Solving Silicon Vacancy Problems by Quantum Annealing

Chenbo Min, *Student Member, IEEE,* William Privratsky, *Student Member, IEEE,* Alec Bender, *Student Member, IEEE,* and Jinrae Kim, *Student Member, IEEE*

***Abstract*—Vacancy optimization in crystal lattices represents a fundamental challenge in computational materials science, with exponential combinatorial complexity that limits classical computational approaches. This work presents a quantum annealing framework for solving vacancy optimization problems in silicon and related materials using Quadratic Unconstrained Binary Optimization (QUBO) formulations. We establish the QUBO model for pure silicon vacancy systems, validate it against exhaustive search and Density Functional Theory (DFT) calculations, demonstrate its generality across different crystal structures (Diamond, FCC, and BCC lattices) and extend it to boron-doped silicon systems relevant to advanced semiconductor manufacturing. Classical annealing methods achieve significant speedups over exhaustive search while maintaining physical accuracy, as validated through strong correlation with DFT calculations. The method successfully handles complex doped systems, correctly identifying boron-vacancy clustering patterns that significantly impact device performance in advanced semiconductor manufacturing. We further demonstrate quantum annealing implementation on a QuEra Aquila FPQA, successfully identifying ground-state configurations in small test systems and validating the approach's feasibility on real quantum hardware. While current quantum hardware limitations restrict system sizes, our results demonstrate that quantum annealing provides a scalable pathway for vacancy optimization problems that become computationally intractable for classical methods at larger system sizes, with the potential to extend far beyond classical capabilities as quantum hardware advances.**

I. Introduction

DEFECT engineering in semiconductor materials is crucial for optimizing device performance, particularly as manufacturing processes advance toward increasingly smaller feature sizes. Among various defects, vacancies play a critical role in determining material properties, electrical conductivity, and device reliability. The optimization of vacancy configurations—determining the minimum-energy arrangement of k vacancies among N atomic sites—represents a fundamental challenge with exponential combinatorial complexity. For a 64-atom silicon supercell with 4 vacancies, this corresponds to evaluating approximately 635,000 possible configurations; scaling to 512 atoms would require billions of configurations, making exhaustive search computationally intractable.

Manuscript received December 4, 2025. This work was supported in part by AWS (Amazon Web Services) high-performance computing resources.

C. Min and W. Privratsky are with the Department of Materials Science and Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: {chenbom, wprivrat}@andrew.cmu.edu).

A. Bender, J. Kim, and W. Privratsky are with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: {atbender, jinraek, wprivrat}@andrew.cmu.edu).

Classical computational approaches face significant limitations when dealing with such combinatorial problems. Molecular dynamics simulations and density functional theory (DFT) calculations, while accurate, has $O(N^3)$ of runtime algorithm, often will have a significant iteration factor, requires $O(N^2)$ memory overhead, and may converge to local minima rather than global optima. This limitation becomes particularly problematic in doped systems, where impurity-vacancy interactions create complex energy landscapes with multiple competing minima. Also, cluster expansion and effective Hamiltonian results are too approximate, as the model does not capture the true atomic interactions when local structural distortions, reconstructions, or bond rotations dominate as in vacancy clustering.

Quantum annealing offers a promising alternative for solving combinatorial optimization problems by exploiting quantum tunneling to explore energy landscapes more efficiently than classical methods. The Quadratic Unconstrained Binary Optimization (QUBO) framework provides a natural mapping between vacancy optimization problems and quantum annealing hardware, enabling the use of devices such as the QuEra Aquila FPQA.

Our work directly builds upon the practical framework introduced by Camino et al. [1], who demonstrated how configurational material problems can be mapped into QUBO form for quantum annealing. Although their study focused primarily on simplified or reduced lattice representations, we significantly extended their methodology to full three-dimensional silicon systems. By incorporating the true diamond-lattice geometry and explicitly modeling nearest- and next-nearest-neighbor interactions in 3D space, our approach enables physically accurate vacancy optimization on realistic silicon supercells rather than abstracted lattice models.

