TAQOS: A Benchmark Protocol for Quantum Optimization Systems

Valentin Gilbert[1 [0009−0001−1004−0204]], Stéphane Louise[1 [0000−0003−4604−6453]], and Renaud Sirdey[1 [0000−0003−4720−9269]]

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

Abstract. The growing availability of quantum computers raises questions about their ability to solve concrete problems. Existing benchmark protocols still lack problem diversity and attempt to summarize quantum advantage in a single metric that measures the quality of found solutions. Unfortunately, the solution quality metric is insufficient for measuring quantum algorithm performance and should be presented along with time and instances coverage metrics. This paper aims to establish the TAQOS protocol to perform a Tight Analysis of Quantum Optimization Systems. The combination of metrics considered by this protocol helps to identify problems and instances liable to produce quantum advantage on Noisy-Intermediate Scale Quantum (NISQ) devices for useful applications. The methodology used for the benchmark process is detailed and an illustrative short case study on the Max-Cut problem is provided.

Keywords: Benchmark protocol · Quantum computing · QAOA · Quantum annealing.

1 Introduction

Quantum manufacturers are currently building chips with several hundred qubits for circuit-based quantum computers and thousands of qubits for quantum annealers. As the NISQ era [20] begins, it remains unclear whether noisy quantum computers will have useful applications in the near term, since quantum error correction codes still require too much qubits to be efficient. Defining whether a quantum algorithm could bring a quantum advantage on a specific task is far from straightforward, as the full quantum stack usually involves complex classic and quantum processing where each subpart constitutes a full research domain. One relevant class of problems that may be subject to quantum advantage are optimization problems that naturally map on Adiabatic Quantum Optimization (AQO) systems. Hybrid quantum algorithms also provide an interesting option to solve optimization problems, especially using the Quantum Approximate Optimization Algorithm (QAOA) [8]. This algorithm exhibits a robust behavior under noisy regime [11, 23] and encouraging theoretical bounds of convergence have been proven for specific problems at fixed depth [4,8]. The plethora of optimization problems being developed and benchmarked using Quantum Annealing (QA) and the QAOA requires a rigorous methodology to report the performance

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13

of these heuristics. To this end, this paper introduces one methodology termed the TAQOS protocol, which performs a Tight Analysis of Quantum Optimization Systems performance. This protocol defines the guidelines and properties of an application-based competitive benchmark. The instances and the source code are available at [10].

1.1 Related Work

Protocols used to benchmark the performance of classical heuristics appeared in the 1970s and provide useful guidelines to produce high-quality classical computer benchmarks. Several best practices and guidelines for evaluating computer performance exist in the current literature [15]. One important approach is found in [3], which splits the performance study into two types of benchmarks. The first type is *competitive benchmark* which aims to directly and quantitatively compare the performance of different algorithms. Whereas the second type, named *descriptive benchmark*, is used to analyze and understand the factors that impact algorithm performances. While *competitive benchmark* should be composed of fast-to-compute unbiased metrics to compare algorithm performance, *descriptive benchmark* can be composed of more complex metrics serving a better understanding of the algorithmic behavior.

As quantum annealers have improved (e.g., D-Wave systems [1]), the scientific community has begun to evaluate their performance against advanced classical heuristics. T. Albash et al. [2] showed that the scaling advantage of QA could outperform well-known classical heuristics such as simulated annealing. Several studies on specific Ising models, such as Spin-Glass and Sherrington-Kirkpatrick models, have shown that quantum annealers could perform better than classical methods on specific cases [12, 19].

Quantum circuit performance evaluation started with randomized benchmarking methods of single-qubit gates circuits. This protocol, presented by E. Knill in [13], was then extended to multi-qubits gates circuits in [17]. Both protocols are scalable as they are strictly based on circuits only using Clifford gates, producing an output distribution that can be known efficiently with a classical computer.

Other studies have tried to define a set of metrics to measure the potential of quantum circuits. The Quantum Volume [6] evaluates the maximum size of a square circuit that can run reliably on a given quantum chip. The Volumetric Benchmark [5] extends this method to rectangle circuits. Both metrics provide insights about the volume (width and depth) that can run reliably on a chip. The precise and costly evaluation of the output distribution (based on Heavy Output Generation) classifies both metric use into *descriptive benchmark*. These metrics do not report on the fidelity nor quality of the output of specific application circuits and are not scalable for such uses.

