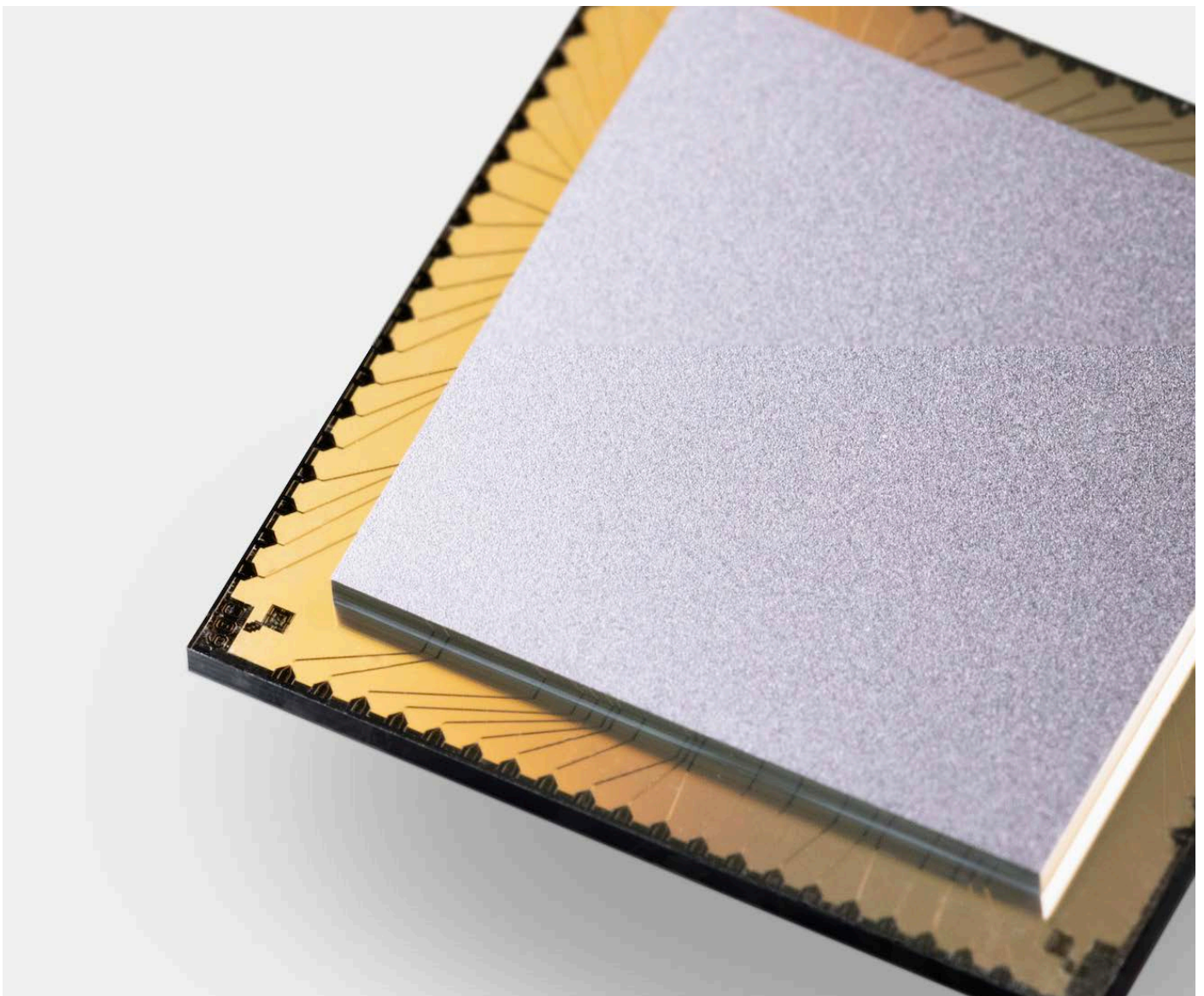




Technology and Performance Benchmarks of IQM's 20-Qubit Quantum Computer

IQM Garnet



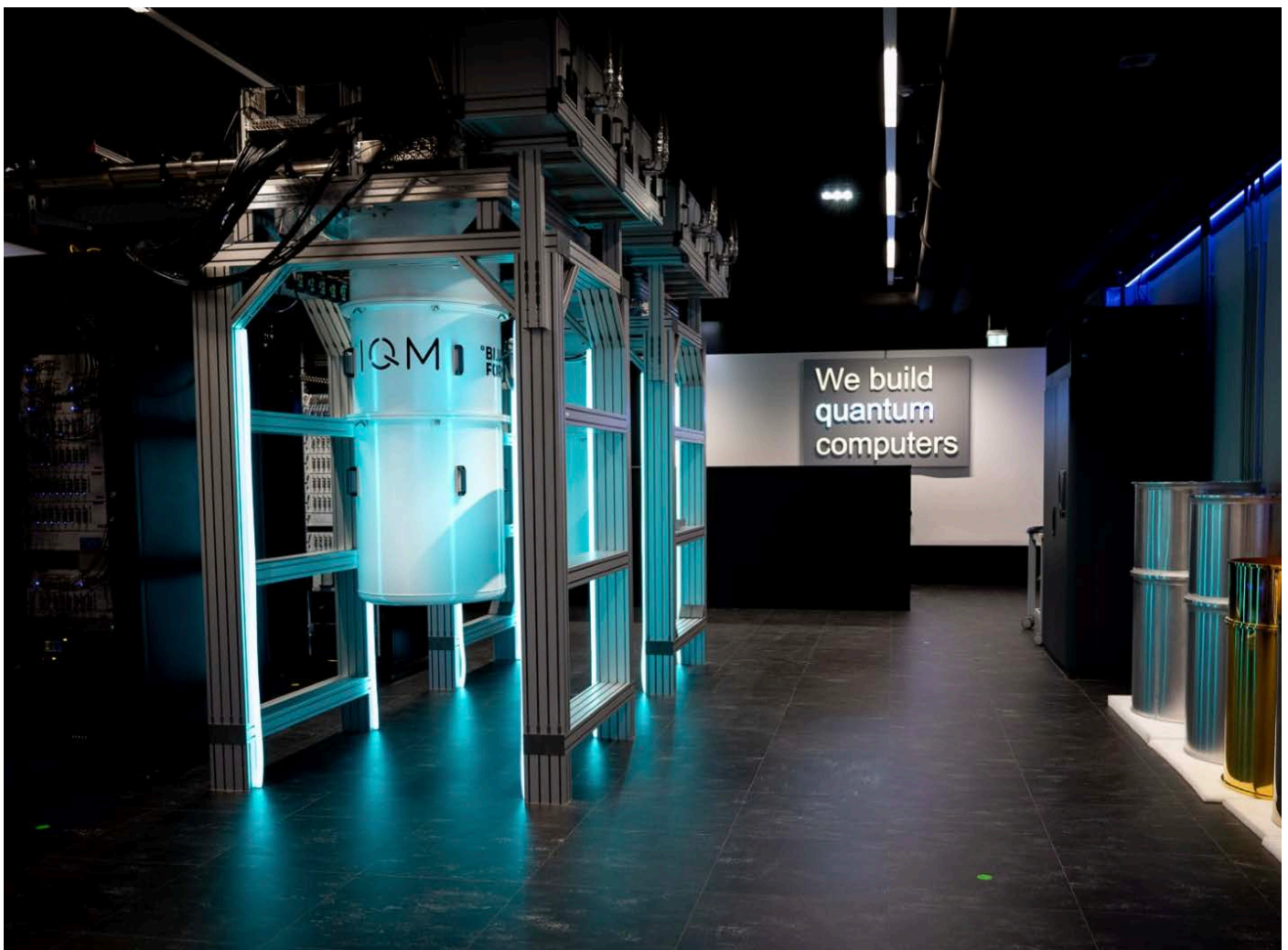
Contributors

Leonid Abdurakhimov, Janos Adam, Hasnain Ahmad, Manuel Algaba, Guillermo Alonso, Ville Bergholm, Rohit Beriwal, Matthias Beuerle, Clinton Bockstiegel, Alessio Calzona, Chun Fai Chan, Daniele Cucurachi, Saga Dahl, Rakhim Davletkaliyev, Olexiy Fedorets, Alejandro Gomez Frieiro, Zheming Gao, Johan Guldmyr, Andrew Guthrie, Juha Hassel, Hermann Heimonen, Johannes Heinsoo, Tuukka Hiltunen, Keiran Holland, Juho Hotari, Hao Hsu, Antti Huhtala, Eric Hyypä, Aleksi Hämäläinen, Joni Ikonen, Sinan Inel, David Janzso, Teemu Jaakkola, Mate Jenei, Shan Jolin, Kristinn Juliusson, Jaakko Jussila, Shabeeb Khalid, Seung-Goo Kim, Miikka Koistinen, Roope Kokkonen, Anton Komlev, Caspar Ockeloen-Korppi, Otto Koskinen, Janne Kotilahti, Toivo Kuisma, Vladimir Kukushkin, Kari Kumpulainen, Ilari Kuronen, Joonas Kylmälä, Niclas Lamponen, Julia Lamprich, Alessandro Landra, Martin Leib, Tianyi Li, Per Liebermann, Aleksi Lintunen, Wei Liu, Jyrge Luus, Fabian Marxer, Arianne Meijer–van de Griend, Kunal Mitra, Jalil Khatibi Moqadam, Jakub Mrozek, Henriikki Mäkynen, Janne Mäntylä, Tiina Naaranoja, Francesco Nappi, Janne Niemi, Lucas Ortega, Mario Palma, Miha Papič, Matti Partanen, Jari Penttilä, Alexander Plyushch, Wei Qiu, Aniket Rath, Kari Repo, Tomi Riipinen, Jussi Ritvas, Pedro Figueroa Romero, Jarkko Ruoho, Jukka Rabinä, Sampo Saarinen, Indrajeet Sagar, Hayk Sargsyan, Matthew Sarsby, Niko Savola, Mykhailo Savytskyi, Ville Selinmaa, Pavel Smirnov, Marco Marín Suárez, Linus Sundström, Sandra Stupićska, Eelis Takala, Ivan Takmakov, Brian Tarasinski, Manish Thapa, Jukka Tiainen, Francesca Tosto, Jani Tuorila, Carlos Valenzuela, David Vasey, Edwin Vehmaanperä, Antti Vepsäläinen, Aapo Vienamo, Panu Vesanen, Alpo Välimaa, Jaap Wesdorp, Nicola Wurz, Elisabeth Wybo, Lily Yang and Ali Yurtalan

Quantum computing holds enormous potential for overcoming some of the fundamental limitations of classical information processing. However, current technological constraints in quality and scalability are preventing us from fully unlocking this potential. Among the various approaches, quantum computing based on superconducting quantum processing units (QPUs) stands out as particularly promising for achieving practical quantum advantage.

