SCIENTIFIC REP⭕RTS

OPEN

Determination and correction of persistent biases in quantum annealers

Alejandro Perdomo-Ortiz[1,2], Bryan O'Gorman[1,3], Joseph Fluegemann[1,4], Rupak Biswas[5] & Vadim N. Smelyanskiy[6]

Calibration of quantum computers is essential to the effective utilisation of their quantum resources. Specifically, the performance of quantum annealers is likely to be significantly impaired by noise in their programmable parameters, effectively misspecification of the computational problem to be solved, often resulting in spurious suboptimal solutions. We developed a strategy to determine and correct persistent, systematic biases between the actual values of the programmable parameters and their user-specified values. We applied the recalibration strategy to two D-Wave Two quantum annealers, one at NASA Ames Research Center in Moffett Field, California, and another at D-Wave Systems in Burnaby, Canada. We show that the recalibration procedure not only reduces the magnitudes of the biases in the programmable parameters but also enhances the performance of the device on a set of random benchmark instances.

Quantum annealing (QA) is a metaheuristic for solving combinatorial optimization problems[1]. The recent introduction of QA hardware by D-Wave Systems[2,3] has invigorated theoretical and experimental research into the computational power and practical implementation challenges of the QA paradigm. Current research studies focus on both fundamental and applied aspects, including application to real-world problems[4–8], criteria for detecting quantum speedup[9], the computational role of quantum tunneling[10], error-supression[11], the relationship between classical simulated annealing and quantum annealing[12–17], spin-glass perspectives on the hardness of computational problems[18–22], and programming strategies that address intrinsic noise[23,24].

The quantum annealers used for this study are of the second generation of D-Wave devices, also called D-Wave Two[2]: one located at NASA Ames Research Center in Moffett Field, California, ("NASA device"), and another located at D-Wave Systems in Burnaby, Canada ("Burnaby device"). These consist of 64 unit cells of a previously characterized eight-qubit unit cell[3,25]. In the NASA and Burnaby devices, post-fabrication characterization determined that only 509 and 424 qubits, respectively, out of the 512 qubit arrays can be reliably used for computation. The array of coupled superconducting flux qubits is, effectively, an artificial Ising spin system with programmable spin-spin couplings and longitudinal and transverse magnetic fields. It is designed to solve instances of the following (NP-hard[26]) classical optimization problem: Given a set of local fields $\{h_i\} \subset [-2, 2]$ and couplings $\{J_{ij}\} \subset [-1, 1]$, find the assignment $\mathbf{s}^* = s_1^* s_2^* \cdots s_n^* \in \{\pm 1\}^n$, that minimizes the objective function,

$$E_{\text{Ising}}(\mathbf{s}) = \sum_{1 \leq i \leq n} h_i s_i + \sum_{1 \leq i < j \leq n} J_{ij} s_i s_j. \tag{1}$$

Finding the optimal $\mathbf{s}^*$ is equivalent to finding the ground state of the corresponding quantum Ising Hamiltonian $H_{\text{Ising}} = \sum_i h_i \sigma_i^z + \sum_{i<j} J_{ij} \sigma_i^z \sigma_j^z$, where σ_i^z is the Pauli z operator acting on the ith spin. More details of QA can be found in the supplementary information.

Currently, D-Wave devices are only calibrated at the level of ensuring that the low-level control circuitry has its intended effect on the physical quantities like current, flux, etc., that it is meant to control[27] (Lanting, T. Private communication, 2015). Early research into the performance of D-Wave devices has indicated the presence of

[1]Quantum Artificial Intelligence Lab., NASA Ames Research Center, Moffett Field, CA 94035, USA. [2]University of California Santa Cruz at NASA Ames Research Center, Moffett Field, CA 94035, USA. [3]SGT Inc., 7701 Greenbelt Rd, Suite 400, Greenbelt, MD 20770, USA. [4]San Jose State Research Foundation at NASA Ames Research Center, Moffett Field, CA 94035, USA. [5]Exploration Technology Directorate, NASA Ames Research Center, Moffett Field, CA 94035. [6]Google, 150 Main St, Venice Beach, CA, 90291 Correspondence and requests for materials should be addressed to A.P.-O. (email: alejandro.perdomoortiz@nasa.gov)

significant imprecision in the setting of the fields that define the problem to be solved, a significant impairment to the successful solution of the problem[13,17,19,23]. Recently, some work has used a phenomenological noise model of the fields $\{h_i\}$ and $\{J_{ij}\}$ in which the distributions of the deviations from the programmed values are given by Gaussians with means zero and standard deviations, respectively, of 0.05 and 0.035 (in units of the maximal J_{ij}), independently instantiated for each qubit and anneal and constant throughout the course of a given anneal. The parameters of the Gaussians were derived by adding in quadrature the variances of several known microscopic sources of noise (Lanting, T. Private communication, 2015). This model has been used in an attempt to explain the failure rate of D-Wave devices as partly due to misspecification of the programmable values. There are many sources of noise in quantum annealers, each with a different effect and time scale, and we address here only one manifestation. The variances we report in this paper are in a sense incomparable to those just mentioned, and relevant only within the context of the experiments described below.