A scalable competitive metric, called the Q-score, has recently emerged to evaluate solutions to the Max-Cut problem [18]. This metric is the first attempt to design a hardware-independent way to measure quantum performance. Fellous et al. [9] introduced a methodology named Metric-Noise-Resource (MNR) to evaluate the ratio between energy consumption and quality of the solution provided by an error-corrected quantum computer. MNR is the first methodology to estimate energy consumption with the launching of the Quantum Energy Initiative.

Finally, several frameworks have been developed to benchmark applications, such as

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13

the QEC-D framework [16] and the QASM Bench [14]. These frameworks provide sets of metrics to test applications but are still dedicated to perform *descriptive benchmark* of small instances by studying the fidelity of the output distribution of the circuit.

2 TAQOS Benchmark Protocol

The TAQOS protocol aims to establish a fair benchmarking protocol to compare quantum algorithms, such as AQO and the QAOA, with classical algorithms. Fig. 1 shows the workflow of the two quantum heuristics. Each dotted box is an abstract description of a computational task. One can perform a factorial study by testing several implementations of a single dotted box and letting the rest of the workflow unchanged. Each of the two quantum heuristics exhibits at least one critical task proven NP-Hard: the Quadratic Unconstrained Binary Optimization (QUBO) problem mapping for AQO (task #2) and the transpilation of the circuit to the hardware topology for the QAOA (task #8). The methods used to select optimized hyper-parameters (task #3, #7, #8 and #10) for the execution of quantum algorithms should be specified. For example, the selection of the chain strength and the unembedding method should be documented for experiments on AQO using D-Wave systems. For the QAOA, the local and global optimization procedure to get an appropriate set of angles should be detailed with processing time spent and termination conditions. For the fairness of benchmark studies, each experiment should analyze the computation time corresponding to each dotted box.

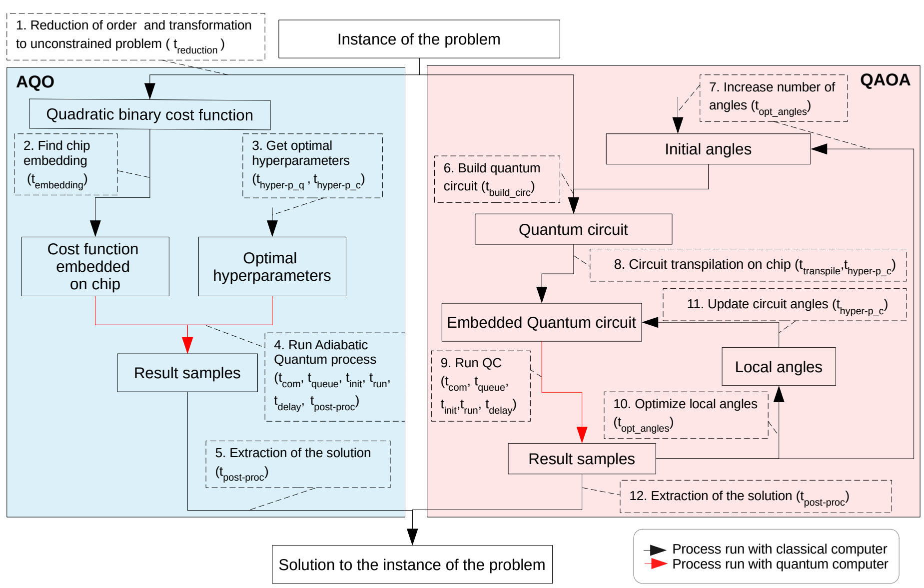


Fig. 1: Workflow of quantum optimization methods: AQO and the QAOA. Each dotted line box defines a processing action with run time variables t_x involved.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13

2.1 Metrics

Competitive benchmark metrics must be scalable, hardware-independent, and efficiently computed. In addition, the metrics must be comparable to the results obtained by state-of-the-art optimization methods. The TAQOS protocol considers three different quantities to evaluate the performance of a quantum solver: the quality of the result, the time to get the solution, and the energy spent by the computer to get the result. A fourth metric evaluates the robustness of the quantum heuristic and computes the coverage of the set of instances. Let $\mathcal{P} = \{P_1, P_2, ..., P_\mathrm{N}\}$ be a set of combinatorial problems and $\mathcal{I}$ the set of all possible instances associated to the problem P_n. Each instance $I_i \in \mathcal{I}$ has a set of solutions S. Let S_q be the subset of solutions found by a quantum computer. The objective function c evaluates the quality of a solution s. Considering a maximization problem, the best solution obtained by a quantum computer has the cost $c^*_\mathrm{q} = \max_{s_i \in S_\mathrm{q}} c(s_i)$ and is denoted s^*_q. Let $c_\mathrm{c}*$ be the cost associated with the best solution obtained by a classical heuristic and c_ref the cost of reference (for example, the best-known solution to a specific problem). The function r_ref evaluates the quality of a solution s_i as a ratio of the reference cost:

