Experimental Investigation of an Eight-Qubit Unit Cell in a Superconducting Optimization Processor

R. Harris, M.W. Johnson, T. Lanting, A.J. Berkley, J. Johansson, P. Bunyk, E. Tollakcheva, E. Ladizinsky, N. Ladizinsky, T. Oh, F. Ciocca, I. Perminov, P. Spear, C. Enderud, C. Rich, S. Uchaikin, M.C. Thom, E.M. Chapple, J. Wang, B. Wilson, M.H.S. Amin, N. Dickson, K. Karimi, B. Macready, C.J.S. Truncik, and G. Rose

D-Wave Systems Inc., 4401 Still Creek Drive, Burnaby BC, Canada, V5C 6G9

(Dated: June 29, 2010)

A superconducting chip containing a regular array of flux qubits, tunable interqubit inductive couplers, an XY-addressable readout system, on-chip programmable magnetic memory, and a sparse network of analog control lines has been studied. The architecture of the chip and the infrastructure used to control it were designed to facilitate the implementation of an adiabatic quantum optimization algorithm. The performance of an eight-qubit unit cell on this chip has been characterized by measuring its success in solving a large set of random Ising spin glass problem instances as a function of temperature. The experimental data are consistent with the predictions of a quantum mechanical model of an eight-qubit system coupled to a thermal environment. These results highlight many of the key practical challenges that we have overcome and those that lie ahead in the quest to realize a functional large scale adiabatic quantum information processor.

I. INTRODUCTION

Of the known paradigms of quantum computing, those related to quantum annealing (QA)1–5 are unique in that they leverage what ought to be a robust natural phenomenon - the tendency for physical systems to seek and remain in a low energy configuration. These methods have received considerable theoretical attention of late with claims and counter-claims concerning their ultimate utility6–10. In contrast, they have received relatively little experimental attention. This is a significant concern as data are needed to help clarify the impact of environmental noise upon such approaches to quantum computing11–19. Experiments have been performed using nuclear magnetic resonance in molecules20, but the prospects for scaling this approach to larger systems are limited. At least one proposal for an adiabatic quantum optimization (AQO) processor based upon superconducting flux qubits exists21,22 and rudimentary experiments have been performed upon related small-scale devices with fixed qubit and coupler parameters23–25. Experimental investigations of QA in large-scale systems have only ever been performed upon solid state samples in which the experimenter has no control over the individual couplings between Ising spins26,27. Consequently, there are many open questions regarding how to implement a large-scale programmable QA information processor in practice.

The purpose of this article is threefold: First, we address some of the practical questions regarding how to design a scalable superconducting AQO information processor. The answers to these questions will then serve as a motivation for the architecture of the device that we have fabricated. Second, we present experimental results from a unit cell on one such chip consisting of eight flux qubits28, sixteen in situ tunable inductive interqubit couplers29, an XY-addressable high fidelity readout architecture30, an array of in situ programmable flux storage devices addressed by a single flux quantum demultiplexing circuit31, and a sparse network of analog control lines. The data demonstrate that the unit cell can be used as a computer for solving Ising spin glass problems. Third, we compare experimental data to the results of numerical simulations in order to highlight the fact that, when run very slowly with respect to the adiabatic limit, the performance of the unit cell is influenced by its tendency to thermalize to an environment.

This article is organized as follows: Section II contains a mathematical description of how the physics of an Ising spin glass can be mapped onto superconducting hardware composed of rf-SQUID flux qubits and couplers. Section III contains a brief review of the hardware that has been fabricated and key calibration data from one of the eight-qubit unit cells. This particular unit cell was used to solve a large number of Ising spin glass problem instances and the results have been summarized in Section IV. A dynamical model of the unit cell coupled to an environment in thermal equilibrium is introduced in Section V. Section VI contains a brief discussion of the key results of this article and presents a series of important open questions that have been motivated by this work. The conclusions are summarized in Section VII. For convenience, we have provided a glossary of abbreviations used in this article in Appendix A. A table of specific example problem instance settings referred to in Sections IV and V has been placed in Appendix B.

II. MAPPING ADIABATIC QUANTUM OPTIMIZATION ONTO HARDWARE

We begin by mapping a particular class of optimization problems onto a scalable processor architecture that uses superconducting flux qubits. Significant emphasis will be placed upon minimizing the number of unique time-dependent control signals, thus making efficient use of the limited number of externally supplied biases that
can possibly be routed to such a processor. On the other
hand, it will be assumed that unique time-independent
control signals can be generated locally on chip using
a scalable form of programmable magnetic memory
(PMM) that has been described in detail elsewhere31.

Let there be an optimization problem of the form
\[ E(\vec{s}) = -\sum_i h_i s_i + \sum_{i,j \neq i} K_{ij} s_i s_j \]  
(1)
where \( s_i = \pm 1 \), \(-1 \leq h_i \leq +1 \), and \(-1 \leq K_{ij} \leq +1 \). For
any given set of \( h_i \) and \( K_{ij} \) there exists at least one optimal
solution \( \vec{s}_{\text{opt}} \) that minimizes the objective \( E \). Finding
\( \vec{s}_{\text{opt}} \) for such a system of coupled variables can be NP-
hard32. Equation (1) can be mapped onto an Ising spin
glass (ISG) Hamiltonian33
\[ \mathcal{H}_{\text{ISG}} \bigg|_{E_0} = -\sum_i h_i \sigma_z^{(i)} + \sum_{i,j > i} K_{ij} \sigma_z^{(i)} \sigma_z^{(j)} \]  
(2)
where \( \sigma_z^{(i)} \) are Pauli matrices acting upon spin \( i \) and
\( E_0 \) is a convenient energy. When cast in this form, \( \vec{s}_{\text{gs}} \)
represents the ground state of the ISG. The objective is
to design a physical system that reliably finds \( \vec{s}_{\text{opt}} \).

One possible means of finding \( \vec{s}_{\text{gs}} \) is via the adiabatic method described in Ref. 2. In an idealized picture of
AQO, the Hamiltonian of the system at arbitrary time \( t \)
can be expressed as that of a quantum Ising spin glass
(QSG). In such a system, there are pairwise magnetic
interactions between spins \( i \) and \( j \propto K_{ij} \) and each spin \( i \)
is subjected to a local longitudinal magnetic field \( h_i \)
and a global transverse magnetic field \( \Gamma(t) \):
\[ \mathcal{H}_{\text{QSG}}(t) = -\sum_i h_i \sigma_z^{(i)} + \sum_{i,j > i} K_{ij} \sigma_z^{(i)} \sigma_z^{(j)} - \Gamma(t) \sum_i \sigma_x^{(i)} \]  
(3)
One finds the lowest energy solution to an optimization
problem encoded in the quantities \( h_i \) and \( K_{ij} \) by guiding
the physical system through an evolution path described
by \( \Gamma(t) \) subject to the following constraints:
\[ \Gamma(0) \gg h_i, K_{ij} \]  
\[ \Gamma(t_f) \ll h_i, K_{ij} \]  
where \( t_f > 0 \) is the run time of the algorithm. Here,
it is assumed that the physical system readily reaches
its ground state \( |\Pi_0 \rangle \) at \( t = 0 \) and, in the best of
circumstances, will remain in the ground state during the
course of the evolution, thus yielding \( |\vec{s}_{\text{gs}} \rangle \) at \( t = t_f \).
Understanding the conditions under which adiabaticity is
violated, thus yielding final states other than \( |\vec{s}_{\text{gs}} \rangle \), is a
matter of considerable debate at the moment15–16.