The presence of systematic biases in quantum annealers has been reported elsewhere[24]. The biases referred to there are fundamentally different in nature from the ones address here, in that the former are collective biases on the qubits of ferromagnetic chains that depend on the strength and topology of the couplings therein and are due to the noise specifically caused by those couplings, and they must be determined anew for each embedding topology used. In this work, we present a methodology for determining, in parallel and using a relatively small amount of total annealing time, the persistent (see Fig. S3), systematic biases in all of the individually available programmable parameters of a quantum annealer. We show that correcting for these biases produces an increase in the quality of solutions found on a set of random benchmark instances. The strategy presented here is the first proposal for a software-level recalibration of the full device by the user, i.e. based only on the data from tailored instances and without access to the low-level control circuitry.

Results

Because actual quantum annealers operate at non-zero temperature, there exists some threshold for the values of the fields $\{h_i\}$ and couplings $\{J_{ij}\}$ below which thermal effects dominate the annealing process. When the strengths of the fields and couplings are set sufficiently small, the probabilities of the final states of the qubits are well described by a Boltzmann distribution. Roughly, the relevant energy scale is given by kT, where k is Boltzmann's constant and T is an effective temperature, not necessarily equal to the device temperature. (Henceforth, we will work in units in which $k = 1$.) By running experiments in this regime, persistent biases in the values of the programmed fields can be uncovered, as described in detail in the Methods section.

In our model, $h_i = h_i^{(p)} + h_i^{(b)}$ is the effective value of the local field of qubit i, where $h_i^{(p)}$ is the user-programmed value and $h_i^{(b)}$ is the $h_i^{(p)}$-independent bias. In an ideal device, $h_i^{(b)} = 0$.

Let $p_i(\uparrow|h_i^{(p)})[p_i(\downarrow|h_i^{(p)})]$ be the probability of qubit i being in the spin-up [spin-down] state at the end of an anneal with the programmed value $h_i^{(p)}$. A completely thermal model for this probability is given by

$$p_i(\uparrow|h_i^{(p)}) = \frac{e^{-\alpha_i(h_i^{(p)},h_i^{(b)},T_i)}}{e^{\alpha_i(h_i^{(p)},h_i^{(b)},T_i)} + e^{-\alpha_i(h_i^{(p)},h_i^{(b)},T_i)}}, \tag{2}$$

where $\alpha_i(h_i^{(p)}, h_i^{(b)}, T_i) \equiv h_i/T_i = (h_i^{(p)} + h_i^{(b)})/T_i$ and T_i is the effective temperature of qubit i. This yields
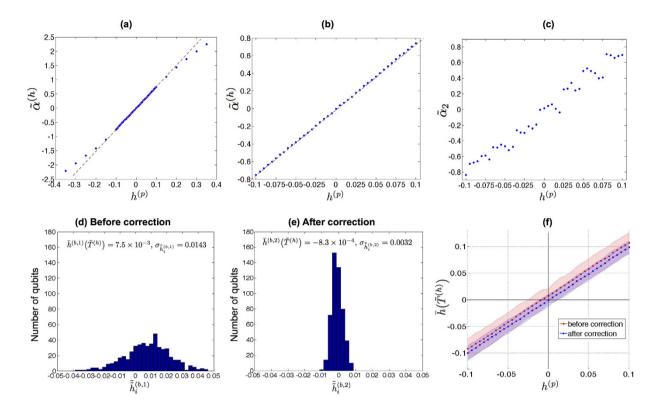
$$\alpha_i(h_i^{(p)}) = \alpha_i(h_i^{(p)}, h_i^{(b)}, T_i) = \frac{1}{2} \ln \frac{1 - p_i(\uparrow|h_i^{(p)})}{p_i(\uparrow|h_i^{(p)})}. \tag{3}$$

(We assume that $h_i^{(b)}$ and T_i are constant at least over the course of the experiment.) More generally, we define the function $\alpha(p) \equiv (1/2)\ln[(1-p)/p]$. Once experimental values of $\alpha_i(h_i^{(p)})$ are obtained for various values of $h_i^{(p)}$, the data are fit to obtain the estimates of $h_i^{(b)}$ and T_i. The biases of the couplings between qubits can be determined in a similar fashion. (See Methods section for more details).

To illustrate the efficacy of our approach in quantum annealers, we applied our method to two D-Wave devices, the NASA device and the Burnaby device.