$$r_\mathrm{ref}(c(s_i), c_\mathrm{ref}) = \frac{c_\mathrm{ref} - c(s_i)}{c_\mathrm{ref}} \tag{1}$$

A negative ratio implies that the quality of the solution s_i is better than the solution of reference. Let the wall clock time associated with the quantum (classical) computation be t_q (t_c) and the energy consumption be e_q (e_c).The following set of inequalities defines a definitive quantum advantage over classical computation for a specific instance:

$$\begin{gathered} r_\mathrm{ref}(c^*_\mathrm{q}, c_\mathrm{ref}) \leq r_\mathrm{ref}(c^*_\mathrm{c}, c_\mathrm{ref}) \\ t^*_\mathrm{q} \leq t^*_\mathrm{c} \\ e^*_\mathrm{q} \leq e^*_\mathrm{c} \end{gathered} \tag{2}$$

Quality metric. The benchmark of NP-Hard problems requires the definition of an efficiently calculated quality metric. We opt to measure the quality of the solution following the recommendations of R. S. Barr et al. [3], taking the Best-Known Solution (BKS) as the solution of reference of cost c_ref. $r_\epsilon(\mathcal{I})$ computes the fraction of instances for which the ratio r_ref is less than ϵ. We define slices with $\epsilon \in [0,1]$, e.g., within 1, 5 or 10% to optimality, to detect sets of instances amenable to produce close-to-optimal results:

$$r_\epsilon(\mathcal{I}) = \frac{|\{I_i \in \mathcal{I} \text{ with } r_\mathrm{ref}(c_\mathrm{q}*, c_\mathrm{ref}) < \epsilon\}|}{|\mathcal{I}|} \tag{3}$$

$r_0(\mathcal{I})$ outputs the ratio of solutions that are better than the solution of reference.
Wall clock time metric. Wall clock time metric should include the whole processing time from the problem formulation to the solution extraction (see Fig. 1). The device setting time of hyper-parameter that require quantum (classical) processing is denoted $t_{\mathrm{hyper-p_q}}$($t_{\mathrm{hyper-p_c}}$). The wall clock processing time of AQO is defined in Eq. 4.

$$\begin{gathered} t_\mathrm{quantum} = (t_\mathrm{init} + t_\mathrm{run} + t_\mathrm{delay}) \times nb_\mathrm{shots} + t_{\mathrm{hyper-p_q}} \\ t_\mathrm{classical} = t_\mathrm{reduction} + t_\mathrm{embedding} + t_{\mathrm{hyper-p_c}} + t_\mathrm{com} + t_\mathrm{queue} + t_\mathrm{post-proc} \\ t_\mathrm{AQO\_wall\_clock} = t_\mathrm{quantum} + t_\mathrm{classical} \end{gathered} \tag{4}$$

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13

The wall clock processing time of the QAOA with local and global optimization of angles can be specified as:

$$\begin{aligned} t_{\text{local\_quantum}} &= (t_{\text{init}} + t_{\text{run}} + t_{\text{delay}}) \times nb_{\text{shots}} \times nb_{\text{local\_opt}} \\ t_{\text{local\_classical}} &= nb_{\text{local\_opt}} \times (t_{\text{com}} + t_{\text{queue}} + t_{\text{hyper-p}_c} + t_{\text{opt\_angles}}) \\ t_{\text{quantum}} &= nb_{\text{global\_opt}} \times t_{\text{local\_quantum}} \\ t_{\text{classical}} &= nb_{\text{global\_opt}} \times (t_{\text{build\_circ}} + t_{\text{compile}} + t_{\text{local\_classical}}) + t_{\text{post-proc}} \\ t_{\text{QAOA\_wall\_clock}} &= t_{\text{quantum}} + t_{\text{classical}} \end{aligned} \tag{5}$$

Energy consumption metric. Quantum computers are deemed less energy-consuming than supercomputers. However, their power consumption is presently not disclosed with enough precision by quantum hardware manufacturers. At this stage, we therefore let energy consumption metrics as perspectives.

