Solving Higher Order Binary Optimization Problems on NISQ Devices: Experiments and Limitations

Valentin Gilbert[1][0009-0001-1004-0204], Julien Rodriguez[1,2][0000-0003-3583-0859]
Stéphane Louise[1][0000-0003-4604-6453], and Renaud Sirdey[1][0000-0003-4720-9269]

Université Paris-Saclay, CEA-List,F-91120[1]
Université de Bordeaux, INRIA[2]
Palaiseau, France
{valentin.gilbert, julien.rodriguez, stephane.louise, renaud.sirdey}@cea.fr

Abstract. With the recent availability of Noisy Intermediate-Scale Quantum devices, the potential of quantum computers to impact the field of combinatorial optimization lies in quantum variational and annealing-based methods. This paper further compares Quantum Annealing (QA) and the Quantum Approximate Optimization Algorithm (QAOA) in solving Higher Order Binary Optimization (HOBO) problems. This case study considers the hypergraph partitioning problem, which is used to generate custom HOBO problems. Our experiments show that D-Wave systems quickly reach limits solving dense HOBO problems. Although the QAOA demonstrates better performance on exact simulations, noisy simulations reveal that the gate error rate should remain under 10^{-5} to match D-Wave systems' performance, considering equal compilation overheads for both device.

Keywords: HOBO · Balanced hypergraph partitioning · Quantum computing · QAOA · Quantum annealing

1 Introduction

As we enter the Noisy Intermediate Scale Quantum (NISQ) era, companies are now building chips that control a few hundred qubits for quantum circuit models and several thousand for quantum annealers. The selection of interesting problems that run successfully on noisy quantum chips is now a key point of interest for researchers and industries alike. As quantum heuristics performance limits are easier to reveal in the higher instance density regime, which either requires qubit duplications or larger circuit depths, we use HOBO problems to generate k-local Hamiltonians of custom density. An experimental study of the impact of the HOBO formulation on the QAOA was done in [5], demonstrating that higher order formulations were favorable to the QAOA. E. Pelofske et al. [11] also compared the ability of QA and the QAOA to solve HOBOs perfectly adapted to `ibm_washington`'s graph connectivity containing cubic interaction terms, showing that current ideal QAOA execution on real hardware could not match QA results quality.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_18

We propose another study case to experimentally evaluate bounds on the error rate that would permit QAOA to beat QA on results quality. As a use case, we generate HOBO formulations from Balanced Hypergraph Partitioning (BHP) problems, which is well-known in combinatorial optimization due to the difficulty of finding a good solution. It consists in dividing the vertices into different subsets, considering a balancing constraint while minimizing the number of hyperedges connecting the partitions. The balancing constraint acts as a global constraint and requires a strong coupling between the variables of the problem. This problem is interesting as its transformation into the Ising model gives a fully connected 2-local Hamiltonian with some k-local terms representing hyperedges. A general formulation for graph bi-partitioning using the Ising model was proposed in [9]. This formulation has been extended to graph k-partitioning in [16], with an experimental comparison between state-of-the-art partitioning methods and the quantum hybrid method *qbsolv* , which seems competitive. H. N. Djidjev et al. [6] are less optimistic and demonstrate that the advantage of the quantum annealer is still limited by the size of the quantum chip. They also underline the importance of accounting for compilation time, which can represent up to 99% of the computation run time for large instances. Recent theoretical results on the limitations of the QAOA on pure k-spin model are available in [2]. The authors show that the QAOA is subject to optimality limitations for any even $k \geq 4$ in the infinite size limit for fixed p. It sets a first theoretical bound, proving that the QAOA may encounter strong limitations in solving HOBO problems.
Our contributions are two-fold. The first one is a recursive formulation of the BHP problem as a HOBO problem. The second contribution is a performance comparison of two quantum heuristics: the QA and the QAOA. In particular, our experiments suggest that noisy QAOA will only compete with D-Wave systems on low density problems if the error rate remains under 10^{-5}.

