Quantum Annealing Stencils with Applications to Fuel Loading of a Nuclear Reactor

Joseph Fustero, Scott Palmtag, Frank Mueller
North Carolina State University, {jgfuster,sppalmta}@ncsu.edu, mueller@cs.ncsu.edu

***Abstract*—A method for mapping quadratic unconstrained binary optimizations expressed as nearest neighbor stencils onto contemporary quantum annealing machines is developed. The method is shown to be scalable in providing higher utilization of annealing hardware resources than prior work. Applying the technique to the problem of determining an effective fuel loading pattern for nuclear reactors shows that densely mapped quantum stencils result in higher fidelity solutions of optimization problems then the sparser default solutions. These results are likely to generalize to quadratic unconstrained binary optimizations that can be expressed as dense quantum stencils, thereby improving optimization results obtained from noisy quantum devices.**

***Index Terms*—quantum annealing, noisy intermediate-scale quantum computing, topology graph embeddings**

I. Introduction

Quantum computing has been realized on a set of first-generation hardware devices with systems that are publicly accessible. Two paradigms dominate the field: (1) gate-based quantum computers from manufacturers such as the IBM Q, Rigetti's Aspen, or IonQ, using superconducting transmons or ion tube technology [1], [2]; and (2) quantum annealing devices from D-Wave using RF-Squids [3]. Both types of systems are available in the cloud and can be programmed using Python, e.g., via IBM's Qiskit in the IBM Q Experience [4], Rigetti's Forest DSK in their Quantum Cloud Services [2], Ion Q's API, and D-wave's Ocean Software [5] accessible via the cloud through D-Wave Leap [6]. The latter three are also accessible via AWS through Amazon Braket [7].

It has been shown that adiabatic quantum computing, in its general form, can solve the same problems as gate-based (universal) quantum computing. However, this requires at least two degrees of freedom for 2-local Hamiltonians [8]–[10]. D-Wave supports 2-local Ising Hamiltonians with a single degree of freedom in their system, which restricts its application area. Nonetheless, computationally hard optimization problems can be expressed as 2-local Isings, which has the potential to solve problems orders of magnitude faster than classical algorithms [11], [12]. A diverse range of applications from computer science to chemistry has been utilizing these capabilities for computationally hard problems [12]–[16]. Any problem first has to be transformed into a 2-local Ising Hamiltonian and then to a quantum unconstrained binary optimization (QUBO) before it is subsequently mapped onto a qubit annealing topology with couplers between qubits [17], [18]. This process of embedding a QUBO representation of dependent variables onto a quantum topology of interconnected qubits is itself a challenging graph mapping problem that is NP hard, such that heuristic algorithms often require long computation time before they find a solution [19]. The scalability of applying optimization problems to quantum annealing is often limited by the inefficiency of such mapping problems.

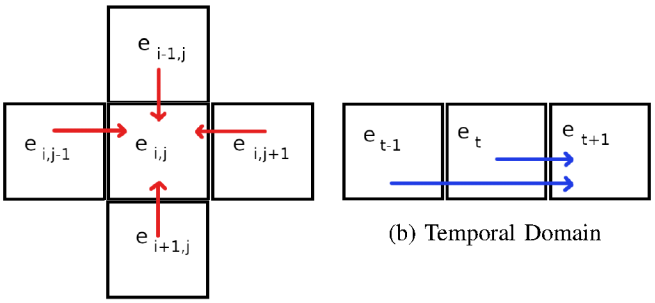
This work contributes a novel approach for embedding QUBO structures into annealer topologies for a specific class of problems. When the Ising formulation itself is highly regular, certain properties of the target embedding graph can be exploited to provide a hierarchical approach. A novel method is develop to specifically target "quantum stencils". First, a local subset of the problem is embedded in a highly connected "unit cell" of the target topology. Second, the higher-order dependencies are systematically resolved by constructing a regular pattern connecting unit cells symmetrically with each other. The key to success is often a dual "unit cell" that fuses these steps together, i.e., unit cells with alternating mappings (or even alternating rows of coupler mappings) are realized. The resulting embedding has the potential to scale up to full topology size depending on the relation of problem shape vs. target topology shape.

This work further contributes a novel application area of quantum annealing. As a nuclear reactor operates, the fuel depletes and a portion of the core must be replaced periodically. During the refueling, the placement and number of fuel assemblies greatly affects the operation and economics of the reactor core. The problem here is to determine the placement of fuel assemblies to meet design constraints and maximize output efficiency. This problem is formulated as a 2-local Ising Hamiltonian in this paper and then transformed into a QUBO before being subjected to quantum annealing to obtain close to optimal results with low computational overhead. The resulting QUBO is exposed to D-Wave's default embedding algorithm, which fails to find a solution. Instead, the newly developed quantum stencil embedding is shown to provide a space efficient solution. Subsequent quantum annealing provides operationally efficient fuel replacement patterns as indicated by the overall energy of the annealing process.

Overall, the novel approach of quantum stencil embeddings has the potential to generalize to other stencil patterns and different target topologies of quantum qubit interconnects. Its application to the fuel loading problem may further provide an alternative to solve another optimization problem with low computational complexity.

II. Quantum Annealing Stencils

The idea of quantum stencils bears a similarity to particle simulations performed via finite differencing. Particle simulations are defined by modeling force interactions between particles in a multi-dimensional space over time. An evolving system of particles can be discretely simulated by expressing these force interactions as a transfer function using the finite difference method. The transfer function has a spatial component, referred to as a stencil (shape), and a temporal component as it evolves over time.



(a) Spatial Domain

(b) Temporal Domain

Fig. 1: Five-point Stencil

Figure 1 depicts such a discretization in a two-dimensional domain with a five-point stencil (spatially) and a three-deep time dependence (temporally). A corresponding transfer function is given by

$$e_{i,j,t+1} = \alpha_{i,j}e_{i,j,t} - \beta_{i,j}e_{i,j,t-1} + f_{i,j} + \gamma_{i-1,j}e_{i-1,j,t} + \gamma_{i+1,j}e_{i+1,j,t} + \gamma_{i,j-1}e_{i,j-1,t} + \gamma_{i,j+1}e_{i,j+1,t},$$

where e denotes the energy of a particle and f denotes an optional external force field discretized over the 2D domain indexed by i, j and time t. More complex stencils in space (d-dimensional with p-point stencils with wider ii/jj ranges) and time (dt-deep dependencies) can be expressed as

$$e_{i,j,t+1} = \alpha_{i,j}e_{i,j,t} - \sum_{tt=t-1}^{t-dt} \beta_{i,j}e_{i,j,tt} + f_{i,j} + \sum_{ii=w(i)} \sum_{jj=h(j)} \gamma_{ii,jj}e_{ii,jj,t},$$

where functions w and h define the width and height of the stencil, optionally with some $\gamma_{ii,jj} = 0$ (e.g., for the five-point case).

We constrain ourselves to the simple case for $d = 2, p = 5, dt = 1$. This is due to practical limitations when considering quantum devices, as discussed later. Further, instead of evolving a particle system for a fixed number of time steps, our objective is to find a stable state with an overall energy close to the minimum. In this sense, we are interested in solving an optimization problem with the objective of finding a (close to) minimal overall energy H of the 2D ($m \times n$) system described by

$$H = min \sum_{i=1}^{m} \sum_{j=1..n} (\alpha_{i,j}e_{i,j} + \sum_{ii=1}^{m} \sum_{jj=1}^{n} \beta_{i,j,ii,jj}e_{i,j}e_{ii,jj}). \quad (1)$$

Minimizing this objective is computationally intractable (NP-hard) in the classical sense [20]. As this objective function describes a two-local Hamiltonian H, our objective is to assess the feasibility to exploit quantum computing to obtain a solution. We can obtain a quantum solution on contemporary quantum devices using the Quantum Approximate Optimization Algorithm (QAOA) on gate-based quantum devices [21], or we can exploit quantum annealing to directly represent Hamiltonians on annealing devices [22]. In this work, we assess the latter as we are interested in solving large optimization problems with many "particles". In fact, adiabatic quantum computing relies on the dynamic evolution of a quantum state exploiting the tunneling effect, where the state is controlled by a Hamiltonian and changes occur adiabatically [23]. Given that quantum devices are fraught with noise today for adiabatic and gate devices [24]–[39], we can hope to obtain a close approximation of a minimum energy state for a given problem, but there is no guarantee that the minimum has been found.

