Achieving fair sampling in quantum annealing

Vaibhaw Kumar¹, Casey Tomlin¹, Curt Nehrkorn¹, Daniel O’Malley², Joseph Dulny III³

¹Booz Allen Hamilton
²Los Alamos National Laboratory

(Dated: July 17, 2020)

Sampling all ground states of a Hamiltonian with equal probability is a desired feature of a sampling algorithm, but recent studies indicate that common variants of transverse field quantum annealing sample the ground state subspace unfairly. In this note, we present perturbation theory arguments suggesting that this deficiency can be corrected by employing reverse annealing-inspired paths. We confirm that this conclusion holds in previously studied models with degeneracy and propose an algorithm that produces the full set of ground states without additional exponential overhead.

Quantum annealing (QA)-based optimization has gained a lot of interest recently, fueled partially by the advent of QA hardware [1]. The usual aim of this technique is to produce a ground state of a Hamiltonian cost function with sufficiently high probability, with less attention paid to the distribution of sampled solutions. In cases where there are several cost function minima, one would hope to find that each is sampled with equal probability across many annealing runs; i.e., one would like a protocol that achieves fair sampling. Fair sampling is crucial for a number of applications. One such example, the set membership problem, ubiquitous in computer science [2,3], involves finding whether an element belongs to a given set. SAT-based probabilistic membership filters [5,6] used to establish set membership require access to not one, but a majority of the solutions of the underlying SAT problem. Other applications for which fair sampling is important include ground state entropy calculations, counting problems [10,11], and machine learning [12,13].

“Vanilla” quantum annealing involves evolution with respect to a Hamiltonian

\[ H(t) = (1 - t/T)H_d + (t/T)H_p, \quad t \in [0, T], \quad (1) \]

where \( T \) is the total evolution time, \( H_p \) is the ‘problem’ Hamiltonian, diagonal in the \( z \)-basis, whose ground states are sought, and \( H_d \) is a ‘driver’ Hamiltonian capable of inducing transitions between \( z \)-eigenstates (usually \( -\sum_i \sigma_i^z \)). For sufficiently large \( T \), the adiabatic theorem [14] promises that the instantaneous ground state of \( H(t) \) is approximately tracked throughout the evolution, so that a \( z \)-basis measurement at \( t = T \) should return a ground state of \( H_p \).

Perturbation theory arguments, simulations of small toy models, and actual QA hardware experiments on relatively large 2D square lattice spin models indicate that this simple protocol fails to sample degenerate ground states fairly [15-21]. In such cases, some of the ground states are “hard” suppressed (the probability of observing these states is approximately zero). Other cases feature “soft” suppression: all ground states are observed with non-zero probability, but some are seen more frequently than others. Soft suppression is usually repairable with a sufficient number of annealing runs and post-processing, but hard suppression is particularly detrimental for certain applications, such as those mentioned above.

In principle, a denser transverse driver can mitigate the sampling bias, but such drivers are difficult both to engineer, and to simulate classically [15] [18]. As observed in [18], even dense drivers cannot completely remove sampling bias, except in the extreme case of a complete graph driver. Another recent proposal [22] based on “extended” quantum annealing [23] also appears to be difficult to realize experimentally.

In this note, we use perturbation theory arguments similar to those used in [18] to suggest a simple solution to the fair sampling problem, based on reverse annealing [24]. In particular, we consider random diagonal perturbations of the final Hamiltonian, which in perturbation theory trivially break degeneracy, leading to a uniform choice of computational basis state from the originally degenerate subspace in each run. We verify that these perturbation theory arguments are borne out via simulations of the small systems described in [18]. Though we do not take into account time-dependent effects in the perturbative arguments, we believe that their inclusion should only serve to mitigate the problem further (via late-time transitions to excited states).