In this article, we explore the technological advancements of IQM Quantum Computers, focusing on both the QPU and the complete full-stack quantum computer. Our spotlight is on a 20-qubit quantum computer, featuring the IQM Garnet QPU, which we will scale up to 150 qubits. Additionally, we share benchmarks for both the QPU and system levels, highlighting achievements such as a median 2-qubit gate fidelity of 99.5% and the genuine entanglement of all 20 qubits in a Greenberger-Horne-Zeilinger (GHZ) state.

”



1. Introduction

The exponential nature of the quantum state, represented by the complex valued 2^N -dimensional state vector of a quantum computer with N qubits leads to the failure of any classical simulation to predict the operation of even moderate-sized systems. For the most powerful classical supercomputing clusters, the limit is today on the order of 50 qubits, above which the simulation becomes unfeasible in humanly accessible computing times [1, 2]. There are many known quantum algorithms today that can exploit the properties of quantum mechanics to overcome some fundamental limitations of classical information processing, enabling computations otherwise impossible or extremely hard [3]. Yet, the immature state of quantum hardware today prevents full exploitation of the potential.

Many variants of superconducting qubits and quantum processors have been introduced to date [4]. In this white paper we present the solutions developed by IQM. The qubits and coupling structures of the QPU are described along with other sub-systems, including the cryogenic system with connectivity, room temperature control electronics, and software. These solutions are employed in IQM systems with up to 150 qubits. High-quality quantum computing systems of this scale have potential for early quantum utility, and they constitute a necessary milestone in the roadmap towards broader quantum advantage. In more detail, we describe the 20-qubit quantum computer representative of IQM core technology choices. To demonstrate the performance enabled by this technology, we present extensive benchmarking results from foundational level fidelity benchmarks to application benchmarks quantifying performance in selected classes of computational tasks.

2. QPU Architecture

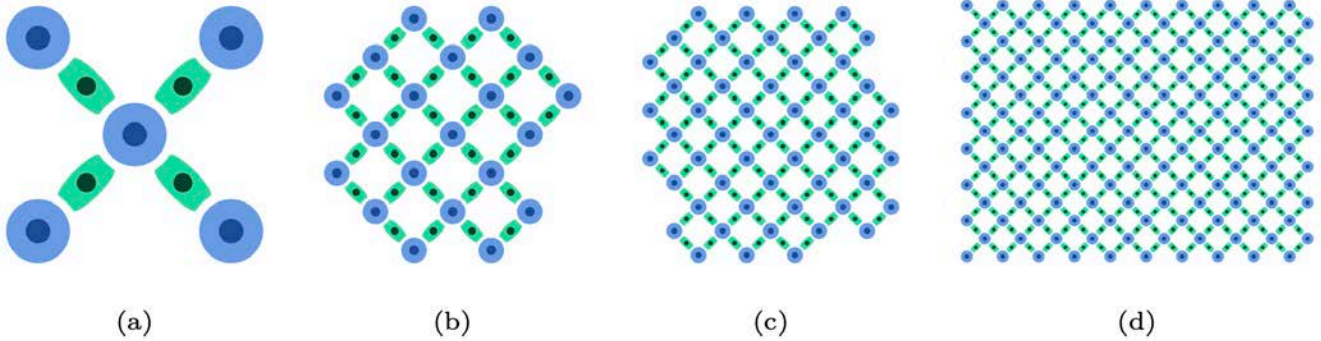


Fig. 1: IQM qubit crystal QPU topologies (a) with 5, (b) with 20, (c) with 54, (d) and with 150 computational qubits (blue) interconnected with tunable couplers (green).

IQM offers QPUs with different architectures. While resonator star architecture provides higher connectivity [5, 6], qubit crystal topology allows more parallelism, further discussed below. Today, IQM offers qubit crystals with computational qubit count from 5 to 150, see Fig. 1, with the focus of this white paper being on the 20-qubit QPU we refer to as IQM Garnet. In IQM qubit crystals, qubits are arranged on a square lattice where the lattice is rotated respect to crystal edge by 45 degrees. Qubit-qubit connectivity is mediated by tunable couplers. Both computational qubits and tunable couplers are based on flux-tunable transmon qubits. Qubit states are probed by frequency-multiplexed dispersive readout.

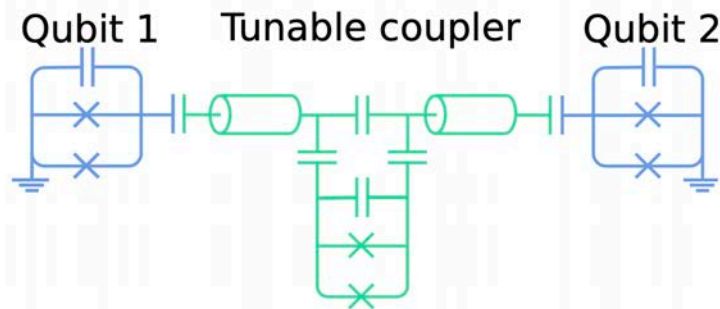


Fig. 2: Circuit representation of the qubit-coupler-qubit unit

2.1 Connectivity

The main way in which gate errors influence the final error of the algorithm is via the algorithm's runtime; in particular, its dependence of the circuit's depth. Within planar geometries, the square lattice with nearest-neighbor connectivity is close to ideal as a general purpose platform providing a balance between parallelism and worst-case routing distance. Consider an N -qubit circuit with commuting 2-qubit gates between each and every qubit, for example the phase separator in a QAOA circuit for a Sherrington-Kirkpatrick spin-glass problem. On an all-to-all connected QPU with no parallelization capabilities the runtime would be proportional to N^2 , while for the fully parallelized one-dimensional QPU the runtime would be proportional to N [7]. The square grid QPU has the same scaling as the one-dimensional QPU with a better prefactor. In contrast, for a sparse circuit, where only few gates can run in parallel due to the nature of the algorithm, like QAOA for MaxCUT on regular graphs, the circuit runtime scales as \sqrt{N} [8, 9] outperforming maximally parallel qubit chain.

The mentioned strengths of square lattice also reduce the gate performance requirement for quantum supremacy demonstrations using random circuit sampling experiments [1] because of the increased tensor-network simulation complexity of this topology compared to, for example a heavy-hex topology [10, 11].

Furthermore, the square array topology is compliant with standard surface code based error correction schemes. For surface code experiments, we have further increased the utility by the choice of angle in between lattice and crystal edge which increases the number of possible weight-4 parity checks for the given qubit number.

2.2 Quantum gates

The QPU and control system architecture support a universal gate-set based on single-qubit rotation gates around X and Y axes, arbitrary-angle virtual-Z gate [12], and the conditional phase gate (CZ). The single-qubit rotation gates are implemented as microwave pulses applied through control lines that are capacitively coupled to the qubit island. The pulse shapes are optimized to minimize the leakage to any of the adjacent qubit states or to the higher excited states of the driven qubit [13].

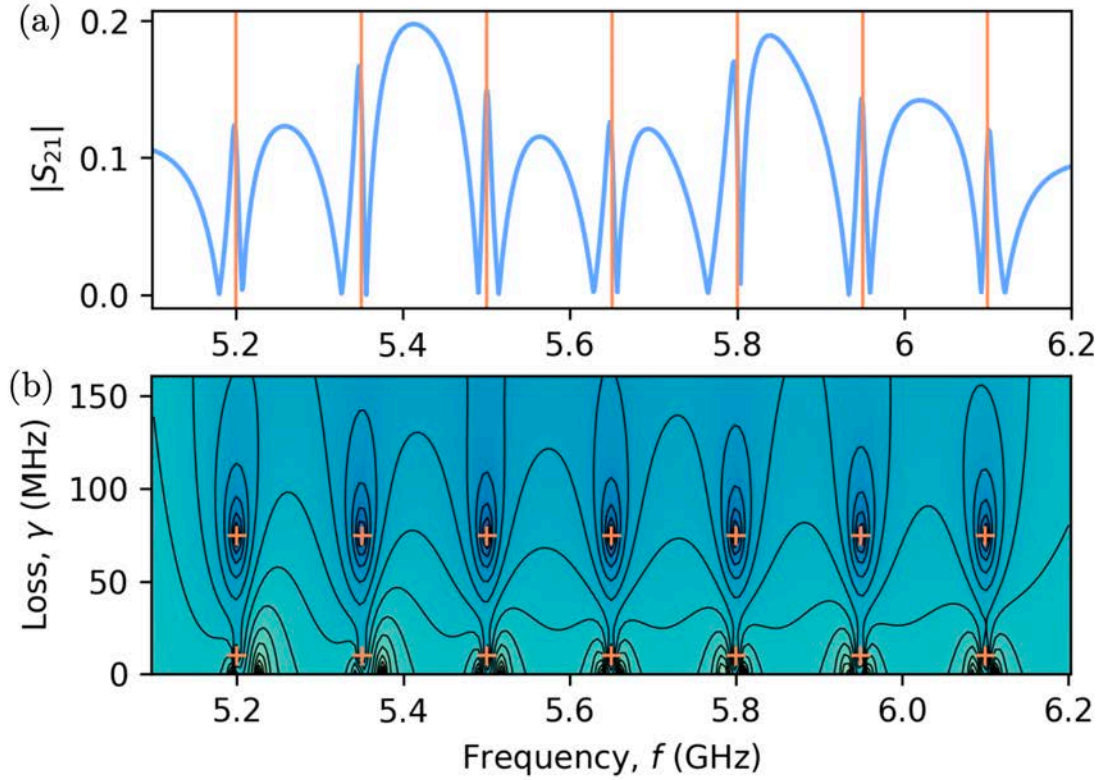


Fig. 3: Designed readout circuit (a) transmission magnitude as a function of frequency $S_{21}(f)$ and (b) impedance magnitude $\log_{10}(|Z_{11}|)$ as a function of complex frequency s , where red crosses indicate target mode locations in frequency and linewidth.

The qubits are coupled to each other through transmon-based couplers that enable strong tunable ZZ -type interaction up to 50 MHz which can be fully turned off during the idling operation [14]. In contrast, static capacitive coupling would limit the effective qubit-qubit interaction on-off ratio to about 50, resulting in either slow gates or unacceptable amount of ZZ -type errors. The CZ-gate is implemented by applying a magnetic flux pulse on the coupler, resulting in a fast gate with a typical duration of 20 ns – 40 ns.

