Calibration and Performance Evaluation of a Superconducting Quantum Processor in an HPC Center

1st Xiaolong Deng
QCT Department
Leibniz Supercomputing Centre
Garching, Germany
xiaolong.deng@lrz.de

2nd Stefan Pogorzalek
Technology Department
IQM Quantum Computers
Munich, Germany
stefan.pogorzalek@meetiqm.com

3rd Florian Vigneau
Technology Department
IQM Quantum Computers
Munich, Germany
florian.vigneau@meetiqm.com

4th Ping Yang
Technology Department
IQM Quantum Computers
Munich, Germany
ping.yang@meetiqm.com

5th Martin Schulz
Chair of Computer Architecture and Parallel Systems
Technical University of Munich
Munich, Germany
schulzm@in.tum.de

6th Laura Schulz
QCT Department
Leibniz Supercomputing Centre
Garching, Germany
schulz@lrz.de

***Abstract*—As quantum computers mature, they migrate from laboratory environments to HPC centers. This movement enables large-scale deployments, greater access to the technology, and deep integration into HPC in the form of quantum acceleration. In laboratory environments, specialists directly control the systems' environments and operations at any time with hands-on access, while HPC centers require remote and autonomous operations with minimal physical contact. The requirement for automation of the calibration process needed by all current quantum systems relies on maximizing their coherence times and fidelities and, with that, their best performance. It is, therefore, of great significance to establish a standardized and automatic calibration process alongside unified evaluation standards for quantum computing performance to evaluate the success of the calibration and operation of the system. In this work, we characterize our in-house superconducting quantum computer, establish an automatic calibration process, and evaluate its performance through quantum volume and an application-specific algorithm. We also analyze readout errors and improve the readout fidelity, leaning on error mitigation.**

***Index Terms*—quantum computer, quantum-HPC, HPC, HPCQC, superconducting circuit, calibration, benchmarking**

I. Introduction

Quantum computing promises to significantly accelerate computational capability, thereby improving results for a specific class of problems, including applications in chemistry, engineering, and finance workloads. Various technologies comprise the portfolio of available quantum systems, including superconducting circuits, trapped ions, neutral atoms, and photonics. To realize their potential, they must integrate into the fundaments of the HPC environment on the hardware side to reach the needed low latencies and on the software side to enable the hybrid execution of these workloads.

This integration necessitates the direct placement and operation of quantum systems in HPC centers. It presents a paradigm shift from their typical operations model as research systems in physics laboratories under the supervision of expert-level lab operators attuned to the technologies' specificities and able to tune and calibrate the systems for maximal performance manually. In HPC centers, however, we require more autonomy and automation in operations, with far less manual intervention and physical access. In particular, regular maintenance and adjustments must happen automatically when triggered by the system infrastructure, necessitating a clearly defined procedure coupled with performance metrics to track the system's quality.

At our center, direct HPC and QC integration is facilitated through the deployment of two superconducting quantum computers: a 5-qubit Quantum Processor Unit (QPU) in operation and a 20-qubit QPU in the installation phase. Additionally, a 20-qubit trapped-ion quantum computer is in the installation phase, and additional technologies are already acquired. These systems must operate in an HPC center under typical environment levels for noise, temperature, and vibration, among others.

All systems are classified as Noisy Intermediate Scale Quantum (NISQ) systems [1], as fault-tolerant quantum computers have yet to be realized. Such systems do not have error correction as in classical computers and are prone to considerable error rates. Although the size is still limited, the number of qubits has increased significantly. As quantum computers move from physics labs to HPC centers, efficiently detecting and correcting error rates in qubits and gates is a core challenge for operators. At the same time, efforts are underway to enhance quantum gate performance and reduce noise to

make quantum hardware more reliable. Maintaining stable and high quantum computing performance for our users is our goal. However, quantum hardware drifts over time, undergoing small physical changes that affect the accuracy of quantum gates. To compensate, quantum systems run characterization and recalibration experiments to adjust gate pulse parameters. Especially for superconducting circuit quantum computers, another factor is that control parameters need to be given for all qubits individually, requiring a lot of step-by-step experimental procedures to revisit their drift. To mitigate this, one of the most essential steps for HPC centers is establishing a standard and efficient recalibration and benchmarking process for quantum systems. Specifically, recalibration must ultimately be automated and triggered by some criteria, and the recalibration time should be optimized to reduce to a point while still maintaining high fidelity. In this paper, we investigate the characterizations of qubits and gates based on our chip and establish a calibration and benchmarking process. This case study will help us understand our strategy's performance and establish a standard process for future chips.