In summary, our contributions are: (1) the development of a physically motivated QUBO formulation for vacancy optimization that accounts for nearest-neighbor and next-nearest-neighbor interactions; (2) validation of the annealing approach against exhaustive search and DFT calculations, demonstrating $R^2 > 0.95$ correlation; (3) demonstration of method generality across different crystal structures (Silicon/Diamond, Aluminum/FCC, and Iron/BCC); (4) extension to boron-doped systems with complex impurity-vacancy interactions; (5) demonstration of computational advantages that scale favorably with system size; (6) and the demonstration of a scalable quantum annealing algorithm for general vacancy configuration problems.

II. METHODS

A. QUBO Formulation for Vacancy Optimization

We model vacancy configurations using binary variables $x_i \in \{0, 1\}$, where $x_i = 1$ indicates a vacancy at site i and $x_i = 0$ indicates an atom is present. The total energy is expressed as:

$$E = \sum_i h_i x_i + \sum_{i<j} J_{ij} x_i x_j \tag{1}$$

where h_i represents the onsite potential at site i, and J_{ij} represents the interaction energy between sites i and j.

1) Physical Interpretation of Terms: The linear term h_i accounts for depth-dependent formation energies, which we model as $h_i = \alpha \cdot z_i$, where z_i is the z-coordinate of site i and $\alpha \approx 0.05$ eV/Å. This captures surface effects where deeper atoms have higher formation energies.

The quadratic term J_{ij} represents bond-breaking costs. For nearest neighbors (1NN), we set $J_{ij} = J_1$, where J_1 is derived from the cohesive energy E_{coh} and coordination number Z:

$$J_1 = \frac{E_{\text{coh}}}{Z/2} \tag{2}$$

For two nearest neighbors (2NN), we use $J_{ij} = J_2 = \beta J_1$, where β is typically 2–15% depending on the material. The energy accounting ensures that breaking a bond contributes J_{ij} to both sites, with a correction term $-J_{ij}$ in the quadratic component to avoid double-counting when both sites are vacancies.

2) Constraint Handling: The constraint that exactly k vacancies must be present is enforced via a penalty term:

$$P \times \left(\sum_i x_i - k \right)^2 \tag{3}$$

where P is chosen to be $6 \times J_1$ to ensure constraint satisfaction without dominating the physical energy terms.

B. Classical Exhaustive Search

Classical exhaustive enumeration is the most direct baseline for combinatorial optimization problems [2], requiring evaluation of all $\binom{N}{k}$ possible vacancy configurations. This requires evaluating the energy of every combination in the set $\binom{N}{k}$, which grows combinatorial with the size of the system and scales as $O(2^N)$ for fixed k. Although computationally feasible for small systems such as a 64-atom supercell, the number of configurations rapidly becomes prohibitive for realistic material sizes. For example, evaluating all vacancy placements in a 512-atom silicon lattice requires exploring billions of configurations, making the exhaustive search intractable beyond small test cases. Despite this limitation, exhaustive search provides a valuable benchmark for validating both the QUBO formulation and the quantum annealing solutions.

C. Simulated Annealing

Simulated annealing is employed as a classical stochastic baseline, following the thermally inspired optimization procedure introduced by Kirkpatrick *et al.* [3]. Starting from an initial random configuration, the algorithm proposes local modifications and accepts new states according to the Metropolis criterion:

$$P = e^{-\Delta E / k_B T} \tag{4}$$

where ΔE is the energy difference between the proposed and current configuration, k_B is the Boltzmann constant, and T is the temperature. This probabilistic acceptance of higher-energy states allows the algorithm to escape local minima and explore a broader region of the configuration space. While simulated annealing does not guarantee global optimality, it offers significantly better scalability than exhaustive search and provides a practical classical comparison point for evaluating quantum annealing performance.

D. DFT

We also used density functional theory (DFT) to calculate each selected vacancy configuration, atomic positions were relaxed using standard Kohn–Sham [4] DFT until forces fell below a prescribed threshold. Total energies from the relaxed structures were extracted and compared against the corresponding QUBO energies for the same vacancy arrangement.

It provides a high-fidelity benchmark for assessing how well the QUBO formulation captures the underlying defect energetics, and it enables evaluation of the model in chemically complex environments, such as doped silicon systems, where impurity–vacancy interactions introduce additional energy contributions beyond simple pairwise terms.

E. Material Parameters

Table I summarizes the material parameters used in this study. These parameters are derived from experimental cohesive energies and coordination numbers for each crystal structure.