In addition to the CZ gate, the architecture supports multiple other gate types, such as the cross-resonance gate [15], the parametric resonance gate [16], or the iSWAP gate [17], which enables choosing the optimal gate-set to fit the particular purpose. In addition, the qubit frequencies can be tuned by current-biasing magnetic flux lines of individual qubits, enabling efficient execution of analog quantum simulations on the QPU [18].

2.3 Qubit readout

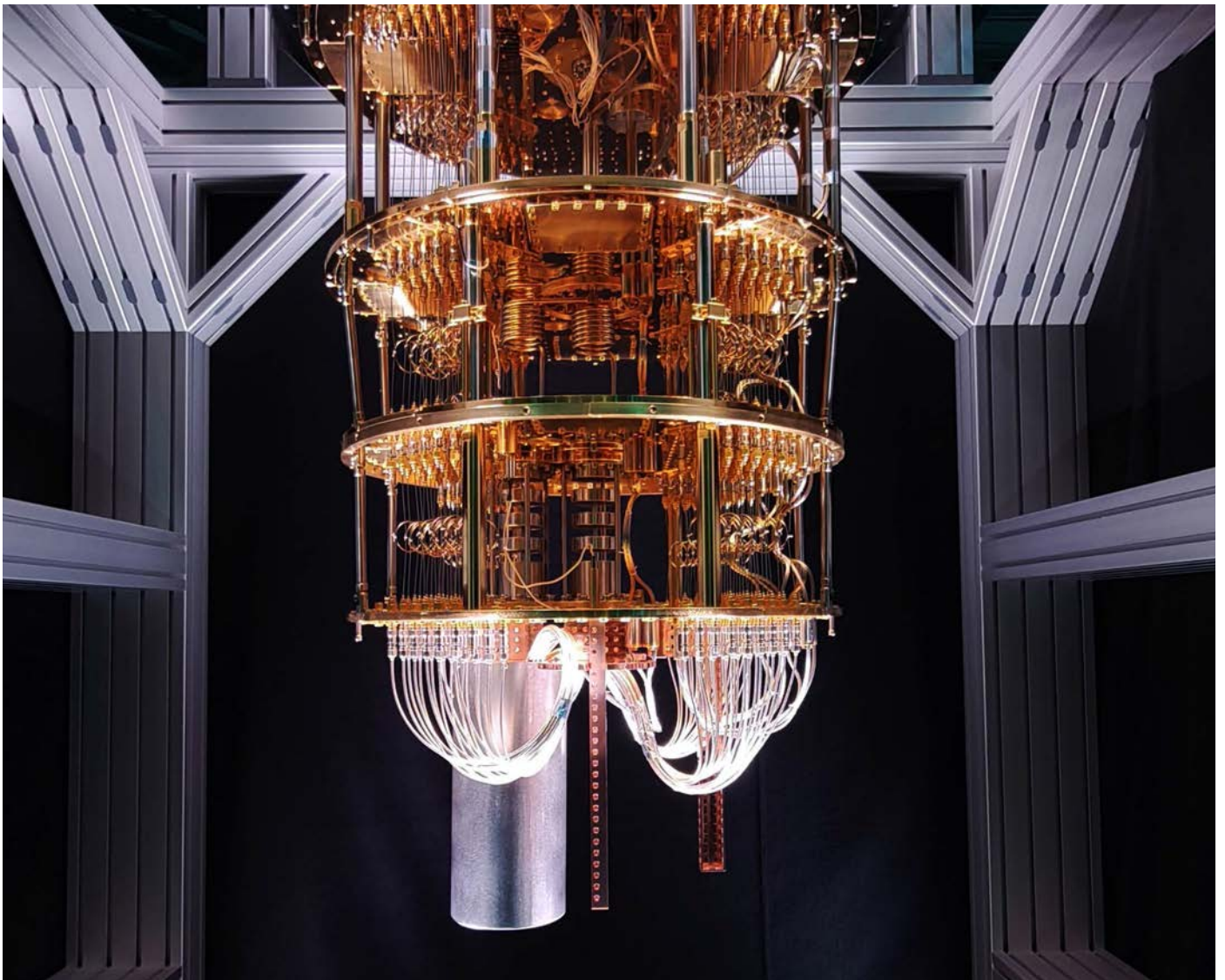
For qubit readout, each qubit has its own readout structure composed of a narrow band readout resonator and a wide band Purcell filter [19], with the former being capacitively coupled to the qubit island. A subset of readout structures is then coupled to a wider passband-like filter embedded in the probe line. The standing wave in the readout probe line enhances the coupling with readout structures and enables faster qubit readout. The variant of IQM Garnet presented in this whitepaper features 3 probe lines with 7, 7 and 6 readout structures coupled to each, see the example spectrum in Fig. 3. Although the transmission spectrum of readout structures is non-trivial, standard qubit state discrimination methods can be used and this architecture has been shown to have low readout crosstalk [19].

2.4 Crosstalk

A key aspect of any scalable architecture is the suppression of crosstalk to enable parallel operation. Above, we already mentioned reduction of readout crosstalk using individual Purcell filters and cancellation of ZZ -interactions between idled neighboring qubits using the tunable couplers. In addition, our tunable coupler design, see Fig. 2, features coupling extenders which feature compact field distribution to reduce qubit drive crosstalk and provide space in the qubit lattice for aforementioned readout structures [14]. Also, the specific topology of the tunable coupler allows them to be operated at a large detuning from the readout structures reducing readout crosstalk further and improving gate fidelity.

2.5 Control routing and fabrication

The drawback of the nearest-neighbor connectivity implemented with tunable couplers is 2.5x increase in the number of tunable transmons from 20 computational qubits to 50 tunable transmons in total. Moreover, each tunable coupler requires a control line for the fast magnetic flux pulses. Adding to the coupler control lines the aforementioned qubit drive, flux and readout lines, IQM Garnet is controlled by 76 control lines, or 3.8 lines per qubit. The requirements for the fabrication technology and system-level complexity is essentially set by the number of physical qubits. The fabrication complexity is addressed by 3D integrated stack featuring routing and qubit chips connected through a superconducting flip-chip technology. IQM QPUs are fabricated in IQM's dedicated fabrication facility in Espoo, Finland capable of high-yield production of superconducting quantum devices [20, 21].



3. System

3.1 General hardware description

The quantum computer is designed with practical considerations, prioritizing features such as noise reduction, minimal vibrations, ease of installation, and efficient deliverability. Internal components like water distribution, power supply, networking, and compressed air facilities are enclosed within the system. The system features a Bluefors XLD dilution refrigerator as the cryogenic host, supporting the QPU with a standard cascade of attenuation and filtering components.

Magnetic shields are in place to protect critical components including the QPU and Traveling Wave Parametric Amplifier (TWPA) [22]. Microwave isolators are further shielded and strategically located above the mixing chamber flange to minimize interference, Fig. 4 shows the cryogenic components. Other aspects of the cryogenic design implement the known best standards to use high purity copper, non-magnetic gold plating, intermediate thermalization in the filter stack, and tidy cable clamping.

The entire system is arranged to fit with a standard row of equipment racking used typically by High Performance Computing (HPC) centers, seen in Fig. 5, and includes itself a standard 19" electronics rack. The measurement rack (Fig. 5a) hosts the IQM Quantum Control System and includes a double conversion Uninterruptible Power Supply (UPS) for a reliable and electrically cleaner power supply. An additional auxiliary electronics rack handles system monitoring, fridge control, and provides a local access terminal for manual control.

To minimize physical presence requirement, the system incorporates remote power management, monitoring, and control of system fans. Secondary access administration channels are integrated for remote management of the servers without physical intervention. Network security is ensured through fully managed network switches and a firewall.

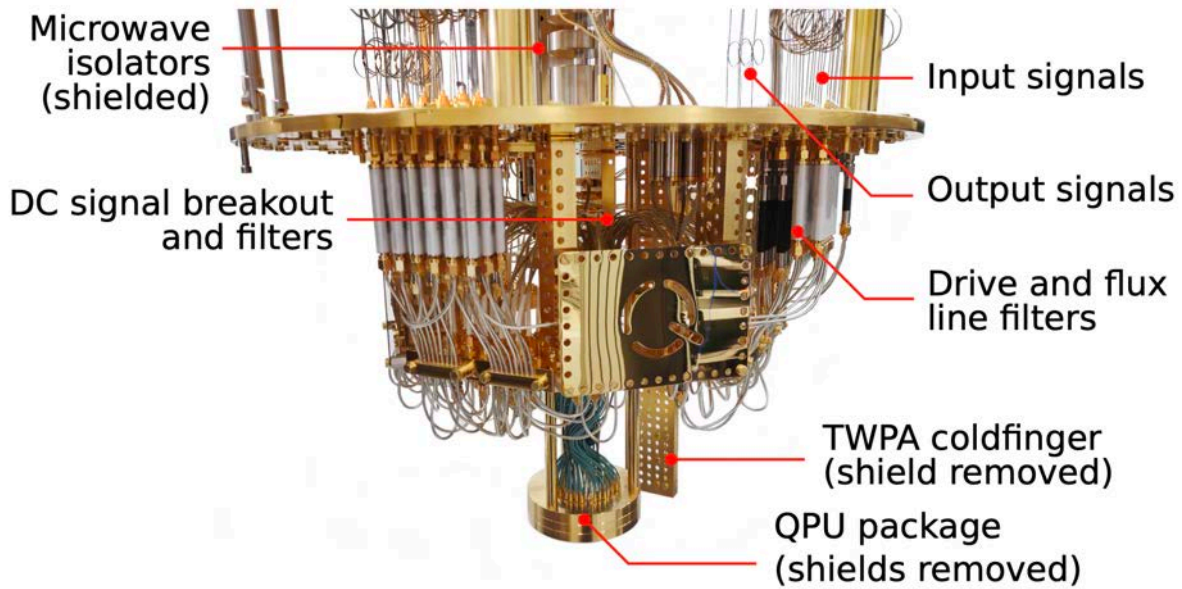


Fig. 4: Cryogenic components inside the host dilution refrigerator

3.2 QPU control electronics

The microwave pulses for single-qubit rotations, the base-band flux pulses for two-qubit gates, and the microwave probe pulses for readout of the qubits are generated by the in-house designed IQM Quantum Control System (QCS). The system is built around PXIe infrastructure which provides modularity, scalability, and high-speed connections between the modules, see Fig. 6. The microwave pulses for single-qubit rotations are generated by direct digital synthesis, allowing a wide bandwidth and avoiding the complexities related to mixer-based solutions. The DC-coupled modules for generating the flux signals include two separate digital-to-analog converters specifically chosen to provide both stable DC bias signals and pulses with a sufficient bandwidth for a rapid and well-controlled ramp-up and ramp-down. The readout module is designed to generate frequency multiplexed probe pulses and process the response for simultaneous readout of up to ten qubits. The real-time control of the executed pulse sequence is implemented with a field-programmable gate array (FPGA) associated with each module of the system. Additionally, IQM QCS implements an advanced clock distribution and timing solution, enabling accurate phase and event synchronization across all the channels in the system. Overall, the IQM QCS has been designed to provide high-quality control signals specific to the chosen QPU architecture in a cost-effective manner.