A. Maximal Five-local Mapping onto D-Wave Chimera Topology

In the following, we further constrain ourselves to the equivalent of five-point stencils, i.e., two-local Hamiltonians, where $ii = i - 1, i + 1, jj = j - 1, j + 1)$ in Eq. 1. These Hamiltonians can be mapped directly onto the Chimera topology of qubits of a D-Wave 2000Q annealing device with up to 2,048 qubits depicted in Fig. 2. Let us first consider D-Wave's 8-qubit "unit cell", which forms a $K(4, 4)$ local bi-partite graph, i.e., two sets (called "parts") of four qubits each that are completely connected with one another across the sets (but not within each set), as depicted in the red box within Fig. 2. Qubits (nodes) have associated weights α and are connected by couplings (edges) to neighboring qubits — in directions north, south, east, west — with strengths β of Eq. 1.

The objective is to provide a dense mapping of a five-point Hamiltonian H onto the Chimera graph that maximizes the number of qubits included in the resulting embedding. To this end, we first consider an embedding of particles with energy e within the $K(4, 4)$ unit cell. We observe that within each part (left set and right set), qubits are not connected. This means that particles have to be represented by qubits of different parts of the unit cell, e.g., alternating left and right. Furthermore, we need to map the five-point stencil shape beyond the unit cell, horizontally (east/west) and vertically (north/south) to other particles, where other unit cells with particle embeddings reside. However, any qubit has either vertical or horizontal connections, but not both. Hence, we result in a technique of transforming H to build "chains" of (physical) qubits that represent a (logical) particle.

Figure 3(a) depicts a unit cell embedding of four particles, each represented by two qubits, in a horizontal subset (shown as a gray subset). This provides the means to fully connect particles with one another within the unit cell, as well as horizontally and vertically with neighboring unit cells within the Chimera graph. Figure 3(b) depicts the five-point stencil

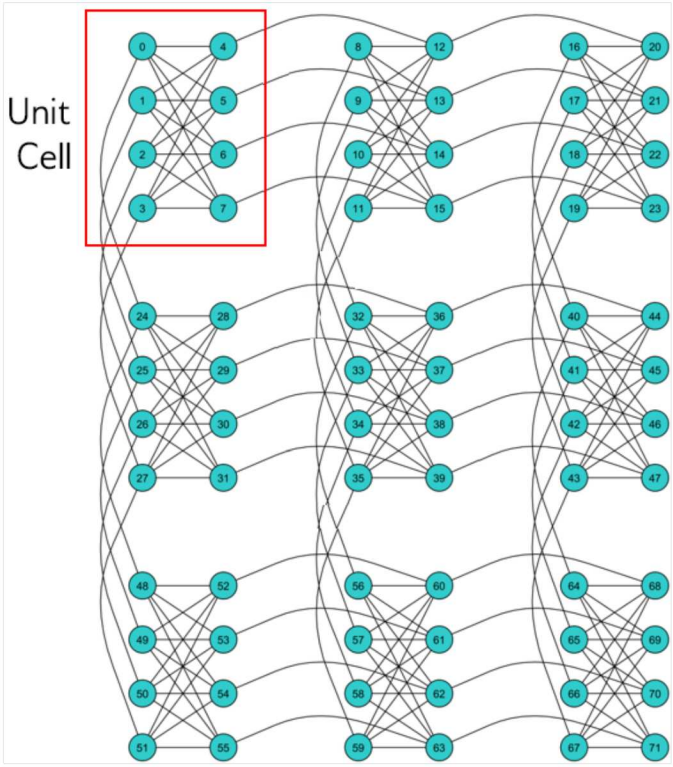
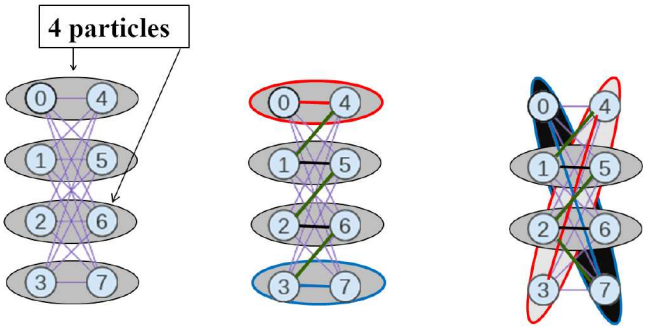


Fig. 2: 3x3x8 Chimera Graph

connections within such a unit cell mapping. Horizontal connections (red/black/blue) denote couplings building the chains to embed a particle within two qubits. Diagonal connections (green) depict the vertical north/south connections between particles.



(a) particle placement (b) intra couplers (c) dual couplers

Fig. 3: Chimera Unit Cell

The challenge now lies in finding a dense mapping to combine unit cells. Horizontally, such connections can be trivially constructed as depicted in Figure 4. Unfortunately, vertical connections between the top and bottom qubits of two adjacent unit cells do not exist. However, top-to-top and bottom-to-bottom connections exist. To exploit these couplings for vertical connections, we provide a "dual" mapping into the unit cell, as depicted in Figure 3(c). The (formally top) red set and (bottom) blue set are swizzled to stretch from top to bottom diagonally crossing within the unit cell. This now provides additional freedom on the connections with vertically neighboring unit cells: Red (formally top) can connect up (north) via qubit 3 (bottom); conversely, blue (formally bottom) can connect down (south) via qubit 0 (top).

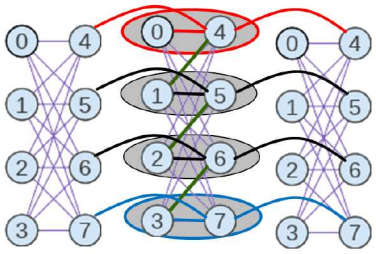


Fig. 4: Horizontal Inter-Cell Couplers

The resulting embedding with connections across unit cells is depicted in Figure 5. We observe that odd rows of unit cells utilize a horizontal internal embedding (from Figure 3(b)) while even rows feature the dual mapping (from Figure 3(c)). The overall embedding utilizes all qubits, i.e., it is maximal in density due to the vertical constraints of the unit cell within (as discussed before).