It was recognized in Ref. 21 that one could implement
a QSG using a network of rf SQUIDs in which some are
designed to fulfill the role of qubits and others to fulfill
the role of tunable inductive interqubit couplers.
The physics of the two lowest lying states of an isolated flux
qubit \( i \) can be captured by an effective Hamiltonian
\[ \mathcal{H}_i = -\frac{1}{2} \left[ \epsilon_i \sigma_z^{(i)} + \Delta q \sigma_x^{(i)} \right] \]  
(4)
where \( \epsilon_i = 2 |I_q^p| (\Phi^x_q - \Phi^y_q) \) and \( \Delta q \) are the bias and
tunneling energy, respectively. Here, \( |I_q^p| \) represents
the magnitude of the supercurrent flowing about the rf-
SQUID loop, \( \Phi^x_q \) is the external flux bias applied to the
qubit loop about which the supercurrent flows and \( \Phi^y_q \)
represents the qubit degeneracy point. Upon comparing
with like symmetry in Hamiltonians (3) and
(4), one can see that \( \Delta q = 2 = E_0(t) \Gamma(t) \). In order to
implement an AQO algorithm in which \( \Gamma(t) \) is altered,
one must provide a means for tuning \( \Delta q \) in situ. This
is accomplished by incorporating at least one dc-
SQUID loop into an rf-SQUID body28,34–37. Since we
have used the compound-compound Josephson junction
(CCJJ) rf SQUID in our experiments28, let the external
flux bias applied to this device be represented by \( \Phi^x_{ccjj} \).
This leads to the substitution \( \Delta q \rightarrow \Delta q (\Phi^x_{ccjj}) \) in Hamiltonian (4). Rather unfortunately, \( |I_q^p| \) is also a function of \( \Phi^x_{ccjj} \) in any compound junction rf-SQUID flux qubit:
\[ |I_q^p| \rightarrow |I_q^p (\Phi^x_{ccjj}) | \]. Therefore, one does not naturally ob-
tain orthogonal control of \( \epsilon_i \) and \( \Delta q \) via \( \Phi^x_q \) and \( \Phi^x_{ccjj} \)
respectively. This drawback is by no means unique to
the CCJJ design, as it occurs in all flux qubits with in
situ tunable \( \Delta q \) that have been reported on in the litera-
ture to date. This intrinsic lack of orthogonal control of
the two defining parameters of the flux qubit, \( |I_q^p| \) and
\( \Delta q \), has driven some of the design choices made in
implementing the QSG discussed herein.

Assuming that one has a set of flux qubits for which
\( |I_q^p (\Phi^x_{ccjj}) | \) and \( \Delta_q (\Phi^x_{ccjj}) \) are nominally identical, a
general time-dependent Hamiltonian for a system of in-
ductively coupled flux qubits can be written as
\[ \mathcal{H}_0(t) = -\frac{1}{2} \sum_i \left[ 2 |I_q^p(t)| (\Phi^x_q(t) - \Phi^y_q(t)) \sigma_z^{(i)} + \Delta_q(t) \sigma_x^{(i)} \right] + \sum_{i,j > i} M_{ij} (t) |I_q^p(t)|^2 \sigma_z^{(i)} \sigma_z^{(j)} \]  
(5)
where the time dependence of the qubit parameters \( |I_q^p| \)
and \( \Delta_q \) is implicitly driven by \( \Phi^x_{ccjj}(t) \). In addi-
tion, there could be time-dependent in the individual qubit flux
biases \( \Phi^x_q(t) \) and in the interqubit effective mutual in-
ductances \( M_{ij} \) via their control biases \( \Phi^x_{ccjj}(t) \).

One of the virtues of the AQO algorithm is its sim-
plicity, which provides some significant advantages when
considering how to implement a processor. Note that
Hamiltonian (5) is a function of a global bias \( \Phi^x_{ccjj}(t) \) to
which all qubits are uniformly subjected. This is
desirable for two reasons: First, Hamiltonian (3) specifies
that the transverse field \( \Delta_q(t)/2 = E_0(t) \Gamma(t) \) be uniform,
which then corresponds to all \( \Delta_q (\Phi^x_{ccjj}) \) being identical.
As a corollary, this renders all \( |I_q^p (\Phi^x_{ccjj}) | \) identical. Sec-
ond, one can provide uniform \( \Phi^x_{ccjj} \) to multiple qubits
using a single global current bias line, as opposed to one
bias line per qubit. This scenario is depicted in Fig. 1,
in which the current bias \( I_{ccjj}(t) \) drives the fluxes \( \Phi^x_{ccjj} \in a
multitude of qubits. In doing so, one substantially
reduces the number of external biases that must be applied

to the processor, thus significantly improving the scaling of this particular architecture. Such design choices are most likely necessary in order to realize multiqubit processors that contain more than 10’s of qubits.

Given the ideal Hamiltonian (3) and the flux qubit-based device Hamiltonian (5), one must now map the problem specification, embodied by \( h_i \) and \( K_{ij} \) in Eq. (1), onto external control parameters. This can be accomplished by scaling Hamiltonian (5) and then performing a term-by-term comparison to Hamiltonian (3). A convenient scaling factor is the interqubit coupling energy when a coupler is set to provide maximum antiferromagnetic (AFM) coupling, \( M_{ij} (\Phi_{co,ij}) = M_{AFM} \). Define this energy scale as \( J_{AFM}(t) = M_{AFM} |I_q(t)|^2 \). Rearranging Hamiltonian (5) yields

\[
\frac{\mathcal{H}_0(t)}{J_{AFM}(t)} = -\sum_i h_i \sigma_z^{(i)} + \sum_{i,j \neq i} K_{ij} \sigma_z^{(i)} \sigma_z^{(j)} - \Gamma(t) \sum_i \sigma_z^{(i)},
\]

(6a)

\[
h_i = \frac{|I_q(t)| (\Phi_{c}^+(t) - \Phi_{0}^+(t))}{M_{AFM} |I_q(t)|^2} = \frac{\Phi_{c}^+(t) - \Phi_{0}^+(t)}{M_{AFM} |I_q(t)|},
\]

(6b)

\[
K_{ij} = \frac{M_{ij}(t)|I_q(t)|^2}{M_{AFM} |I_q(t)|^2} = \frac{M_{ij}(t)}{M_{AFM}},
\]

(6c)

\[
\Gamma(t) = \frac{\Delta_q(t)}{2J_{AFM}(t)}.
\]

(6d)

In order to solve a particular optimization problem, \( h_i \) and \( K_{ij} \) must be time-independent. According to Eq. (6c), one must hold all \( M_{ij} \) constant during operation. This is convenient as it obviates the application of individually tailored time-dependent flux bias signals to each interqubit coupler. Rather, one need only apply a static control signal \( \Phi_{co,ij} \) to each coupler, as depicted in Fig. 1, that can be provided by PMM. On the other hand, according to Eq. (6b), one must apply time-dependent qubit flux biases of the form

\[
\Phi_{q}^+(t) = \Phi_{0}^+ + h_i \times M_{AFM} |I_q|^2
\]

(7)

to render \( h_i \) time-independent. Thus, it is necessary to provide a custom-tailored time-dependent control signal plus a static offset to every qubit. The static component \( \Phi_{0}^+ \) can be provided by PMM. As for the time-dependent component, providing these signals with one external bias per qubit would not constitute a scalable approach for building a multiqubit processor. Rather, one can take advantage of the fact that, according to Eq. (7), all qubits must receive a control signal with the same time-dependent shape but with custom-tailored time-independent scale factors \( \propto h_i \). A scalable architecture for providing these signals is depicted in Fig. 1 and has been further expounded upon in Ref. 31. Here, a single global current bias \( I_{q}(t) = \alpha |I_q(t)| \), where \( \alpha \) is a convenient scale factor, is coupled to multiple qubits via in situ tunable mutual inductances of magnitude \( M_i = h_i M_{AFM}/\alpha \). The very same type of device that is used to provide in situ tunable interqubit coupling can be retooled to provide coupling between flux qubits and a global bias line. Moreover, each \( M_i \) can be controlled with a static flux bias \( \Phi_{i}^+ \) provided by PMM. We will hereafter refer to this architecture as persistent current (\(|I_q|\)) compensation, as it is a means of compensating for changing \(|I_q(t)| \) such that the ISG problem specified by \( h_i \) and \( K_{ij} \) remains on target throughout annealing.

