A Review and Collection of Metrics and Benchmarks for Quantum Computers: definitions, methodologies and software

Deep Lall[1], Abhishek Agarwal[1], Weixi Zhang[1,2], Lachlan Lindoy[1], Tobias Lindström[1], Stephanie Webster[1], Simon Hall[1], Nicholas Chancellor[3,4], Petros Wallden[2], Raul Garcia-Patron[2,5], Elham Kashefi[2,6], Viv Kendon[7], Jonathan Pritchard[7], Alessandro Rossi[1,7], Animesh Datta[8], Theodoros Kapourniotis[9], Konstantinos Georgopoulos[9], Ivan Rungger[1,10]

[1]National Physical Laboratory, Hampton Road, Teddington, Middlesex, TW11 0LW, United Kingdom | [2]School of Informatics, QSL, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, United Kingdom | [3]School of Computing, Newcastle | University, 1 Science Square, Newcastle upon Tyne NE4 5TG, United Kingdom | [4]Department of Physics, Joint Quantum Centre, Durham University, South Road, Durham DH1 3LE, United Kingdom | [5]Phasecraft Ltd., London, United Kingdom | [6]Sorbonne Université, CNRS, LIP6, 75005 Paris, France | [7]Department of Physics, University of Strathclyde, Glasgow G4 0NG, United Kingdom | [8]Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom | [9]National Quantum Computing Centre, Didcot OX11 0QX, United Kingdom | [10]Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, United Kingdom

Introduction

Numerous metrics to benchmark the performance of quantum computers, ranging from their individual hardware components to entire applications, have been proposed over the years. Navigating the resulting extensive literature can be overwhelming. Objective comparisons are further hampered in practice as different variations of the same metric are used, and the data disclosed together with a reported metric value is often not sufficient to reproduce the measurements.

This article addresses these challenges by providing a review of metrics and benchmarks for quantum computers and

1. a comprehensive collection of benchmarks allowing holistic comparisons of quantum computers,
2. a consistent format of the definitions across all metrics including a transparent description of the methodology and of the main assumptions and limitations, and
3. a reproducible approach by linking the metrics to open-source software, available at qcmet.npl.co.uk, used to evaluate them.

The benchmarks span aspects of a device such as architectural properties, quality, speed, and stability, and they also include metrics tailored to several non-gate-based architectures.

Generic metric entry

Metric title

Short definition of the metric.

Description

More detailed description of the metric.

Measurement procedure

Description of the methodology to measure the metric.

Assumptions and limitations

Where relevant, a list of assumptions underpinning the metric is provided, as well as practical limitations.

Source code

Where available, this section provides links to free open source libraries suitable for measuring the metric.

Metric references

At the end of each entry, a short list of relevant literature is provided.

Example (randomized benchmarking)

M3.3. Clifford randomized benchmarking average gate error

The Clifford randomized benchmarking (RB) average gate error metric provides an estimate of the average gate error of a set of single- and multi-qubit Clifford gates in a quantum computer [81, 82].

Description

Source code

A tutorial demonstrating how to calculate the Clifford RB average gate error for a single qubit is provided in the QCMet software tutorial under

`qcvv_execution_quality_metrics/randomized_benchmarking/clifford_randomized_benchmarking`

README.md

Clifford Randomized Benchmarking

This directory contains the code for Clifford randomized benchmarking (RB) average gate error metric. This metric provides an estimate of the average gate error of a set of single- or multi-qubit Clifford gates in a quantum computer. To run this metric, please use the `clifford_rb.ipynb` file.

Parameters

To run the RB protocol, you will need to run the `clifford_rb.ipynb` file.

There are parameters that can be adjusted, such as:

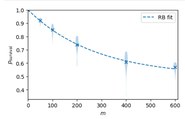
- `n_qubits` - the number of qubits to run RB on. Choose from 1 or 2
- `n_cliffords` - the list of the number of Clifford gates, i.e., a list of circuit depths
- `samples_per_depth` - the number of randomised circuits for each number of Clifford gates

Python tutorial

```
In [18]:
rb_exp_1q = RandomizedBenchmarking(n_qubits=1, n_cliffords=[0, 50, 100, 200, 400, 600], se
all_circs = rb_exp_1q.generate_all_circuits()
all_counts = {}
shots = 10240
for depth, circs in all_circs.items():
    t_circs = qiskit.transpile(circs, basis_gates = ['cx', 'id', 'rz', 'sx'])
    submitter.submit_circuits(shots, qasm_strs=[c.qasm() for c in t_circs], skip_asking=Tr
    all_counts[depth] = [submitter.convert_counts_to_qiskit(counts) for counts in submitte
rb_exp_1q.load_counts(all_counts, shots = shots)

In [19]:
# a_0 = np.power(alpha, a) + b_0
alpha, a_0, b_0 = 0.99,1,0
initial_guess = [alpha, a_0, b_0]
rb_exp_1q.get_fit(initial_guess = initial_guess)
average_gate_error = rb_exp_1q.get_average_gate_error()
fig = plt.figure(figsize=(5,3))
rb_exp_1q.plot_rb_results()
plt.legend()
plt.show()

Out [19]:
The Clifford randomized benchmarking average gate error is 0.0018205512168484738
```