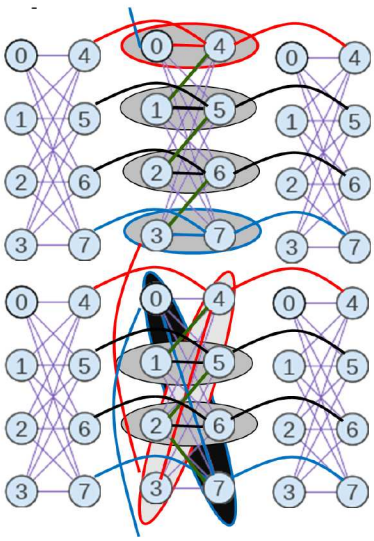
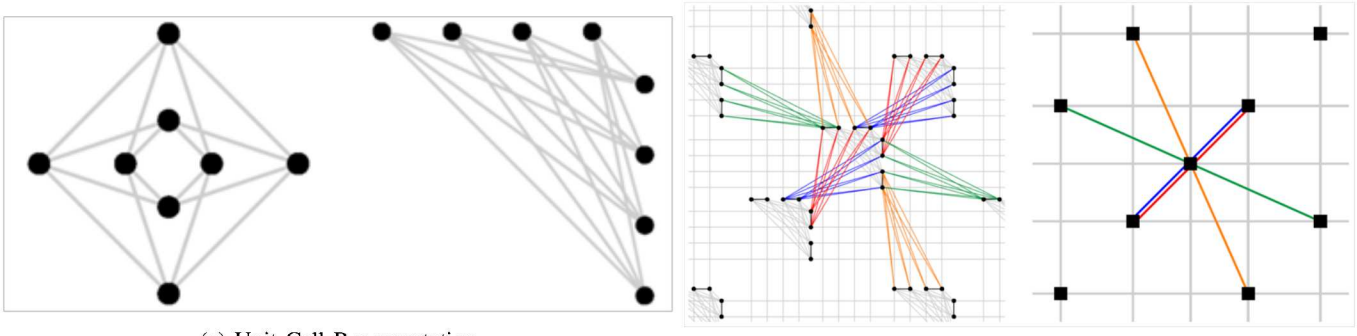


Fig. 5: Vertical Inter-Cell Couplers

B. *Five-local Mapping onto D-Wave Pegasus Topology*

D-Wave's Advantage QPUs (with up to 5,640 qubits) utilize a Pegasus graph depicted in Figure 6 with its components. The unit cell (Figure 6a) has additional local connections within pairs of cells (black edges in Figure 6b). However, there are no lateral edges connecting four qubits between cells; instead, edges are crossing (e.g., red and blue). This prevents a one-to-one mapping between logical particles and physical qubits locally within the unit cell, i.e., the crossings would result in reversed north/south (or east/west) connections. Thus, we can



(a) Unit Cell Representation (b) Inter-Cell Couplings (c) Compressed Abstraction

Fig. 6: Pegasus Graph

again only represent four particles using chaining within a cell, just as in the Chimera topology. (Had red and blue edges not been crossing, eight particles would have fit into one cell.)

Our mapping here is for 2×2 particles 2D grid-wise connected (pink edges) using local cell edges (see Fig. 7). Additionally, chains (black edges) extend the east-facing particles to qubits but swizzle their position internally. Between cells, east/west connections now are realized via red/blue edges to utilize global couplings for the lateral links. Pairs of global north/south connections utilize green/orange edges to medial cells via global couplings. These color codings correspond the those in Figure 6b. When considering the east/west axis (diagonal row) as a reference, odd rows use green/orange north/south edges, respectively, while even rows have the opposite pattern of orange/green for north/south, respectively (figure omitted). This provides a larger area that the 2D grid can be projected onto within the Pegasus graph.

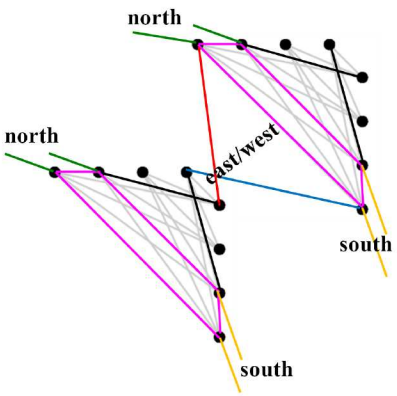


Fig. 7: Pegasus Inter-Cell Embedding

C. Comparing Topologies

Figure 8 depicts a $5 \times 5 \times 3 \times 8$ Pegasus graph in terms of the compressed units from Figure 6(c). The mapping onto a diagonal placement constrains the largest 2D grid within the actual $15 \times 15 \times 3 \times 8$ Pegasus graph (9X that of the depicted figure) for the current D-Wave devices to 42×34 (not using any unit cells outside this square), i.e., $1,428$ particles for a 38% utilization of qubits within the device. The largest quadratic problem that can be represented has $36 \times 36 = 1,296$ particles.

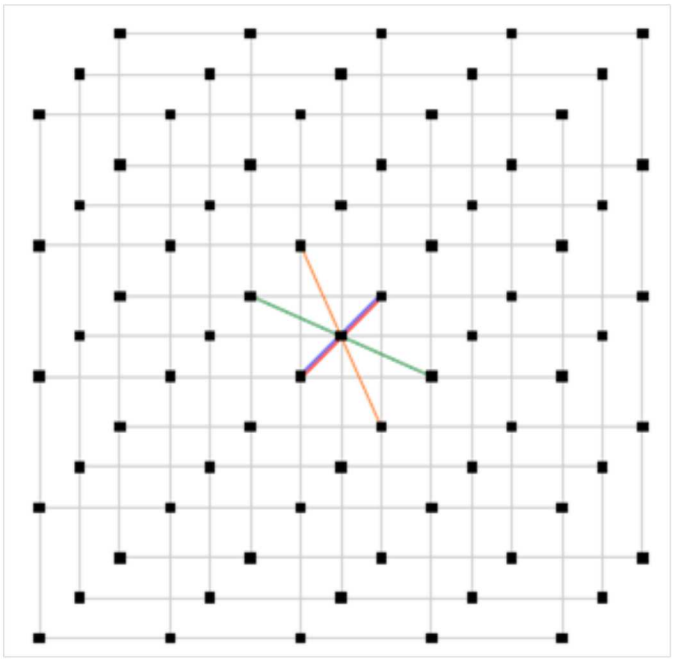


Fig. 8: 5x5x3x8 Pegasus Graph of Compressed Units

In contrast, the Chimera topology can accommodate $64 \times 16 \times 4 = 1,024$ particles in the said shape, where each unit cell holds 4 particles and all qubits are utilized; the largest quadratic problem that can be represented is $16x16 = 256$. An alternate Chimera embedding can be constructed, where the unit cell holds a square of 2×2 particles, which allows for a $32 \times 32 = 1,024$ square problem representation (also relying on a dual cell), but the Pegasus graph still outperforms in terms of problem size by a factor of $\approx$1.4X for a rectangular grid and $\approx$1.3X for a quadratic one. Considering that the Pegasus topology has 2.75X the number of qubits, these factors are rather moderate, which can again be attributed to the red/blue crossings preventing a more efficient 2D layout.

Chimera may result in higher qubit utilization (100% of the qubits), yet Pegasus uses fewer physical qubits as not all

cells are utilized and 2 qubits per cell remain unused. For noisy hardware devices, utilizing fewer resources to represent a problem has the potential to reduce error, which would give Pegasus an advantage. However, if problem domains are not quadratic but resemble a rhombus relative to the diagonals of the Pegasus graph, all cells may even be used for a total qubit utilization of 75%. Notice that the four additional local couplings within the unit graph were key for Pegasus to provide an advantage over Chimera.

III. Application to Optimizing Fuel Loading within Nuclear Reactors

We apply our quantum stencil embedding method to a real-world problem encountered in optimization of fuel loading patterns in nuclear reactors. We first describe the problem, and then provide a QUBO formulation of it.

A. The Fuel Loading Problem for Pressurized Water Reactors

When refueling nuclear reactors, the location of fuel assemblies within the core determines the power distribution in the core, the cycle lifetime, and the economics of the fuel cycle. Efficient core designs, that meet all of the design constraints and reduce the number of fresh fuel assemblies needed, can save the utility millions of dollars in fuel costs at each cycle.

For this study, we consider a reactor core based on the APR1400 pressurized water reactor (PWR) design [40]. The core consists of 241 fuel assemblies arranged on a circular grid, as shown in Figure 9. The core is designed with quarter-core rotational symmetry, so only 61 of the 241 possible locations are unique.