Determination of the h biases.
Figure 1(a,b) shows two windows of $\widetilde{\alpha}^{(h)}(h^{(p)})$ for one experiment on the NASA device. Figure 1(b) shows the range $h^{(p)} \in [-0.1, 0.1]$ used in the calculation of the biases. There, the linearity of $\widetilde{\alpha}^{(h)}$, and thus the accuracy of the thermal model, is evident; this is typical of all the experiments reported. Figure 1(a) shows a wider window $h^{(p)} \in [-0.35, 0.35]$, where the nonlinearity outside of the former range is evident, indicating the failure of the thermal model for larger magnitudes of $h^{(p)}$, where annealing dynamics start to dominate thermal dynamics. Figure 1(c) shows $\widetilde{\alpha}_2(h^{(p)})$, revealing the limited resolution 0.025 of the digital-to-analog converters (DACs) used to implement h_2. Such stepping behavior is typical for all of the qubits.

Narrowing of the bias distribution. To show the correctability of the persistent biases, we ran the experiment described above repeatedly, each time attempting to correct the biases using estimates thereof from the prior iteration. Let $h_i^{(b,k)}[\widetilde{T}^{(h,k)}]$ be the experimentally determined value of the bias $h_i^{(b)}$[median temperature $\widetilde{T}^{(h)}$]based on the k-iteration; by convention, set $h_i^{(b,0)} = 0$. Let $h^{(p,0)}$ be the desired programmed value. In the k-th iteration, we set

Figure 1. Detection and correction of systematic biases. (**a**) From a single experiment without correction, the median quantity $\widetilde{\alpha}^{(h)}(h^{(\mathrm{p})})$, over the qubits, of the quantities $\{\widetilde{\alpha}_i(h^{(\mathrm{p})})\}$ for various values of $h^{(\mathrm{p})}$ in $[-0.35, 0.35]$. (**b**) Same as (**a**), but only using values of $h^{(\mathrm{p})}$ in $[-0.1, 0.1]$. Note the tightness of the fit to a line, indicating the validity of the thermal model. (**c**) From the same experiment as in (**b**), the quantity $\widetilde{\alpha}_2(h^{(\mathrm{p})})$ for a single, typical qubit, 2. Note the step function resulting from the limited precision of the digital-to-analog converter used to control the field. (**d,e**) The estimated biases $\{\widetilde{\widetilde{h}}_i^{(b,k)}\}$ from two experiments $k = 1, 2$. The first is without correction; the second was corrected using the biases estimated from the first. Note that the distribution significantly more narrowed and centered near zero after correction. (**f**) The average over the qubits of the quantites $\{\widetilde{h}_i\}$ for different values of the programmed $h^{(\mathrm{p})}$, with error bands given by the standard deviation. As for (**d,e**), note the narrowing and centering of the distributions.

$$h_i^{(\mathrm{p})} = h_i^{(\mathrm{p},k)} = h^{(\mathrm{p},0)} - \sum_{k'=1}^{k} h_i^{(\mathrm{b},k')}. \tag{4}$$

Figure 1(d,e) show the distribution of $\{\widetilde{h}_i^{(b,k)}\} \equiv \{\widetilde{\alpha}_i^{(b,k)} \widetilde{T}^{(h,k)}\}$ for the first and second iterations, i.e. before and after the recalibration procedure. The narrowing of the distribution is a clear indication that the procedure is working to remove the biases.

Further support for the success of recalibration is provided by looking at the distribution over the qubits of the success probabilities for all values of $h^{(\mathrm{p})}$. Figure 1(f) shows a uniform reduction of the variance over the qubits in the values of $\widetilde{h}$. Each point is the mean $\overline{h}$ over the qubits of $\{\widetilde{h}_i\}$ for the corresponding $h^{(\mathrm{p})}$, and the shaded region indicates the standard deviaton. The narrowing of the distribution is clear evidence that the recalibration procedure not only narrows the distribution of the biases [Fig. 1(d,e)], as reflected in the shift of the mean, but also reduces the variance for all values of $h^{(\mathrm{p})}$.

Determination of the *J* biases. In the data presented here, the *J* biases were determined using (7). The programmable value of the local fields $\{h_i^{(\mathrm{p})}\}$ were uniformly set to zero (see Methods section for more details).

Figure 2(a) shows the median quantity $\widetilde{\alpha}^{(J)}$ for evenly spaced values of $J^{(\mathrm{p})}$ in $[-0.1, 0.1]$, as well as a line fit thereto. The closeness of the fit of the line confirms the accuracy of the thermal model as for $\widetilde{\alpha}^{(h)}$.