To summarize up to this point, a prescription for implementing AQO to solve ISG problems using a network of inductively coupled CCJJ rf-SQUID flux qubits has been presented. A problem specified by a set of \( h_i \) and \( K_{ij} \) can be embedded in the hardware using time-independent interqubit couplings controlled by PMM and time-dependent qubit flux and CCJJ biases. The qubit flux bias signals can be supplied using a combination of static flux offsets provided by PMM and a single global signal \( I_{q}(t) \) that is applied to each qubit through in situ tunable couplers that are also controlled by PMM. The CCJJ bias can, in principle, also be provided to all qubits simultaneously using a single global control signal.
III. DEVICE ARCHITECTURE AND CALIBRATION

With the mapping of the AQO algorithm onto the scalable architecture completed, we now turn to a high level description of a superconducting chip whose architecture embodies that algorithm. All of the principal components of the processor, namely the qubits \(q\), couplers \(c\), readout \(r\), and PMM \(m\) have been described in detail in other publications. As such, we will only provide brief summaries of the important points as pertaining to the functioning of the collective system herein.

As stated previously, we have incorporated CCJJ rf-SQUID flux qubits in our design. A schematic of an isolated flux qubit with the two external bias controls relevant for this study is shown in Fig. 2(a). This particular qubit is robust against fabrication variations in the Josephson-junction critical currents and facilitates the homogenization of the net critical current among a population of such qubits. This device also contains an inductance \(L\) tuner that can be used to compensate for variations of qubit inductance due to fabrication and from tuning the interqubit couplers. To each qubit we have added an \(|I_p|\) compensator, as introduced in Sect. II. We have provided PMM to flux bias the two minor lobes of each CCJJ, the \(L\) tuner, the \(|I_p|\) compensator, and the qubit body \(\Phi_0\) for all qubits on the chip.

The chip that was used for this study was composed of sixteen eight-qubit unit cells that were tiled on a \(4 \times 4\) square grid. To limit the scope of this article, we focus upon a single unit cell near the center of the chip. A discussion of the complete processor, with multiple unit cells acting in concert, will be reserved for a future publication. A schematic layout of the unit cell is shown in Fig. 2(b). The qubits are depicted as extended horizontal and vertical loops with 16 compound Josephson junction (CJJ) couplers located at the intersections of the qubits. Additional couplers at both extrema of the qubit bodies connect them to qubits in neighboring unit cells to the left, right, top and bottom. These latter couplers were set to zero coupling to isolate the single unit cell for this study. Each qubit was connected to its own quantum flux parametron (QFP) enabled readout. Not shown in Fig. 2(b) are the PMM elements, the demultiplexing tree for addressing the PMM, or the analog bias lines used to calibrate and operate the unit cell.

A simplified representation of the hardware connectivity is depicted in Fig. 2(c). Here, qubits and couplers correspond to the vertices and edges of a graph. As with the unit cell schematic, this eight-vertex graph can be tiled to the left, right, top, and bottom in order to generate larger graphs. Each vertex is connected to a minimum (maximum) of 4 (6) other vertices, depending upon the number and arrangement of unit cells used to form a larger graph. This hardware does not provide full connectivity within a population of \(N\) qubits in which each physical qubit is connected to \(N-1\) other physical qubits. This limitation can potentially be overcome, at the cost of reducing the number of unique vertices < \(N\), by using a ferromagnetically coupled chain of physical qubits to form a single logical qubit, as suggested in Ref. 22.

The chip used in these experiments was fabricated on an oxidized Si wafer with Nb/Al/Al\(_2\)O\(_3\)/Nb trilayer junctions and four Nb wiring layers separated by planarized
plasma enhanced chemical vapor deposited SiO$_2$. A scanning electron micrograph of the fabrication cross-section can be found in Fig. 4(b) of Ref. 28. The Nb metal layers are referred to as BASE, WIRA, WIRB and WIRC, from bottom to top, respectively. Flux qubit wiring was primarily located in WIRB and consisted of 2-μm-wide leads arranged as an approximately 900-μm-long differential microstrip located 200 nm above a ground plane in WIRA. Coupler wiring was primarily located in WIRC, stacked on top of the qubit wiring to provide inductive coupling. PMM flux storage loops were implemented as stacked spirals of 13-20 turns of 0.25-μm-wide wiring with 0.25 μm separation in BASE and WIRA (WIRB). Stored flux was picked up by one-turn washers in WIRB (WIRA) and fed into transformers for flux-biasing devices. External control lines were mostly located in BASE and WIRA. Resistors that were used in the PMM demultiplexing circuit were made from a TiPt layer referred to as RESI. All of these control elements resided below a ground plane in WIRC. The ground planes under the qubits and over the PMM/external control lines were electrically connected using extended vias in WIRB so as to form a nearly continuous superconducting shield between the analog devices on top and the bias circuitry below. Transformers for biasing qubits, couplers, QFPs and dc SQUIDs were enclosed in superconducting boxes with BASE and WIRC forming the top and bottom, respectively, and vertical walls formed by extended vias in WIRA and WIRB. Minimally sized openings were placed in the vertical walls through which the bias and target device wiring passed at opposing ends of each box. This design reduced most on-chip parasitic crosstalk to a negligible level.

An optical image of a unit cell completed through the processing of WIRB is shown in Fig. 3(a). One can discern the trenches in which the qubit wiring resides, where the bottom is formed by ground plane in WIRA that is electrically connected to thick vertical walls that are formed by extended vias in WIRB. A number of PMM flux storage and pickup loops are visible as spirals and washers, respectively. The PMM circuitry is not visible in a completed chip as it resides below patches of ground plane in WIRC that are electrically connected at their perimeter to the top of the vertical walls defining the qubit trenches. Thus the qubits (PMM) reside above (below) the nearly contiguous contiguous shielding layer in a completed chip. Note that much of the coupler wiring is absent from this image as it resides in WIRC atop the qubit wiring in WIRB.

Tiling of the eight-qubit unit cell to make a larger processor is explicitly demonstrated in Fig. 3(b). One can discern a 4 × 4 array of eight-qubit unit cells. The particular unit cell used in this study was in the third column and third row. Outside of the field of this image are four blocks of on-chip hysteretic dc SQUIDs, with pickup loops oriented in-plane and orthogonal to the wafer, that are used as magnetometers to measure local fields during active field compensation, as well as wiring channels that run to wirebonding pads. There are 128 qubits, 352 couplers and 128 readouts on this processor. These analog components would require a total of 1632 unique biases to operate if no effort was made to develop a scalable control architecture. It would be impractical to build such a processor with so many independent external bi-
ases due to circuit density and wirebonding constraints. However, our complete processor, consisting of the analog components enumerated above and 992 PMM elements, requires only 84 pairs of differentially-driven external biases to calibrate and operate.