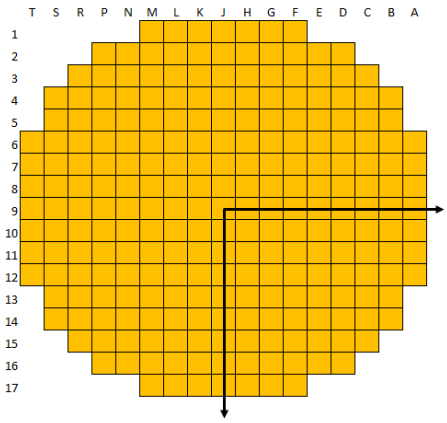


Fig. 9: APR1400 reactor core layout.

We will consider a simplified problem where there are only two fresh fuel types allowed in the core, either low enrichment or high enrichment. In addition, we will assume that the center assembly is always low enrichment, so there are a total of 60 degrees of freedom in our problem. It is desirable to have an even split of 30 low and 30 high enrichment fuel assemblies to support subsequent fuel cycles. (In the next cycle, the low enriched bundles will be discharged, and new high enriched bundles will be added.) Even with the constraint of only using fresh fuel with two enrichments, there are over 10^{17} potential valid loading patterns to evaluate. Optimization algorithms, such as Simulated Annealing [41], have been applied to solve this problem, but significant room for improvement exists.

The selection of a loading pattern is motivated by a desire to 1) minimize the maximum local power peaking factor (with a hard upper limit of 2.2) and 2) maximize the cycle length. Thus, the cost function that the loading pattern seeks to minimize can be stated as

$$E = -0.12Bu + F_q + 2H(F_q - 2.2),$$

where H is the Heaviside step function, Bu is the cycle exposure in units of $\frac{GWd}{MT}$ and F_q is the maximum local power peaking factor.

The SIMULATE-3 reactor core simulator [42] is a high-fidelity multiphysics solver that can be used to model the reactor and determine the cost function for each loading pattern. However, given the large number of available loading patterns, it is not feasible to perform a brute force search of all possible loading patterns; thus, optimization methods are desired.

The specific physics of loading pattern optimization for light water reactors (LWRs) is a good match for stencil optimization problems. By formulating it as a 2D Ising model, with only nearest-neighbor connectivity and at most 2$^{\text{nd}}$ order terms, it can map efficiently to D-Wave. The sparse connectivity and lack of need for representation of higher order terms allows for the entire 60 degrees of freedom to be mapped to the D-Wave architecture.

B. Ising Formulation

The 2D Ising model is considered a potentially appropriate model for the problem due the short neutron mean free path in LWRs. A fuel assembly is approximately 10 mean free paths across, so the interactions between assemblies are dominated by short-term effects. Thus, ground state solutions generated on D-Wave may correspond to optimal loading patterns as determined by the core simulator.

The 2D Ising model energy (physics sign convention) is

$$E = -\sum_{i}^{N} h_i s_i - \sum_{i<j}^{N} J_{ij} s_i s_j.$$

Reformulated as a 2D Ising problem with an inverted sign convention, D-Wave can solve for this problem as

$$E = \sum_{i}^{N} h_i s_i + \sum_{i<j}^{N} J_{ij} s_i s_j.$$

The 2D Ising to be embedded onto D-Wave was generated using several steps. First, 15,000 random loading patterns with 30-30 fuel assembly design splits (30 high enrichment and 30 low enrichment fuel assemblies) in the quarter-core geometry were processed in SIMULATE-3 for a single cycle depletion calculation. Next, 3,000 target loading patterns with

30-30 fuel assembly splits were simulated with SIMULATE-3. In the case of the targeted samples, each loading pattern differed by a small (2, 4, 6, 8 or 10) or large (50, 52, 54, 56, 58 or 60) number of assembly placements compared to the previous sample. The targeted samples allow for the problem ruggedness to estimate for an arbitrary number of spin flips (assembly swaps). Next, all 18,000 samples were used to train the 2D Ising model. The training algorithm implemented was an adaptation from [43]. This training algorithm seeks to reconstruct the Boltzmann distribution of spin configurations from which the samples were drawn.

The Boltzmann distribution requires a minimal ground state value of $E = 0$. Thus, the initial cost function must be adjusted. The left-hand-side of the cost function appears Gaussian while the right-hand-side is skewed. It was assumed that the ground state is reached for the value of the cost function corresponding to 10^{-17}. The ground state of the initial cost function was found to be -3.99. Thus, the cost function was adjusted to

$$E = -0.12Bu + F_q + 2H(F_q - 2.2) + 3.99.$$

The probability distribution of the updated cost function can be seen in the kernel density estimate (KDE) depicted in Figure 10. Due to the large uncertainty in estimating the ground state value, future approaches to solving the loading optimization problem should not attempt to reconstruct the Boltzmann distribution.

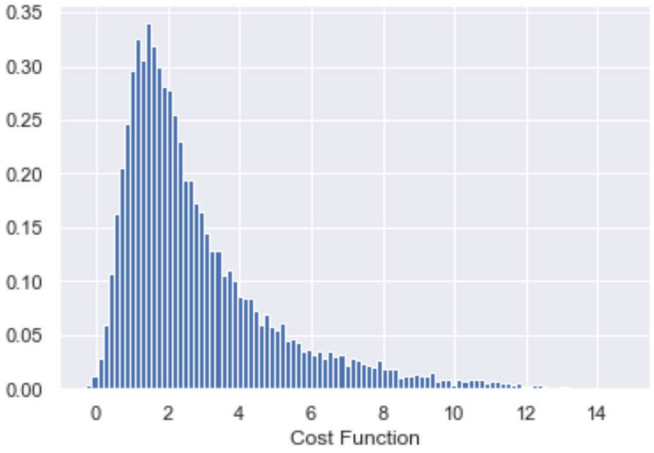


Fig. 10: The cost function KDE is skewed right but approximately Gaussian prior to reaching its maximum between 1 and 2.

Next, each of the 18,000 samples was compared to each other in terms of 1) the number of spin flips (i.e., the number of differences in assembly placement in the input) and 2) the difference in SIMULATE-3 prediction of an output (total cost function or its cycle exposure and maximum local peaking power constituents, respectively, depicted in Figures 11 and 12).

Typical values of the cycle exposure are on the order of $20\frac{GWd}{MT}$, which corresponds to approximately 18-24 months.

