indeed the origin of the asymptotic behavior illustrated in Fig. 2b. Notice in Fig. 3 that the quantum energy eigenvalues (solid lines) at $s = s^*$ are not very different from the classical ones (dashed lines), although they are very different earlier. This means that the final probability distribution should also be close to the classical Boltzmann distribution, although the dynamics of the system during the evolution are very different from classical.

It has been conjectured that quantum annealing can become advantageous over thermal annealing or SA for problems that have thin but tall barriers in their energy landscapes. This is certainly a dynamical effect and not an equilibrium property. If the source of a possible quantum speedup is dynamical, as conjectured, then searching for the speedup requires access to such dynamics. In general, an equilibrium population is only a function of the energy eigenvalues and therefore is independent of the dynamics. For a quasistatic quantum annealer with a sharp freeze-out at $s = s^*$, the final ground state population will be close to $P_0^B(s^*)$, which is only a function of $E_a(s^*)$. It is clear that such a Boltzmann probability density it is not possible to determine how fast this distribution is established. For example, consider a hypothetical quantum annealer that can return a Boltzmann distribution of $H(s^*)$ within a constant time $t_a = O(1)$ (independent of $N$) that is much faster than any other classical solvers and obviously scales better than all of them. Since sampling from a Boltzmann distribution is a very hard computational problem, the hypothetical quantum annealer (although may never exist) can provide an incredible quantum speedup for such a hard problem. Nevertheless, when we assess its power based on a suboptimal TTS graph, it may scale worse than some optimized classical solvers, merely due to the scaling of $P_0^B(N)$, and we may conclude that quantum speedup is ruled out.

Now, let us return to our original question of whether it is possible to determine a bound for quantum speedup from suboptimal data. More precisely, if the measured slope in Fig. 1 provides a lower bound for the optimal slope as it appears in that figure, then one may know with certainty when quantum speedup is not possible. If the measured slope is already too large (i.e., larger than that obtained from classical solvers), then the optimal slope would be even larger and therefore the optimized quantum annealer would never be able to provide quantum speedup. On the other hand, if the measured slope is smaller than that obtained from classical solvers, then no conclusion can be made because there is always a possibility that the optimal slope be larger than the classical ones. In other words, it is possible to “rule out” quantum speedup from a suboptimal performance, but it is not possible to prove it. Based on what we discussed before, if the suboptimal probabilities do not provide any information about the quantum dynamics, then it should not be possible to even put a bound on the quantum speedup, which seems to be contradicting the above argument. To resolve this contradiction, we notice that the situation depicted in Fig. 1 is based on assumptions which may not hold in general for QA. For example, it is assumed that the slope of each curve at a fixed $t_a$ is a monotonically increasing function of $N$, or that $TTS$ increases with $t_a$ at small $N$, but decreases at large $N$. It is also assumed (implicitly in Fig. 1 but explicitly in Fig. 2b) that the
probability $P_0$ is a monotonically increasing function of $t_a$. Most of these assumptions are based on observations from classical solvers such as SQA (see e.g., supplementary information of Ref. [8]) and may even apply to most classical solvers. Establishing that they also hold in general for QA requires a proof. It is, however, possible to disprove them by a counterexample, as we do next.

As a counterexample, we examine the last assumption mentioned above, i.e., $P_0(t_a)$ is a monotonically increasing function. We calculate $P_0$ as a function of $t_a$ for the 16 qubit instance of Fig. 3 using the Redfield model that we used before with the same realistic parameters. Circles (squares) in Fig. 4 are the results of master equation calculations with (without) coupling to the environment. It is clear that the open system probability behaves completely non-monotonically. For short annealing times (the gray region in Fig. 4), the environment does not have enough time to excite the system from the ground state. Therefore the open system’s ground state probability coincides with that of the closed system. At longer annealing times, the environment starts to populate the excited states and the probabilities decrease with $t_a$ (pink region). The dynamics are non-equilibrium in this region. At longer annealing times the system starts observing quasistatic evolution and an equilibrated final population. The ground state probability increases linearly with $\ln(t_a)$ as predicted in Eq. (8).