In particular, this paper makes the following contributions:

- We have characterized the in-house 5-qubit QPU (Q5) by standard quality metrics, such as coherence times, gate fidelity, and readout fidelity.
- We have developed an automated recalibration process for our Q5 chip, enabling its autonomous operation.
- To verify the performance of the calibrated QPU, we have implemented quantum volume and variational quantum eigensolver algorithms on our QPU, made measurements, and analyzed the performance.
- We have investigated the readout errors in our QPU chip and improved the performance through error mitigation.

The rest of the paper is outlined as follows. Section II provides an overview of superconducting qubits and describes how to characterize the qubits. Section III outlines the benchmarking approaches and discusses three concrete algorithms: randomized benchmarking, quantum volume, and variational quantum eigensolver for our QPU. Section IV details the handling of errors. In the first subsection, we describe our calibration process and discuss how to automate and optimize it. Then, in the second subsection, we explain the readout errors in our QPU. Section V presents experimental results on the QPU and evaluates the QPU's performance. Section VI concludes the paper.

II. Characterization of the QPU

In this work, we focus on superconducting circuit quantum computers. Superconducting circuits are relatively easy to operate and scale well, but are prone to noise and crosstalk on the circuits. Superconducting qubits are based on superconducting LC circuits with Josephson junctions. According to different physical properties, qubits can be designed as either charge qubits, flux qubits or phase qubits. Charge qubits utilize electronic charges in metallic islands, while Flux qubits consider the clockwise or counter-clockwise direction of current circulation in a superconducting loop, and Phase qubits use the difference of the superconducting phase in the Josephson junctions. The transmon qubit used in our QPU is a charge qubit consisting of two capacitor structures connected by Josephson junctions. When using tuneable transmons, the Josephson junctions and the metal form a superconducting loop allowing an external magnetic flux, induced by an electrical current, to control the frequency of the qubit. An advantage of transmon qubits is that the large capacitance exponentially reduces sensitivity to charge noise, but, on the other hand, algebraically reduce the anharmonicity [2].

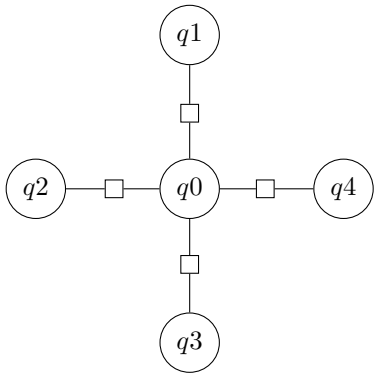


Fig. 1. Our Q5 chip implements a star topology. Circular nodes are indexed by qubit id and square nodes are tuneable couplers. Tunable couplers mediate interactions between qubits by adjusting their frequency. In particular, they can turn ON the qubit-qubit coupling during two-qubit gate execution.

Performing quantum operations on a qubit and reading out its state is achieved by coupling the qubit to LC electrical circuits [3]. Typically, superconducting transmission line resonators are used to couple the transmon to obtain a strong enough energy exchange. Our QPU uses tuenable couplers to reduce the execution time and to increase the fidelity of two qubit gates [4], see its topology in Fig. 1. In quantum mechanics, such a system can be described by the Jaynes-Cummings Hamiltonian, and in the simplest case is analytically solvable under some approximations. Typically, quantum computing is performed in dispersive regimes, where the qubit is non-resonantly coupled to a resonator and the difference between their frequencies is much greater than their coupling strength.

A. Rabi Experiment

When an external driving microwave pulse matches the frequency of the qubit, the quantum state of the driven qubit can oscillate between the two energy levels $|1\rangle$ and $|0\rangle$ (on the computational basis). This is called Rabi oscillation. By measuring the probability of the qubit staying in $|1\rangle$ or $|0\rangle$ one knows its quantum state. When a Rabi oscillation occurs, adjusting the time is enough to obtain different probabilities. For a specific time one has a π pulse that drives the qubit from $|0\rangle$ to $|1\rangle$ or $|1\rangle$ to $|0\rangle$, while halfway through that particular time there is a $\pi/2$ pulse, which takes the qubit from $|0\rangle$ to the superposition state $\frac{1}{\sqrt{2}}(|0\rangle + \mathrm{i}|1\rangle)$. In quantum circuits,

the above processes can be written as $|0\rangle$ —[$R_x(\pi)$]— and $|0\rangle$ —[$R_x(\pi/2)$]—, respectively. The well-known Hadamard (H) gate on $|0\rangle$ is $|0\rangle$ —[$R_y(\pi/2)$]—. In Rabi experiments one can not only change the length of microwave pulses, but also the amplitude and phase of microwave pulses to control the qubit state.