Figure 2(b), analogous to Fig. S2(b), shows the standard deviation $\sigma_{\overline{J}_i}$ over the qubits of the estimates $\{\overline{J}_{i,j}^{(k)}\} = \{\widetilde{\alpha}_{i,j}^{(k)} \overline{T}^{(J,k)}\}$ of $\{J_{i,j}\}$ for different values of $J^{(\mathrm{p})}$, for each of the three iterations of experiments as described above. The estimates $\{\overline{J}_{ij}^{(k)}\}$ were calculated independently for each iteration k using its mean effective temperature $\overline{T}^{(J,k)}$. There is a significant change from the first iteration of corrections, but then the standard deviation remains approximately the same after the second iteration. Considering the average over the values of $J^{(\mathrm{p})}$, the overall
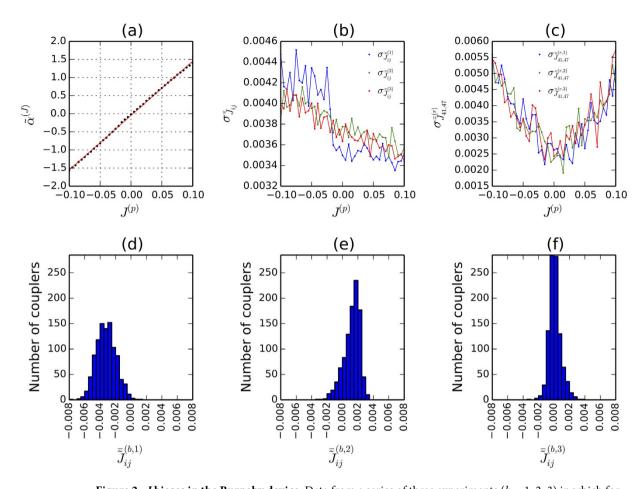
Figure 2. J biases in the Burnaby device. Data from a series of three experiments ($k = 1, 2, 3$) in which for each experiment the sums of biases estimated in the previous ones are subtracted from the original programmed values, using Eq. 5. The first experiment is without any correction, and the second and third use increasingly accurate corrections. All quantities are calculated using the mean $\overline{T}^{(J,k)}$ of the qubit temperatures, calculated indepently in each experiment. (**a**) From only the first experiment, the median quantity $\widetilde{\alpha}^{(J)}(J^{(\mathrm{p})})$, over the couplers, of the quantities $\{\widetilde{\alpha}_{ij}(J^{(\mathrm{p})})\}$ for evenly spaced values of $J^{(\mathrm{p})}$ in $[-0.1, 0.1]$. (**b**) The standard deviation over the couplers of the estimated $\overline{\overline{J}}_{ij}^{(k)}$ at each value of the original $J^{(p,0)}$.

(**c**) For a single, typical coupler, (41, 47), the standard deviation over 100 runs of the estimates $\{\overline{\overline{J}}_{41,47}^{(r)}\}$ of $J_{41,47}$ versus the original $J^{(p,0)}$. (**d–f**) Residual biases $\overline{\overline{J}}_{ij}^{(b,k)}$ estimated from each of the experiments.

variance is about the same before and after correction, yet is much more uniform after correction, which we consider beneficial.

Figure 2(c), analogous to Fig. S1(c), shows the standard deviation $\sigma_{\overline{\overline{J}}_{41,47}^{(r,k)}}$ of the estimates $\{\overline{\overline{J}}_{41,47}^{(k)}\}$. The correction seems to have no effect in this regard, and the there is a consistent increase in the variance with increasing magnitude of the programmed value $J^{(\mathrm{p})}$. A similar phenomenon occurred in the analogous h data.

Define $J_{ij}^{(b,k)}$ to be the estimate $\widetilde{\alpha}_{ij}^{(b,k)}\overline{T}^{(J,k)}$ for $J_{ij}^{(b)}$ using the data from the k-th iteration, with $J_{ij}^{(b,0)} = 0$ by convention. In the kth iteration, we set the programmed values of the couplers to

$$J_{ij}^{(\mathrm{p})} = J_{ij}^{(p,k)} = J^{(p,0)} - \sum_{k'=1}^{k} J_{ij}^{(b,k')}, \tag{5}$$

i.e. by subtracting the sums of the residual biases from the prior iterations from the desired values. Figure 2(d–f) show the narrowing of the distribution of residual J biases with correction.

Unlike the case for the h biases, for which data indicate that the distribution is essentially converged after a single iteration, here we see two new phenomena. First, the distribution continues to narrow between the second and third iterations. Second, the distribution of the residual biases from the second iteration, while narrower than that from the first, is not centered around zero. We believe this is due to overcorrection; that is, the estimates of the biases from the first iteration have a high degree of uncertainty, and so simply subtracting their values from the intended value introduces some amount of bias itself. This is consistent with the overall small magnitudes of the J biases relative to those of the h biases, especially as compared to the corresponding noise levels. This overcorrection can

Range r_J	1	2	4	8	16
Greedy	0.58	0.63	0.59	0.68	0.53
Elite mean	0.65	0.65	0.73	0.72	0.67

Table 1. Comparison of performance with and with *h*-correction on benchmarks. The probability that correcting for *h* biases (using data from a single experiment) improved performance on 100 random instances from an ensemble parameterized by the range of values *r*. Performance was compared according to two metrics: greedy comparison of the energies and degeneracies, and comparison of the elite mean score function.

be mitigated by weighting the correction in a way that accounts for the uncertainty in the estimate using Bayesian reasoning. We will explore these ideas further in future work.