The chip was enclosed in a superconducting Al shield ($T_c \approx 1.2 \text{ K}$) that was inside two coaxial $\mu$-metal shields. All three shielding layers and the sample were thermalized to the mixing chamber of a pulse tube dilution refrigerator. The refrigerator itself was located inside two room temperature coaxial $\mu$-metal shields. The magnetic field in the vicinity of the chip was minimized at 3.9 K using triple-axis Cu compensation coils located outside of the Al shield. The array of on-chip dc-SQUID magnetometers surrounding the qubit circuitry was used to infer the magnetic field vector near the processor. The current in the compensation coils was controlled via feedback implemented in software. Once the magnitude of the three-dimensional magnetic field vector had been minimized, the chip was thermally cycled to $\gtrsim 9.5 \text{ K}$ to expel trapped flux from the Nb chip ($T_c \approx 9.1 \text{ K}$) and then cooled to base temperature $T \approx 19 \text{ mK}$. Thereafter, the current applied to the compensation coils was removed. The residual magnetic field was estimated to be $\sim 2 \text{ nT}$ ($3 \text{ nT}$) perpendicular (parallel) to the plane of the chip.

External biases applied to the chip were provided by a custom-built 14-bit 128-channel differentially driven programmable 50 MHz arbitrary waveform generator. The signals generated at room temperature were sent into the refrigerator using individually shielded twisted pair wires that were connected to a combination of lumped-element and copper powder filters at the mixing chamber. To minimize the effect of environmental noise and since the AQO algorithm does not require resonant excitation of any element of the circuit, we chose to restrict the bandwidth of these bias lines to 5 MHz. Signals were then routed onto Sn traces on a printed circuit board ($T_c \approx 3.7 \text{ K}$) and passed onto the chip via Al wirebonds.

A summary of the flux qubit parameters that are relevant for this study is presented in Fig. 4. Data have been plotted versus normalized flux, where $\Phi_0 \equiv h/2e$ is the flux quantum. The reader is referred to Ref. 28 for a detailed description of the CCJJ rf-SQUID flux qubit and the experimental methods used to obtain $|I_q^g(\Phi_{ccjj})|$. We have obtained $\Delta_q(\Phi_{ccjj})$ using two methods. First, we used the single qubit Landau-Zener (1QLZ) method\(^3^9\) to yield $2 \times 10^6 \text{ Hz} \lesssim \Delta_q/h \lesssim 10^7 \text{ Hz}$. The lower and upper bound were set by experimental run time constraints and the bandwidth of our external control lines, respectively. Second, we ran two-qubit Landau-Zener (2QLZ) experiments on pairs of qubits that were strongly coupled with $M_{ij} = -2.82 \text{ pF}$. The Hamiltonian for a pair of qubits, subject to zero net flux bias, can be expressed as

$$\mathcal{H}_{ij} = -\frac{1}{2} \left[ \Delta_i \sigma_x^{(i)} + \Delta_j \sigma_x^{(j)} \right] + M_{ij} \left| I_p^i \right| \left| I_p^j \right| \sigma_z^{(i)} \sigma_z^{(j)} . \quad (8)$$

Data from a 2QLZ experiment yields a spectral gap $g$, which can be compared to that obtained by solving for the eigenenergies of Hamiltonian (8). The range of $g$ that could be inferred from measurements was also limited by the run time and bandwidth constraints cited above. In the first variant of 2QLZ, we locked the target $\Phi_{ccjj}$-biases of the qubits together. Assuming that $\Delta_i = \Delta_j = \Delta_q$, one can use Hamiltonian (6a) to solve for $\Delta_q(\Phi_{ccjj})$ given the independently calibrated $g(\Phi_{ccjj})$, $|I_p^i(\Phi_{ccjj})|$, and $|I_p^j(\Phi_{ccjj})|$. This procedure proved effective for extracting $5 \times 10^7 \text{ Hz} \lesssim \Delta_q/h \lesssim 2 \times 10^8 \text{ Hz}$. In the second variant of 2QLZ, we targeted $\Phi_{ccjj}$-biases of qubit $i$ (a bias for which we have calibrated $\Delta_i$ using the 1QLZ method) and scanned the target $\Phi_{ccjj}$ of qubit $j$. We then used Hamiltonian (6a) to solve for $\Delta_j = \Delta_q(\Phi_{ccjj})$ given the independently calibrated $g(\Phi_{ccjj})$, $\Delta_i(-0.6235 \Phi_0)$, $|I_p^i(-0.6235 \Phi_0)|$, and $|I_p^j(\Phi_{ccjj})|$. This procedure proved effective for extracting $4 \times 10^8 \text{ Hz} \lesssim \Delta_q/h \lesssim 3 \times 10^9 \text{ Hz}$.

Prior to collecting the data shown in Fig. 4, we had fully exercised the CCJJ minor lobe and L-tuner biases.
on each of these qubits in order to homogenize \( |I_p^j| \) at \( \Phi_{ccjj}^e / \Phi_0 = -0.6146 \) and \(-0.6325\) among the eight qubits. Residual discrepancies between \( |I_p^j| \) and \( \Delta_q \) of different qubits at arbitrary \( \Phi_{ccjj}^e \) were attributed to small variations in the mutual inductance between the global CCJJ current bias line and the individual qubits. Given that the experimental data agree with a multilevel quantum mechanical of the CCJJ rf SQUID whose input parameters were independently calibrated, we are confident in our identification of these devices as flux qubits.\(^{28}\)

The interqubit couplers and \( |I_p^j| \) compensators on this chip were calibrated according to the methods described in Ref. 29. The experimental data are summarized in Fig. 5. From the set of calibrations for all 352 interqubit couplers on the chip, we determined that the smallest maximum AFM coupling was \( M_{AFM} = 1.88 \pm 0.03 \) pH, which was readily achievable for all sixteen couplers within the unit cell used in this study. We chose this value of \( M_{AFM} \) to define the energy \( J_{AFM} \) introduced in Sect. II. Likewise, from the complete set of 128 \( |I_p^j| \) compensators on this chip we determined that the maximum usable AFM coupling was \( \max |M_i| = 0.100 \pm 0.002 \) pH. We found it convenient to use the scale factor \( \alpha = 2 \times M_{AFM} / \max |M_i| = 37.6 \) to scale the \( I_p^j \) compensation signals, as discussed in Sect. II.

\[ M_{ij}(\Phi_{ccjj}^e / \Phi_0) \]

\[ M_i(\Phi_{ccjj}^e / \Phi_0) \]

\[ M_i(\Phi_{ccjj}^e / \Phi_0) \]

FIG. 5: (Color online) (a) Mutual inductance \( M_{ij} \) versus coupler flux bias \( \Phi_{ccjj}^e \) for all sixteen interqubit couplers in the unit cell. Solid curves are the theoretical \( M_{ij} \) obtained using the mean of the individual coupler fit parameters. (b) Mutual inductance \( M_i \) versus \( |I_p^j| \)-compensator flux bias \( \Phi_{ccjj}^e \). Solid curve is the theoretical \( M_i \) obtained using the mean of the individual \( |I_p^j| \)-compensator fit parameters.

The theoretical predictions from a quantum mechanical Hamiltonian describing the CCJJ rf SQUID, as given by Eqs. (5a)→(5d) in Ref. 28, are shown as solid curves in Fig. 4. The model results were obtained using the means of the independently calibrated CCJJ rf SQUID parameters, namely the body inductance \( L_{body} = 339 \pm 3 \) pH, CCJJ loop inductance \( L_{ccjj} = 26 \pm 1 \) pH, critical current \( I_c = 3.355 \pm 0.040 \) μA and capacitance \( C_q = 182 \pm 4 \) fF.