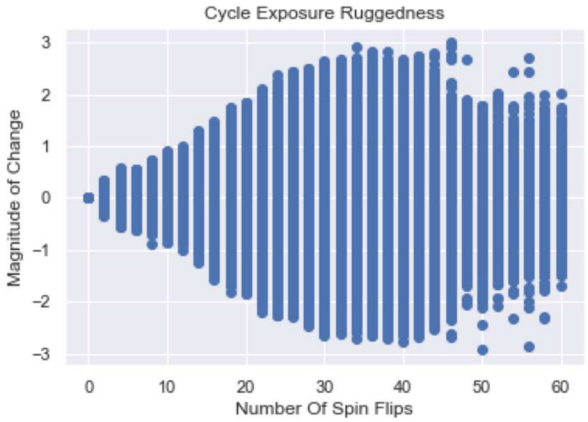


Fig. 11: Ruggedness of the cycle exposure

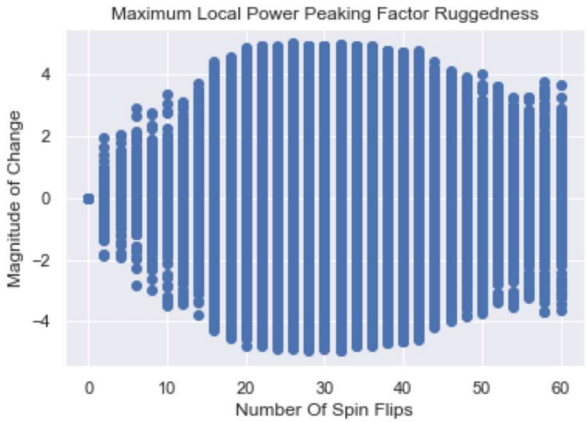


Fig. 12: Ruggedness of the maximum local power peaking factor

Typical values of the maximum local power peaking factor (F_q) are 2. Thus, while from Figures 11 and 12 both metrics show comparable changes as a function of the number of spin flips, the maximum local power peaking factor is far more sensitive to changes in the loading pattern. This is expected as the cycle exposure represents a globally averaged value (across all assemblies) while the power peaking is a local quantity. In general, from both Figures 11 and 12, the magnitude in the change of each metric increases with a larger number of spin flips between loading patterns, but there is large variation in this behavior. Thus, is it feasible to claim that the problem's cost function landscape is favorable to quantum annealing. The problem is defined as many local minima divided by thin energy barriers of arbitrary height. The cost function is nearly random with respect to the input loading pattern. In this case, the problem ruggedness is largely driven by the ruggedness of the local maximum power peaking. Local values (maximum values) are typically used when evaluating the safety limits of an engineering design.

Once the cost function has been updated, the values of the qubit biases and qubit coupling biases can be computed

from a modified training algorithm. The (modified, re-weighs to posterior Boltzmann distribution) training algorithm (which assumes the physics sign convention) seeks to minimize what is known as an interaction screening operator (ISO, S) at each vertex given the set of $M = 18,000$ samples.

$$S_i(J_i, h_i) = \sum_{m=0}^{M} e^{-\beta E_m} e^{-\sum_{j \neq i} J_{ij} s_i s_j - h_i s_i}$$

$$(\hat{J}_i, \hat{h}_i) = [\ln S_i(J_i, h_i) + \lambda \|J_i\|_1]$$

An analysis of the original algorithm showed that the closer the tuning parameter λ was to 1, the fewer samples the algorithm would need to ensure accuracy, but that the closer the tuning parameter was to 0, the more likely the quadratic coefficients would not vanish. It was determined that a selection of $\lambda = 0$ was appropriate: An analysis of the first 3,000 samples followed by analyses with each additional batch of 3,000 samples showed that for this particular model, the values were somewhat insensitive to the number of samples. Additionally, raising the tuning parameter even as high as $\lambda = 0.1$ caused the coupling coefficients to vanish. This was deemed unacceptable due to the desire to capture the interaction effect between adjacent assemblies.

For the training process, conventions of $s = -1$ for the lower-enrichment fuel assembly and $s = +1$ for the higher-enrichment fuel assembly were assumed. For inverse temperatures of $\beta = 1$ and $\beta = 10$ the 2D Ising was trained. These relatively low values for β were selected in order that the recovered Boltzmann distribution have physical meaning. In reality, the appearance of a spin configuration is more likely for those closer to the ground state energy. Only when the temperature is very high, is a random sampling procedure justified. Random sampling must be performed as the associated energy of a loading pattern is not known a priori. In order to account for the observed differences in energy levels, the factor $e^{-\beta E_m}$ was added to the original ISO term.

For $\beta = 1$ the magnitude of the error in the 2D Ising surrogate cost function compared with SIMULATE-3 was on average 2.44 while for $\beta = 10$ it was 5.78. In addition to the larger error, a cursory examination of the 2D Ising model map for $\beta = 10$ appears to show saturation/clipping of parameter values. An optimal value for the inverse temperature was searched for. For $\beta = 2.5$ the average error magnitude was 1.15.

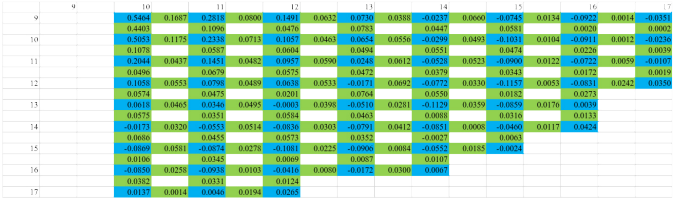


Fig. 13: Inverted 2D Ising model, $\beta = 1$

Figures 14 and 16, which reflect the shape of the bottom-right quartile of the fuel core, represent the 2D Ising surrogate

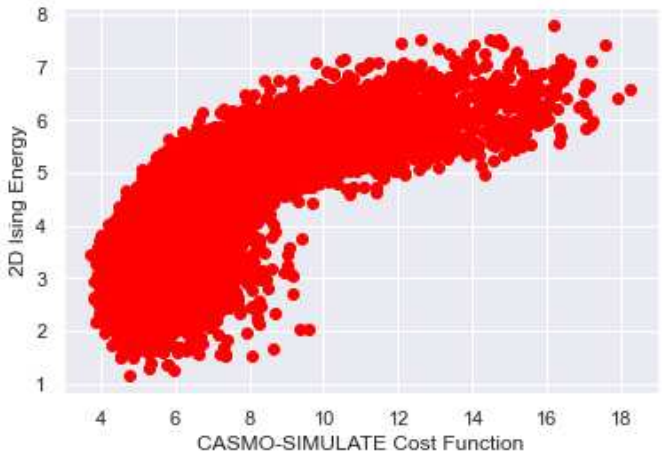


Fig. 14: 2D Ising surrogate (for SIMULATE-3) accuracy, $\beta = 1$

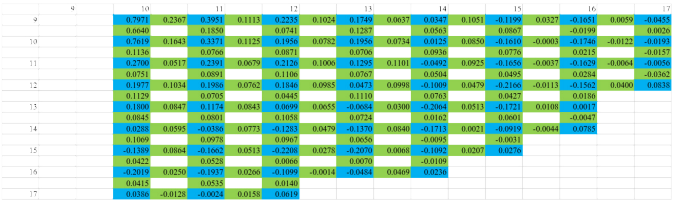


Fig. 15: Figure 6. Inverted 2D Ising model, $\beta = 2.5$

prediction (cell weights blue, inter-cell couplings green) compared to actual SIMULATE-3 result while Figures 13 and 15 present a graphical representation of the 2D Ising model map with the signs flipped (to the D-Wave sign convention) from the traditional physics convention. From Figures 14 and 16 both the low-fidelity and high-fidelity results are correlated up to a SIMULATE-3 cost function value of about 9 and then afterwards they remain correlated, but less so. This change in behavior may be due to imposing a hard constraint of

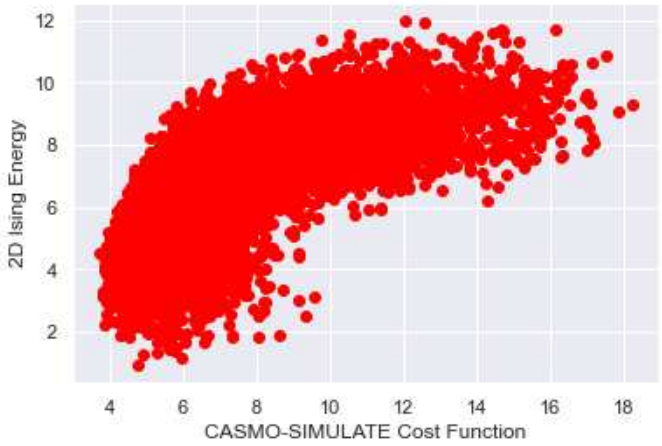


Fig. 16: Figure 7. 2D Ising surrogate (for SIMULATE-3) accuracy, $\beta = 2.5$

$F_q = 2.2$. Lastly, for both inverse temperature values the surrogate tends to underpredict the actual value of the cost function. Thus, the inability to estimate the cost function ground state prevents the construction of a high accuracy surrogate for lower energies.