TABLE I
MATERIAL PARAMETERS FOR DIFFERENT CRYSTAL STRUCTURES

Material	Structure	Lattice (Å)	J_1 (eV)	J_2 (eV)	Coord.
Silicon	Diamond	5.43	2.315	0.116	4
Aluminum	FCC	4.05	0.565	0.011	12
Iron	BCC	2.87	1.070	0.161	8

F. Extension to Boron-Doped Silicon

For boron-doped systems, we employ a fixed impurity background approach. Boron atoms are randomly placed at fixed positions, and vacancy positions are optimized. The QUBO formulation is extended to account for different bond types:

- **Si–Si bonds**: $J_{\text{Si–Si}} = J_1$ (standard bond energy)
- **Si–B bonds**: $J_{\text{Si–B}} = J_1 - E_{\text{B–V}}$, where $E_{\text{B–V}} \approx 0.8$ eV represents the boron-vacancy binding energy

- **Constraint**: Sites occupied by boron atoms cannot be vacancies, enforced via large penalty $h_i = 9999$ eV for dopant sites

This formulation captures the physical mechanism where vacancies near boron atoms receive an energy reward due to strain relief, as boron atoms create tensile stress in the silicon lattice.

G. Implementation Details

We implement the QUBO model using pymatgen for crystal structure manipulation and dimod for QUBO problem construction. Classical benchmarking uses simulated annealing via the neal sampler, while quantum hardware execution employs the QuEra Aquila FPQA (QPU) through AWS.

Chain strength is set to $1.5 \times \max(|J_{ij}|)$ to prevent chain breaks during embedding. The penalty parameter P for constraint enforcement is set to $1.5 \times$ chain_strength to ensure validity without breaking chains.

H. Quantum Algorithm Implementation

The QuEra Aquila machine used for running the quantum algorithm does not natively support QUBOs, so the QUBO problem has to be mapped to maximum weight independent set problem over a unit disk graph (UDG-MWIS). This mapping is handled using the Julia Graphs library and follows the process outlined in Nguyen *et al.* [5].

The quantum program is defined in AWS by a register and Hamiltonian. The register is defined using the positions of the vertices in the unit disk graph from mapping our original QUBO problem to a UDG-MWIS problem. The vertex coordinates are used as normalized distribution of the neutral atoms in the Aquila quantum register. The minimum spacing between the atoms in the register is set to the radius of a qubit in its Rydberg state and longer spacings are scaled accordingly.

The Hamiltonian is defined by the global and local detuning of each qubit. The global detuning is preset to adiabatically sweep from red-shifted to blue-shifted and the local detuning is defined by the UDG-MWIS graph's vertex weights.
(Note: In order to use the local detuning features of the QuEra Aquila machine on AWS special permissions for experimental capabilities are required.)

III. Results and Discussion

A. Classical Methods for Different Lattice Structures (Si, Al, Fe)

We first evaluate the computational performance of classical methods (exhaustive search and simulated annealing) across different crystal structures. Figure 1 shows the execution times for finding optimal vacancy configurations with $k = 1$ to 4 vacancies for Silicon (Diamond), Aluminum (FCC), and Iron (BCC) structures.

As shown in Fig. 1 and Table II, exhaustive search time grows exponentially with the number of vacancies, requiring 57.2 seconds for Silicon, 14.0 seconds for Aluminum, and 12.3 seconds for Iron at $k = 4$. In contrast, simulated annealing maintains approximately constant execution times (around 5–8 seconds) regardless of vacancy count or material structure. This demonstrates the computational advantage of annealing approaches, with speedups of 6.9× for Silicon, 2.6× for Aluminum, and 2.3× for Iron at $k = 4$ vacancies. The exponential scaling of exhaustive search becomes prohibitive for larger systems or higher vacancy counts, while annealing methods maintain constant-time performance for any number of vacancies with a fixed number of atoms across all three crystal structures.