2 Problem Formulation

The formulation of the BHP problem is an extension of a previous work based on hypergraph bi-partitioning [13]. A Hypergraph is a generalization of a graph where hyperedges can be connected to one or more vertices. Let $\mathcal{H} \overset{\text{def}}{=} (\mathcal{V},\mathcal{E})$ the hypergraph defined from a set of vertices $\mathcal{V}$ and a set of hyperedges $\mathcal{E}$. A k-partition Π of $\mathcal{H}$ is a splitting of $\mathcal{V}$ into k vertex subsets π_i with $1 \leq i \leq k$, called parts, such that : (i) each part π_i respects the capacity constraint : $\forall i, |\pi_i| \leq \frac{|\mathcal{V}|}{k}$; (ii) all parts are pairwise disjoint : $\forall i,j \; i \neq j, \pi_i \cap \pi_j \; = \; \emptyset$; the union of all parts is equal to $\mathcal{V}$: $\bigcup_i \pi_i = \mathcal{V}$. A cut for a k-partition Π of $\mathcal{H}$ is the union of hyperedges that contain at least two vertices in different parts and the cut-size f_c is the number of cut edges. Our formulation minimizes the *min-cut* metric with a balanced constraint. Considering k the final number of partitions, at a given level of recursion, the capacity constraint for one recursion splitting the vertices v into 2 sub-parts $x_v \in \{0,1\}$ is:

$$H_A = \left(\sum_{v \in \mathcal{V}} \omega_v x_v - \left\lceil \left\lfloor \frac{k}{2} \right\rfloor / k \times \Omega(\mathcal{V}) \right\rceil \right)^2 \tag{1}$$

where ω_v is the weight of each node v and $\Omega(\mathcal{V})=\sum_{v\in\mathcal{V}}\omega_v$. $\Omega(\mathcal{V})=|\mathcal{V}|$ for unweighted graphs. This expression weights the number of nodes that should appear in π_0 and π_1 according to their total weight. The second component of the cost function is used to minimize the *min-cut* metric f_c:

$$H_B=\sum_{e\in\mathcal{E}}\left(\omega_e\times\left(1-\prod_{v\in e}x_v-\prod_{v\in e}(1-x_v)\right)\right) \tag{2}$$

ω_e corresponds to the weight of the hyperedge e. The weight ω_e is added to f_c when $\forall v\in e,\exists v'\in e$ with $v'\neq v$ and $x_v\neq x_{v'}$. The objectives H_A and H_B are then gathered to create the final objective to minimize. Coefficients A and B are real numbers and are used to weight each objective:

$$C(x)=AH_A+BH_B \tag{3}$$

The reader can refer to the method described in the paper of Lucas et al. [9] to set the coefficients A and B. The upper formulation is only valid for a recursive k-partitioning algorithm. If the formulation was for a k-direct partitioning, it would be possible to encode vertex affectation to each partition using logarithmic k-partition encoding, as for coloring problems [15].

3 Experimental Setup

We solve HOBO problems using two quantum optimization methods: QA and the QAOA. The metric used for comparison is the energy gap Δ_E^*, which is the difference between the energy of the ground state (classically exhaustively computed) and the mean energy of the expectation value.

Our work is based on hypergraphs composed of 10 nodes and 15 hyperedges. A first set of instances is composed of k-uniform hypergraphs with $k\in\{2,3,4\}$. The parameter k is limited to 4 to avoid trivial solutions. For each value of k, 15 instances are randomly generated.

3.1 Setup of D-Wave Systems

D-Wave processors [3] are designed to minimize an Ising cost function H taking an input vector $s=(s_1,s_2,...,s_n)$ with $s_i\in\{+1,-1\}$ where h_i and J_{ij} are real numbers.

$$H(s_1,s_2,...,s_n)=-\sum_{i=1}^{n}h_is_i-\sum_{i<j}^{n}J_{ij}s_is_j \tag{4}$$

The translation between QUBO and Ising cost function is straightforward with a simple variable change $x_i=\frac{1-s_i}{2}$.

Experiments were done on the most recent chip *Advantage*2_*prototype*1.1, which produced the best results minimizing the Ising cost functions. D-Wave systems require a QUBO formulation of the initial problem. The transformation used to convert

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_18

HOBO to QUBO is done using Rosenberg reduction [14] to quadratize the terms of the cost function.