Coverage metric. The last metric evaluates the coverage of the set $\mathcal{I}$. Classical studies based on Algorithm Selection Problem [21] demonstrated an existing link between the instance structure and the relative performance of specific heuristics [7,22]. Combined with the quality metric, the coverage metric evaluates the robustness of the heuristic. We follow the work of I. Dunning [7] and compute a set of metrics specific to one optimization problem (e.g., the density of an instance for a problem based on graphs). The coverage c_ϵ of a metric f is an interval at fixed ϵ:

$$c_\epsilon(f, I_i) = [f(I_i) - \epsilon, f(I_i) + \epsilon] \cap [0,1] \tag{6}$$

The whole coverage of a metric f on a set of instances $\mathcal{I}$ is:

$$C_\epsilon(f, \mathcal{I}) = \bigcup_{I_i \in \mathcal{I}} c_\epsilon(f, I_i) \tag{7}$$

These four metrics define the building blocks of the TAQOS protocol. An illustration of their use is presented in the next section.

2.2 Use Case on the Max-Cut Problem

This section presents a case study of the TAQOS protocol on the Max-Cut problem. The Max-Cut formulation is very close to the Ising model problem, easily mapped on existing qubit interconnects. Moreover, the classical community has studied this problem well, with several open-source implementations of heuristics (e.g., the MQLib [7]). Let $\mathcal{G} \stackrel{\text{def}}{=} (\mathcal{V}, \mathcal{E})$ the graph with a set of vertices $\mathcal{V}$ and a set of edges $\mathcal{E}$. The maximum cut of a graph is the partition of its vertices into two subsets S and T such that the number of edges shared by S and T is maximum. The cost function to be maximized is $C(\mathcal{G}) = -\sum_{i,j \in \mathcal{E}} \omega_{ij} s_i s_j$ with $s_i, s_j = \pm 1$. The problem is turned into a minimization problem by changing the sign of ω_{ij}.

Our instances of the Max-Cut problem are generated from the topology of four D-Wave systems. Random ω_{ij} coefficients are drawn from the set $\{+1, -1\}$ with same probability. Each instance is strongly favorable to D-Wave systems as it perfectly maps the topology of the quantum chip. However, the generated Ising Spin-glass

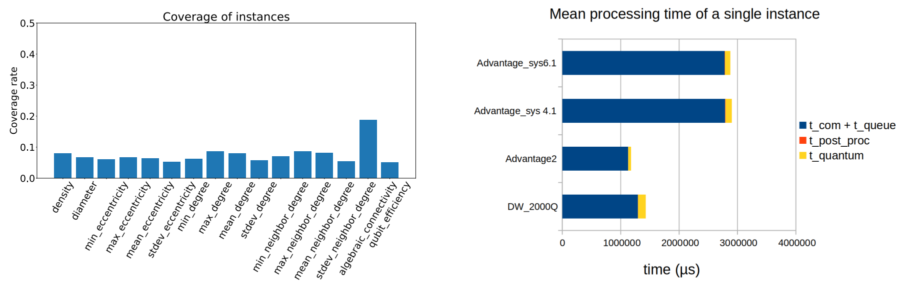
ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13

problem is still hard to solve for classical heuristics. Results are presented in Table 1. The benchmark is done on 30 instances for each graph, considering D-Wave solutions as reference solutions. The annealing time is set to 100 µs and the sampling is done over 256 shots (nb_{shots}). We did not tune the gauge inversion or pausing times. The D-Wave's performance is compared with algorithms from the MQLib [7] that constitute state-of-the-art methods used to solve the Max-Cut problem. Each classical algorithm is run over three time periods (1 s, 10 s, 100 s) on a single processor Intel® Core™ i7-6600U 2.6GHz. The metric c_ϵ is measured for $\epsilon \in \{0, 0.01, 0.05, 0.1\}$. For large graphs (i.e., Chimera and Pegasus) D-Wave annealers constantly outperform classical heuristics, even with less run time. The classical heuristics perform well on smaller graph (Zephyr) and outperform some reference solutions found by D-Wave, even with less run time. However, the competitive performance of the D-Wave systems must be interpreted considering the coverage rate of tested instances, shown in the Fig. 2a. These four graphs cover a very small range of graph-specific coverage metrics (less than 10% for almost every metric with $\epsilon = 0.05$). The run time of D-Wave systems is low because the set of instances, owing to their topology, avoids time-consuming operations such as reduction, embedding and hyper-parameter settings (see Fig. 2b). This use case shows the importance of being transparent about experiments done on quantum devices. The topology of instances strongly impacts the quality of the results

Table 1: Performance comparison between quantum and classical algorithms used to solve the Max-Cut problem on four different graphs tailored for D-Wave's quantum chips topology. Results are averaged over 30 instances for each graph. Green cells underline best classical runs for each time frame:{1, 10, 100}.