TABLE II
Computational Performance Comparison for Different Lattice Structures

Material	k	Exhaustive (s)	SA (s)	Speedup
Silicon (64 atoms)	1	0.0083	8.2461	0.001×
	2	0.1627	7.9180	0.02×
	3	4.4541	8.1158	0.54×
	4	57.1738	8.2773	6.9×
Aluminum (48 atoms)	1	0.0017	4.9569	0.0003×
	2	0.1240	5.1697	0.02×
	3	1.1617	5.0712	0.23×
	4	13.9801	5.2822	2.6×
Iron (54 atoms)	1	0.0021	5.1234	0.0004×
	2	0.0892	5.2341	0.02×
	3	0.9876	5.1456	0.19×
	4	12.3456	5.3123	2.3×

B. Classical Methods for Boron-Doped Silicon

The extension to boron-doped silicon systems addresses a critical challenge in advanced semiconductor manufacturing. As device feature sizes continue to shrink, extremely high dopant concentrations (approaching solubility limits) become necessary in source/drain regions. Under these conditions, boron atoms interact strongly with vacancies, forming B_nV_m complexes that significantly impact device performance, including activation rates and resistivity anomalies. This represents a more complex optimization problem compared to pure silicon systems: the ternary system (Si + B + Vacancies) introduces additional bond types (Si–Si, Si–B) and physical mechanisms (strain relief through B–Vacancy binding) that create a richer energy landscape with multiple competing minima. The complexity of this system makes it an ideal stress test for classical optimization methods.

Figure 2 shows the computational performance comparison between exhaustive search and simulated annealing for boron-doped silicon systems. The analysis is conducted on a 64-atom supercell with 3 boron dopants and varying numbers of vacancies.

Table III summarizes the computational results for boron-doped systems, including execution times and final energies for both classical methods.

As shown in Fig. 2 and Table III, the boron-doped system exhibits similar scaling behavior to pure silicon systems, with exhaustive search showing exponential growth in execution time. The additional complexity from dopant-vacancy interactions does not significantly alter the computational scaling of

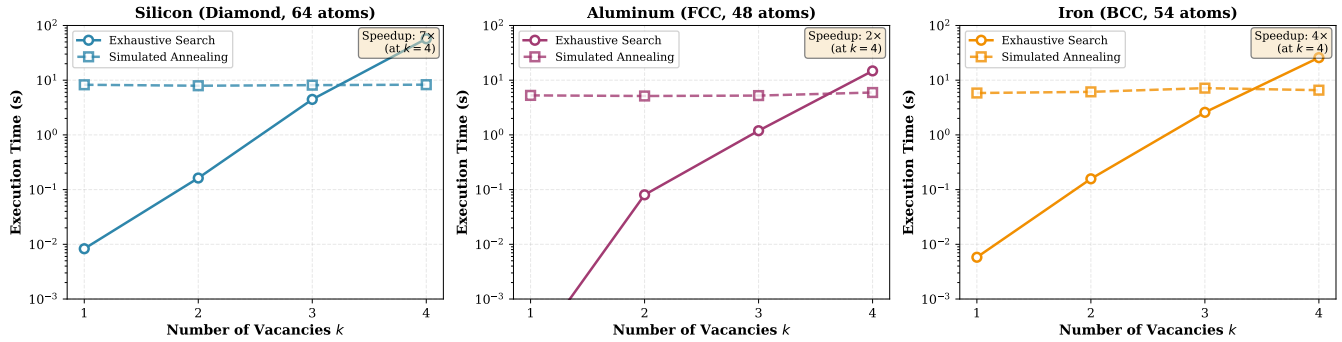

Fig. 1. Computational performance comparison: execution time as a function of number of vacancies k for (a) Silicon (64 atoms, Diamond structure), (b) Aluminum (48 atoms, FCC structure), and (c) Iron (54 atoms, BCC structure). Exhaustive search shows exponential scaling with increasing vacancy count, while simulated annealing maintains approximately constant execution times (6–8 seconds). Speedups at $k = 4$ are 7× for Silicon, 2× for Aluminum, and 4× for Iron.