B. Coherence Time: T_1, T_2

Qubits inevitably exchange energy with the external environment. T_1 is used to describe the decay time of a qubit relaxing from the state $|1\rangle$ to the state $|0\rangle$ or by absorbing energy from $|0\rangle$ to $|1\rangle$. T_1 time can be measured by the Rabi experiment, as introduced above: a) use a π pulse to excite the qubit from $|0\rangle$ to $|1\rangle$, b) wait for a period time of t, and then c) measure the probability of being at $|1\rangle$. This probability is described by the formula $P(|1\rangle) = \exp(-t/T_1)$, where $1/T_1$ is the relaxation rate. By measuring the probabilities at different times and one can get the T_1 time by fitting the above formula.

As a second metric, the T_2 time describes the loss of coherence of the qubit being in a superposition state. Compared with the T_1 relaxation process on the longitudinal axis of the Bloch sphere, the $T2$ process can be imagined as an expansion on the horizontal axis of the Bloch sphere, for instance from $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ to $|-\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$ or to the mixture of $|+\rangle$ and $|-\rangle$ with equal probabilities. The T_2 time can be measured in the following way: a) prepare the qubit into the superposition state $|+\rangle$ using a $\pi/2$ pulse, b) wait for a period of time t, c) apply a second $\pi/2$ pulse to the qubit, and then d) measure the probability of being in the state $|0\rangle$. Over a long period of time, the measurement is expected to be closer to $1/2$. Similar to the relaxation $T1$, the amplitude damping also follows an exponential decay $P(|0\rangle) = \exp(-t/T_2)$. An exponential fitting gives the T_2 time.

C. Gate Fidelity

Quantum gates are the basic building blocks of quantum circuits. Generally speaking, single-qubit gates create a superposition of individual qubits, while two-qubit gates create entanglement. The single-qubit gate corresponds to a rotation on the Bloch sphere and is realized by a microwave pulse. The two-qubit gates in our Q5 system is a controlled-Z gate (CZ), which is implemented by bringing the $|11\rangle$ state into resonance with $|20\rangle$ state, where $|2\rangle$ is the second excitation level just above $|1\rangle$ [3]. After a period of time (equivalent to π over the detuning coupling strength between $|1\rangle$ and $|2\rangle$) and performing a 2π rotation, the state $|11\rangle$ will get a phase -1. Thus, this process implements a CZ gate.

Gate fidelity is one of the most important quality metrics for a QPU. Randomized benchmarking methods are commonly used in experiments to characterize the error rates of qubit gates [5], [6]. A sequence of randomly generated Cliffords gates on the given qubits is followed by its inverse so that the entire unitary is ideally the identity. Gate fidelity will be obtained from the survival probability of the initial state on the circuit. Details will be shown in Section III-A.

D. Readout Fidelity

In the dispersive regime, the frequency of the resonator depends on the qubit state. Readout measurements exploit this property to measure the qubit state. This task can be achieved by applying a resonant microwave pulse to the resonator. Then, recording the phase and amplitude of the signal propagating back allows to determine the qubit state.

Each individual measurement can be represented by a point in the in-phase and quadrature plane of the transmitted field. The histogram of the measurements forms two distinguishable clusters, corresponding to $|0\rangle$ and $|1\rangle$, each cluster having a Gaussian distribution. A threshold is introduced that separates the two clusters, which classifies future measurement shots as $|0\rangle$ or $|1\rangle$.