On the other hand, from Figures 13 and 15, while the 2D Ising does not capture the problem in great detail, some of its qualitative features align with traditional loading pattern designs. The qubit coupling terms are nearly strictly positive. Thus, the reactor system is by and large antiferromagnetic. The coupling terms also tend towards 1 closer to the center of the reactor core. This is consistent with traditional loading pattern design. The reactor core being antiferromagnetic corresponds to dissimilar assemblies being placed adjacent to each other — a checkerboard design. Checkerboard patterns are also typically preferable closer to the center of the core, away from the problem boundary. Checkerboard patterns help to reduce power peaking, which is especially of concern closer to the center of the core as the boundary effect of some assemblies not being entirely surrounded by other assemblies vanishes. Additionally, from both Figures 13 and 15 the qubit bias terms are positive at the center of the core, trends towards 0 and the negative further from the center of the core and then trends towards flipping back to positive on the outermost periphery of the core. The linear coefficient behavior on the outermost ring of the reactor indicates a preference for lower-enrichment fuel at that location. This is as expected in order to improve (i.e., reduce) the neutron leakage (and thus increase the cycle length). Additionally, the behavior on the interior of the core, favoring lower-enrichment closer to the center and favoring higher-enrichment fuel closer to the core periphery (but not on the periphery itself), is consistent with traditional IN-OUT loading patterns. These patterns balance the higher-enrichment fuel assemblies with increased geometric attenuation further from the core center. This allows again for the flattening of the core power profile.

C. QUBO Formulation

The trained 2D Ising coefficients must be converted to the correct quadratic unconstrained binary optimization (QUBO) problem format. First, as stated before, all of the signs on both linear and quadratic coefficients are flipped. Next, all of the quadratic coefficients are generated:

$$Q_{ij} = 4J_{ij}$$

Then (slightly more complex) all of the linear coefficients are generated:

$$Q_i = 2h_i - \frac{1}{2}\sum_{i<j} Q_i$$

Now, the QUBO problem may be solved:

$$f(x) = \sum_i Q_i x_i + \sum_{i<j} Q_{ij} x_i x_j$$

This conversion is fairly straightforward. Discrete problems with higher-order terms or denser connectivity may also be converted to QUBOs with an increased number of qubits. As earlier, future attempts to the loading pattern optimization problem should use higher order terms to increase surrogate fidelity. Additionally, the D-Wave solution will be unconstrained with no guarantee of a 30-30 split in the fuel assembly designs. In order to introduce an energy penalty for designs that do not meet this constraint, a fully-connected graph would be required [44]. However, we decided to assess if training the 2D Ising model with only 30-30 batch splits would implicitly guide ground state solutions towards this constraint.

IV. Implementation

Our embeddings of the quantum annealing stencils for the reactor core loading problem were realized on a D-Wave 2000Q device (DW_2000Q_6). This specific device, as many others, has manufacturing defects in terms of dead qubits, specifically seven qubits at the time of experimentation. These qubits and any of their couplers need to be excluded from an embedding. By default, D-Wave uses the minor miner algorithm to find an embedding on the subset of good qubits [18]. Unfortunately, this algorithm failed to produce any valid mapping for our fuel loading QUBO of 60 fuel assemblies. The minor miner algorithm uses a set of heuristics with seed randomization to greedily place edges and qubits one by one such that "some space" is left for future placement of more edges/qubits. Apparently, this results in mapping attempts that are too sparse to fit within the subset of good qubits on the DW_2000Q_6 device.

Our embedding explicitly utilizes a dense and dual mapping into the unit cell, and then globally connects unit cells of four particles with one another. This local/global embedding combined with the idea of dual embeddings is key to success. It is also an indication that D-Wave may benefit from geometric embeddings within local unit cells before global embeddings are considered for structured QUBOs, i.e., where two-local interactions follow a common pattern.

Our embedding further takes into account the list of dead qubits available via the Ocean API. We search for a consecutive range of rows within the Chimera graph with only good qubits large enough to fit our dense embedding. A 8x9 embedding (or 2x9 in terms of unit cell rows/columns) suffices to fit our problem. For our dense mapping, a single dead qubit renders an entire cell with its corresponding row and column useless. In other words, routing around dead qubits (and their unit cell) is infeasible with our dense Chimera embedding as we utilize all qubits of a unit cell.

In case of the DW_2000Q_6 device, three pairs of dead qubits belong to the same cell for a total of four unusable cells and two failed cells in the same row. This amounts to a total loss of three rows and four columns or $(3+4)\times 16 - 4 = 63$ cells with eight qubits each, i.e., almost 25% of their cells cannot be utilized.

In general, sparser embeddings with symmetry and duality can be constructed to address this problem. In fact, our Pegasus

mapping is sparser as it only utilizes 75% of the qubits in a unit cell, and these qubits can be relocated within a cell, i.e., cells with one or sometimes two dead qubits can still be used. However, if multiple global connectors are affected, routing around an entire unit cell has prohibitive cost for our relatively dense embeddings, and symmetry would render an entire cell row and column useless. Notice that alternate algorithms exist to consider dead qubits and couplers, but they tend to consider specialized graphs rather symmetric graph structures [45].

V. Evaluation

We performed evaluations of our models on the D-Wave 2000Q platform with the inverted 2D Ising model ($\beta = 1$) formulated as a QUBO. We specified the maximum number of samples (10,000) for the experiment and then searched the result space for ≈50%/50% load patterns (equivalent to 30/30 ± 1 low/high enriched assemblies). These runs provided 396, 302, 244, and 188 results, for repeated runs, with lowest energies of -4.9958, -5.0531, -5.0219, -5.0485, respectively. These best balanced loading patterns in terms of power efficiency of the nuclear reactor have very similar shape, depicted as a 2D layout in Fig. 17, where 1/0 denote low/high fuel enrichments.

```
1 1 1 1 1 0 0 0 1
1 1 1 1 1 0 0 0 0
1 1 1 1 1 0 0 0 0
1 1 1 0 0 0 0 1
1 1 1 0 0 0 0 1
0 1 1 0 1 0 0
0 0 1 0 0 1
0 0 0 1
```

(a) E=-4.9958

```
1 1 1 1 1 1 0 0 0
1 1 1 1 1 0 0 0 0
1 1 1 1 1 1 0 1 0
1 1 1 1 1 0 0 0
0 1 1 0 0 0 1 1
0 0 1 0 0 0 0
0 0 0 0 1 1
0 0 0 0
```

(b) E=-5.0531

Fig. 17: Lowest Energy for 50%/50% Patterns of 2 Runs

In general, the energy of the QUBO solution varies between -1.5704 and -5.3252, with the lowest energy pattern depicted in Figure 18, which has an unbalanced load pattern. The idea of balancing patterns is an economical one to aid in the fuel loading of subsequent cycles. All patterns have in common a concentration of low enriched assemblies in the center of the core (in the upper left corner of the quarter core). The placement of low enriched fuel in the center will tend to decrease the core power peaking (F_q in the cost function).

```
1 1 1 1 1 1 0 0 1
1 1 1 1 1 0 0 0 1
1 1 1 1 1 0 0 0 1
1 1 1 1 1 0 1 0
1 1 1 1 1 0 0 0
1 0 0 0 0 0 0
0 0 0 0 1 1
0 0 0 1
```

Fig. 18: Pattern for a 34/26 split with E=-5.0531

VI. Related Work

Early work on qubit embeddings generically considered simple graph topologies [46] and resulted in the development of an embedding algorithm, TRIAD, specifically geared at triangular clique embeddings [17]. D-Wave's minor miner embedding algorithm places an edge/vertex at a time leaving "enough space" for future placements, combined with initial randomization, to provide a sparse enough partial mapping that future placements are likely to succeed [18]. Other work exploits maximal minor embeddings to support cliques, which is very specific in terms of combining dense local and global constraints, and has even been developed to "patch" connections around dead qubits [45]. Another approach utilizes a greedy algorithm that considers a biclique virtual topology abstraction as a mapping target, which realizes odd cycles cleverly that can only be realized with ancillas for Chimera graphs [47]. The Pegasus graph supports odd cycles without ancillas, i.e., it is not clear if the biclique virtual topology provides a benefit here. In contrast to these approaches, our work focuses an a dense hierarchical embedding strategy, where local unit graphs are first regularized, often with dual mappings, before they a stitched together globally. It is not greedy but rather follows a constructive structural design pattern that scales well.