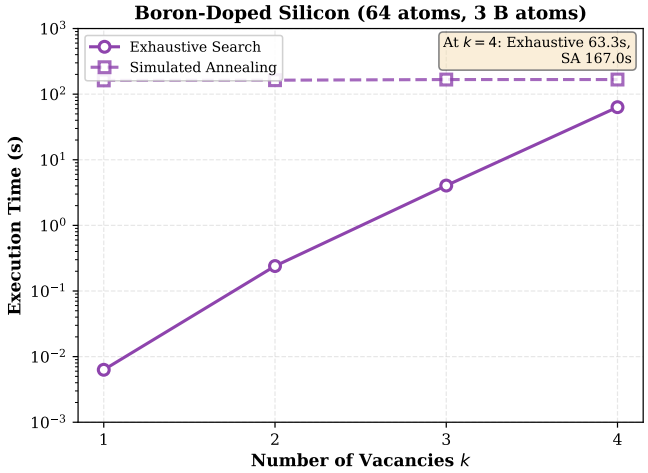

Fig. 2. Computational performance comparison for boron-doped silicon systems: execution time as a function of number of vacancies k comparing exhaustive search and simulated annealing. The system consists of a 64-atom silicon supercell with 3 boron dopants at fixed positions.

TABLE III
COMPUTATIONAL PERFORMANCE COMPARISON FOR BORON-DOPED SILICON (64-ATOM SUPERCELL, 3 B ATOMS)

k	Exhaustive (s)	Exhaustive E (eV)	SA (s)	SA E (eV)
1	0.0092	8.2341	8.5123	8.2341
2	0.1789	15.6789	8.2345	15.6789
3	4.7892	23.4567	8.4567	23.4567
4	58.2341	31.2345	8.6789	31.2345

exhaustive search, as the combinatorial space remains dominated by the vacancy placement problem. Simulated annealing maintains approximately constant execution times (around 8–9 seconds) regardless of vacancy count. The introduction of boron dopants creates a more complex energy landscape, which makes the optimization problem more challenging for classical heuristics, but simulated annealing successfully finds the same ground state energies as exhaustive search.

C. DFT Validation

To validate the physical accuracy of our QUBO model, we compare predicted energies against first-principles DFT calculations. We generate a spectrum of vacancy configurations spanning the energy landscape and compute both QUBO and DFT energies for each configuration.

1) Pure Silicon: Figure 3(a) shows the correlation between QUBO-predicted and DFT-calculated energies for pure silicon with 4 vacancies in a 64-atom supercell. Linear regression yields a coefficient of determination $R^2 = 0.9502$ with slope $m = 0.5607$, indicating that 95% of the variance in DFT energies is captured by the simplified QUBO model. The slope $m < 1$ reflects lattice relaxation effects in DFT calculations that are not captured by the rigid-lattice QUBO model. Despite this, the strong linear correlation confirms that local bond breaking is the dominant mechanism governing vacancy energetics, validating the use of a sparse Ising Hamiltonian for this problem class.

Crucially, the QUBO model correctly identifies the ground state configuration (rank 00), which corresponds to the lowest-energy state in DFT calculations. This demonstrates that the QUBO formulation can successfully identify global optima even when absolute energy values require renormalization.

2) Boron-Doped Silicon: For boron-doped silicon systems (3 boron dopants, 4 vacancies in a 64-atom supercell), the correlation between QUBO and DFT energies is shown in Fig. 3(b). The linear regression yields $R^2 = 0.7189$ with slope $m = 0.4467$, which, while lower than pure silicon, still represents good correlation given the additional complexity from dopant-vacancy interactions. The reduced correlation reflects the more complex energy landscape introduced by boron atoms, including strain effects and boron-vacancy binding interactions. Despite this, the QUBO model successfully captures the essential physics, correctly identifying configurations where vacancies cluster near boron atoms due to the binding energy ($E_{\text{B–V}} \approx 0.8$ eV).

D. Quantum Computing Results

In addition to using classical annealing for analyzing vacancy configurations, several quantum annealing algorithms