Perturbation theory: Let \( s := t/T \) be a dimensionless time parameter. Towards the end of the anneal, where \( 1 - s \ll 1 \) (or rather when \( s \approx 1 - \lambda \) with \( \lambda \ll 1 \)), \( H(1 - \lambda) \) can be viewed as a perturbation of the problem Hamiltonian \( H_p \) by the driver \( H_d \)

\[ H(1 - \lambda) = H_p + \lambda H_d + O(\lambda^2) \quad (2) \]

If \( H_p \) has \( m \) degenerate (computational basis) ground states \( |g_i\rangle, i \in \{1, \ldots, m\} \) with energy \( E^{(0)} \), switching on the perturbation will result in the splitting of those

\[ \lambda \ll 1 \]

At this stage we are assuming exact adiabatic evolution and ignoring time-dependent effects.
states into \( m \) eigenstates of \( H(1 - \lambda) \) with distinct energies\(^2\). Perturbation theory posits that the energies \( E_k(\lambda) \) and eigenstates \( |k\rangle \), \( k \in \{1, \ldots, m\} \) of \( H(1 - \lambda) \) can be expressed as a power series in \( \lambda \):

\[
E_k(\lambda) = E(0) + \lambda E(1) + O(\lambda^2) 
\]

Inserting these Ansätze into the (time-independent) Schrödinger equation, one can compute corrections to the states and energies order by order in \( \lambda \) in terms of quantities known from the unperturbed system \( \mathbb{H} \).

For the resulting analysis to remain consistent, one must identify a “good” basis for the degenerate ground state subspace that diagonalizes a given perturbation. More specifically, suitable \( \beta_k \) that define good basis states

\[
|k\rangle = |k(0)\rangle + \lambda |k(1)\rangle + O(\lambda^2) 
\]

As pointed out in [18], perturbation theory provides a simple explanation as to why some ground states are hard-suppressed in QA: Assuming exact adiabaticity, the corresponding eigenstate \( |k\rangle = \sum |g_i\rangle \beta_i \) dictates the sampling probabilities in the computational basis; in particular, \( \beta_i = 0 \) leads to hard suppression of the \( i \)th degenerate ground state of \( H_p \) since the state being tracked by adiabatic evolution has no support on \( |g_i\rangle \) near the end of the anneal. This suppression is caused by the sparse nature of the standard hypercube driver \( V = -\sum \sigma^x \) as perturbation, and the set of ground states of \( H_p \). Occasionally one can get lucky when using \( V \), with all ground states reachable from each other by single bit flips, in which case soft suppression is a worst-case scenario. But this is not the case in general, even with denser drivers—only for the complete graph driver are we guaranteed that the late-time instantaneous ground state has support on all ground states of \( H_p \). Unfortunately, engineering anything close to the complete graph driver seems prohibitively difficult. This compels us to search for alternative drivers that provide support for a would-be suppressed \( |g_i\rangle \), at least in a perturbative analysis, if we seek to cure this suppression.

\(^2\) If the degeneracy is not completely lifted by the perturbation at first order, some of these eigenstates will share the same energy, but in the protocol described below this is rare.
θ(m log m) runs are required.

Reverse annealing: Feasibly engineerable time-dependent Hamiltonians that feature terms like \( H_z \) as perturbations of \( H_p \) at late times are present in so-called reverse annealing schemes \([20,28]\), which fall under a broader class of protocols that employ non-convex combinations of initial and final Hamiltonians. Here we consider Hamiltonians of the type

\[ H(s) = D(s)H_d + sH_p + Z(s)H_z, \]

where again \( H_d = -\sum_s \sigma_i^z \) is the standard hypercube driver, \( H_p \) is the problem Hamiltonian, and \( D(s) \) and \( Z(s) \) are schedules which we specify further below. For our perturbation theory arguments, it is only necessary that

\[ D(1 - \lambda) \approx O(\lambda^3), \]
\[ Z(1 - \lambda) \approx O(\lambda). \]