Finally, to characterize the readout fidelity of the unit cell, we set all interqubit couplers to zero coupling and then sequentially prepared the eight qubits into each of the \( 2^8 = 256 \) possible final spin configurations and recording the number of correct reads for 1000 repetitions. (b) Probability of observing a given number of errors within 1000 repetitions, as taken from the data in (a). Data have been fit to a Poisson distribution with mean number of errors \( \lambda = 0.36 \).

\[ \text{Number Correct} \]

\[ \text{Probability} \]

\[ \lambda = 0.36 \]

FIG. 6: (Color online) Verification of readout fidelity for the eight-qubit unit cell. (a) Example results from sequentially preparing the eight qubits into each of the \( 2^8 = 256 \) possible final spin configurations and recording the number of correct reads for 1000 repetitions. (b) Probability of observing a given number of errors within 1000 repetitions, as taken from the data in (a). Data have been fit to a Poisson distribution with mean number of errors \( \lambda = 0.36 \).
binary word $|s\rangle$ into a decimal number $|\text{State}\rangle$, where

$$\text{State} = \sum_{i=0}^{7} \frac{s_i + 1}{2} 2^{-i}.$$  

In the majority of instances the readout yielded the expected result 1000 times. We never observed more than 3 errors per 1000 repetitions. Multiple iterations of this check revealed that results for those states that appear to be less than perfect in Fig. 6(a), such as State=143, were not reproducible. Further analysis indicated that these infrequent errors were not due to problems in preparing $|s\rangle$. Rather, the errors were most likely generated by the statistical nature of the dc SQUID switching times. Note that while other researchers have chosen to focus upon particular classes of Ising spin problems, such as random 3-SAT, we have found our random ISG problem set to be of great utility as it encompasses a broad range of problem classes that could be posed to an AQO processor in practice.

**IV. EXPERIMENTAL PROCESSOR PERFORMANCE**

To test the performance of the eight-qubit unit cell, we generated a set of 800 unique random Ising problem instances whose $h_i$ and $K_{ij}$ were drawn from a distribution of 15 evenly spaced values within the set

$$h_i, K_{ij} \in [-1 - 6/7 \ldots + 6/7 + 1].$$  

We then took advantage of the symmetry of the physical layout of the qubits and couplers shown in Fig. 2b and generated 8 permutations of each of the 800 progenitor problems by rotating and reflecting their $h_i$ and $K_{ij}$ embeddings about the center of the unit cell. The result was a test set of 6400 problem instances, each requiring a unique embedding, that were then posed to the processor. To compensate, the PMM controlling the qubit flux offset $\Phi_0^s$ was updated.

With the details concerning the PMM aside, the analog control signal sequence for implementing the AQO algorithm introduced in Sect. II was quite simple, as depicted in Fig. 7. During the annealing phase there were only two time-dependent signals applied to the entire unit cell: the CCJJ control signal $\Phi_{\text{ccjj}}^x(t)$ and the $|I_g|^2$-compensation signal $I_g(t)$. For simplicity, we chose to use a relatively slow 100-µs-long linear ramp of $\Phi_{\text{ccjj}}^x(t)$ through the region relevant for qubit operation between $\Phi_{\text{ccjj}}^x/\Phi_0 = -0.590$ and $\Phi_{\text{ccjj}}^x/\Phi_0 = -0.650$, where the superscripts $m$ and $b$ denote biases for which the CCJJ rf-SQUID potential energy is deep in the monostable and bistable regime, respectively. The aforementioned ramp time was sufficiently long to avoid the loss of ground state probability to Landau-Zener transitions at anticrossings. Experimental data and numerical simulations support-
ing this latter statement will be discussed in Sect. V. The time-dependent wave form \( I_q(t) = \alpha |\Psi_q^\alpha(t)| \), with \( \alpha = 37.6 \), was constructed by sampling the modeled \( |\Psi_q^\alpha(\Phi_{ccjj}^\alpha(t))| \), shown as the solid curve in Fig. 4(a), at 200 points between \( \Phi_{ccjj}^\alpha \) and \( \Phi_{ccjj}^\beta \).

At the end of the annealing phase, the final task that involved time-dependent wave forms was that of reading out the state of the unit cell. The states of all eight qubits were simultaneously latched by raising the QFP tunnel barriers in unison using their global bias \( \Phi_{batch}^q(t) \). Thereafter, the dc-SQUID current biases and flux biases were modulated to serially read out the state of each QFP. In Fig. 7, only those current and flux bias wave forms for operating the readout connected to \( q_0 \) have been depicted. The entire wave form sequence, including programming of the 992 PMM elements, a 1 ms cooling period following PMM programming, and 1000 repetitions of the anneal and read sequence depicted in Fig. 7 was completed in \( \approx 2.5 \text{ s} \).

The key experimental parameter that was varied in this study was the temperature of the processor \( T \). We calibrated \( T \) in situ using macroscopic resonant tunneling (MRT) measurements versus \( \Phi_q^x \) at \( \Phi_{ccjj}/\Phi_0 = -0.6238 \) using a dwell time of 1 ms on qubits \( q_0 \) and \( q_4 \). At this particular CCJJ bias, the qubits reached thermal equilibrium within the specified dwell time and the resultant population statistics versus \( \Phi_q^x \) could be fit to the form

\[
P_{MRT} = \frac{1}{2} \left[ 1 + \tanh \left( \frac{2 |I_q^p| (\Phi_q^x - \Phi_0)}{2k_BT} \right) \right],
\]

where \( |I_q^p| = 1.26 \pm 0.02 \mu\text{A} \) was obtained from the independent calibration shown in Fig. 4(a) and \( \Phi_0^q \) was the independently calibrated qubit flux offset. This then left only one free parameter, namely \( T \).

The unit cell was tested using the set of 6400 problem instances at three values of \( T \). Each instance was run 1000 times and the probabilities of observing each of the 256 possible \( |s\rangle \) were recorded. Given the small size of the problems it was trivial to independently determine \( |s_{gs}\rangle \). Histograms of the probability of observing the ground state \( P_{gs} \) are shown in Fig. 8. The output of numerical simulations shown in these plots will be discussed in Sect. V. Here, the vertical axes correspond to the number of problem instances for which the hardware returned \( P_{gs} \) within the bounds of each bin of the histogram. The data taken at \( T = 20 \text{ mK} \), see Fig. 8(a), show an obvious spike in the highest bin, \( 0.95 < P_{gs} \leq 1 \), and a tail that appears to be exponentially suppressed at lower \( P_{gs} \). For \( T = 35 \text{ mK} \), see Fig. 8(b), the spike in the highest bin appears diminished and a broad hump in the vicinity of \( P_{gs} \approx 0.6 \) is observed. Finally, for \( T = 50 \text{ mK} \), see Fig. 8(c), the high \( P_{gs} \) spike has been flattened and the broad hump has shifted to slightly lower \( P_{gs} \approx 0.5 \). Nonetheless, for all three values of \( T \), \( P_{gs} > 0.1 \) for all instances.