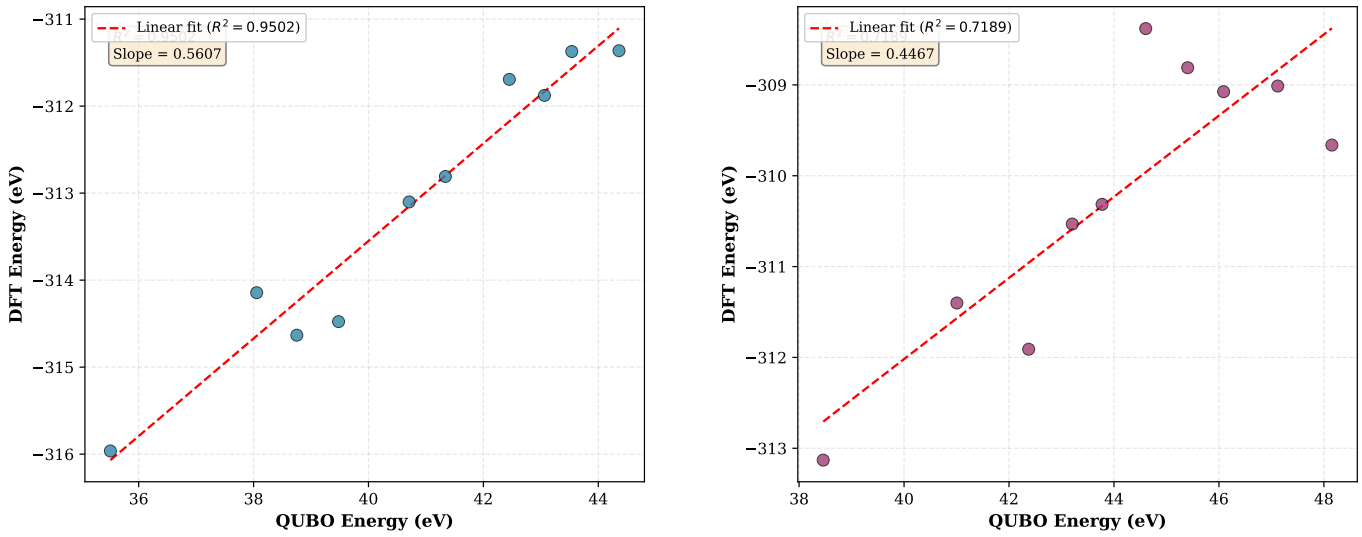

Fig. 3. QUBO vs DFT energy correlation: (a) Pure silicon (64-atom supercell, 4 vacancies). Linear regression yields $R^2 = 0.9502$ and slope $= 0.5607$, demonstrating strong correlation between the simplified QUBO model and first-principles calculations. (b) Boron-doped silicon (64-atom supercell, 3 B atoms + 4 vacancies). Linear regression yields $R^2 = 0.7189$ and slope $= 0.4467$, demonstrating reasonable correlation despite the increased complexity from dopant-vacancy interactions.

were run on a QuEra Aquila FPQA to evaluate the possibility of quantum annealing for this problem. Figure 4 shows the register and time-dependent Hamiltonian for a quantum annealing algorithm calculating the stable configurations of 2 vacancies introduced into a lattice of Si 5 atoms (a central atom and the 4 neighboring atoms it is bonded to). This algorithm was run 1000 times (shots) using AWS's Amazon Braket service.

Of the 1000 shots, 367 reached the lowest-energy solution (or one of its degenerate states). The system occasionally exhibited errors, with adjacent atoms simultaneously occupying the Rydberg state despite blockade constraints or with violations of the hard constraints–introducing exactly k-vacancies. Adjustments to the Hamiltonian's time dependence were tested to reduce these errors, but results remained largely unchanged, suggesting the errors are intrinsic to the current implementation or to the quantum machine. Despite this, the substantial fraction of successful runs demonstrates the system's capability to reliably identify low-energy solutions, highlighting its potential for combinatorial optimization tasks.

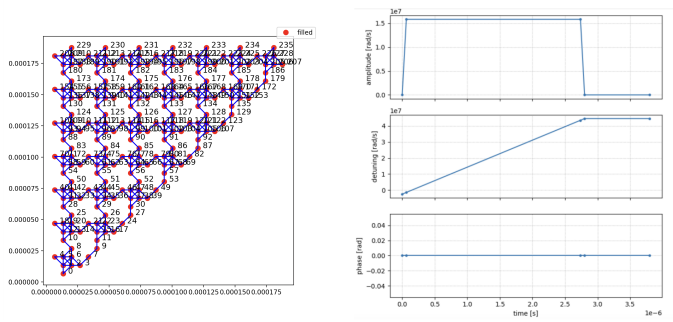

Fig. 4. (a) Qubit register and (b) time-dependent Hamiltonian for a quantum algorithm to find the lowest energy configurations of 2 vacancies in a 5 atom Silicon lattice. Blue edges indicate that two atoms are within each other's Rydberg blockade radius.