Quantum solvers	Wall clock time (s)	Chimera graph DW_2000Q $\lvert\mathcal{V}\rvert$:2041 $\lvert\mathcal{E}\rvert$:5974				Pegasus graph Adv4.1 $\lvert\mathcal{V}\rvert$:5621 $\lvert\mathcal{E}\rvert$:40279				Pegasus graph Adv6.1 $\lvert\mathcal{V}\rvert$:5616 $\lvert\mathcal{E}\rvert$:40135				Zephyr graph Adv2 $\lvert\mathcal{V}\rvert$:563 $\lvert\mathcal{E}\rvert$:4790			
DW2000Q	1.43	c_{ref}				/				/				/			
Adv4.1	2.90	/				c_{ref}				/				/			
Adv6.1	2.88	/				/				c_{ref}				/			
Adv2	1.18	/				/				/				c_{ref}			
Classical		r_0	$r_{0.01}$	$r_{0.05}$	$r_{0.1}$	r_0	$r_{0.01}$	$r_{0.05}$	$r_{0.1}$	r_0	$r_{0.01}$	$r_{0.05}$	$r_{0.1}$	r_0	$r_{0.01}$	$r_{0.05}$	$r_{0.1}$
Random	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.03
DUARTE 2005	1	0	0	0.97	1	0	0	0.9	1	0	0	0.7	1	0.1	0.3	1	1
	10	0	0	1	1	0	0	1	1	0	0	1	1	0.23	0.47	1	1
	100	0	0	1	1	0	0	1	1	0	0	1	1	0.4	0.6	1	1
FESTA 2002 GPR	1	0	0	0	1	0	0	0	0.37	0	0	0	0.53	0	0.17	0.93	1
	10	0	0	0.9	1	0	0	0	1	0	0	0	0.9	0.3	0.5	1	1
	100	0	0	1	1	0	0	0	1	0	0	0	1	0.3	0.53	1	1
FESTA 2002 GVNS	1	0	0	0	1	0	0	0.03	1	0	0	0	1	0	0.07	0.97	1
	10	0	0	0	1	0	0	0.03	1	0	0	0.07	1	0.1	0.3	1	1
	100	0	0	0	1	0	0	0.33	1	0	0	0.23	1	0.17	0.5	1	1
FESTA 2002 GVNSPR	1	0	0	0.03	1	0	0	0.03	1	0	0	0	1	0.07	0.27	1	1
	10	0	0	1	1	0	0	0.17	1	0	0	0.2	1	0.3	0.6	1	1
	100	0	0	1	1	0	0	0.5	1	0	0	0.37	1	0.4	0.67	1	1

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13



(a) Coverage of instances (b) Quantum heuristics run time

Fig. 2: (a) Shows the coverage rate of the set of evaluated instances. Coverage rates are computed from normalized graph metrics such as density, diameter, eccentricity, etc. The last metric measures the qubit mapping efficiency. The minimum, maximum, mean and standard deviation are available to study the distribution of these metrics. The coverage rate corresponds to the total length of intervals in $C_{0.05}(f,I)$. (b) Details the run time of quantum heuristics. Each time is averaged over the 30 instances. $t_{\mathrm{reduction}}$, $t_{\mathrm{embedding}}$ and $t_{\mathrm{hyper-p_q}}$ are set to 0 as the study does not require any of the corresponding computational task.

returned by quantum devices. The coverage metric quantifies its robustness and can be used to identify classes of instances producing high-quality results on quantum devices.

3 Conclusion

This paper has introduced the TAQOS benchmark methodology, which fairly compares classical and quantum heuristics performance. TAQOS is a scalable framework of metrics that analyses the trade-offs between quality and robustness. It constitutes a competitive methodology to benchmark hybrid algorithms such as AQO and the QAOA. It uses field-proven metrics to compare quantum to classical results obtained with existing benchmark methodologies. This paper illustrated the application of the TAQOS protocol on the Max-Cut problem in a favorable context for D-Wave systems and showed that performance reports should consider instances set coverage to avoid misleading conclusions. The use case illustrates the ability of TAQOS to gauge the fairness of quantum optimization experiments. In particular, this allows us to separate the experiments favorable to some quantum hardware from the more generic experiments that would manifest a real and robust quantum advantage. Future studies will be done on other optimization problems (especially Higher Order Binary Optimization problems and the TSP). This future work will provide insight into problems and instance properties that might benefit from a quantum advantage.