Conclusions

Quantum computing technology has reached the stage where a number of methods for performance characterization are backed by a large body of real-world implementation and use, as well as by theoretical proofs. These mature benchmarking methods will benefit from commonly agreed-upon approaches. Indeed, such agreement is the only way to fairly, unambiguously, and objectively benchmark quantum computers across manufacturers. However, there are also benchmarking techniques that are still significantly evolving and for which standardization would be premature.

We identify five areas where international standardization working groups could be established, namely:

i) the identification and agreement on the categories of metrics that comprehensively benchmark device performance;
ii) the identification and agreement on a set of well-established metrics that together comprehensively benchmark performance;
iii) the identification of metrics specific to hardware platforms, including non-gate-based quantum computers;
iv) inter-laboratory comparison studies to develop best practice guides for measurement methodology; and
v) agreement on what data and software should be reported together with a metric value to ensure trust, transparency and reproducibility.

We provide potential routes to advancing these areas, thereby contributing to accelerate the progress of quantum computing hardware towards quantum advantage.

Performance Metrics for Quantum Computers

M1. HARDWARE ARCHITECTURE PROPERTIES

M1.1. Number of usable qubits
M1.2. Pairwise connectivity
M1.3. Native gate set
M1.4. Capability to perform mid-circuit measurements

M2. - M5. QUALITY METRICS

HARDWARE MANUFACTURER → *END-USER*

M2. QUBIT QUALITY METRICS

M2.1. Qubit relaxation time (T_1)
M2.2. Qubit dephasing time (T_2)
M2.3. Idle qubit purity oscillation frequency

M3. GATE EXECUTION QUALITY METRICS

M3.1. Gate set tomography-based process fidelity
M3.2. Diamond norm of a quantum gate
M3.3. Clifford randomized benchmarking average gate error
M3.4. Interleaved Clifford randomized benchmarking gate error
M3.5. Cycle-benchmarking composite process fidelity
M3.6. Over- or under-rotation angle
M3.7. State preparation and measurement fidelity

M4. CIRCUIT EXECUTION QUALITY METRICS

M4.1. Quantum volume
M4.2. Mirrored circuits average polarization
M4.3. Algorithmic qubits
M4.4. Upper bound on the variation distance

M5. WELL-STUDIED TASK EXECUTION QUALITY METRICS

M5.1. Variational Quantum Eigensolver metric
M5.2. Quantum Approximate Optimization Algorithm metric
M5.3. Fermi-Hubbard model simulation metric
M5.4. Quantum Fourier Transform metric

HARDWARE COMPONENTS → *APPLICATIONS*

M6. SPEED METRICS

M6.1. Time taken to execute a general single- or multi-qubit gate
M6.2. Time to measure qubits
M6.3. Time to reset qubits
M6.4. Overall device speed on reference tasks

M7. STABILITY METRICS

M7.1. Standard deviation of a specified metric evaluated over a time interval

M8. - M10. NON-GATE-BASED QC METRICS

M8. METRICS FOR QUANTUM ANNEALERS

M8.1. Single qubit control errors
M8.2. Size of largest mappable fully connected problem
M8.3. Dimensionless sample temperature

M9. METRICS FOR BOSON SAMPLING DEVICES

M9.1. Hardware characterization and model as metrics
M9.2. Quantum advantage demonstration as metric

M10. METRICS FOR NEUTRAL ATOM DEVICES

M10.1. Analogue process fidelity
M10.2. Trap lifetime
M10.3. Reconfigurable connectivity

arXiv:2502.06717

Ackgnolwedgements: This project was funded and supported by the UK National Quantum Computer Centre [NQCC200921], which is a UKRI Centre and part of the UK National Quantum Technologies Programme (NQTP). JP acknowledges support from EPSRC (Grant No. EP/T005386/1). AR acknowledges support from UKRI (Grant No. MR/T041110/1). We thank the Collaborative Computational Project Quantum Computing (ccp-qc, EPSRC Grant No. EP/T026715/2), the EPSRC Quantum Computing and Simulation Hub (EP/T001062/1), the UKRI ExCALIBUR project QEVEC (EP/W00772X/2), and the Superconducting Quantum Materials and Systems Center (SQMS, DE-AC02-07CH11359), for facilitating collaboration and discussion

www.npl.co.uk

© NPL Management Limited, 2023