Effect of correction on the performance of benchmark problems. Ultimately, the goal of calibration is to optimize the performance of a quantum annealer on problems of computational interest. It is not clear a priori that the biases present in one- and two-qubit experiments are the same as those present in anneals involving hundreds of qubits. Even if they were, their estimation would be of no practical value unless their correction improves performance. To address this, we tested the effect of correcting the *h* biases on the performance of the quantum annealer at NASA Ames, using the same parameterized random ensemble of instances used in a previous study benchmarking a D-Wave quantum annealer[9]. As in those studies, *r* is a parameter that tunes the difficulty of the average instance (the larger the *r* the more difficult the average instance). We expected the recalibration to have a major positive impact in the harder family of instances.

For each instance, the uncorrected and corrected results were compared using two methods, a "greedy" one and the elite mean. The results of the comparison are summarized in Table 1, showing the proportion of instances, for each *r*, for which the correction improved the performance, using each of the two comparison methods described above. The data set indicates that correcting for the *h* biases improves performance according to these two reasonable metrics. At a large enough range *r*, however, even correction of the biases is not enough. In this limit of large *r*, the spacing of 1/*r* between the different specifiedd *J* values is beyond the precision of the device and poorly resolved. In this limit inherent fluctuations lead to almost zero success probabilities due to problem misspecification, i.e., the device is finding the solution to another problem different from the one indicated. This would explain the possible pattern seen in the elite mean comparison (Table 1) that the advantage of correction peaks seems to peak at the level of *r* considered to correspond to the precision limit of the device. (That such a pattern is not as apparent in the greedy comparison is easily explained by natural noisiness of that comparison method, especially for instances with extremely low success probability as was the case here).

Discussion

Disentangling the mutual effect of the *h* and *J* biases on each other by alternating between iterations of the iterations of *h* and *J* experiments (as opposed to doing each alone as reported) will likely lead to more accurate estimates of each individually. Lastly, the risk of overcorrection can be mitigated by weighting the correction by the degree of certainty of the estimate of the bias to be corrected.

Although we focused initially on a standard random ensemble of Ising instances for benchmarking the performance of quantum annealers, the effect of correcting biases should be greatest on instances whose ground states are most sensitive to misspecification of the programmable parameters[20,21].

There is reason to suspect that correction will also have a beneficial effect in reducing the effect of gauge selection on success probability. While there are other suspected reasons for the effect of gauge selection (which would be non-existent in an ideal device), biases such as the ones corrected here could be one of the leading factors. The effect of gauge selection is significant, sometimes leading to an orders-of-magnitude difference in the success probabilities, and so this is a promising application for bias correction.

Importantly, while the *J* biases determined here are in general smaller than the *h* biases, numerical studies indicate that often instances are more sensitive to misspecification in the *J* parameters than in the *h* parameters[20].

The methods presented here complement a growing suite of tools for optimal programming of quantum annealers[23,24], tuning the performance thereof to cope with the intrinsic noise in current and future physical implementations.

Methods

Calculation of the *h* biases. To experimentally determine the biases $\{h_i^{(b)}\}$, we set all $J_{ij}^{(p)} = 0$ and initially assume the effect of nonzero $\{J_{ij}^{(b)}\}$ to be negligible. We therefore ran the experiments for all qubits within a given device simultaneously, with the same value of $h_i^{(p)} = h^{(p)}$ for every working qubit. Each value of $h^{(p)}$ was run 100 times, where each run consisted of 1,000 annealing cycles. The probability $p_i^{(r)}(\uparrow)$ was calculated for each run *r*, and the median probability $\tilde{p}_i(\uparrow|h_i^{(p)})$ taken over the 100 $\{p_i^{(r)}(\uparrow|h_i^{(p)})\}$ calculated. From this, we define the "median" $\tilde{\alpha}_i(h_i^{(p)}) = \alpha(\tilde{p}_i(\uparrow|h_i^{(p)})) = 0.5 \ln[1 - \tilde{p}_i(\uparrow|h_i^{(p)})] - 0.5 \ln \tilde{p}_i(\uparrow|h_i^{(p)})$. Technically, this is a slight abuse of terminology; while the median of $\{\alpha(p_i^{(r)})\}$ is almost the same as $\alpha(\tilde{p}_i)$ because the function $\alpha(p)$ is monotonic, the two quantities can differ slightly in the case of an even number of values.

Since $\alpha_i = h_i^{(p)}/T_i + h_i^{(b)}/T_i$, for each qubit, we fit a line to $\{(h_i^{(p)}, \tilde{\alpha}_i)\}$ by minimizing the quadratic loss; the resulting slope gives us an estimated inverse qubit temperature $\tilde{\beta}_i = 1/\tilde{T}_i$, and from the intercept $\tilde{\alpha}_i^{(b)}$ we can determine the bias in several ways.