VII. Conclusion

This work contributes a novel, scalable method for mapping QUBOs onto topological qubit structures of quantum annealing devices, theoretically up to full topology size. Optimization problems specified as quantum stencils are shown to be efficiently mapped onto contemporary quantum annealing devices.

The work further provides a solution to the fuel loading problem of nuclear reactors. When formulated as an optimization problem matching a two-local Ising Hamiltonian, a QUBO formulation is derived that matches the quantum stencil shape. Experiments with a 2000Q D-Wave device show that efficient solutions for such problem domains can be obtained with our topological embedding method while default embedding algorithms fail to provide a solution.

Quantum stencils are particularly appealing for regular, repeating global structures. Global embeddings benefit from geometric similarity to the problem shape, otherwise density is sacrificed but flexibility is gained to route around dead qubits, as seen for the Pegasus topology. These findings provide inspiration for improving D-Wave's current embedding approach, i.e., by providing a more constructive geometric approach as an alternative. The findings may also provide inspiration for consideration of future graph topologies of RF-Squids or even for connecting superconducting transmons, e.g., by trading off the amount of crossing and non-crossing couplers, which support remote, multi-hop chains vs. local connectivity, respectively.

VIII. Future Work

Quantum stencils generalize in that they provide embeddings for QUBOs of arbitrary problem domains. The nuclear reactor fuel loading problem itself features a 2D problem abstraction that provides a natural representation mapping onto D-Wave topologies. 3D problems would be harder to map to current topologies, i.e., a 3D interconnect would be preferable, particularly one that extends beyond a depth of three as given by the Pegasus graph. Qubit utilization could further be improved for 2D and 3D mappings with small changes in the intra- and inter-cell routing, but without increasing the degree of connectivity per qubit (see Section II). Furthermore, the fuel loading problem is simplified as global constraints are omitted in the quantum model and only checked classically within the solution space provided by the quantum device. Supporting global constraints natively within the quantum device would require a second interconnect, where most (if not all) qubits are connected to a single "central" qubit, which may not easily be realized with superconducting transmons, but might be more feasible with ion traps.

Further work must take place to evaluate the methodology established here with traditional approaches. First, it should be remembered that the 2D Ising is ultimately a surrogate for the real problem and that eventually, all solutions to the 2D Ising should be evaluated using the higher-fidelity SIMULATE-3 code system. Next, alternate solvers to the 2D Ising should be investigated and have their solution quality compared to that of D-Wave, given an equal number of generated solutions [48]. With consideration given to the cost of creation of the 2D Ising surrogate, 2D Ising solvers should be compared with traditional engineering approaches that directly manipulate the SIMULATE-3 input and do not require the creation of a 2D Ising surrogate model.

Extensions to other types of reactor designs should also be investigated. While it may be obvious that D-Wave's performance in designing loading patterns for MOX-fueled reactors would be equal to its performance for UO_2-fueled reactors as presented here, it is not. The effect of vanishing quadratic (coupling) coefficients could be significant due to the smaller neutron mean free path in MOX-fueled reactors. Additionally, denser connectivity (such as diagonal interactions between assemblies) and higher-order terms (such as interactions centered at the corners rather than the sides of assemblies that may require quadratic or cubic terms, which need to be decomposed to quadratic terms) should be considered in surrogate modeling fidelity. Lastly, efforts should be made to investigate reactor re-loading problems requiring a very dense connectivity (such as in many fast reactor designs, where the neutron mean free path is on the order of the size of fuel assemblies) and reloading patterns that feature more than two fuel levels (this is more realistic), which requires solving the generalized Potts model [49]. Modeling of boiling water reactors (BWRs) on D-Wave will likely require a computational cost intermediate of PWRs and fast reactors due to the still small neutron mean free path combined with the fuel assemblies being of larger number and smaller size.

Acknowledgment

This work was supported in part by the following awards: NSF 1917383, 1818914, 2120757 and LANL subcontract 725530.

References

[1] IBM, "IBM Q Experience," https://quantumexperience.ng.bluemix.net/qx.
[2] "Rigetti forest," 2018. [Online]. Available: https://www.rigetti.com/forest
[3] S. Boixo, T. Albash, F. M. Spedalieri, N. Chancellor, and D. A. Lidar, "Experimental signature of programmable quantum annealing," *arXiv:1212.1739*, 2012. [Online]. Available: http://arxiv.org/abs/1212.1739
[4] H. Abraham, AduOffei, I. Y. Akhalwaya, and et al., "Qiskit: An open-source framework for quantum computing," 2019. [Online]. Available: qiskit.org
[5] D-Wave, "D-wave's ocean software," https://ocean.dwavesys.com/.
[6] ——, "D-wave leap," https://www.dwavesys.com/take-leap.
[7] AWS, "Amazon braket," https://aws.amazon.com/braket/.
[8] D. Aharonov, W. van Dam, J. Kempe, Z. Landau, S. Lloyd, and O. Regev, "Adiabatic quantum computation is equivalent to standard quantum computation," *ArXiv e-prints*, May 2004. [Online]. Available: https://arxiv.org/abs/quant-ph/0405098
[9] ——, "Adiabatic quantum computation is equivalent to standard quantum computation," *SIAM J. Comput.*, vol. 37, no. 1, pp. 166–194, Apr. 2007. [Online]. Available: http://dx.doi.org/10.1137/S0097539705447323
[10] D. Bacon, S. T. Flammia, and G. M. Crosswhite, "Adiabatic quantum transistors," 2012.
[11] J. King, S. Yarkoni, J. Raymond, I. Ozfidan, A. D. King, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, "Quantum annealing amid local ruggedness and global frustration," *Journal of the Physical Society of Japan*, vol. 88, no. 6, p. 061007, 2019.
[12] V. S. Denchev, S. Boixo, S. V. Isakov, N. Ding, R. Babbush, V. Smelyanskiy, J. Martinis, and H. Neven, "What is the computational value of finite-range tunneling?" *Phys. Rev. X*, vol. 6, p. 031015, Aug 2016. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevX.6.031015
[13] A. Lucas, "Ising formulations of many np problems," *Frontiers in Physics*, vol. 2, p. 5, 2014. [Online]. Available: https://www.frontiersin.org/article/10.3389/fphy.2014.00005
[14] E. G. Rieffel, D. Venturelli, B. O'Gorman, M. Do, E. M. Prystay, and V. Smelyanskiy, "A case study in programming a quantum annealer for hard operational planning problems," *Quantum Inf. Process.*, vol. 14, no. 1, pp. 1–36, 2015. [Online]. Available: https://doi.org/10.1007/s11128-014-0892-x
[15] H. Neven, G. Rose, and W. G. Macready, "Image recognition with an adiabatic quantum computer i. mapping to quadratic unconstrained binary optimization," 2008.
[16] I. Kassal, J. D. Whitfield, A. Perdomo-Ortiz, M.-H. Yung, and A. Aspuru-Guzik, "Simulating chemistry using quantum computers," *Annual Review of Physical Chemistry*, vol. 62, no. 1, pp. 185–207, 2011.
[17] V. Choi, "Minor-embedding in adiabatic quantum computation: II. minor-universal graph design," *Quantum Inf. Process.*, vol. 10, no. 3, pp. 343–353, 2011. [Online]. Available: https://doi.org/10.1007/s11128-010-0200-3
[18] J. Cai, W. G. Macready, and A. Roy, "A practical heuristic for finding graph minors," 2014.
[19] T. Humble, A. McCaskey, J. Schrock, H. Seddiqi, K. A. Britt, and N. Imam, "Performance models for split-execution computing systems," *arXiv:1607.01084*, 07 2016. [Online]. Available: http://arxiv.org/abs/1607.01084
[20] J. Kempe, A. Kitaev, and O. Regev, "The complexity of the local hamiltonian problem," *SIAM J. Comput.*, vol. 35, no. 5, p. 1070–1097, May 2006. [Online]. Available: https://doi.org/10.1137/S0097539704445226
[21] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014.
[22] A. Condello and contributors, "dwave-system," 2018.