To gain further insight into the source of the structure of the histograms shown in Fig. 8, we selected an exemplary problem instance from each of the three features in the \( T = 35 \text{ mK} \) results: Example A from the spike (\( P_{gs} \sim 1 \)), Example B from the broad hump (\( P_{gs} \sim 0.6 \)), and Example C from the tail (\( P_{gs} \sim 0.4 \)). The instance settings \( h_i \) and \( K_{ij} \) for these three examples are given in Appendix B. We then ran each of these example instances 1000 times at a series of temperatures 20 mK \( \leq T < 100 \text{ mK} \). A summary of the observed population statistics \( P \) is shown in Fig. 9. The output of numerical simulations shown in these plots will be discussed in Sect. V. For Examples A, B, and C, a total of 28, 13, and 38 states were observed above the read-

![FIG. 8: (Color online) Histograms of the experimentally observed and the simulated probability of finding the system in the ground state \( P_{gs} \) from a test involving 6400 unique random ISG instances at three different temperatures: (a) \( T = 20 \text{ mK} \), (b) \( T = 35 \text{ mK} \), and (c) \( T = 50 \text{ mK} \).](image)
out resolution threshold at \( T \approx 91.7 \text{ mK} \), respectively. We have plotted the data for only those states that were observed with \( P > 0.01 \) at \( T = 91.7 \text{ mK} \).

The results shown in Fig. 9 indicate that a small number of \( |\bar{s}\rangle \) are returned for any given problem instance and that the probability of observing those states evolves smoothly with \( T \). To assist in determining why particular states were observed, we have evaluated the objective function \( E(\bar{s}) \), given by Eq. (1), for Examples A, B, and C using the values of \( h_i \) and \( K_{ij} \) provided in Appendix B.

The results have been summarized in Fig. 10. This plot reveals that the ground state \( |\bar{s}_{gs}\rangle = |101\rangle \), \( |200\rangle \), and \( |97\rangle \) for Example A, B, and C, respectively. Likewise, the first excited state(s) can be identified as \( |\bar{s}\rangle = |69\rangle \), \( |232\rangle \), and the degenerate pair \((|54\rangle,|101\rangle)\) for Example A, B, and C, respectively. From a comparison of Figs. 9 and 10 one can make two important observations: First, for all three examples the order of the observed states, from most to least probable, \textit{exactly} matches the order when ranked by ascending \( E(\bar{s}) \). Thus the ground state is the most probable state, the first excited state is the second most probable, and so forth. The only exception is the order of \( |240\rangle \) and \((|45\rangle,|100\rangle)\) in Example A, though this discrepancy is within the noise. Second, there is a preferred order for some of the pairs of degenerate excited states such as \((|54\rangle,|101\rangle)\) and \((|190\rangle,|117\rangle)\) for Example C. While the preference for one over the other appears to be subtle, it is nonetheless very reproducible.

To summarize this section of the paper, we have tested an eight-qubit unit cell inside a superconducting processor that embodies an AQO algorithm. The unit cell has been tested by running 6400 random ISG problem instances at three temperatures and three representative instances at multiple temperatures. With repeated measurements, the hardware returns a small number of final spin configurations, with the ground state always being the most probable result. The remaining states that are observed correspond to low lying excited states.

V. SIMULATED PROCESSOR PERFORMANCE

The objective of this section of the paper is to present a physical picture that naturally explains the final distribution of \( |\bar{s}\rangle \) that was obtained from any given ISG problem that was posed to the hardware. This picture relies upon having independently calibrated all qubit, coupler, and \( |I_p\rangle \)-compensator parameters that enter into Hamiltonian (6a) and then adding interactions between all eight qubits and an environment. It will be demonstrated that the inclusion of quantum mechanical relaxation\(^\text{46}\) is sufficient to explain the observed distributions.
We have chosen to model this system by neglecting all but the two lowest lying states in each CCJJ rf SQUID (qubit approximation). Furthermore, we have also assumed that the qubit parameters \( P(\Phi_{ccjj}^x) \) and \( \Delta_q(\Phi_{ccjj}^x) \) are the same for all qubits and are given by the solid curves in Fig. 4. This latter assumption would not be justified without having first leveraged all of the features of the CCJJ rf SQUID so as to synchronize their qubit parameters to a reasonable level. The low energy physics of the ideal closed system will then be dictated by Hamiltonian (6a).

Instantaneous eigenspectra as a function of annealing parameter \( \Phi_{ccjj}^x \) for example problem instances A, B, and C, as generated using Hamiltonian (6a), are shown in Fig. 11. Note the negative sign in front of the independent variable in these plots, which has been inserted such that annealing progresses from left to right. The eigenenergies \( E \) have been plotted relative to the ground state energy \( E_0 \) such that the ground state is coincident with the horizontal axis. The lowest 16 levels have been plotted for Examples A and B and the lowest 32 levels for Example C. In all cases the eigenspectra have been truncated above the second excited state of an isolated CCJJ rf SQUID, here denoted as \( E_{rfs} \), as Hamiltonian (6a) is not the correct representation of the physics of the closed system beyond this energy. This latter point is of significant practical interest and will receive more attention in Sect. VI. The initial states have been identified on the left of each plot. Here, the ground state is denoted as 

\[
|\Xi_0\rangle = \frac{1}{16} \prod_{l=0}^{7} (-1)^l + 1
\]

and is a product of the isolated qubit ground states. There is an octet of first excited states, denoted as \( |\Xi_{1,m}\rangle \), which are product states of all qubits in their ground state except a single qubit \( m \) in its excited state, \((-1)^m + 1\rangle_{m}\rangle / \sqrt{2}\). Analogously, there are 64 second excited states, denoted as \( |\Xi_{2,m,n}\rangle \), in which all but two qubits are found in their ground state. In all cases, the excited states drop in energy with increasing annealing parameter until they reach a minimum in the vicinity of \( \Phi_{ccjj}^x / \Phi_0 \sim -0.615 \). Each of the minima correspond to a point in the annealing process at which the ground state, \( |\Xi_0\rangle \), sheds a localized spin configuration \( |\bar{s}\rangle \) from the superposition, thus eventually localizing in the solution state \( |\bar{s}_{gs}\rangle \). Likewise, the excited states become localized in configurations \( |\bar{s}\rangle \neq |\bar{s}_{gs}\rangle \) as they are ejected from the ground state, and then increase in energy to the right of the minima. In a large-scale QSG, this process would give rise to a second-order phase transition between a ground state possessing a high degree of symmetry (a superposition of all \( |\bar{s}\rangle \)) to the left of the minima and a ground state with broken symmetry (localized in \( |\bar{s}_{gs}\rangle \)) to the right of the minima. Those \( |\bar{s}\rangle \) that were observed with \( P > 0.01 \) at \( T = 91.7 \) mK, as indicated in Fig. 9, have been labeled on the right of each corresponding plot in Fig. 11.

Each plot in Fig. 11 has been marked with a vertical dashed line at \( \Phi_{ccjj}^x / \Phi_0 = -0.622 \) to roughly delineate where in the annealing process the population statistics appear to freeze. As a crude approximation, one can take the energies of the excited states along this vertical line and calculate Boltzmann occupation factors that roughly agree with the results presented in Fig. 9. However, such a simplistic picture cannot explain the very reproducible
disparity between the observed probabilities of excited states that are degenerate along the vertical line.

To begin the discussion of our dynamical model, it is prudent to briefly address what is not the source of either the structure of the $P_{gs}$ histograms in Fig. 8 or the nonzero probability of observing excited states in Fig. 9. We argue that none of these observations can be explained by a violation of the adiabatic condition. First, we have studied all 800 progenitor random ISG problems by numerically solving a dynamical model the system described by Hamiltonian (6a) in the absence of environmental noise and subject to the wave forms depicted in Fig. 7. These simulations revealed that the 100 $\mu$s CCJJ bias ramp time was sufficiently long as to deem the evolution of the closed system adiabatic, as they always returned $P_{gs} = 1$ to within numerical error. Second, we have experimentally studied Examples A, B, and C at multiple CCJJ bias ramp times from 10 $\mu$s to 1 ms and observed relatively little change in the outcome. Bandwidth limits on our apparatus prevented us from studying the system subject to shorter ramp times.