One way is to simply use the fitted parameters as is: $\tilde{h}_i^{(b)} = \tilde{\alpha}_i^{(b)} \tilde{T}_i$. Experimental data, however, indicate that the estimates of the qubit "temperatures" calculated as above are not exactly that, but include in their calculation effects other than that due to true variation in temperature between the qubits. Some estimate of a uniform device temperature should therefore be used. (See Sec. III in the SI for more detail.) In our experiments, we used two different quantities. The first is the "mean temperature", $\overline{T}^{(h)} = \frac{1}{n}\sum_{i=1}^{n} \tilde{T}_i$, where $\tilde{T}_i = 1/\tilde{\beta}_i$ and n is the number of (working) qubits. The second is the "median temperature" $\tilde{T}^{(h)}$, called thus not because it is the median of $\{\tilde{T}_i\}$ but because it is calculated by taking the inverse of the slope of the line fit to the points $\{(h^{(p)}, \tilde{\alpha}^{(h)})\}$, where $\tilde{\alpha}^{(h)}(h^{(p)})$ is defined as the median over the qubits of $\{\tilde{\alpha}_i(h^{(p)})\}$. In practice, the quantities $\tilde{T}^{(h)}$ and $\overline{T}^{(h)}$ are effectively the same.

Calculation of the J biases. Let $J_{ij} = J_{ij}^{(p)} + J_{ij}^{(b)}$ be the effective value of the coupling between qubits i and j and $p_{ij}(\uparrow\uparrow|h_i^{(p)}, h_i^{(b)}, h_j^{(p)}, h_j^{(b)}, J_{ij}^{(p)}, J_{ij}^{(b)})$ be the probability of qubits i and j both being in the spin-up state, where $J_{ij}^{(p)}$ is the programmed value of the coupler and $J_{ij}^{(b)}$ the bias, analogous to $h_i^{(p)}$ and $h_i^{(b)}$, respectively. Here, we set $h_i^{(p)} = h_j^{(p)} = 0$, and write simply $p_{ij}(\uparrow\uparrow|J_{ij}^{(p)}) \equiv p_{ij}(\uparrow\uparrow|h_i^{(b)}, h_j^{(b)}, J_{ij}^{(p)}, J_{ij}^{(b)})$. Other probabilities $p_{ij}(\cdot\cdot|J_{ij}^{(p)})$, and combinations thereof such as $p_{ij}(\uparrow\uparrow\vee\downarrow\downarrow|J_{ij}^{(p)}) \equiv p_{ij}(\uparrow\uparrow|J_{ij}^{(p)}) + p_{ij}(\downarrow\downarrow|J_{ij}^{(p)})$, are analogously denoted.

One approach to determine the bias $J_{ij}^{(b)}$ is to naively assume that $h_i^{(b)} = h_j^{(b)} = 0$, in which case the thermal distribution is modeled by

$$p_{ij}(\uparrow\uparrow\vee\downarrow\downarrow|J_{ij}^{(p)}) = \frac{2e^{-\alpha_{ij}(J_{ij}^{(p)})}}{2e^{\alpha_{ij}(J_{ij}^{(p)})} + 2e^{-\alpha_{ij}(J_{ij}^{(p)})}} \tag{6}$$

where $\alpha_{ij}(J_{ij}^{(p)}) \equiv J_{ij}/T_{ij} = (J_{ij}^{(p)} + J_{ij}^{(b)})/T_{ij}$. For concision, we leave the dependence of p_{ij} and α_{ij} on h_i, h_j, $J_{ij}^{(b)}$, and T_{ij} implicit. Similarly to the case for h_i, this yields

$$\alpha_{ij}(J_{ij}^{(p)}) = \frac{1}{2} \ln \frac{1 - p_{ij}(\uparrow\uparrow\vee\downarrow\downarrow)}{p_{ij}(\uparrow\uparrow\vee\downarrow\downarrow)}. \tag{7}$$

For a given value of $J^{(p)}$, the experiment was run in six batches. In each batch, the programmed coupling $J_{ij}^{(p)}$ was uniformly set to $J^{(p)}$ for each coupler of a pairwise disjoint subset of all of the couplers, and for each of the rest $J_{ij}^{(p)}$ was set to zero. Over the six batches, each coupling $J_{ij}^{(p)}$ was set to $J^{(p)}$ exactly once. As for the h biases, each value of $J^{(p)}$ was run 100 times (for each coupler), with each run consisting of 1,000 annealing cycles. The median probability $\tilde{p}_{ij}(\uparrow\uparrow\vee\downarrow\downarrow|J_{ij}^{(p)})$ was taken over the 100 $\{p_{ij}^{(r)}\}$ calculated from the runs, from which we calculate the "median" $\tilde{\alpha}_{ij}(J_{ij}^{(p)}) = \alpha(\tilde{p}_{ij}(\uparrow\uparrow\vee\downarrow\downarrow|J_{ij}^{(p)}))$. For each coupler, a line was fit to $\{(J_{ij}^{(p)}, \tilde{\alpha}_{ij})\}$, yielding a slope $\tilde{\beta}_{ij} = 1/\tilde{T}_{ij}$ and an intercept $\tilde{\alpha}_{ij}^{(b)}$. As for h, we define the mean temperature $\overline{T}^{(J)} = \frac{1}{m}\sum_{\{i,j\}} \tilde{T}_{ij}$, where m is the number of couplers.