We should emphasize that this nonmonotonic behavior can be more generic than the chosen example. In QA, the system starts in the ground states and therefore the thermal excitation populates the excited states from bottom up (unlike the top down process in SA). By decreasing the annealing time beyond the relaxation time, the thermal process will not have enough time to excite the system, thus the ground state probability will increase until the point where the annealing becomes so fast that the non-adiabatic excitations start dominating the error.

Of course, the details of this behavior is strongly dependent on the energy spectrum, the form of eigenstates, and the strength of coupling to the environment. Similar nonmonotonic behavior has recently been observed in SQA [25].

If in Fig. 1 we allow a non-monotonic dependence of $P_0$ on $t_a$, we may face a very different situation. For example, if we allow the short $t_a$ probabilities exceed the asymptotic probability in Fig. 2a, then we may obtain Fig. 5a, which leads to a $TTS$ plotted in Fig. 5b. It is evident that the situation depicted in Fig. 5b is very different from that in Fig. 1 and therefore the same conclusion on the relative slopes cannot be reached. Indeed, the plotted optimal line in Fig. 5b has a smaller slope than the asymptotic one. It should be emphasized that we do not claim that the behavior depicted in Fig. 5 is generic for quantum annealers. Other situations may arise that could be different from both Figs. 1 and 5. The only point we want to make is that an information obtained in a quasistatic suboptimal regime cannot be extrapolated to faster annealing times.

Since quantum Monte Carlo (QMC) simulation is designed to provide correct samples for an equilibrated quantum system, correlation with QMC, equilibrated at the correct freeze-out point and the right temperature is expected. Reference [12] provides indications of such correlations, although QMC was used as an annealing algorithm (SQA) with time-dependent Hamiltonian (not at a fixed $S^*$), and the temperature used was smaller than the actual temperature. This may explain why correlations did not exist when considering excited states [12]. Correlation with SA is also possible (see e.g., Ref. [9]), but it requires quantum eigenenergies that coincide with the classical ones, which is not always the case. Other semiclassical models may also correlate with a quasistatic quantum annealer (see e.g., Ref. [27]) if they can return the same equilibrated population. These correlations do not mean that the dynamics of the quantum annealer is the same as, or can be simulated by, these algorithms, but it requires quantum eigenenergies that coincide with the classical ones, which is not always the case. Other semiclassical models may also correlate with a quasistatic quantum annealer (see e.g., Ref. [27]) if they can return the same equilibrated population. These correlations do not mean that the dynamics of the quantum annealer is the same as, or can be simulated by, these algorithms, because as we discussed above such dynamics are inaccessible.

Finally, it is important to mention that the final population of a quantum annealer is not always an equilib-
rium population, even at long annealing times. Since the relaxation rates between different energy levels change differently with time, they may freeze at different points during the annealing, leading to a distributed freeze-out (a wide freezing region in Fig. [3]). In such cases, it is not possible to identify a single freeze-out point at which the equilibrium population gives the final population. Such situations are likely to happen when the problem has a complicated landscape with numerous valleys and large barriers between them. These problems are expected to have final populations that are more sensitive to the annealing time and do not correlate with equilibrated QMC simulations. In Ref. [26], it was shown that random problems on the Chimera graph have no spin glass phase transition at any nonzero temperature. The lack of spin glass phase transition is an indication that the classical energy landscape of the problem is not very complex, thus equilibration may be easy. This signifies the importance of the problem selection for any exploration of the role of quantum dynamics in quantum annealing, as pointed out in Ref. [26].

To summarize, we have shown that a quasistatic evolution due to a long annealing time can mask the underlying quantum dynamics in a quantum annealer. The final population of such an annealer is likely to be close to a Boltzmann distribution at a single freeze-out point. It is therefore expected to correlate well with a quantum Monte Carlo simulation equilibrated at the same point and at the correct temperature. Moreover, correlation with a proper simulated annealing is also possible if the energy eigenvalues of the Hamiltonian happen to be close to the classical ones at the freeze-out point. These correlations are signatures of a quasistatic behavior and do not mean that the dynamics of the quantum annealer can be simulated by such algorithms. The lack of access to the relevant quantum dynamics can make the search for quantum speedup in a suboptimal quantum annealer infeasible. Therefore, the optimization procedure proposed in Ref. [5] is necessary for any reliable conclusion about quantum speedup. We should emphasize that our argument does not undermine the importance of benchmarking in assessing the performance of a quantum annealer, even if suboptimal, nor do we claim that a quasistatic quantum annealer is not useful. Indeed, providing samples from a Boltzmann distribution is a very hard problem with many applications especially in machine learning. However, different benchmarking strategies are required to assess QA for such applications.