We now turn to a description of the open system model that we have investigated. We wish to first deal with the physics of relaxation and then discuss the effects of dephasing thereafter. Furthermore, we must carefully distinguish between AQO, in which the evolution of the closed system can be described as being adiabatic, and QA, in which the coupled qubit system exchanges energy with its environment. The Hamiltonian for this open system can be written as

$$\mathcal{H}(t) = \mathcal{H}_0(t) + \frac{1}{2} \sum_{k=0}^{7} 2 \left| I^p_q(t) \right| Q_i(t) \sigma_z^{(i)} ,$$  

where $\mathcal{H}_0(t)$ is given by Hamiltonian (6a) and $Q_i(t)$ is an operator that acts upon the environment seen by qubit $i$ that has units of flux. Note the factor of $\left| I^p_q(t) \right|$ preceding each $Q_i(t)$, which captures the fact that as annealing progresses that the coupling between qubit $i$ and environmental flux noise will change. The noise spectral density for the environment surrounding qubit $i$ as a function of angular frequency $\omega$ can be expressed as

$$S_i(\omega) = (2 \left| I^p_q(0) \right|)^2 \int dt' e^{i\omega t'} \langle Q_i(t') Q_i(0) \rangle .$$  

For Markovian noise, the fluctuation-dissipation theorem allows one to write

$$S_i(\omega) = \frac{2\text{Im}\{L_0(\omega)\} (2 \left| I^p_q(t) \right|)^2 \hbar \omega}{1 - \exp \left( -\frac{\hbar \omega}{k_B T} \right)} ,$$  

where $L_0(\omega)$ can be thought of as an inductance per unit bandwidth that is imparted to qubit $i$ by its local environment. At a more fundamental level, $L_0$ must be proportional to the environment’s complex magnetic susceptibility summed over all modes.

While there is general consensus that the flux noise power spectral density in superconducting flux qubits is $1/\omega$ in nature at very low frequency, relatively little is known about the form of $S_i(\omega)$ that is appropriate at the high energies required to induce transitions in these qubits. In lieu of a more detailed picture of $S_i(\omega)$, we have chosen to follow a simple but physically reasonable approach by taking the Ohmic approximation in which $\text{Im}\{L_0(\omega)\}$ is treated as a constant. This simple quantum mechanical model of an environment is sufficient for our purposes as it provides an asymmetric $S_i(\omega)$ with respect to $\omega$. Consequently, the fluctuation-dissipation theorem ensures that there are thermally occupied modes of the environment that can fuel excitations of the eight-qubit system, thus compromising $P_{gs}$.

To model the dynamics of the open system embodied by Hamiltonian (10), we have constructed a numerical model based upon a generalized Bloch-Redfield formalism. In this model, one integrates a set of coupled first order differential equations for the elements of the eight-qubit system’s density matrix $\rho(t)$, as given by

$$\partial_t \rho_{mn} = -i\omega_{mn} \rho_{mn} - \sum_{k,l} (R_{mnlk} + M_{mnlk}) \rho_{kl} ,$$  

$$M_{mnlk} = -\delta_{nk} \langle l | \partial_t | m \rangle - \delta_{ml} \langle n | \partial_t | k \rangle ,$$

where $R_{mnlk}$ is the so-called Redfield tensor which is constructed from matrix elements of the flux operators between the eight-qubit states $|m\rangle$ and the $S_i(\omega)$ of the individual qubits. These matrix elements, which vary proportional to the number of qubit flips between a pair of states, are what facilitates thermalization of the eight-qubit system. In order to make the numerical work manageable on a personal computer, we kept only the ten lowest lying states. We have verified that the output of this model in the limit of small $\text{Im}\{L_0\}$ matches that of numerically solving the Schrodinger equation.

Our numerical model of the open system contains only one free parameter, namely, $\text{Im}\{L_0\}$. One can generate a crude estimate of this parameter based upon the width $W$ of a single qubit lowest order MRT peak deep in the incoherent regime. From Eq. (9) of Ref. 51,

$$W^2 = \int d\omega S_i(\omega) .$$

For the qubits in the circuit studied in this paper, experimental results yielded Gaussian-shaped MRT peaks of width $W/2 \left| I^p_q \right| \sim 80 \mu \Phi_0$ at $T = 20 \text{ mK}$. Low frequency drift measurements, of the type described in Ref. 52, revealed low frequency flux noise with a one-sided power spectral density that could be fit to the

$$S = \frac{2\text{Im}\{L_0(\omega)\} (2 \left| I^p_q(t) \right|)^2 \hbar \omega}{1 - \exp \left( -\frac{\hbar \omega}{k_B T} \right)} .$$
form $A^2/f$ with $A \sim 1\mu\Phi_0/\sqrt{\text{Hz}}$. Integrating this 1/f power spectrum over a nominal ten decades in frequency yields a root mean squared flux noise on the order of $W_{\text{LF}}/2|I_q^2| \sim 5\mu\Phi_0$. Therefore, there was an additional $(W - W_{\text{LF}})/2|I_q^2| \sim 75\mu\Phi_0$ of integrated flux noise that was not captured by the 1/f portion of the noise spectrum. Assuming that the remainder of $W/2|I_q^2|$ can be attributed to a small white flux noise level that is integrated out to a cutoff $f_c = k_B T/h \sim 460 \text{ MHz}$, we estimate that the amplitude of the one-sided flux noise spectral density will be on the order of $\delta\Phi_n \sim 3.5 \mu\Phi_0/\sqrt{\text{Hz}}$. This white noise level is well below the detection limits realized in all low frequency flux noise measurements on SQUIDs reported in the literature to date. Experimental evidence of an even lower white flux noise level in a 3-junction rf SQUID flux qubit has been reported in Ref. 53. In terms of an environmental inductance, our estimate of $\delta\Phi_n$ translates into $\text{Im}\{L_0\} = 47 \text{ yH/Hz}$.

The results of numerical simulations of Eq. (13a) with the estimated $\text{Im}\{L_0\}$ and the problem instance settings for Examples A, B, and C as a function of $T$ are summarized in Fig. 9. Here, the coloring and format of the curves have been chosen to correspond to the color and fill of the data symbols in those plots, with hollow (solid) points corresponding to dashed (solid) curves. While the match between theory and experiment is not perfect, it is nonetheless clear that the generalized Bloch-Redfield model does provide a reasonable prediction of the populations for all three Examples. We believe that most of the discrepancies can be explained by slight errors in problem specification, as will be discussed below, and therefore did not pursue any fitting of these data by changing $\text{Im}\{L_0\}$. Very importantly, we draw attention to the splitting of the population statistics between the pairs of degenerate excited states ($|54\rangle,|101\rangle$) and ($|190\rangle,|117\rangle$) for Example C. These features arise quite naturally from the model and are a result of differences in the transition rates between states during the course of annealing, as governed by the quantum mechanical matrix elements of the flux operators between pairs of initial and final states.

In addition to the three examples cited above, we have simulated all 6400 problem instances using our model and constructed $P_{gs}$ histograms to compare with the data in Fig. 8. The results of these simulations are summarized in those plots. These simulations did indeed yield the expected general features, specifically the loss of counts from the highest 0.95 < $P_{gs}$ ≤ 1 bin to the broad hump centered about $P_{gs} \sim 0.6$ with increasing $T$, and an exponentially suppressed tail at low $P_{gs}$. However, the simulations also yielded more small-scale structure in the histograms than was observed in experiment.