A more accurate estimate for $J_{ij}^{(b)}$ can be obtained by considering nonzero $h_i^{(b)}$ and $h_j^{(b)}$ (but still setting $h_i^{(p)} = h_j^{(p)} = 0$). Let

$$Z \equiv \sum_{s_i, s_j \in \{\pm 1\}} e^{(s_i h_i^{(b)} + s_j h_j^{(b)} + s_i s_j (J_{ij}^{(p)} + J_{ij}^{(b)}))/T_{ij}} \tag{8}$$

be the partition function. Then

$$\frac{p_{ij}(\uparrow\uparrow)p_{ij}(\downarrow\downarrow)}{p_{ij}(\uparrow\downarrow)p_{ij}(\downarrow\uparrow)} \tag{9}$$

$$= \frac{e^{-(h_i^{(b)} + h_j^{(b)} + J_{ij})/T_{ij}}e^{-(-h_i^{(b)} - h_j^{(b)} + J_{ij})/T_{ij}}/Z^2}{e^{-(h_i^{(b)} - h_j^{(b)} - J_{ij})/T_{ij}}e^{-(-h_i^{(b)} + h_j^{(b)} - J_{ij})/T_{ij}}/Z^2} \tag{10}$$

$$= e^{-4J_{ij}/T_{ij}}, \tag{11}$$

or

$$\alpha_{ij}(J_{ij}^{(p)}, J_{ij}^{(b)}, T_{ij}) = \frac{1}{4} \ln \frac{p_{ij}(\uparrow\downarrow)p_{ij}(\downarrow\uparrow)}{p_{ij}(\uparrow\uparrow)p_{ij}(\downarrow\downarrow)}. \tag{12}$$

Note that the assumption that $h_i^{(b)} = h_j^{(b)} = 0$ implies that $p_{ij}(\uparrow\uparrow) = p_{ij}(\downarrow\downarrow)$ and $p_{ij}(\uparrow\downarrow) = p_{ij}(\downarrow\uparrow)$, and that in this case (12) reduces to (7).

Benchmark studies. To generate a single instance with range r, the local fields $\{h_i\}$ were uniformly set to zero, and each available J_{ij} was independently and uniformly selected from $\{-r, -r+1, \ldots, -1, 0, 1, \ldots, r\}$. The resulting instance was then scaled by the overall factor $0.9/r$ so that the largest magnitude $|J_{ij}|$ was 0.9. (This was

necessary, rather than scaling to 1 as in previous studies, to allow for consistency with future experiments in which the J biases are corrected.) 100 such instances were generated and run twice with 1,000 annealing cycles for each of the same (uniformly randomly generated) 10 gauges. In all runs, $\{J_{ij}\}$ were programmed as in the instances. For the first set of runs, which we call "uncorrected", the $\{h_i\}$ were also programmed as in the instances, i.e. to zero. For the other set, which we call "h-corrected", the local fields were programmed to the inverse of the biases computed via experiments as in Eq. 4.

Greedy and elite mean metrics. The greedy comparison is as follows: the energies of all states returned were computed, and those for all gauges were grouped together. Whichever method (uncorrected or corrected) returned the lower minimum energy was deemed to have performed better. If the minimum energies were the same, the tie was broken by the number of times that energy was returned. If this number was the same, the method with the second-lowest energy was deemed to have performed better, with ties broken by the number of times the second-lowest energy was returned, and so on. The "elite mean" score function[23], a quantity previously introduced to allow comparison of the performance of different programming parameters in quantum annealers when the success probabilities are too low (and thus noisy), is defined as the mean energy of the "elite" states, i.e. those with the lowest energies. The elite mean is parameterized by the fraction of energies over which to take the mean; here we use 2%. For $r = 1$ and $r = 2$, there were 6 and 2 instances, respectively, for which the elite mean comparison was tied, all but one due to success probabilities greater than 2% for both the corrected and uncorrected experiments.