[23] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, "Quantum computation by adiabatic evolution," 2000.

[24] M.-H. Yung, "Thermal noise on adiabatic quantum computation," 2008.

[25] E. Wilson, S. Singh, and F. Mueller, "Just-in-time quantum circuit transpilation reduces noise," in *IEEE International Conference on Quantum Computing and Engineering (QCE)*, Oct. 2020.

[26] P. Murali, N. M. Linke, M. Martonosi, A. J. Abhari, N. H. Nguyen, and C. H. Alderete, "Full-stack, real-system quantum computer studies: Architectural comparisons and design insights," in *International Symposium on Computer Architecture*, 2019, pp. 527–540.

[27] P. Murali, J. M. Baker, A. J. Abhari, F. T. Chong, and M. Martonosi, "Noise-adaptive compiler mappings for noisy intermediate-scale quantum computers," *arXiv preprint arXiv:1901.11054*, 2019.

[28] A. JavadiAbhari, S. Patil, D. Kudrow, J. Heckey, A. Lvov, F. T. Chong, and M. Martonosi, "ScaffCC: A framework for compilation and analysis of quantum computing programs," in *Proceedings of the 11th ACM Conference on Computing Frontiers*, ser. CF '14. New York, NY, USA: ACM, 2014, pp. 1:1–1:10. [Online]. Available: http://doi.acm.org/10.1145/2597917.2597939

[29] J. Preskill, "Quantum computing in the NISQ era and beyond," *arXiv:1801.00862v3*, vol. [quant-ph], 2018.

[30] L. Bishop and J. Gambetta, "Reduction and/or mitigation of crosstalk in quantum bit gates," 2019, US Patent App. 15/721,194.

[31] D. C. McKay, T. Alexander, L. Bello, M. J. Biercuk, L. Bishop, J. Chen, J. M. Chow, A. D. Córcoles, D. Egger, S. Filipp, J. Gomez, M. Hush, A. Javadi-Abhari, D. Moreda, P. Nation, B. Paulovicks, E. Winston, C. J. Wood, J. Wootton, and J. M. Gambetta, "Qiskit backend specifications for OpenQASM and OpenPulse experiments," *preprint arXiv:1809.03452*, 2018.

[32] A. Zulehner, A. Paler, and R. Wille, "An efficient methodology for mapping quantum circuits to the IBM QX architectures," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2018.

[33] S. S. Tannu and M. K. Qureshi, "Ensemble of diverse mappings: Improving reliability of quantum computers by orchestrating dissimilar mistakes," in *International Symposium on Microarchitecture*, 2019, pp. 253–265.

[34] ——, "Mitigating measurement errors in quantum computers by exploiting state-dependent bias," in *International Symposium on Microarchitecture*, 2019, pp. 279–290.

[35] ——, "Not all qubits are created equal: A case for variability-aware policies for NISQ-era quantum computers," in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, 2019.

[36] G. Li, Y. Ding, and Y. Xie, "Tackling the qubit mapping problem for NISQ-era quantum devices," in *Proceedings of the Twenty-Fourth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '19. New York, NY, USA: ACM, 2019, pp. 1001–1014. [Online]. Available: http://doi.acm.org/10.1145/3297858.3304023

[37] ——, "Towards efficient superconducting quantum processor architecture design," in *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems*, ser. ASPLOS '20. To appear: ACM, 2020.

[38] R. Wille, O. Keszocze, M. Walter, P. Rohrs, A. Chattopadhyay, and R. Drechsler, "Look-ahead schemes for nearest neighbor optimization of 1D and 2D quantum circuits," in *2016 21st Asia and South Pacific Design Automation Conference (ASP-DAC)*. IEEE, 2016, pp. 292–297.

[39] M. Y. Siraichi, V. F. d. Santos, S. Collange, and F. M. Q. Pereira, "Qubit allocation," in *Proceedings of the 2018 International Symposium on Code Generation and Optimization*. ACM, 2018, pp. 113–125.

[40] "APR1400 design control document tier 2," Korea Electric Power Corporation, APR1400-K-X-FS-14002-NP, Dec. 2014.

[41] D. J. Kropaczek and R. Walden, "Large-scale application of the constraint annealing method for pressurized water reactor core design optimization," *Nuclear Science and Engineering*, vol. 193, no. 5, pp. 523–536, 2019. [Online]. Available: https://doi.org/10.1080/00295639.2018.1550970

[42] K. R. Rempe, K. S. Smith, and A. F. Henry, "Simulate-3 pin power reconstruction: Methodology and benchmarking," *Nuclear Science and Engineering*, vol. 103, no. 4, pp. 334–342, 1989. [Online]. Available: https://doi.org/10.13182/NSE89-A23686

[43] A. Y. Lokhov, M. Vuffray, S. Misra, and M. Chertkov, "Optimal structure and parameter learning of ising models," *CoRR*, vol. abs/1612.05024, 2016. [Online]. Available: http://arxiv.org/abs/1612.05024

[44] E. Grant, T. S. Humble, and B. Stump, "Benchmarking quantum annealing controls with portfolio optimization," *Physical Review Applied*, vol. 15, no. 1, Jan 2021. [Online]. Available: http://dx.doi.org/10.1103/PhysRevApplied.15.014012

[45] C. Klymko, B. D. Sullivan, and T. S. Humble, "Adiabatic quantum programming: minor embedding with hard faults," *Quantum Inf. Process.*, vol. 13, no. 3, pp. 709–729, 2014. [Online]. Available: https://doi.org/10.1007/s11128-013-0683-9

[46] V. Choi, "Minor-embedding in adiabatic quantum computation: I. the parameter setting problem," *Quantum Inf. Process.*, vol. 7, no. 5, pp. 193–209, 2008. [Online]. Available: https://doi.org/10.1007/s11128-008-0082-9

[47] T. D. Goodrich, B. D. Sullivan, and T. S. Humble, "Optimizing adiabatic quantum program compilation using a graph-theoretic framework," *Quantum Inf. Process.*, vol. 17, no. 5, p. 118, 2018. [Online]. Available: https://doi.org/10.1007/s11128-018-1863-4

[48] J. King, Y. Sheir, J. Raymond, I. Ozfidan, A. D. King, M. M. Nevisi, J. P. Hilton, and C. C. McGeoch, "Quantum annealing amid local ruggedness and global frustration," *Journal of the Physical Society of Japan*, vol. 88, no. 6, Mar 2019. [Online]. Available: https://doi.org/10.7566/JPSJ.88.061007

[49] S. Okada, M. Ohzeki, and K. Tanaka, "Efficient quantum and simulated annealing of potts models using a half-hot constraint," *Journal of the Physical Society of Japan*, vol. 89, no. 9, Jul 2020. [Online]. Available: https://doi.org/10.7566/JPSJ.89.094801