In addition to the IQM QCS, the control electronics setup includes an in-house designed IQM Microwave Generator (MWG) for the continuous wave pump signals for the TWPA, and a Qblox D5a module for DC flux currents for qubits which are operated at fixed frequency throughout computation.

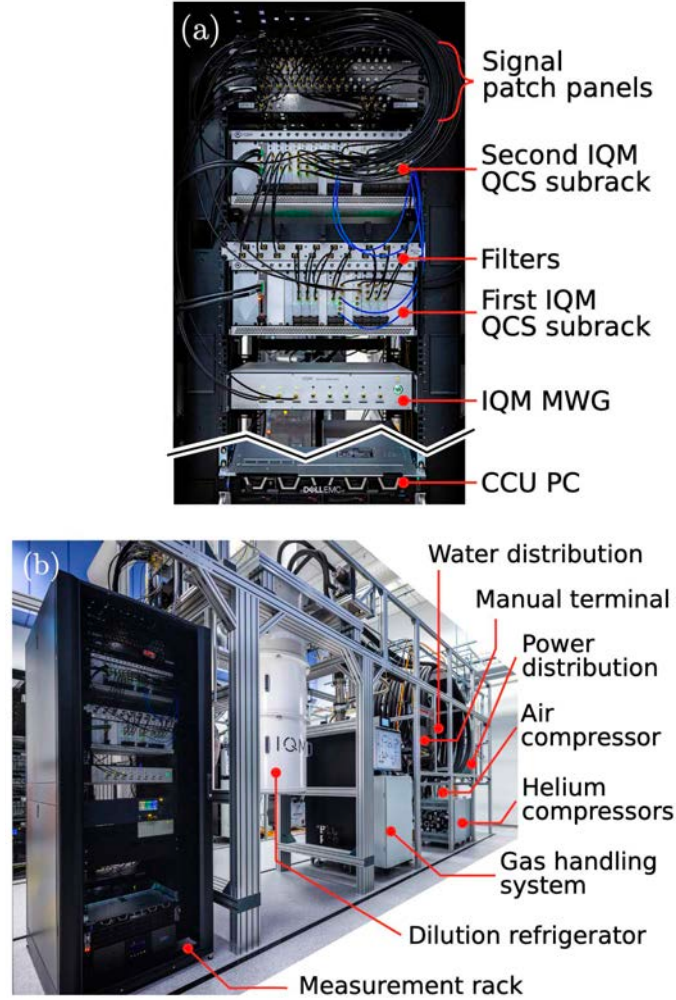


Fig. 5: Photograph of the system without covers

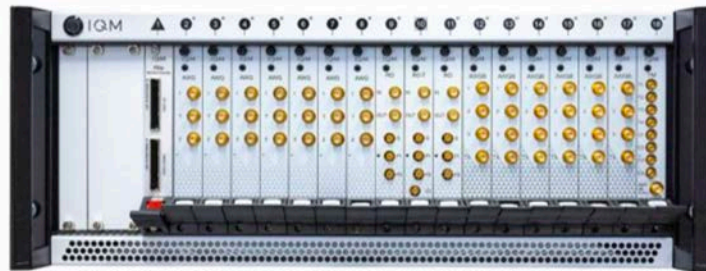


Fig. 6: Photograph of a fully populated subrack of IQM QCS electronics.

3.3 QC control software

The control software stack is divided into multiple functional layers presented in Fig. 7. It can be configured based on the required level of access. The software modules and their roles are described in more detail in the following subsections.

IQM QCCSW is delivered as a Debian package, enveloping application images, helper scripts and configuration files, and a `systemd` service. By packaging the images and relevant files with Debian package manager, installation and up-keeping become easy, offering easy upgrades and configuration management. Containerisation brings in capabilities like resource allocation and consolidating dependency. Coupled with robust service management capabilities like boot process optimisation, service isolation, and automatic restarts, the software stack ensures reliability, security and resilience.

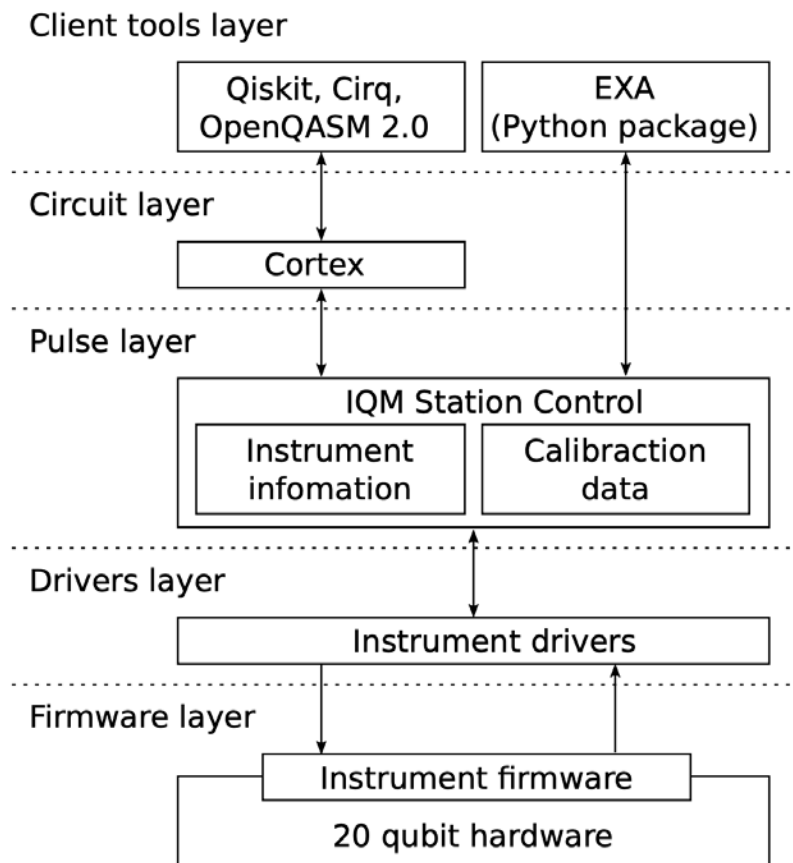


Fig. 7: The software layers and modules of our quantum computer control software stack.

3.3.1 Cortex

Cortex is the highest level of abstraction in the QCCSW stack, that allows running quantum algorithms on IQM quantum computer. It is built to enable quantum computation for the end user, rather than experimenting with the behaviour of underlying elements of the quantum computer itself. Cortex users can define and execute the quantum algorithms, expressed as quantum circuits using application-level frameworks like Qiskit, Cirq and OpenQASM 2.0. The input to the IQM quantum computer is a computational job with one or more quantum circuits, the shot count, and any possible parameters defined by the client tool. The job is then queued for execution and the results of the circuit measurements can be retrieved when the job is completed.

3.3.2 EXA

EXA is a comprehensive, customisable Python library for characterisation, calibration and control of IQM quantum computers. The central feature of EXA is the Experiment class, which gives the user pulse-level access to the quantum computer and combines functionalities such as waveform control, execution flow, data manipulation, analysis and presentation. It allows the user to change any parameter of the system and defines sensible defaults which can be easily expanded or overridden. The EXA Experiment Library enables calibration of quantum computers, greatly simplifying measurement and control processes. It helps automate standard procedures, reduce repetitive tasks, and develop entirely new experiments with only minimal amounts of new code.

3.3.3 IQM Station Control

IQM Station Control is responsible for low-level functionalities like instrument parameters, hardware drivers and persistence of raw data and metadata. It abstracts out the hardware details from higher-level components such as EXA and Cortex.

Station Control service's non-RESTful JSON (JavaScript Object Notation) HTTP (HyperText Transfer Protocol) interface handles communication with Cortex and EXA. Parameter sweeps and pulse schedule executions can be carried out using the endpoints provided by this interface. The user does not need to directly interact with the service normally. Device-specific drivers help Station Control to encapsulate the details of each instrument, including its low-level communication protocol.

4. Performance benchmarking

As the name suggests, quantum computers are built for practical computing purposes. As practical quantum advantage is yet to be achieved, and it is unclear where it will first be discovered, the field measures quantum computer performance through a variety of benchmarking metrics [23, 24]. At IQM we have taken a four-fold approach to benchmarking:

- Foundational operation level
- System total operation level
- Fundamental physics based
- Application benchmarks

With these benchmarks, we can understand the operation of the device on the component level (foundational operation level) while making sure we can track the performance of the device as a whole (system total operation level). The fundamental physics and application benchmarks are needed to ensure the metrics are relevant to end applications allowing for improving our quantum computers over time. We provide benchmarking results for IQM Garnet on all four levels below.



4.1 Foundational operation level benchmarks

High-performance quantum computing builds on low error rates of the gate-level operations. As discussed in Section 2.4, a key enabler for low QPU-level error rates in parallel operation is low crosstalk. We define flux and drive crosstalk by

$$C_{\text{Flux}} = 20 \log_{10} \frac{\partial \Phi_i / \partial I_j}{\partial \Phi_j / \partial I_j}, \quad (1)$$

$$C_{\text{Drive}} = 20 \log_{10} \frac{\partial \Omega_i / \partial V_j}{\partial \Omega_j / \partial V_j} \quad (2)$$

correspondingly where Φ_i is the flux bias, I_i is the flux line current, Ω_i is the Rabi rate for resonant drive, and V_i drive pulse amplitude applied on the drive line of qubit i . On IQM Garnet devices, we have measured median crosstalk level of $C_{\text{Flux}} = -70$ dB and $C_{\text{Drive}} = -48$ dB, see Fig. 8a. The observed median values for flux crosstalk are more than 3 dB, and for drive crosstalk are more than 6 dB lower compared to other recently published QPU with a similar architecture [25] and more favorably compared to most older designs and to other quantum computing platforms [25]. In a smaller demonstrator system with wirebond-free package 9 dB lower drive crosstalk is demonstrated, indicating possible future gain by upgrading to the next generation packaging solution [26].

We obtain average single qubit gate fidelity by averaging over Clifford group in a randomized benchmarking [27] experiment and normalize the error per native gate. For characterizing CZ gate error we use interleaved randomized benchmarking [28]. We characterize gates in parallel in distance-two groups, where qubits are separated by two idling couplers and one idling qubit and obtain median single qubit gate error of 9×10^{-4} , CZ error of 5×10^{-3} and two qubit Clifford error of 2×10^{-2} , see Fig. 8b. Benchmarking the two qubit gates in distance-one groups results in a reduction of mean fidelity by 0.6 for CZ gates and 3 percentage points for two-qubit Clifford gates.