References

1. Kadowaki, T. & Nishimori, H. Quantum annealing in the transverse ising model. *Phys. Rev. E* **58,** 5355–5363 (1998).
2. Bunyk, P. *et al.* Architectural considerations in the design of a superconducting quantum annealing processor. *Applied Superconductivity, IEEE Transactions on* **24,** 1–10 (2014).
3. Johnson, M. W. *et al.* Quantum annealing with manufactured spins. *Nature* **473,** 194–198 (2011).
4. Perdomo-Ortiz, A., Dickson, N., Drew-Brook, M., Rose, G. & Aspuru-Guzik, A. Finding low-energy conformations of lattice protein models by quantum annealing. *Sci. Rep.* **2,** 571 (2012).
5. Gaitan, F. & Clark, L. Ramsey numbers and adiabatic quantum computing. *Phys. Rev. Lett.* **108,** 010501 (2012).
6. Perdomo-Ortiz, A., Fluegemann, J., Narasimhan, S., Biswas, R. & Smelyanskiy, V. N. A quantum annealing approach for fault detection and diagnosis of graph-based systems. *Eur. Phys. J. Special Topics* **224,** 131–148 (2015).
7. Rieffel, E. G. *et al.* A case study in programming a quantum annealer for hard operational planning problems. *Quantum Information Processing* **14,** 1–36 (2015).
8. O'Gorman, B., Babbush, R., Perdomo-Ortiz, A., Aspuru-Guzik, A. & Smelyanskiy, V. Bayesian network structure learning using quantum annealing. *Eur. Phys. J. Special Topics* **224,** 163–188 (2015).
9. Ronnow, T. F. *et al.* Defining and detecting quantum speedup. *Science* **345,** 420–424 (2014).
10. Boixo, S. *et al.* Computational role of multiqubit tunneling in a quantum annealer. *arXiv:1502.05754* (2015).
11. Pudenz, K. L., Albash, T. & Lidar, D. A. Error-corrected quantum annealing with hundreds of qubits. *Nat Commun* **5** (2014).
12. Boixo, S., Albash, T., Spedalieri, F. M., Chancellor, N. & Lidar, D. A. Experimental signature of programmable quantum annealing. *Nat Commun* **4** (2013).
13. Boixo, S. *et al.* Evidence for quantum annealing with more than one hundred qubits. *Nature Physics* **10,** 218–224 (2014).
14. Shin, S. W., Smith, G., Smolin, J. A. & Vazirani, U. Comment on "distinguishing classical and quantum models for the d-wave device". *arXiv:1404.6499v2* (2014).
15. Albash, T., Rnnow, T. F., Troyer, M. & Lidar, D. A. Reexamining classical and quantum models for the d-wave one processor. *Eur. Phys. J. Special Topics* **224,** 111–129 (2015).
16. Hen, I. *et al.* Probing for quantum speedup in spin glass problems with planted solutions. *arXiv:1502.01663* (2015).
17. King, A. D. Performance of a quantum annealer on range-limited constraint satisfaction problems. *arXiv:1502.02098* (2015).
18. Katzgraber, H. G., Hamze, F. & Andrist, R. S. Glassy chimeras could be blind to quantum speedup: Designing better benchmarks for quantum annealing machines. *Phys. Rev. X* **4,** 021008 (2014).
19. Venturelli, D. *et al.* Quantum optimization of fully-connected spin glasses. *arXiv:1406.7553* (2014).
20. Zhu, Z., Ochoa, A. J., Schnabel, S., Hamze, F. & Katzgraber, H. G. Best-case performance of quantum annealers on native spin-glass benchmarks: How chaos can affect success probabilities. *arXiv:1505.02278v1* (2015).
21. Martin-Mayor, V. & Hen, I. Unraveling quantum annealers using classical hardness. *arXiv:1502.02494* (2015).
22. Katzgraber, H. G., Hamze, F., Zhu, Z., Ochoa, A. J. & Munoz-Bauza, H. Seeking quantum speedup through spin glasses: The good, the bad, and the ugly. *arXiv:1505.01545* (2015).
23. Perdomo-Ortiz, A., Fluegemann, J., Biswas, R. & Smelyanskiy, V. N. A performance estimator for quantum annealers: Gauge selection and parameter setting. *arXiv:1503.01083* (2015).
24. King, A. D. & McGeoch, C. C. Algorithm engineering for a quantum annealing platform. *arXiv:1410.2628* (2014).
25. Harris, R. *et al.* Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor. *Phys. Rev. B.* **82,** 024511 (2010).
26. Barahona, F. On the computational complexity of ising spin glass models. *Journal of Physics A: Mathematical and General* **15,** 3241 (1982).
27. Harris, R. *et al.* Experimental investigation of an eight-qubit unit cell in a superconducting optimization processor. *Phys. Rev. B* **82,** 024511 (2010).

Author Contributions

A.P.-O., B.O., J.F., R.B. and V.N.S. contributed to the ideas presented in the paper. A.P.-O. and B.O. designed and ran the experiments and wrote the manuscript. All the authors revised the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Perdomo-Ortiz, A. *et al.* Determination and correction of persistent biases in quantum annealers. *Sci. Rep.* **6**, 18628; doi: 10.1038/srep18628 (2016).