We used the heuristic presented in [4] to map the QUBO on the D-Wave Quantum Process Unit. Our experiments consider average-quality embedding to avoid bias by selecting only the best embeddings over multiple tries. Majority voting is used during the post-processing phase to determine the final value of each variable. We do not use further specific processing such as spin reversal technique [12] or pausing time.

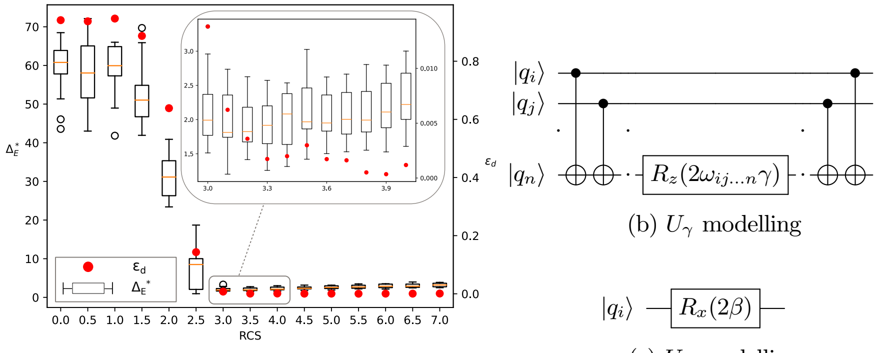
For each group of HOBO problems transformed to QUBO problems, we numerically study their optimal chain strength cs using a factor called Relative Chain Strength [17]:

$$cs = RCS \times max(\{h_i\} \cup \{J_{ij}\}) \tag{5}$$

The sampling of different values for the chain strength experimentally determines the optimal RCS factor. Fig. 1 a. shows that a phase transition occurs when the RCS factor becomes sufficient, leading to a significant improvement. The duplication error rate of qubits measured when the majority vote occurs follows the same phase transition. This evaluation is repeated for each k and optimal average values RCS^* are presented in Table 1.

3.2 Setup of the QAOA

The QAOA [7] circuit is built from the Hamiltonian derived from the HOBO cost function. Unlike D-Wave systems, the k-local Hamiltoninan can be implemented by the QAOA without quadratization. Fig. 1 b. and c. show the unitary implementation corresponding to both problem and mixing Hamiltonian. We perform perfect and noisy simulations of these quantum circuits using the IBM Qiskit library [10]. The *Aer*



(b) U_γ modelling

(c) U_β modelling

(a) RCS parameter sampling

Fig. 1: Parameter settings of QA and the QAOA. (a) shows the RCS parameter sampling for 2-uniform instances with impact on the energy gap Δ_E^* and duplication error rate on qubits ϵ_d. (b) shows the implementation of each $\omega_{ij...n}\sigma_i\sigma_j...\sigma_n$ term derived from the HOBO cost function terms. (c) shows U_β implementation with β rotation around the X-axis since the domain is not restricted.

simulator is used to perform the simulations, which offers a nice trade-off between execution speed and quality of the results. We use Qiskit Pauli error model for noisy simulation with the same error rate ϵ for bit-flip and phase-flip. We assume that the initialization and measurement of qubits are noiseless. Noisy simulations of quantum circuits are executed on different topologies to analyze the benefits of each chip density D. We study 3 different topologies: one which is fully connected, another one based on IBM's *ibmq_guadalupe* heavy-hex topology with a cycle layout of 12 qubits ($D=0.17$). The last topology comes from *sycamore* chip and is a grid layout of 12 qubits ($D=0.26$). The mapping of circuits on topologies that are not fully connected requires additional SWAP gates added by the Qiskit transpiler, which generates gate depth overheads. The QAOA experiment is done from $p=1$ to $p=30$. At each step $p=i$, a local optimizer is used to find the optimal set of angles $\vec{\gamma_p^*}=(\gamma_1,...,\gamma_p)$, $\vec{\beta_p^*}=(\beta_1,...,\beta_p)$. We use *Nelder-Mead* optimization method [8] with a maximal number of function evaluations (i.e. quantum circuit execution) set to 300. The concentration of good parameters at p-depth $(\vec{\gamma_p^*},\vec{\beta_p^*})$ for small values of p has been analytically proven in [1]. L. Zhou et al. [18] introduced an optimization method based on discrete sine and cosine transform that benefits from this parameter concentration. The authors call it FOURIER[q, R], and use it to initialize angles $(\gamma_{p+1},\ \beta_{p+1})$ from the sets $\vec{\gamma_p},\vec{\beta_p}$. The variable q specifies the length of the vector of frequencies. We consider the case when $q=p$, meaning that q parameter grows with p when a pair of angles is added. R parameter is the number of local optima calculated at each level p. Following their notations, we use FOURIER[∞,10] global optimization method for each QAOA simulation. For each experiment, we set $\gamma\in[0,2\pi]$ and $\beta\in[0,\pi]$.