We define readout error as the probability of not discriminating qubit to be in the prepared state without correcting the for state preparation errors [29]. We observe median readout error for simultaneous QPU readout to be 3×10^{-2} .

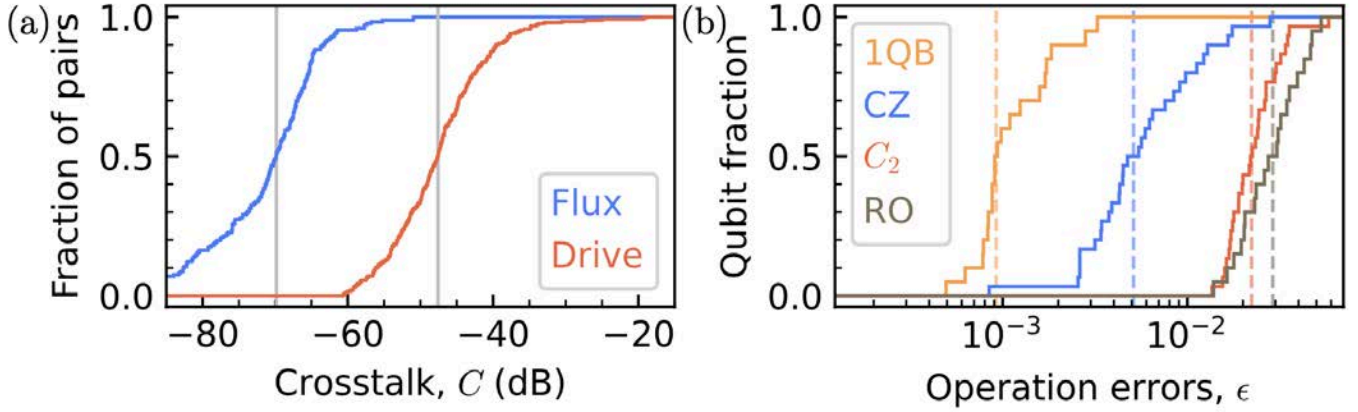


Fig. 8: (a) Cumulative distribution of drive and flux control crosstalk between pairs of operation target and spectating qubits. (b) Cumulative distribution of average single-qubit gate (1QB), native two qubit gate (CZ), composite two-qubit Clifford operation (C_2), and readout (RO) operation errors for IQM Garnet QPU components, with vertical lines indicating median values.

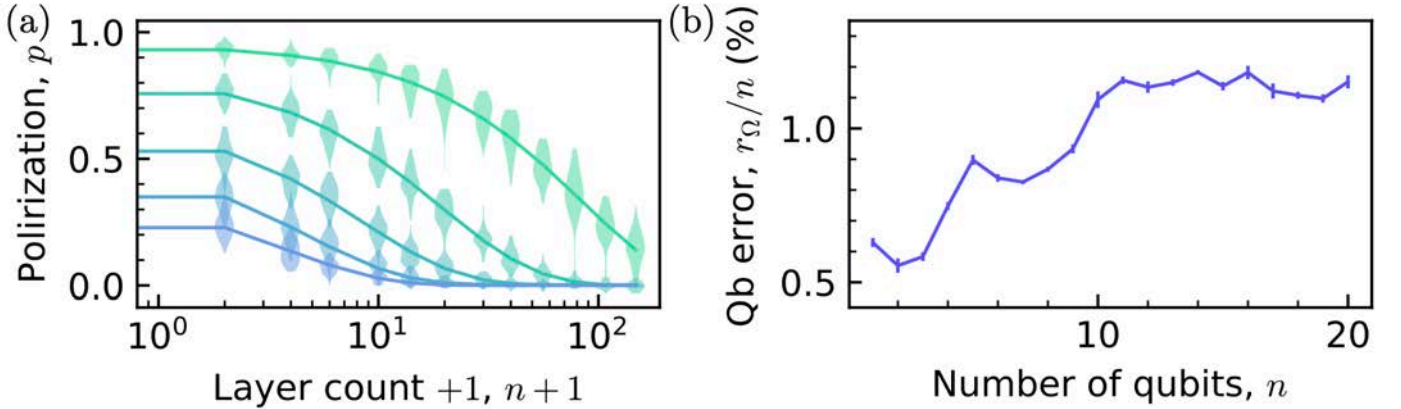


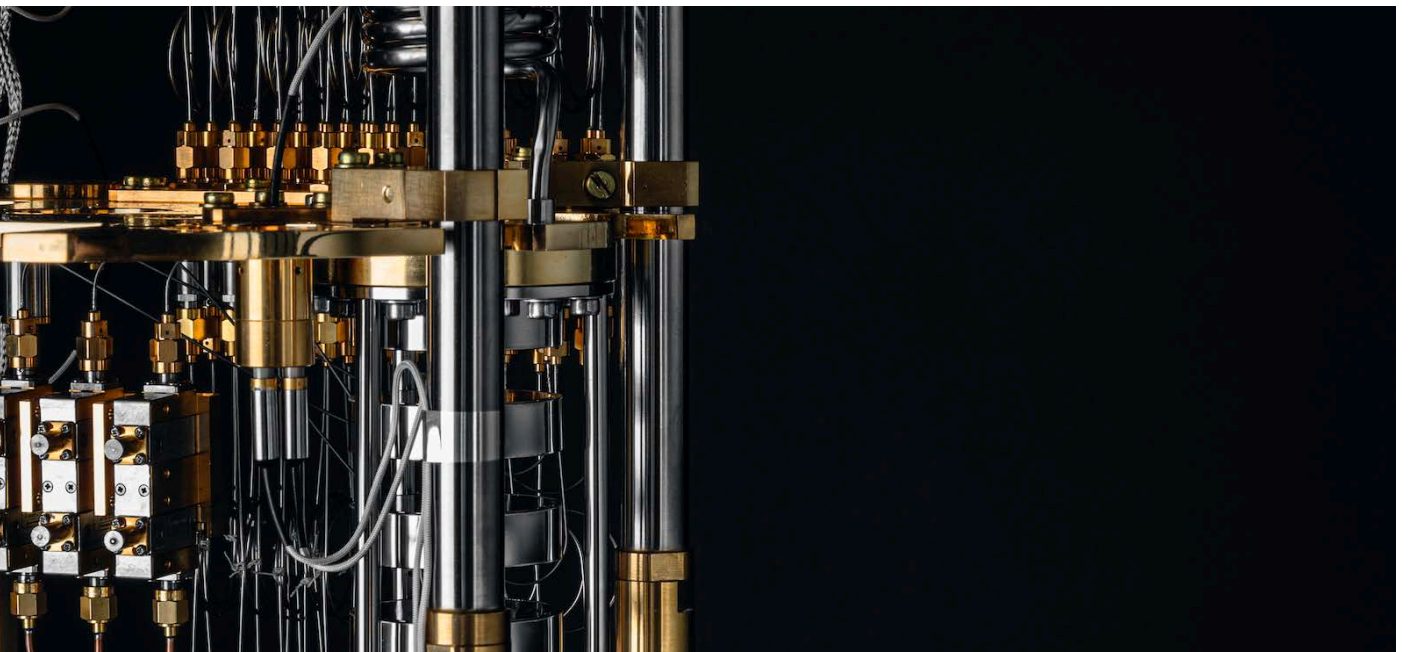
Fig. 9: (a) State polarization distributions (violins) after d layers of random gates in MRB experiments together with exponential fits for $n = 2, 6, 10, 16, 20$ qubits (blue to green). (b) Error per gate layer per qubit extracted from the fitted decay constants like in the first subfigure.

4.2 System level benchmarks

The system level benchmarks measure the joint operation of several qubits or the whole processor in one go. They measure how the entire device operates, and provide information on the strengths and weaknesses of the specific calibration and parts of the quantum computer. Typically they do not however directly predict performance on practical tasks quantum computers will be used for. All the methods presented in this section are based on random quantum circuits.

4.2.1 Mirror randomized benchmarking

Randomized mirror circuits combine a mirrored structure with randomized compiling to enable scalable and robust randomized benchmarking (RB) of both Clifford and universal multi-qubit gate sets without the classical computation overhead [30, 31]. We carried out MRB protocol by interleaving layers with gatesets $\mathbb{G}_1 = \mathbb{C}_1$, where \mathbb{C}_1 is the one-qubit Clifford group, and $\mathbb{G}_2 = \{CZ\}$ with a uniform CZ-gate density $\xi = 1/2$ and a probability distribution Ω over an n -qubit layer set. We observe expected decay of state polarization with increasing gate layer count, see Fig. 9a. By fitting exponential decay models and comparing decay constants normalized by qubit count we see evidence of minor increase of error per layer per qubit $r_{\Omega,perQ} = 1 - (1 - r_{\Omega})^{1/n} \approx r_{\Omega}/n$ with increased number of qubits, see Fig. 9b. This degradation is like due to the fact that larger benchmark has to include qubit pairs with sub-median performance as well as some crosstalk effects.



4.2.2 Quantum Volume and volumetric benchmarking

To get a broader overview of the ability of the device to run circuits of various types and sizes, and to measure the joint scaling of fidelity and qubit count, we have performed volumetric benchmarking [32, 33] and measured the Quantum Volume of IQM Garnet [34]. Quantum volume (QV) is measured by performing “square shaped” circuits where the depth of the circuit equals the width (number of qubits), such that each layer contains a maximal number of random $SU(4)$ two-qubit unitaries between random pairs of qubits. Volumetric benchmarking framework was used here to measure the output fidelities of random circuits with a varying number of qubits and depths.

The result of a successful $QV = 2^5$ benchmark experiment is shown in Fig. 10a, with average heavy-output probabilities converging to a value greater than $2/3$ within two standard deviations. In Fig. 10b, we show the volumetric benchmark background of average fidelities of sets of 40 random circuits with layers of gates consisting of uniformly sampled single-qubit Clifford gates and CZ gates, with a CZ gate density of $\xi = 1/4$. The data surrounded by black borders are measured data, while the other boxes are extrapolated fidelities based on the measured data.