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13

References

1. D-wave systems, d-wave ocean sdk, (2022) release 6.0.1, https://github.com/dwavesystems/dwave-ocean-sdk
2. Albash, T., Lidar, D.A.: Demonstration of a scaling advantage for a quantum annealer over simulated annealing. Physical Review X **8**(3), 031016 (2018)
3. Barr, R.S., Golden, B.L., Kelly, J.P., Resende, M.G.C., Stewart, W.R.: Designing and reporting on computational experiments with heuristic methods. Journal of Heuristics **1**(1), 9–32 (Sep 1995)
4. Basso, J., et al.: The quantum approximate optimization algorithm at high depth for maxcut on large-girth regular graphs and the sherrington-kirkpatrick model. arXiv preprint arXiv:2110.14206 (2021)
5. Blume-Kohout, R., Young, K.C.: A volumetric framework for quantum computer benchmarks. Quantum **4**, 362 (Nov 2020)
6. Cross, A.W., Bishop, L.S., Sheldon, S., Nation, P.D., Gambetta, J.M.: Validating quantum computers using randomized model circuits. Phys. Rev. A **100**, 032328 (Sep 2019)
7. Dunning, I., Gupta, S., Silberholz, J.: What works best when? a systematic evaluation of heuristics for max-cut and QUBO. INFORMS Journal on Computing **30**(3) (2018)
8. Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (2014)
9. Fellous Asiani, M., et al.: Optimizing resource efficiencies for scalable full-stack quantum computers. arXiv preprint arXiv:2209.05469 (2022)
10. Gilbert, V.: Taqos. https://github.com/CEA-LIST/Quantum-Benchmark-CEA-LIST/tree/main/TAQOS (2023)
11. Harrigan, M.P., et al.: Quantum approximate optimization of non-planar graph problems on a planar superconducting processor. Nature Physics **17**(3), 332–336 (2021)
12. Hen, I., Job, J., Albash, T., Rønnow, T.F., et al.: Probing for quantum speedup in spin-glass problems with planted solutions. Phys. Rev. A **92**, 042325 (Oct 2015)
13. Knill, E., et al.: Randomized benchmarking of quantum gates. Phys. Rev. A **77**, 012307 (Jan 2008)
14. Li, A., Stein, S., Krishnamoorthy, S., Ang, J.: Qasmbench: A low-level qasm benchmark suite for nisq evaluation and simulation. arXiv preprint arXiv:2005.13018 (2020)
15. Lilja, D.J.: Measuring computer performance: a practitioner's guide. Cambridge university press (2005)
16. Lubinski, T., Coffrin, C., McGeoch, C., Sathe, et al.: Optimization applications as quantum performance benchmarks. arXiv preprint arXiv:2302.02278 (2023)
17. Magesan, E., Gambetta, J.M., Emerson, J.: Scalable and robust randomized benchmarking of quantum processes. Phys. Rev. Lett. **106**, 180504 (May 2011)
18. Martiel, S., Ayral, T., Allouche, C.: Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way. IEEE Transactions on Quantum Engineering **2**, 1–11 (2021)
19. Oshiyama, H., Ohzeki, M.: Benchmark of quantum-inspired heuristic solvers for quadratic unconstrained binary optimization. Scientific Reports **12**(1) (Feb 2022)
20. Preskill, J.: Quantum computing in the nisq era and beyond. Quantum **2**, 79 (2018)
21. Rice, J.R.: The algorithm selection problem. In: Advances in Computers, pp. 65–118. Elsevier (1976)
22. Smith-Miles, K., Baatar, D., Wreford, B., Lewis, R.: Towards objective measures of algorithm performance across instance space. Computers and Operations Research **45**, 12–24 (May 2014)
23. Xue, C., Chen, Z.Y., Wu, Y.C., Guo, G.P.: Effects of quantum noise on quantum approximate optimization algorithm. Chinese Physics Letters **38**(3), 030302 (2021)

ICCS Camera Ready Version 2023
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-36030-5_13