4 Results

The impact of HOBO problem's density on QA and the QAOA is shown in Fig. 2. It compares D-Wave *Advantage2_prototype1.1* and the QAOA ability to find optimal solutions to HOBO problems generated from k-uniform hypergraphs. The perfect simulation of the QAOA surpasses D-Wave systems on 2-uniform (3-uniform) hypergraphs when $p=24$ ($p=22$). It shows that the increase in the cardinality of the hyperedges severely limits the performance of the D-Wave quantum computer, which becomes highly inefficient for cardinalities greater than 4. This performance loss is caused by the Rosenberg decomposition coupled with the qubits duplication needed for mapping the problem on D-Wave chip's topology. These two processing steps multiply the required physical qubits by 581% for 4-uniform hypergraphs for D-Wave systems (see Table 1). Comparatively, 4-uniform hypergraphs only increase QAOA depth by 155% on the cycle topology compared to a fully connected topology. The overheads difference is less important for 2-uniform hypergraphs (175% qubit overhead for D-Wave against 173% gate depth overhead on the cycle topology). One can observe that there is no significant difference in circuit depth overheads between cycle and grid layouts, meaning that the Qiskit transpiler algorithm doesn't fully take advantage of the higher connectivity of the grid layout.

We further study the group of 2-uniform hypergraphs and estimate a noise rate threshold of single and double qubit gates that would permit the QAOA to reach

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_18

BHP instances		D-Wave system Advantage2 prototype 1.1							
	#terms	RCS^*	#var QR			#qubits			
			min	max	mean	min	max	mean	ratio
2-uniform	55	3.6	10	10	10	17	18	17.5	175%
3-uniform	55	3.0	10	10	10	17	18	17.5	175%
4-uniform	119.6	0.5	24	29	26.8	48	70	58.1	581%

BHP instances	QAOA Circuit depth p=1										
	complete topology			cycle topology				grid topology			
	min	max	mean	min	max	mean	ratio	min	max	mean	ratio
2-uniform	96	126	110.6	167	220	191.3	173%	178	220	196.9	178%
3-uniform	105	150	129.2	174	234	206	159%	185	264	213.9	166%
4-uniform	281	362	323.2	438	581	499	155%	400	485	439	135%

Table 1: BHP problem instances description. Each set of instances is composed of 15 hypergraphs having 10 nodes and 15 edges. The (RCS^*) is calculated for each set. The first table shows the overheads of physical qubits needed by D-Wave systems. The second table shows the overheads of gate depth for each topology compared with a fully connected topology. Green cells highlight smallest overheads.