It was hypothesized that slight errors in embedding problem instances in the hardware could have been responsible for the absence of fine structure in the experimental histograms in Fig. 8. The PMM elements responsible for nulling small qubit flux offsets are on-chip digital-to-analog flux converters whose least significant bit (LSB) weights were measured to be $\delta\Phi_{LSB} \approx 38 \mu\Phi_0$. Therefore, we cannot null qubit degeneracy point offsets $\Phi_0$ to better than $\pm\delta\Phi_{LSB}/2$ using PMM. Furthermore, the PMM elements controlling the coupler and $|I_q^2|$-compensator flux biases, $\Phi_{co,ij}$ and $\Phi_{r,ij}$, had LSB weights $\delta\Phi_{LSB} \approx \delta\Phi_{LSB} \approx 60 \mu\Phi_0$. This latter quantity limits the precision to which $K_{ij} \propto M_{ij}$ and $h_{i} \propto M_{i}$ can be specified, respectively, thereby distorting problem instances. We have attempted to simulate the effects of the finite precision of our PMM elements on all 6400 problem instances. It was observed that such embedding errors only marginally smoothed out the fine structure in the simulated histograms.

A second potential source of error in embedding problem instances onto the hardware was that due to imperfect setting of the CCJJ control biases that are used to null the effects of Josephson junction variations. Such imperfections give rise to an apparent flux offset in a qubit that varies monotonically with the annealing control, meaning that $\Phi_0 \rightarrow \Phi_0(\Phi_{ccjj}^x)$ in Eq. (6b), as given by Eq. (B4c) of Ref. 28. Our single-qubit calibration procedures null $\Phi_0(\Phi_{ccjj}^x) \rightarrow \Phi_0(\Phi_{ccjj}^x)$ to less than 20 $\mu\Phi_0$, where the biases $\Phi_{ccjj}^x$ and $\Phi_{ccjj}^x$ straddle the region that is critical for annealing, as depicted in Fig. 7. Thus, while such residual imperfections will inevitably be present, we nonetheless consider their impact to be subdominant to the errors introduced by the finite precision of the PMM.

Our multiqubit generalized Bloch-Redfield model invokes Ohmic noise spectral densities that extend to high energy, thus providing a mechanism by which each individual qubit can exchange energy with its environment. Within the context of gate model quantum computation (GMQC), high frequency noise of this type gives rise to the timescale $T_1$ over which a qubit prepared in its excited state would relax towards thermal equilibrium. In many respects, QA is compromised by the reverse process in which a qubit prepared in its ground state is excited by the environment over a timescale $e^{-\gamma}/k_B T$, as dictated by the Einstein relation between upward and downward transition rates, where $E_q = \sqrt{\Delta_2^2 + \Delta_2^2}$ is the excitation energy of a qubit for arbitrary bias conditions. This implies that QA is somewhat more robust against the effects of thermalization than GMQC, although our numerical simulations clearly indicate that the ground state probability $P_{gs}$ is reduced by these $T_1$-related processes. Nonetheless, theoretical studies such as Ref. 50 have concluded that the computation time need not be severely compromised by the presence of such an environment.

One of the key oversights of our generalized Bloch-Redfield model is the role of non-Ohmic low frequency flux noise, particularly noise spectral densities of the form $S_\omega(\omega) \propto 1/\omega^\eta \ (\eta > 0)$ that are commonly referred to as 1/f noise. Within the context of GMQC, low frequency flux noise adiabatically manipulates qubit biases, thus leading to dephasing over a timescale frequently denoted as $T_\varphi$. From the perspective of QA, it is more convenient to envision low frequency flux noise as being responsible for the generation of random flux offsets within each
VI. DISCUSSION AND OPEN QUESTIONS

The experimental results in Sect. IV and the modeling in Sect. V present a compelling physical picture of how the state of our prototype processor evolves when run slowly in comparison to the adiabatic limit. These results have prompted some interesting conclusions and a series of open questions that we will briefly address below.

Our study has demonstrated that the most probable result from the hardware, when run slowly, is invariably the ground state. The remaining probability is distributed between low lying states that are roughly within an energy window \(O(k_B T)\) above the energy of the ground state at the point in the annealing process where the population statistics freeze. Depending upon the nature of the optimization problem, such low lying states may still constitute acceptable solutions. This observation suggests two possible modes of running a QA processor in practice:

- Run a given optimization problem a statistically large number of times. Take all \(|\tilde{s}\rangle\) that are observed with probability greater than the readout resolution threshold and calculate \(E(\tilde{s})\) using Eq. (1). Take whatever \(|\tilde{s}\rangle\) provides the lowest \(E\) as the solution.
- Run a given optimization problem once. Take the output \(|\tilde{s}\rangle\) and calculate \(E(\tilde{s})\) using Eq. (1). If \(E\) is less than some user-defined threshold, then accept the solution. If not, then iterate.

Note that had our processor been operated in the first mode, then it would have returned the correct answer to all 6400 problem instances.

We now turn to the open questions that have been motivated by this work:

**What happens when the CCJJ bias ramp time is reduced?** Given the limited bandwidth of our external bias lines, we were unable to elicit a substantial change in population statistics. Thus, we were restricted to studying the regime in which the thermalization times were much shorter than the annealing time. Consequently, we can only claim to have demonstrated QA, not AQC. Efforts are underway to build a new apparatus that will allow us to probe the regime in which thermalization times exceed the annealing time in an eight-qubit processor.

**How does the performance scale with problem size?** The eight-qubit unit cell is too small to be used to address issues concerning scaling. The intention of this study was to provide a basic demonstration of what we believe to be a complete set of essential ingredients needed for building a scalable QA processor acting in concert. We will reserve a discussion of the operation of larger portions of a complete 128-qubit chip to a future publication.

**How does one implement error correction in QA?** The work presented in this article is in much the same spirit as that of Ref. 55 - operate a small scale device that ‘looks’ like a basic quantum information processor and run it using the simplest algorithms known. We have made no attempts beyond statistical sampling to implement any form of error correction in these particular experiments. This is an active area of research at the moment.

To what precision must the qubit parameters \(|I_q^p\rangle\) and \(\Delta_q\) be synchronized as a function of \(\Phi_{ccjj}^x(t)\)? Implicit in
many models of AQO is the assumption that all of the qubits are identical throughout the annealing process. An examination of the data in Fig. 4 reveals that while $|I^x_p|$ is very uniform as a function of annealing parameter $\Phi_{ccjj}$, there are clear discrepancies between values of $\Delta_q$ at the same bias $\Phi_{ccjj}$. While we have demonstrated that our modest eight-qubit unit cell is capable of solving small-scale ISG problem instances, it is neither clear to what degree the asynchronization of $\Delta_q(t)$ may have impeded its performance nor what the implications of such asynchronization are for large-scale systems.

How do the higher excited states of the CCJJ rf SQUIDs affect performance? To our knowledge, this issue has not been addressed in the literature. We strongly advise against literal interpretations of many-qubit eigen-spectra to energies that exceed that of the second excited state of an rf-SQUID qubit. This has consequences for studies regarding the ultimate utility of QA. Higher excited states provide fast interwell relaxation mechanisms that could limit failures during a computation.

VII. CONCLUSIONS

An eight-qubit unit cell that is part of a superconducting optimization processor has been experimentally investigated. This processor makes use of several scalable elements, including on-chip programmable flux sources, XY-addressable high fidelity readouts, and the use of a limited number of global analog control lines to provide a variety of time-dependent control signals to multiple devices. The processor was tested using a large set of randomly generated Ising spin glass problem instances. The experimental results were shown to be consistent with the predictions of a quantum mechanical theory in which the individual qubits are coupled to a thermal environment.

We thank J. Hilton, F. Brito, K. Pudenz, S. Han, A. Kleinsasser, and G. Kerber for useful discussions.