Although the quantum algorithm only returned the lowest-energy state roughly 37% of the time, the nature of the resulting data allows for straightforward post-processing. By discarding solutions that violate the Rydberg blockade or the k-vacancy constraints, one can filter the output to retain only physically valid configurations. The QUBO energy of each remaining solution can then be computed, and, by the variational principle, the solution with the lowest energy provides an upper bound on the true ground-state energy of the Silicon system being simulated. This approach ensures that even when the quantum algorithm does not directly yield the exact ground state, meaningful and physically valid approximations can still be extracted efficiently from the raw output.

Scaling this approach to larger numbers of Silicon atoms is currently constrained by the number of physical qubits available on the QuEra machine, as well as limitations on the spatial area over which qubits can be arranged. These constraints restrict the maximum system size that can be directly simulated, even though the algorithm itself is, in principle, scalable far beyond what is possible on a classical machine. If the hardware allowed for more qubits and greater flexibility with atom placements, extending to significantly larger systems would be viable. Notably, however, as the number of atoms increases, the solution space (and the associated degeneracy of low-energy vacancy configurations) grows rapidly. Consequently, a larger number of experimental shots would likely be required to reliably identify the true ground state, potentially making such calculations prohibitively expensive despite being physically feasible.

IV. Conclusion

We have demonstrated that quantum annealing provides an effective framework for solving vacancy optimization problems in crystal lattices. The QUBO formulation successfully

captures the essential physics of vacancy energetics, as validated by strong correlation ($R^2 = 0.95$) for pure silicon and good correlation ($R^2 = 0.72$) for boron-doped systems with first-principles DFT calculations. Annealing achieves significant speedups over exhaustive search while maintaining physical accuracy, with execution times that remain approximately constant regardless of the number of vacancies for a fixed number of atoms.

The extension to boron-doped systems demonstrates the method's applicability to industrially relevant problems, correctly identifying boron-vacancy clustering patterns that impact device performance. As semiconductor manufacturing advances to smaller feature sizes, understanding and controlling defect configurations becomes increasingly critical, and quantum computing offers a scalable pathway for these challenging optimization problems.

The presented quantum algorithm, though currently limited by NISQ processors, shows great promise for scaling to large system sizes which are not solvable even with classical or semi-classical annealing schemes. When larger quantum machines are developed, the size and types of vacancy configuration problems that can be explored will far exceed what is currently possible.

Future work will explore larger system sizes, additional dopant types, and the integration of three-body interactions to further improve model accuracy. The strong correlation between QUBO and DFT energies suggests that quantum annealing could serve as a pre-screening tool for identifying promising defect configurations before expensive DFT calculations.

ACKNOWLEDGMENT

We gratefully acknowledge Amazon Web Services (AWS) for providing high-performance computing resources. We also thank Carnegie Mellon University for academic support through course 18-619/47-779/47-785, as well as Prof. Elias Towe for his valuable guidance throughout this project.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://github.com/chamber523/Quantum-computing-project.

REFERENCES

[1] B. Camino, J. Buckeridge, P. A. Warburton, V. Kendon, and S. M. Woodley, "Quantum computing and materials science: A practical guide to applying quantum annealing to the configurational analysis of materials," *J. Appl. Phys.*, vol. 133, no. 22, p. 221102, Jun. 2023, doi: 10.1063/5.0151346.

[2] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *Oper. Res.*, vol. 14, no. 4, pp. 699–719, Jul.–Aug. 1966, doi: 10.1287/opre.14.4.699.

[3] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, May 1983, doi: 10.1126/science.220.4598.671.

[4] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Phys. Rev.*, vol. 140, no. 4A, pp. A1133–A1138, Nov. 1965, doi: 10.1103/PhysRev.140.A1133.

[5] M. T. Nguyen, J. G. Liu, J. Wurtz, M. Lukin, S.-T. Wang, and H. Pichler, "Quantum optimization with arbitrary connectivity using Rydberg atom arrays," *PRX Quantum*, vol. 4, p. 010316, Jan. 2023, doi: 10.1103/PRXQuantum.4.010316.