D-Wave performance. Optimal angles are considered to be known at each p-layer. The simulation is done with optimal angles ($\vec{\gamma_p^{opt}}$,$\vec{\beta_p^{opt}}$) found by the FOURIER[∞,10] method on *Aer* simulator. Fig. 3 shows the simulation of the QAOA considering various qubit layouts compared to D-Wave systems' best performance on the same instances. The QAOA simulation reaches the best expectation value found by D-Wave *Advantage2_prototype1.1* at $p = 30$ with $\epsilon = 10^{-5}$. Under this threshold, the QAOA becomes inefficient at $p \approx 10$ for $\epsilon = 10^{-3}$ and $p \approx 27$ for $\epsilon = 10^{-4}$. Curves on Fig. 3 a) and b) demonstrate lots of fluctuation, reminiscent of the noise impact

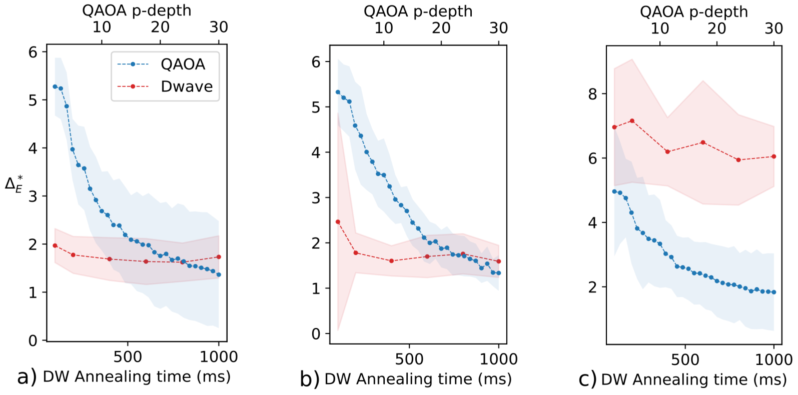


Fig. 2: Quantum heuristics performance solving the BPH problem, using the energy gap Δ_E^* as performance measure. Graphs (a), (b) and (c) respectively show the performance of D-Wave *Advantage2_prototype1.1* and QAOA on 2, 3 and 4-uniform hypergraphs. The shaded area represents the standard deviation of each curve.

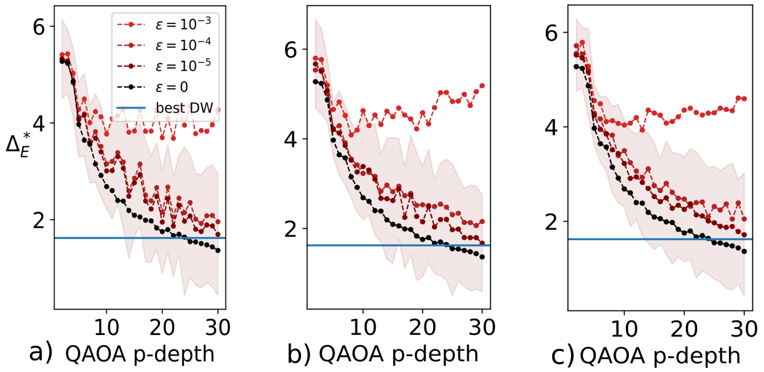


Fig. 3: Simulation of bi-partitioning 2-uniform hypergraphs using QAOA (red) and D-Wave (blue). a) , b) and c) respectively show noisy simulations on fully connected, cycle, and grid layouts. The shaded area represents the standard deviation at $\epsilon = 10^{-5}$.

on the optimization landscape, even when perfect angles are already known. This last experiment can be considered equally favorable to D-Wave and the QAOA, implying approximately 175% overheads for each physical implementation: 175% qubits overheads for D-Wave systems, versus 173% and 178% depth overheads for cycle and grid topologies. Considering an equal overhead produced by the compilation step, this experiment sets a first bound on noise rate to allow the QAOA to reach D-Wave best available systems, which is 10^{-5}.

5 Conclusion

This work proposes a general approach to compare the performance of D-Wave systems with the QAOA solving HOBO problems. We proposed a method to generate HOBO cost functions from BHP problems with various densities. The higher density regime, illustrated by HOBO problems with many terms, identifies the performance limitations of the QAOA and D-Wave systems. The former is limited by the noisy implementation of gates and the latter by its sparse topology. Even if the QAOA reaches D-Wave systems performance on perfect simulations for low-density problems, the variational heuristic gets rapidly stuck on noisy simulations. Our experiment suggests that a single quantum gate error rate $\epsilon < 10^{-5}$ would permit the QAOA to reach D-Wave systems performances, when the compilation overheads are the same for QA and the QAOA. This bound could be improved with more experiments and a larger set of instances. Current circuit chip designers are approaching this threshold with superconducting systems having $\epsilon \approx 10^{-3}$ and ion-based qubits systems having $\epsilon \approx 10^{-4}$ for single-qubit gates. On the other hand, dense HOBO problems represent hard instances for D-Wave systems, implying the use of quadratic reduction techniques and qubits duplications. Future work will investigate the density threshold for which the performance of the QAOA and QA crosses.