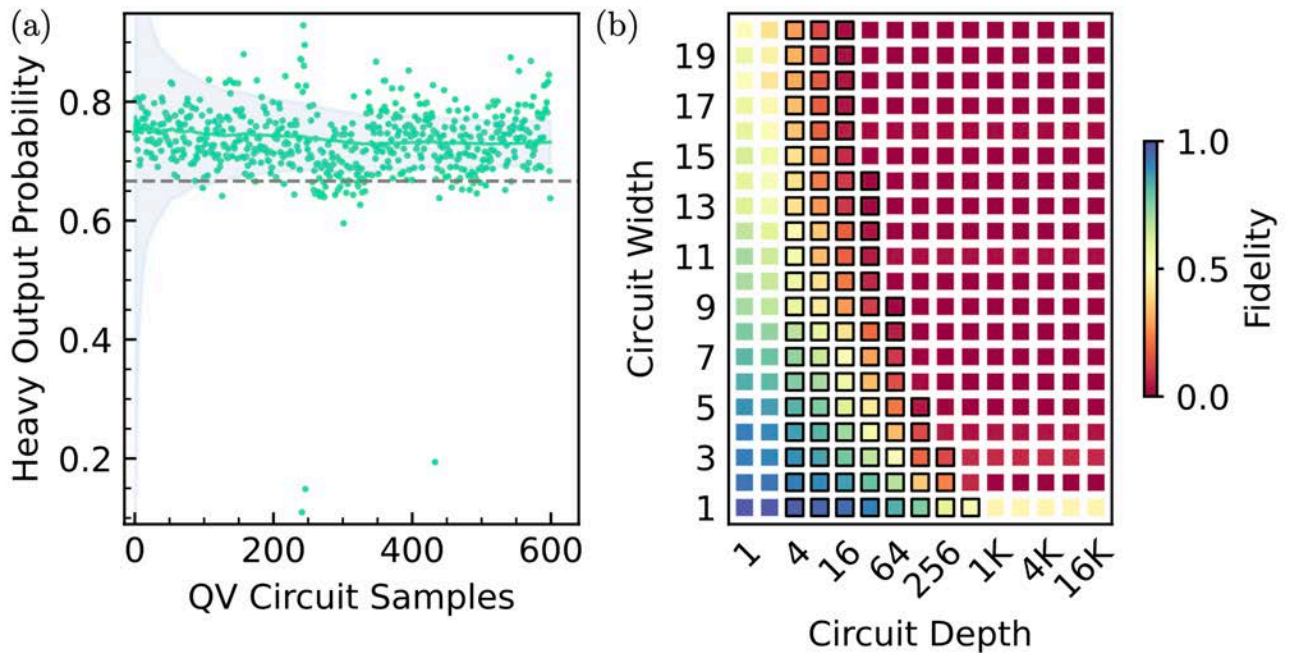


Fig. 10: (a) Observed heavy-output probability for circuit samples for test size $QV = 2^5 = 32$. (b) Volumetric benchmarking fidelity using random circuits. The data surrounded by black borders are measured data, while the other boxes are extrapolated fidelities based on the measured data.

4.2.3 Circuit Layer Operations Per Second (CLOPS)

Circuit Layer Operations Per Second (CLOPS) measure the execution speed of the random circuits used to measure quantum volume of the device [35]. CLOPS also include a measure of the feedback time between the control computer and the quantum computer that is typical for variational algorithms by including parameter updates to the random circuits. The fast execution of circuits that is measured by CLOPS is crucially important for the usage of quantum computers, as a faster through-put rate enables applying more and heavier error-mitigation techniques. These are essential for high-quality computing and research results using current and near-term quantum computers. Recently there has been a new hardware-aware definition of CLOPS that only applies gates between qubits that are connected on the hardware [36], and in this context the quantity reported here is called the virtual CLOPS.

CLOPS is computed as

$$\text{CLOPS} = \frac{M \times K \times S \times D}{T} \quad (3)$$

where:

- M = number of random circuit templates = 100
- K = number of variational parameter updates = 10
- S = number of shots = 100
- D = number of QV layers = $\log_2 QV$
- T = time taken

We have measured a CLOPS of 2600 on IQM Garnet. The result reflects partially the quantum volume, as indicated by D in the formula, but mainly the speed of the execution. Factors that determine it are control electronics, software and signal latency communicating between the quantum and classical computers. Section 3 describes the control electronics and software that enable the fast execution of quantum circuits on IQM Garnet.

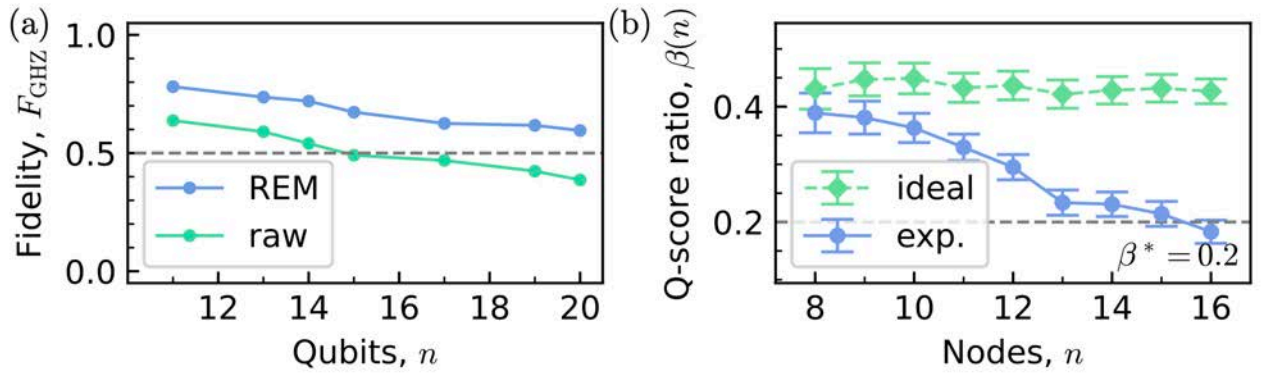


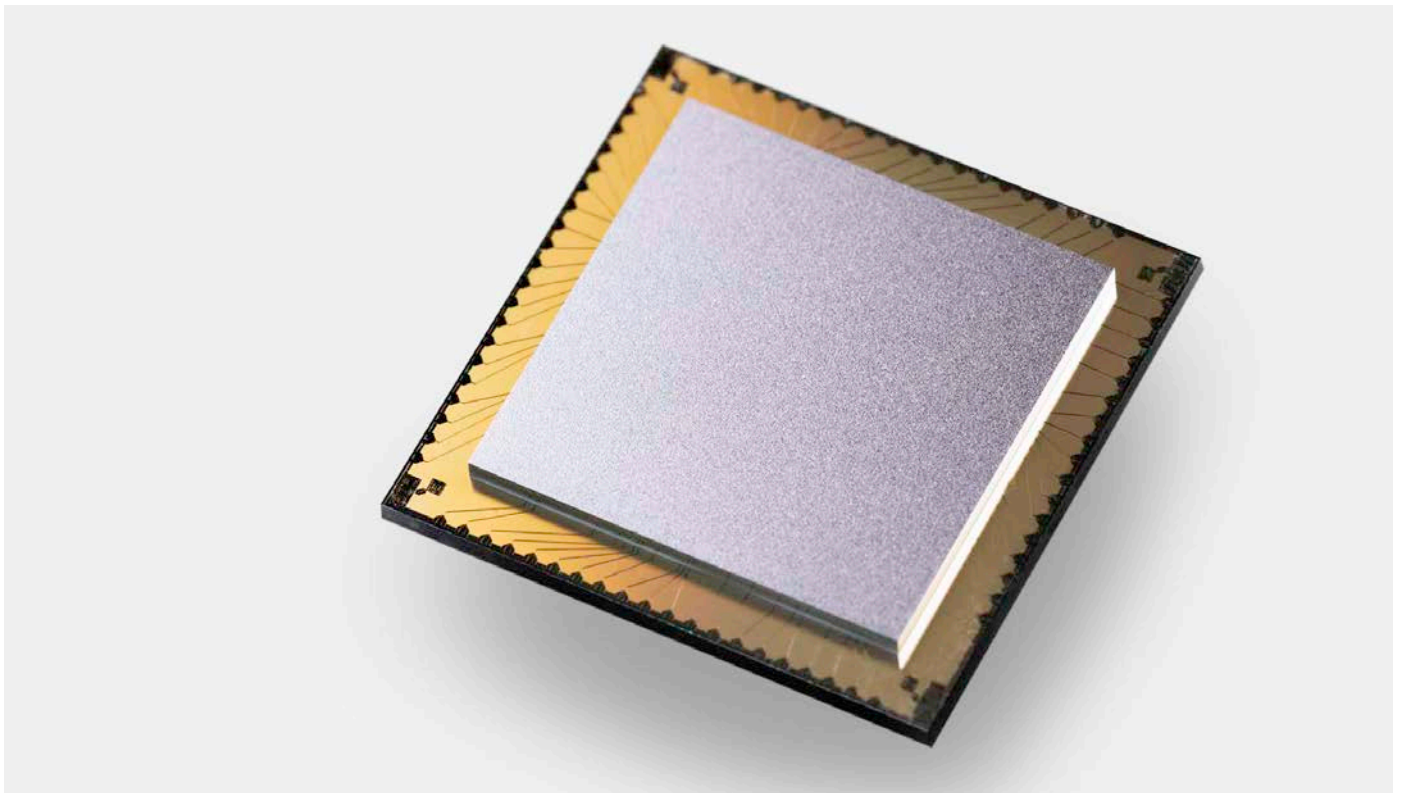
Fig. 11: (a) GHZ-state fidelity F_{GHZ} across the number of entangled qubits with readout read-out-error mitigation (REM, blue), without REM (green) and fidelity threshold $F_{\text{GHZ}} > 0.5$ for genuine multi-qubit entanglement (gray dashed). (b) Measured Q-score ratio $\beta(n)$ (blue) with one virtual node as a function of node count n on $n - 1$ qubits compared to threshold of 0.2 (gray dashed), indicating Q-score of 15, and expected outcome for ideal quantum computer (green).



4.3 Fundamental physics benchmarks

4.3.1 GHZ state creation

With the system-level benchmarks performed, it is time to start looking at the quantumness properties of IQM Garnet. A foundational requirement for useful quantum computing that has potential to go beyond classical capabilities is to entangle all qubits of the processor into a genuinely multi-qubit entangled (GME) state. The typical way to demonstrate GME is to prepare a GHZ state [37]. A GHZ state fidelity $F_{\text{GHZ}} > 0.5$ is a witness for GME [38] and can be measured using the method of multiple quantum coherences [39, 40]. Fig. 11a reports GHZ state fidelities measured on IQM Garnet with and without readout-error mitigation (REM). The results without REM show that GME can be certified up to 14 qubits according to the strictest standards of fundamental physics. A quantum computer however typically does not need to prepare entangled states for the sake of it, but to use them as a resource in computation. In a computational scenario measurement is deferred to the end of the algorithm, so we are interested in the GHZ state fidelity pre-measurement. This is what we access by applying readout-error mitigation using the `mthree` package [41]. The readout-error mitigated results show that we have prepared a 20-qubit genuinely multi-qubit entangled GHZ state with fidelity $F_{\text{GHZ}20} = 0.62$, verifying the non-classical nature of the entire IQM Garnet system.