The reverse annealing paradigm would further specify

\[ D(\lambda) \approx O(\lambda), \]
\[ Z(\lambda) \approx 1 - O(\lambda), \]

initializing the system in the ground state of \( H_z \). The ground state \( |\gamma_{\min}\rangle \) of \( H_z \) has eigenvalue \( \mu(|\gamma_{\min}\rangle) = \sum_{i=0}^{n-1} (-1)^{\gamma_{\min}(i)} c_i \) where \( (-1)^{\gamma_{\min}(i)} c_i / |c_i| = -1 \) for all \( i \). These late-time conditions guarantee that the first-order perturbative corrections to the \( H_p \) eigenstates vanish, but (except in very rare cases) the eigenvalues are split, resulting in a single \( H_p \) ground state providing the entire support of the instantaneous eigenstate.

Though perturbation theory is rather limited in what it can reliably say about the entire evolution (or the general utility of quantum annealing), its conclusions are corroborated by the simulated models considered here, which we describe below.

\[ 3 \text{ See the “coupon collector’s problem.” In the appendix we discuss the situation in which the number of ground states is not known in advance.} \]

\[ 4 \text{ This can also be relaxed to } D(1 - \lambda) \approx O(\lambda^2) \text{ resulting in } O(\lambda) \text{ rotations of instantaneous eigenstates in the } H_p \text{ ground state subspace, which does not affect our general arguments.} \]
either 2 (dark red edges), 1 (orange edges), −1 (light blue edges), or −2 (dark blue edges). (a)-(c) have been shown to exhibit hard suppression while (d) features soft suppression.

**FIG. 3.** The instances we consider. The coupling values are either 2 (dark red edges), 1 (orange edges), −1 (light blue edges), or −2 (dark blue edges).

**Simulations:** The perturbative arguments we have made so far only hold in the adiabatic limit. When time-dependence is taken into account and spectral gaps necessarily close, diabatic transitions will be induced. Though we do not provide an analytic argument, these transitions would presumably only further democratize sampling from the ground state subspace given the convergence of eigenvalues at late times. Indeed, the numerical simulations performed here appear to substantiate this intuition.

In Figure 4, we plot squared magnitude of each $H_p$ ground state coefficient as a function of the total annealing time $T$, obtained by integrating the Schrödinger equation in Qutip [29, 30]. For the reverse annealing runs, we plot the average of these probabilities over multiple trials, where each trial corresponds to a randomly generated $H_z$, and we observe that this protocol does improve the sampling bias. In the case of the piecewise-defined schedule $D_{pw}$ where $H_z$ is the sole first-order perturbation to $H_p$, at late times, we see that for all models, states that were suppressed by vanilla annealing now provide roughly equal support to the wave function for large enough $T$. We also observe that the quadratic driver schedule $D_{q}$, in which both $H_d$ and $H_z$ contribute at first order, is less effective at removing the bias in models (a) and (b).

**FIG. 4.** Measurement probability versus total annealing time $T$ for each of the ground states and schedules. Reverse annealing measurement probabilities were averaged over 32, 32, 64, and 16 trials for the models (a), (b), (c), and (d), respectively, where each trial corresponds to a randomly generated $H_z$. The labels correspond to the model labels in Figure 3. Red and blue curves show (up to symmetry) the probability of individual $H_p$ ground states. The black curves show the sum of probabilities over all ground states.

In the more realistic scenario where one does not know the dimension of the ground state subspace a priori, we give an algorithm in the appendix that will return the full set of ground states (up to some specified failure tolerance), assuming the annealing protocol described here uniformly samples only the ground states.

**Remarks:** We have presented a very simple protocol, motivated by time-independent perturbation theory, that is able to cure a seemingly hard-wired tendency towards biased sampling afflicting vanilla quantum annealing. While it is well-known that the complete graph driver ensures fair sampling, no sparser driver can provide this guarantee in general, and in any case denser drivers seem extremely difficult to realize experimentally. On the other hand, since our method is very similar to existing reverse annealing techniques that have experimental realizations, it is also a practical solution to the unfair sampling problem.
We thank Michael Jarret for scathing comments on an initial draft and fruitful discussions.

