

Metriq: Community-driven quantum computing benchmarks

Vincent Russo vincent@unitary.fund



TQCI Seminar, 11 May 2023 Teratec Quantum Computing Initiative Thales Research & Technology



Teratec Quantum Computing Initiatiue

THE TERM



WILLIAM ZENG, PHD



NATHAN SHAMMAH, PHD



BEN CASTANON

Ryan LaRose, PhD Now at EPFL→ MSU





ANDREA MARI, PHD



MISTY WAHL



NATE STEMEN



DAN STRANO

VINCENT RUSSO, PHD



FRANCES POBLETE



Collaboration with Da Vinci Labs

Unitary Fund to open first European office in Tours (France) in collaboration with Da Vinci Labs









https://unitary.fund/posts/eu.html

• Why do we want to benchmark?

• Why do we want to benchmark?

Answer the question:

"How does QC Platform X running Software Stack Y perform on Workload Z and how has that changed over time?"

• Why do we want to benchmark?

Answer the question:

"How does QC Platform X running Software Stack Y perform on Workload Z and how has that changed over time?"

• What makes a good benchmark?

• Why do we want to benchmark?

Answer the question:

"How does QC Platform X running Software Stack Y perform on Workload Z and how has that changed over time?"

- What makes a good benchmark?
 - Reproducible¹
 - Scalable²
 - Application-centric²
 - Hardware-agnostic²

¹Dasgupta, Samudra, and Travis S. Humble. "Characterizing the stability of nisq devices." *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 2020. ²Martiel, Simon, Thomas Ayral, and Cyril Allouche. "Benchmarking quantum co-processors in an application-centric, hardware-agnostic and scalable way." *arXiv preprint arXiv:2102.12973* (2021).

AI/ML community benchmarks

	← → C ⁱ	withcode.com/sota		··· 🖂 🕁	in m 🔿 🖪 🔲 🧏 🖪	
	E IQT P					. ⊗ Ξ
MLPerf Training Inference Cet Involved About GitHub Forum	[IIII] Search for papers, code and tasks	Q Browse Sta	te-of-the-Art Methods	Trends About RC202	10 🎐 🏪 Log In/i	/Register
Overview						
Rules Reference implementations	Browse State-of-the-Art					
How to submit NAL Do wf						
Results MLPEFT			Follow on 🕑 Twitter for updates			
Use results Fair and useful benchmarks for measuring training and inference	Computer Vision					
performance of ML hardware, software, and services.	E 66 benchmarks	Lassification	Object Detection Let 139 benchmarks 943 papers with code	Image Generation Et 121 benchmarks 458 papers with code	Pose Estimation Let 99 benchmarks 432 papers with code	
	> See all 907 tasks					
	Natural Language Processing					
What's New 10/11/20: MLPerf Inference v0.7 results are available. 7/29/20: MLPerf Training v0.7 results are available. 11/c/19: MLPerf Inference v0.5 results are available. 7/10/19: MLPerf Inference v0.5 launched. Submissions due 10/11. Results public 11/6.	49 benchmarks 91 papers with cole	Language Language Language Modelling Language Modelling 16 benchmarks 871 papers with code	Question Answering Le 59 benchmarks 765 papers with code	E 46 benchmarks 540 papers with code	Text Generation Let 44 benchmarks 366 papers with code	
4.14/13: ML/Perf Training vo. Busurbs due 5/24. 12/2/14: ML/Perf Training vo.5 launched. Results due 11/9. 5/2/18: MLPerf Training vo.5 launched. Results due 11/9.	Medical					
https://rejort.org/training-overview/	ttps://paperswithcode.com/sola	Drug Discovery	Lesion Segmentation	Brain Tumor Segmentation	COVID-19 Diagnosis	

Benchmarking is (almost) non-existent

- No **repository of state-of-the-art benchmarking results** ala "MLPerf" and "Papers with Code" in machine learning.
- QC Benchmarking today does not compare apples to apples, unlike machine learning
 - No standard benchmarks (except QV and randomized benchmarking)
 - No standard workloads
 - No standard interpretation of results
 - No standard ways of compiling or use of error mitigation
 - 0 ...



API

Community-driven quantum computing benchmarks, metriq.info



github.com/unitaryfund/metrig-client

Metriq updates: Community development, benchmark pipeline

- Latest Release: v.0.7.1
- First Metriq Hackathon was on March 24, 25, 2023: 23 new account sign-ups!
- Added a pipeline for Metriq to run and provide cloud benchmarks, (under active development as a 2023 priority)



the parent/child task hierarchy through top-level task categories. Search tasks Featured **Ouantum volume** 73 21 25 Quantum volume (QV) is a benchmark for guantum computing hardware. It expresses the maximum size (quantum circuit depth \$n\$ times number of gubits \$n\$) of square guantum circuits that can be implemented to at least \$2/3\$ success rate by a quantum computer. The "Quantum volume" chart shows \$2^n\$, whereas \$n\$, is shown in the "Log quantum volume". An additional chart shows the total number of "Qubits" in the chip. Quantum volume **Chart Metric:** Label arXiv ID Logarithmic: Log base: State-of-the-art Historical state-of-the-art labels All (±95% CI, when provided) Heavy bitstring sampling Quantinuum System Model H1-1 14 Heavy bitstring sampling Heavy bitstring sampling Heavy bitstring sampling Quantinuum System Model H1-2 arXiv:2110.14808 Heavy bitstring sampling Quantinuum System Model H1-1 8 Heavy bitstring sampling arXiv:2003.0119 System Model H1-1 arXiv:1811.129

Tasks are workloads of interest performed on a quantum computer. Search the task hierarchy to see charts of comparative performance across methods, see our submitter leader board and featured task charts, or click into

Walk-through

Metriq: https://metriq.info/







Thank you!

Contact: vincent@unitary.fund



Metriq's purpose: A platform to enable QC benchmarking

- We can answer the question: *How does QC Platform X running Software Stack Y perform on Workload Z, and how has that changed over time?*
- Developers and users can submit benchmarking results frictionlessly, via web form or Python client, which wrap our REST API.
- Metriq will become a focal point for a large number of short-term and long-term benchmarking efforts sprouting up in the quantum ecosystem.
- **Reproducibility** of benchmarking results in the space will become a habit.
- The long-term cost of deploying quantum computing will go down.

Dasgupta, Samudra, and Travis S. Humble. "Characterizing the stability of nisq devices." 2020 IEEE International Conference on Quantum Computing and Engineering (QCE). IEEE, 2020. Martiel, Simon, Thomas Ayral, and Cyril Allouche. "Benchmarking quantum co-processors in an application-centric, hardware-agnostic and scalable way." *arXiv preprint arXiv:2102.12973* (2021). Application-Oriented Performance Benchmarks for Quantum Computing <u>https://github.com/SRI-International/QC-App-Oriented-Benchmarks</u>