4.4 Application benchmarks

4.4.1 Q-Score: solving Max-Cut with QAOA

Finally, we want to see how IQM Garnet performs in providing solutions to computational problems. To showcase performance in this arena we choose to show results from computing the solutions to the maximum-cut combinatorial optimisation problem using the Q-score benchmark [42]. IQM Garnet has been shown to pass the Q-score benchmark up to problem-size 15, as shown in 11b. Q-score is defined on random Erdős-Rényi graphs with edge probability $p = 0.5$ for each connection. The Q-score test is passed if $\beta(n) \geq 0.2$, where $\beta(n)$ measures the fraction of the optimality gap between a random guess and the optimal solution to the problem graph on n qubits the quantum computers has captured. Due to the accumulation of noise, the Q-score test becomes increasingly hard with growing problem size.

We measure our Q-score performance with a depth $p = 1$ QAOA circuits [43, 44] execution, where we find the optimal circuit parameters by evaluating analytical formulas [45]. This method is enabled by the fact that local expectation values evaluated on the $p = 1$ QAOA state only depend on a reduced number of qubits. Such angle optimization procedure is classically efficient, and ensures the optimal use of quantum resources. We also increase the score by 1 through use of the virtual node technique [46, 47] that introduces an overhead of only a few single-qubit Z -gates.

To further improve performance, we execute the quantum circuits on an optimal layout using the method introduced in [48]. Here, in a pre-processing step, we optimise a noise score for each of the layouts based on the gates count of the circuit and the single- and two-qubit-gate error rates of the hardware. We use readout-error mitigation implemented in the `mthree` package [41], which allows us to leave out readout errors from the noise score.

5. Summary and outlook

We have presented a comprehensive set of benchmarking data for a quantum computing system built on the IQM Garnet, a 20-qubit quantum processor with a square lattice topology and a dedicated tunable coupler for high-quality two-qubit gate operations. These technical solutions are representative of IQM's qubit crystal QPU family.

The five-qubit QPU is the core of the IQM Spark system, which serves as a cost-effective, on-premises product designed for research and educational purposes [47]. IQM qubit crystals, detailed in Section 2, power the IQM Radiance integrated quantum computing systems, showcasing the state of the art commercially available, on-premises superconducting quantum computers. These systems are instrumental in exploring the boundaries of quantum computation, including applications approaching quantum utility and advantage, and in accelerating classical high-performance computing.

Notably, the IQM Garnet is also part of IQM's cloud offering, IQM Resonance [49]. The performance levels we have described reflect the current achievements with today's integrated superconducting quantum computing systems. In the near future, we expect improvements through enhanced control optimization, better coherence, and incremental system and QPU enhancements. We plan to integrate recent advancements in control optimization [13] and aim to improve the relaxation time (T_1) from approximately 40 μs to over 100 μs , a target we have already achieved in test devices [50].



9. References

- 1 Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., Biswas, R., Boixo, S., Brandao, F.G.S.L., Buell, D.A., Burkett, B., Chen, Y., Chen, Z., Chiaro, B., Collins, R., Courtney, W., Dunsworth, A., Farhi, E., Foxen, B., Fowler, A., Gidney, C., Giustina, M., Graff, R., Guerin, K., Habegger, S., Harrigan, M.P., Hartmann, M.J., Ho, A., Hoffmann, M., Huang, T., Humble, T.S., Isakov, S.V., Jeffrey, E., Jiang, Z., Kafri, D., Kechedzhi, K., Kelly, J., Klimov, P.V., Knysh, S., Korotkov, A., Kostritsa, F., Landhuis, D., Lindmark, M., Lucero, E., Lyakh, D., Mandrà, S., McClean, J.R., McEwen, M., Megrant, A., Mi, X., Michielsen, K., Mohseni, M., Mutus, J., Naaman, O., Neeley, M., Neill, C., Niu, M.Y., Ostby, E., Petukhov, A., Platt, J.C., Quintana, C., Riefel, E.G., Roushan, P., Rubin, N.C., Sank, D., Satzinger, K.J., Smelyanskiy, V., Sung, K.J., Trevithick, M.D., Vainsencher, A., Villalonga, B., White, T., Yao, Z.J., Yeh, P., Zalcman, A., Neven, H., Martinis, J.M.: Quantum supremacy using a programmable superconducting processor. *Nature* 574(7779), 505–510 (2019) <https://doi.org/10.1038/s41586-019-1666-5>
- 2 Wu, Y., Bao, W.-S., Cao, S., Chen, F., Chen, M.-C., Chen, X., Chung, T.-H., Deng, H., Du, Y., Fan, D., Gong, M., Guo, C., Guo, C., Guo, S., Han, L., Hong, L., Huang, H.-L., Huo, Y.-H., Li, L., Li, N., Li, S., Li, Y., Liang, F., Lin, C., Lin, J., Qian, H., Qiao, D., Rong, H., Su, H., Sun, L., Wang, L., Wang, S., Wu, D., Xu, Y., Yan, K., Yang, W., Yang, Y., Ye, Y., Yin, J., Ying, C., Yu, J., Zha, C., Zhang, C., Zhang, H., Zhang, K., Zhang, Y., Zhao, H., Zhao, Y., Zhou, L., Zhu, Q., Lu, C.-Y., Peng, C.-Z., Zhu, X., Pan, J.-W.: Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.* 127, 180501 (2021) <https://doi.org/10.1103/PhysRevLett.127.180501>
- 3 Montanaro, A.: Quantum algorithms: an overview. *npj Quantum Information* 2(1), 15023 (2016) <https://doi.org/10.1038/npjqi.2015.23>
- 4 Kjaergaard, M., Schwartz, M.E., Braumüller, J., Krantz, P., Wang, J.I.-J., Gustavsson, S., Oliver, W.D.: Superconducting qubits: Current state of play. *Annual Review of Condensed Matter Physics* 11(1), 369–395 (2020) <https://doi.org/10.1146/annurev-conmatphys-031119-050605>
- 5 Kindel, W., Ku, H.-S., Gusenkova, D., Wurz, N., Lamprich, J., Verjauw, J., Giri, R., Renger, M., Pogorzalek, S., Vigneau, F., et al.: Scaling qubit-resonator architectures for co-design quantum computing. *Bulletin of the American Physical Society* (2024)
- 6 Algaba, M.G., Ponce-Martinez, M., Munuera-Javaloy, C., Pina-Canelles, V., Thapa, M.J., Taketani, B.G., Leib, M., Vega, I., Casanova, J., Heimonen, H.: Co-design quantum simulation of nanoscale nmr. *Phys. Rev. Res.* 4, 043089 (2022) <https://doi.org/10.1103/PhysRevResearch.4.043089>
- 7 Weidenfeller, J., Valor, L.C., Gacon, J., Tornow, C., Bello, L., Woerner, S., Egger, D.J.: Scaling of the quantum approximate optimization algorithm on superconducting qubit based hardware. *Quantum* 6, 870 (2022) <https://doi.org/10.22331/q-2022-12-07-870>
- 8 Steiger, D.S., Häner, T., Troyer, M.: Advantages of a modular high-level quantum programming framework. *Microprocessors and Microsystems* 66, 81–89 (2019) <https://doi.org/10.1016/j.micpro.2019.02.003>
- 9 Kotil, A., Simkovic, F., Leib, M.: Improved Qubit Routing for QAOA Circuits. *arXiv* (2023). <https://doi.org/10.48550/arXiv.2312.15982>
- 10 Datta, K., Kole, A., Sengupta, I., Drechsler, R.: Mapping Quantum Circuits to 2-Dimensional Quantum Architectures. *Gesellschaft für Informatik, Bonn* (2022). https://doi.org/10.18420/inf2022_94
- 11 Tindall, J., Fishman, M., Stoudenmire, E.M., Sels, D.: Efficient tensor network simulation of IBM's eagle kicked ising experiment. *PRX Quantum* 5, 010308 (2024) <https://doi.org/10.1103/PRXQuantum.5.010308>
- 12 McKay, D.C., Wood, C.J., Sheldon, S., Chow, J.M., Gambetta, J.M.: Efficient z gates for quantum computing. *Phys. Rev. A* 96, 022330 (2017) <https://doi.org/10.1103/PhysRevA.96.022330>
- 13 Hyppä, E., Vepsäläinen, A., Papič, M., Chan, C.F., Inel, S., Landra, A., Liu, W., Luus, J., Marxer, F., Ockeloen-Korppi, C., Orbell, S., Tarasinski, B., Heinsoo, J.: Reducing leakage of single-qubit gates for superconducting quantum processors using analytical control pulse envelopes (2024). <https://doi.org/10.48550/arXiv.2402.17757>
- 14 Marxer, F., Vepsäläinen, A., Jolin, S.W., Tuorila, J., Landra, A., Ockeloen-Korppi, C., Liu, W., Ahonen, O., Auer, A., Belzane, L., Bergholm, V., Chan, C.F., Chan, K.W., Hiltunen, T., Hotari, J., Hyppä, E., Ikonen, J., Janzso, D., Koistinen, M., Kotilahti, J., Li, T., Luus, J., Papič, M., Partanen, M., Rabinä, J., Rosti, J., Savitskyi, M., Seppälä, M., Sevriuk, V., Takala, E., Tarasinski, B., Thapa, M.J., Tosto, F., Vorobeva, N., Yu, L., Tan, K.Y., Hassel, J., Möttönen, M., Heinsoo, J.: Long-distance transmon coupler with cz-gate fidelity above 99.8%. *PRX Quantum* 4, 010314 (2023) <https://doi.org/10.1103/PRXQuantum.4.010314>
- 15 Sheldon, S., Magesan, E., Chow, J.M., Gambetta, J.M.: Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Phys. Rev. A* 93, 060302 (2016) <https://doi.org/10.1103/PhysRevA.93.060302>