[1] Zhengbing Bian, Fabian Chudak, William G Macready, and Geordie Rose. The Ising model: teaching an old problem new tricks. *D-wave systems*, 2, 2010.

[2] Andrei Broder and Michael Mitzenmacher. Network applications of bloom filters: A survey. *Internet mathematics*, 1(4):485–509, 2004.

[3] Lukas Radvilavicius, L Marozas, and Antanas Cenys. Overview of real-time antivirus scanning engines. *Journal of Engineering Science & Technology Review*, 5(1), 2012.

[4] Sasu Tarkoma, Christian Esteve Rothenberg, and Eemil Lagerspetz. Theory and practice of bloom filters for distributed systems. *IEEE Communications Surveys & Tutorials*, 14(1):131–155, 2011.

[5] Sean A Weaver, Katrina J Ray, Victor W Marek, Andrew J Mayer, and Alden K Walker. Satisfiability-based set membership filters. *Journal on Satisfiability, Boolean Modeling and Computation*, 8(3-4):129–148, 2012.

[6] Thomas J. Schaefer. The complexity of satisfiability problems. In *Proceedings of the Tenth Annual ACM Symposium on Theory of Computing*, STOC 78, page 216226, New York, NY, USA, 1978. Association for Computing Machinery.

[7] Adam Douglas, Andrew D King, and Jack Raymond. Constructing sat filters with a quantum annealer. In *International Conference on Theory and Applications of Satisfiability Testing*, pages 104–120. Springer, 2015.

[8] Chao Fang, Zheng Zhu, and Helmut G Katzgraber. NaeSAT-based probabilistic membership filters. *arXiv preprint arXiv:1801.06233*, 2018.

[9] Marlon Azinović, Daniel Herr, Bettina Heim, Ethan Brown, and Matthias Troyer. Assessment of quantum annealing for the construction of satisfiability filters. *SciPost Physics*, 2(2):013, 2017.

[10] Carla P Gomes, Ashish Sabharwal, and Bart Selman. Model counting: A new strategy for obtaining good bounds. In *AAAI*, pages 54–61, 2006.

[11] Parikshit Gopalan, Adam Klivans, Raghu Meka, Daniel Štefankovic, Santosh Vempala, and Eric Vigoda. An fptas for # knapsack and related counting problems. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 817–826. IEEE, 2011.

[12] Geoffrey E Hinton. Training products of experts by reversing quantum annealing for the fully-connected p-spin model. *arXiv preprint arXiv:1906.10889*, 2019.

[13] Andris Ambainis and Oded Regev. An elementary proof of the quantum adiabatic theorem. *arXiv preprint quant-ph/0411152*, 2004.

[14] Yoshiki Matsuda, Hidetoshi Nishimori, and Helmut G Katzgraber. Ground-state statistics from annealing algorithms: quantum versus classical approaches. *New Journal of Physics*, 11(7):073021, 2009.

[15] Tameem Albash, Walter Vinci, Anurag Mishra, Paul A Warburton, and Daniel A Lidar. Consistency tests of classical and quantum models for a quantum annealer. *Physical Review A*, 91(4):042314, 2015.

[16] Sergio Boixo, Tameem Albash, Federico M Spedalieri, Nicholas Chancellor, and Daniel A Lidar. Experimental signature of programmable quantum annealing. *Nature communications*, 4:2067, 2013.

[17] Mario S Kötz, Guglielmo Mazzola, Andrew J Ochoa, Helmut G Katzgraber, and Matthias Troyer. Uncertain fate of fair sampling in quantum annealing. *Physical Review A*, 100(3):030303, 2019.

[18] Andrew D King, Emile Hoskinson, Trevor Lanting, Evgeny Andriyash, and Mohammad H Amin. Degeneracy, degree, and heavy tails in quantum annealing. *Physical Review A*, 93(5):052320, 2016.

[19] Brian Hu Zhang, Gene Wagenbreth, Victor Martin-Mayor, and Itay Hen. Advantages of unfair quantum ground-state sampling. *Scientific reports*, 7(1):1–12, 2017.