Acknowledgment

The work presented in this paper has been supported by AIDAS - AI, Data Analytics and Scalable Simulation - which is a Joint Virtual Laboratory gathering the Forschungszentrum Jülich (FZJ) and the French Alternative Energies and Atomic Energy Commission (CEA). We thank D. Vert for useful advice and fruitful discussions.

References

1. Akshay, V., Rabinovich, D., Campos, E., Biamonte, J.: Parameter concentrations in quantum approximate optimization. Physical Review A **104**(1), L010401 (2021)
2. Basso, J., Gamarnik, D., et al.: Performance and limitations of the qaoa at constant levels on large sparse hypergraphs and spin glass models. arXiv preprint arXiv:2204.10306 (2022)
3. Bunyk, P., et al.: Architectural considerations in the design of a superconducting quantum annealing processor. IEEE Transactions on Applied Superconductivity **24**(4), 1–10 (2014)
4. Cai, J., Macready, W.G., Roy, A.: A practical heuristic for finding graph minors. arXiv preprint arXiv:1406.2741 (2014)
5. Campbell, C., Dahl, E.: Qaoa of the highest order. In: 2022 IEEE 19th International Conference on Software Architecture Companion (ICSA-C). pp. 141–146. IEEE (2022)
6. Djidjev, H.N., Chapuis, G., Hahn, G., Rizk, G.: Efficient combinatorial optimization using quantum annealing. arXiv preprint arXiv:1801.08653 (2018)
7. Farhi, E., et al.: A quantum approximate optimization algorithm (2014)
8. Gao, F., Han, L.: Implementing the nelder-mead simplex algorithm with adaptive parameters. Computational Optimization and Applications **51**(1), 259–277 (2012)
9. Lucas, A.: Ising formulations of many np problems. Frontiers in physics p. 5 (2014)
10. MD SAJID, A., et al.: Qiskit: An open-source framework for quantum computing (2022). https://doi.org/10.5281/zenodo.2573505
11. Pelofske, E., Bärtschi, A., Eidenbenz, S.: Quantum annealing vs. qaoa: 127 qubit higher-order ising problems on nisq computers. arXiv preprint arXiv:2301.00520 (2023)
12. Pudenz, K.L.: Parameter setting for quantum annealers. In: 2016 IEEE High Performance Extreme Computing Conference (HPEC). pp. 1–6 (2016)
13. Rodriguez, J.: Quantum algorithms for hypergraph bi-partitioning. In: 23ème congrès annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision. INSA Lyon, Villeurbanne - Lyon, France (Feb 2022), https://hal.archives-ouvertes.fr/hal-03595234
14. Rosenberg, I.G.: Reduction of bivalent maximization to the quadratic case. Cahiers du Centre d'Etudes de Recherche Operationnelle **17**, 71–74 (1975)
15. Tabi, Z., et al.: Quantum optimization for the graph coloring problem with space-efficient embedding. In: 2020 IEEE International Conference on Quantum Computing and Engineering (QCE). pp. 56–62. IEEE (2020)
16. Ushijima-Mwesigwa, H., Negre, C.F., Mniszewski, S.M.: Graph partitioning using quantum annealing on the d-wave system. In: Proceedings of the Second International Workshop on Post Moores Era Supercomputing. pp. 22–29 (2017)
17. Willsch, D., et al.: Benchmarking advantage and d-wave 2000q quantum annealers with exact cover problems. Quantum Information Processing **21**(4), 1–22 (2022)
18. Zhou, L., Wang, S.T., Choi, S., Pichler, H., Lukin, M.D.: Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. Physical Review X **10**(2), 021067 (2020)