- 16 Sete, E.A., Didier, N., Chen, A.Q., Kulshreshtha, S., Manenti, R., Poletto, S.: Parametric-resonance entangling gates with a tunable coupler. *Phys. Rev. Appl.* 16, 024050 (2021) <https://doi.org/10.1103/PhysRevApplied.16.024050>
- 17 Sung, Y., Ding, L., Braumüller, J., Vepsäläinen, A., Kannan, B., Kjaergaard, M., Greene, A., Samach, G.O., McNally, C., Kim, D., Melville, A., Niedzielski, B.M., Schwartz, M.E., Yoder, J.L., Orlando, T.P., Gustavsson, S., Oliver, W.D.: Realization of high-fidelity CZ and ZZ-free iSWAP gates with a tunable coupler. *Phys. Rev. X* 11, 021058 (2021) <https://doi.org/10.1103/PhysRevX.11.021058>
- 18 Karamlou, A.H., Braumüller, J., Yanay, Y., Di Paolo, A., Harrington, P.M., Kannan, B., Kim, D., Kjaergaard, M., Melville, A., Muschinske, S., *et al.*: Quantum transport and localization in 1d and 2d tight-binding lattices. *npj Quantum Information* 8(1), 35 (2022) <https://doi.org/10.1038/s41534-022-00528-0>
- 19 Heinsoo, J., Andersen, C.K., Remm, A., Krinner, S., Walter, T., Salathé, Y., Gasparinetti, S., Besse, J.-C., Potočník, A., Wallraff, A., Eichler, C.: Rapid high-fidelity multiplexed readout of superconducting qubits. *Phys. Rev. Appl.* 10, 034040 (2018) <https://doi.org/10.1103/PhysRevApplied.10.034040>
- 20 Liu, W., Li, T., Venkatesh, M., Chan, C.F., Marxer, F., Jenei, M., Yang, L., Sevriuk, V., Ahmad, H., Kokkonen, R., *et al.*: Development of the quantum processing units (QPUs) at IQM Finland Oy. *Bulletin of the American Physical Society* 66 (2021)
- 21 Liu, W., Luus, J., Kim, S.-G., Plyushch, A., Qiu, W., Li, T., Ockeloen-Korppi, C., Landra, A., Abdurakhimov, L., Zhu, L., *et al.*: Realizing scalable quantum processing units (QPUs): Achievements and prospects from IQM Finland Oy. *Bulletin of the American Physical Society* (2024)
- 22 Perelshtein, M.R., Petrovnikov, K.V., Vesterinen, V., Hamedani Raja, S., Lilja, I., Will, M., Savin, A., Simbierowicz, S., Jabdaraghi, R.N., Lehtinen, J.S., Grönberg, L., Hassel, J., Prunnila, M.P., Govenius, J., Paraoanu, G.S., Hakonen, P.J.: Broadband continuous-variable entanglement generation using a Kerr-free Josephson metamaterial. *Phys. Rev. Appl.* 18, 024063 (2022) <https://doi.org/10.1103/PhysRevApplied.18.024063>
- 23 Kliesch, M., Roth, I.: Theory of quantum system certification. *PRX Quantum* 2(1), 010201 (2021) <https://doi.org/10.1103/PRXQuantum.2.010201>
- 24 Eisert, J., Hangleiter, D., Walk, N., Roth, I., Markham, D., Parekh, R., Chabaud, U., Kashefi, E.: Quantum certification and benchmarking. *Nature Reviews Physics* 2(7), 382–390 (2020) <https://doi.org/10.1038/s42254-020-0186-4>
- 25 Kosen, S., Li, H.-X., Rommel, M., Rehammar, R., Caputo, M., Grönberg, L., Fernández-Pendás, J., Kockum, A.F., Biznárová, J., Chen, L., Križan, C., Nylander, A., Osman, A., Roudsari, A.F., Shiri, D., Tancredi, G., Govenius, J., Bylander, J.: Signal crosstalk in a flip-chip quantum processor (2024). <https://doi.org/10.48550/arXiv.2403.00285>
- 26 Spring, P.A., Cao, S., Tsunoda, T., Campanaro, G., Fasciati, S., Wills, J., Bakr, M., Chidambaram, V., Shteynas, B., Carpenter, L., Gow, P., Gates, J., Vlastakis, B., Leek, P.J.: High coherence and low cross-talk in a tileable 3D integrated superconducting circuit architecture. *Science Advances* 8(16), 6698 (2022) <https://doi.org/10.1126/sciadv.abl6698>
- 27 Emerson, J., Alicki, R., Życzkowski, K.: Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics* 7(10), 347–352 (2005) <https://doi.org/10.1088/1464-4266/7/10/021>
- 28 Magesan, E., Gambetta, J.M., Johnson, B.R., Ryan, C.A., Chow, J.M., Merkel, S.T., Silva, M.P., Keefe, G.A., Rothwell, M.B., Ohki, T.A., Ketchen, M.B., Steffen, M.: Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Phys. Rev. Lett.* 109, 080505 (2012) <https://doi.org/10.1103/PhysRevLett.109.080505>
- 29 Touzard, S., Kou, A., Frattini, N.E., Sivak, V.V., Puri, S., Grimm, A., Frunzio, L., Shankar, S., Devoret, M.H.: Gated conditional displacement readout of superconducting qubits. *Phys. Rev. Lett.* 122, 080502 (2019) <https://doi.org/10.1103/PhysRevLett.122.080502>
- 30 Proctor, T., Seritan, S., Rudinger, K., Nielsen, E., Blume-Kohout, R., Young, K.: Scalable randomized benchmarking of quantum computers using mirror circuits. *Phys. Rev. Lett.* 129, 150502 (2022) <https://doi.org/10.1103/PhysRevLett.129.150502>
- 31 Hines, J., Lu, M., Naik, R.K., Hashim, A., Ville, J.-L., Mitchell, B., Kriekebaum, J.M., Santiago, D.I., Seritan, S., Nielsen, E., Blume-Kohout, R., Young, K., Siddiqi, I., Whaley, B., Proctor, T.: Demonstrating scalable randomized benchmarking of universal gate sets. *Phys. Rev. X* 13, 041030 (2023) <https://doi.org/10.1103/PhysRevX.13.041030>
- 32 Blume-Kohout, R., Young, K.C.: A volumetric framework for quantum computer benchmarks. *Quantum* 4, 362 (2020) <https://doi.org/10.22331/q-2020-11-15-362>
- 33 Proctor, T., Rudinger, K., Young, K., Nielsen, E., Blume-Kohout, R.: Measuring the capabilities of quantum computers. *Nature Physics* 18(1), 75–79 (2022) <https://doi.org/10.1038/s41567-021-01409-7>

- 34 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 (2019) <https://doi.org/10.1103/PhysRevA.100.032328>
- 35 Wack, A., Paik, H., Javadi-Abhari, A., Jurcevic, P., Faro, I., Gambetta, J.M., Johnson, B.R.: Quality, Speed, and Scale: three key attributes to measure the performance of near-term quantum computers (2021). <https://doi.org/10.48550/arXiv.2110.14108>
- 36 Updating how we measure quantum quality and speed — IBM Quantum Computing Blog. <https://www.ibm.com/quantum/blog/quantum-metric-layer-fidelity>. [Online; accessed 18-04-2024]
- 37 Greenberger, D.M., Horne, M.A., Zeilinger, A.: In: Kafatos, M. (ed.) *Going Beyond Bell's Theorem*, pp. 69–72. Springer, Dordrecht (1989). https://doi.org/10.1007/978-94-017-0849-4_10
- 38 Wei, T.-C., Goldbart, P.M.: Geometric measure of entanglement and applications to bipartite and multipartite quantum states. *Phys. Rev. A* 68, 042307 (2003) <https://doi.org/10.1103/PhysRevA.68.042307>
- 39 Wei, K.X., Lauer, I., Srinivasan, S., Sundaresan, N., McClure, D.T., Toyli, D., McKay, D.C., Gambetta, J.M., Sheldon, S.: Verifying multipartite entangled Greenberger-Horne-Zeilinger states via multiple quantum coherences. *Phys. Rev. A* 101, 032343 (2020) <https://doi.org/10.1103/PhysRevA.101.032343>
- 40 Baum, J., Munowitz, M., Garroway, A.N., Pines, A.: Multiple-quantum dynamics in solid state NMR. *The Journal of Chemical Physics* 83(5), 2015–2025 (1985) <https://doi.org/10.1063/1.449344>
- 41 Nation, P.D., Kang, H., Sundaresan, N., Gambetta, J.M.: Scalable mitigation of measurement errors on quantum computers. *PRX Quantum* 2, 040326 (2021) <https://doi.org/10.1103/PRXQuantum.2.040326>
- 42 Martiel, S., Ayril, T., Allouche, C.: Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way. *IEEE Transactions on Quantum Engineering* 2, 1–11 (2021) <https://doi.org/10.1109/TQE.2021.3090207>
- 43 Farhi, E., Goldstone, J., Gutmann, S.: A quantum approximate optimization algorithm (2014). <https://doi.org/10.48550/arXiv.1411.4028>
- 44 Bharti, K., Cervera-Lierta, A., Kyaw, T.H., Haug, T., Alperin-Lea, S., Anand, A., Degroote, M., Heimonen, H., Kottmann, J.S., Menke, T., Mok, W.-K., Sim, S., Kwek, L.-C., Aspuru-Guzik, A.: Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* 94, 015004 (2022) <https://doi.org/10.1103/RevModPhys.94.015004>
- 45 Ozaeta, A., Dam, W., McMahon, P.L.: Expectation values from the single-layer quantum approximate optimization algorithm on ising problems. *Quantum Sci. Technol.* 7 045036 (2022) 7(4), 045036 (2020) <https://doi.org/10.1088/2058-9565/ac9013> [arXiv:2012.03421](https://arxiv.org/abs/2012.03421) [quant-ph]
- 46 Bravyi, S., Kliesch, A., Koenig, R., Tang, E.: Obstacles to variational quantum optimization from symmetry protection. *Physical review letters* 125(26), 260505 (2020)
- 47 Rönkkö, J., Ahonen, O., Bergholm, V., Calzona, A., Geresdi, A., Heimonen, H., Heinsoo, J., Milchakov, V., Pogorzalek, S., Sarsby, M., Savitskyi, M., Seegerer, S., IV, F., Sriluckshmy, P.V., Vesanen, P.T., Nakahara, M.: On-premises superconducting quantum computer for education and research. *EPJ Quantum technology* 11, 32 (2024) <https://doi.org/10.1140/epjqt/s40507-024-00243-z> [arXiv:2402.07315](https://arxiv.org/abs/2402.07315)
- 48 Nation, P.D., Treinish, M.: Suppressing quantum circuit errors due to system variability. *PRX Quantum* 4, 010327 (2023) <https://doi.org/10.1103/PRXQuantum.4.010327>
- 49 IQM Resonance — meetiqm.com. <https://www.meetiqm.com/products/iqm-resonance>. [Online; accessed 18-04-2024]
- 50 Heinsoo, J.: Gate performance statistics from quantum processing units with tens of qubits. *Bulletin of the American Physical Society* (2024)