[20] Salvatore Mandra, Zheng Zhu, and Helmut G Katzgraber. Exponentially biased ground-state sampling of quantum annealing machines with transverse-field driving hamiltonians. *Physical review letters*, 118(7):070502, 2017.

[21] Masayuki Yamamoto, Masayuki Ohzeki, and Kazuyuki Tanaka. Fair sampling by simulated annealing on quantum annealer. *Journal of the Physical Society of Japan*, 89(2):025002, 2020.

[22] Rolando D Somma, Cristian D Batista, and Gerardo Ortiz. Quantum approach to classical statistical mechanics. *Physical review letters*, 99(3):030603, 2007.

[23] D-Wave. Reverse quantum annealing for local refinement of solutions, 9 2017.

[24] Barton Zwiebach. Perturbation theory.

[25] Alejandro Perdomo-Ortiz, Salvador E Venegas-Andraca, and John Winn. The shape boltzmann machine: a strong model of object shape. *International Journal of Computer Vision*, 107(2):155–176, 2014.

[26] Salil Mayor, and Itay Hen. Advantages of unfair quantum ground-state sampling. *Scientific reports*, 7(1):1–12, 2017.

[27] Masaki Ohkuwa, Hidetoshi Nishimori, and Daniel A Lidar. Reverse annealing for the fully connected p-spin model. *Physical Review A*, 98(2):022314, 2018.

[28] Yu Yamashiro, Masaki Ohkuwa, Hidetoshi Nishimori, and Daniel A Lidar. Dynamics of reverse annealing for the fully-connected p-spin model. *arXiv preprint arXiv:1906.10889*, 2019.

[29] J Robert Johansson, Paul D Nation, and Franco Nori. Qutip 2: A python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 184(4):1234–1240, 2013.

[30] J Robert Johansson, PD Nation, and Franco Nori. Qutip: An open-source python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 189(5):1760–1772, 2012.
Appendix: An algorithm for recovering all ground states

We give an algorithm for generating (up to some failure rate) the set of all degenerate ground states \( G \), assuming each is returned uniformly at random by the annealing procedure described in the main text. The key quantity is an estimate of the number of trials required to guarantee with high probability that all ground states have been observed. This is equivalent to the problem of determining the total number \( m \) of coupons in the coupon collector’s problem. For fixed \( m \) we have the following bound for the number of trials \( T \) required to collect all \( m \) coupons:

\[
P[T > m \log (m/\epsilon)] \leq \epsilon \quad \text{(A.1)}
\]

The algorithm below is composed of possibly several rounds of sampling, each round corresponding to an assumed \( m \), and we use this inequality to bound the failure rate of the entire procedure. Specifically, if one asks for a success rate of \( 1 - \epsilon \) after \( r \) rounds of sampling, one needs

\[
T \geq \lceil m \log (rm/\epsilon) \rceil = T(m, r, \epsilon) \quad \text{(A.2)}
\]

samples per round. By iteratively doubling a guess for \( m \) as unseen states are sampled, we have that the maximum number of rounds is \( n-1 \) (for an \( n \)-qubit system; we begin with \( m = 2 \)).

---

**Algorithm 1 GetGroundStates**

**Input:** System size \( n \), failure tolerance \( \epsilon \)

**Output:** Set of ground states \( G \)

1: \( m \leftarrow 2 \)
2: \( T \leftarrow T(m, n, \epsilon) \)
3: \( G \leftarrow \emptyset \)
4: \( t \leftarrow 0 \)
5: while \( t \leq T \) do
6: \( g \leftarrow \text{AnnealOnce}() \)
7: \( G \leftarrow G \cup \{g\} \)
8: \( t \leftarrow t + 1 \)
9: if \( |G| > m \) then
10: \( m \leftarrow 2m \)
11: \( T \leftarrow T(m, n, \epsilon) \)
12: \( t \leftarrow 0 \)
13: return \( G \)