We thank T. Albash, E. Andriyash, S. Boixo, I. Hen, S. Isakov, W. Kaminsky, A. King, D. Lidar, C. McGeoch, H. Neven, J. Raymond, J. Rolfe, A. Roy, A. Smirnov, and R. de Sousa, for useful discussions and comments on the manuscript.

[1] A. B. Finnila, M. A. Gomez, C. Sebenik, C. Stenson, and J. D. Doll, Chemical Phys. Lett. 219, 343 (1994).
[2] T. Kadowaki and H. Nishimori, Phys. Rev. E 58, 5355 (1998).
[3] G. E. Santoro, R. Martońák, E. Tosatti, and R. Car, Science 295, 2427 (2002).
[4] J. Brooke, D. Bitko, T. F., Rosenbaum, and G. Aeppli, Science 284, 779 (1999).
[5] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, Science 292, 472 (2001).
[6] M. W. Johnson et al., Nature (London) 473, 194 (2011).
[7] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar, Nat. Commun. 4, 2067 (2013).
[8] T. F. Rønnow, Z. Wang, J. Job, S. Boixo, S. V. Isakov, D. Wecker, J. M. Martinis, D. A. Lidar, and M. Troyer, Science 345, 420 (2014).
[9] I. Hen, J. Job, T. Albash, T. F. Rønnow, M. Troyer, D. Lidar, eprint arXiv:1502.01663.
[10] V. Martin-Mayor, I. Hen, eprint arXiv:1502.02494.
[11] W. Vinci, T. Albash, A. Mishra, P. A. Warburton, D. A. Lidar, eprint arXiv:1408.3228.
[12] S. Boixo, T. F. Rønnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, M. Troyer, Nature Phys. 10, 218 (2014).
[13] T. Albash, T. F. Rønnow, M. Troyer, D. A. Lidar, EPJ-ST 224, 111 (2015).
[14] S. Boixo, V. N. Smelyanskiy, A. Shabani, S. V. Isakov, M. Dykman, V. S. Denchev, M. Amin, A. Smirnov, M. Mohseni, and H. Neven, eprint arXiv:1502.05754 long version: arXiv:1411.4036 (2014).
[15] A. Perdomo-Ortiz, J. Fliegemann, R. Biswas, V. N. Smelyanskiy, eprint arXiv:1503.01083.
[16] D. Venturelli, S. Mandra, S. Knyshev, B. O’Gorman, R. Biswas, V. Smelyanskiy.
[17] T. Lanting et al., Phys. Rev. X, 4, 021041 (2014).
[18] A. King, eprint arXiv:1502.02098.
[19] R. Harris et al., Phys. Rev. B 82, 024511 (2010).
[20] M. H. S. Amin, C. J. S. Truncik, and D. V. Averin, Phys. Rev. A 80, 022303 (2009).
[21] The detailed form of these functions does not affect our argument here. For a typical example see Ref. [17].
[22] Ohmic noise coefficient is extracted from the macroscopic resonant tunneling experiment described in T. Lanting, M. H. S. Amin, M. W. Johnson, F. Altomare, A. J. Berkley, S. Gildert, R. Harris, J. Johansson, P. Bunyk, E. Ladizinsky, E. Tolkacheva, and D. V. Averin, Phys. Rev. B 83, 180502(R) (2011)
[23] S. Ashhab, J. R. Johansson, Franco Nori, Phys. Rev. A 74, 052330 (2006).
[24] E. Cossor, E. Farhi, C. Yen-Yu Lin, H.-H. Lin, P. Shor, eprint: arXiv:1401.7320.
[25] D. S. Steiger, private communication.
[26] H. G. Katzgraber, F. Hamze, and R. S. Andrist, Phys. Rev. X 4, 021008 (2014).
[27] S. W. Shin, G. Smith, J. A. Smolin, U. Vazirani, eprint arXiv:1401.7087.
[28] This is 1/N-times smaller than the total computation time if only a single-core processor is used. This factor is incorporated to include the possibility of classical parallelism (for more detailed discussion see Ref. [14].)