III. Benchmarking Approaches

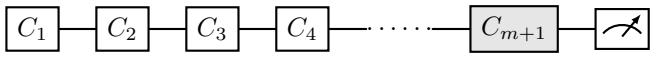
After characterizing the QPU, we need to quantify the capabilities of the quantum computer. Each metric may reflect an aspect of the QPU's quibt or gate performance. However, there is no unique method to determine the overall quality of the quantum computer. Therefore, it is important to establish standard quantum benchmarks to better evaluate the performance. Many quantum benchmarks have been proposed to evalute different aspects of the quantum computer. One approach is to make predictions based on physical models and statistics, such as characterization of qubits and gates. The statistical measurements of randomized benchmarking protocols are also predictive. With statistical methods based on randomized circuits one can either measure the cross-entropy between the observed bit string and the expected probabilities of those strings from simulation, or evaluate the largest randomized circuits of equal width and depth, or measure an entire series of Clifford gates using a certain protocol. Another direction of benchmarking is to run a set of application-based algorithms, for instance, to check the popular quantum algorithms such as a quantum Fourier transform [7], a quantum phase estimation [7], the Shor algorithm [8], the Grover algorithm [9], quantum approximate optimization algorithms [10], variational quantum eigensolvers [11], etc. These algorithms are measured by accuracy and precision of results, execution time, etc. The metrics can be used not only to compare the performance of different QPUs' but also to trigger a re-calibration process, if necessary.

In this paper, we discuss three statistical benchmarking approaches, i.e., randomized benchmarking (RB) [12], gate set tomography [13] and quantum volume (QV) [14], as well as an application implementating variational quantum eigenvolver (VQE) [11]. The RB method is used to determine average gate fidelity, QST is used to predict a best-fit model for the gate set, QV is used to estimate an overall capability of the QPU,

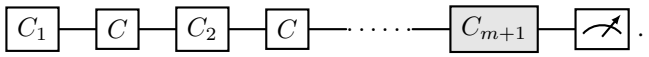
while VQE is used to test application-specific performance on the QPU.

A. Randomized Benchmarking (RB)

Decoherence of quantum states and various types of noise degrade the quality of the gate. Accurate estimation of gate error is critical. To measure the average performance of a gate, one can exploit the concept of random circuits and statistical measurements after sufficient trials. RB is one such experimental protocol for evaluating the quality of quantum gates. The protocol works by operating a sequence of Clifford gates and then their inverse sequences so that the survival probability of the original state should ideally be 1. In the standard RB protocol [5] the sequence of gates from the Clifford group is randomly generated as



where C_{m+1} inverts the entire sequence to ideally obtain an identity. Average gate fidelity can be determined using the standard RB methods. However, in order to determine the fidelity of a specific gate C, we use the interleaved RB calculation [6], where a randomly generated sequence of Clifford gates is interleaved with the desired C gate, e.g.

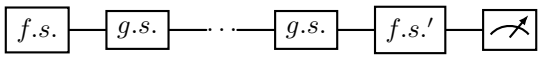


Finally, there is the inverse of the entire sequence C_{m+1} before measurements. Under the gate-independent noise assumption, the average survival probability p_m of a RB process with sequence length m can be fitted as $p_m = A + Bf^m$, where A and B depend on state preparation and measurement error, and f represents the quality metric of the gates. The parameters f is related to the average fidelity in a straightforward manner.

B. Gate Set Tomography (GST)

Quantum state tomography (QST) is to reconstruct the density matrix ρ of a state, generally represented in Pauli basis I, X, Y, Z. For example, for a single qubit, one only needs to measure the three expectations of Pauli operators to obtain the state $\rho = (I + \langle X\rangle X + \langle Y\rangle Y + \langle Z\rangle Z)/2$. Quantum process tomography (QPT) is similar but reconstructs specific gates rather than the state. QST and QPT usually assume that the state preparation and measurement (SPAM) is ideal. As an improvement, gate set tomography [13], [15]–[17] takes into account of the actual SPAM noise to better characterize the noise mechanism of the gate set. The recently proposed Lindblad tomography (LT) [18] also employs the SPAM error and several assumptions (Markov channel, constant SPAM error and perfect single qubit gate) and estimates the Hamiltonian and Lindbladian operator. Ref. [18] demonstrated that crosstalk resources in a quantum superconducting processor can be identified through LT.

GST selects a model parameterized by physical gates in the gate set and repeatedly executes a certain number of structured circuits that are sensitive to the parameters of the model. The estimate of the gate set can then be approximated by optimizing the likelihood of model simulations and experimental data. The gate sequence diagram is as follows:



where $f.s.$ ($f.s.'$) represents a fiducial sub-circuit used for state preparation (measurement) in a particular basis, while $g.s.$ represents a germ sub-circuit that is repeated for a certain number of times so that the gate set parameters sensitive to coherent errors can be amplified. Then one makes a noisy model simulate the data on the QPU. The estimated quantum processes can be obtained by optimizing the cost function using the maximum likelihood method.

In contrast to RB assessing the average fidelity, the actual noise processes from SPAM in GST provides more information for modelling QPUs. The experiments of GST and RB of single-qubit gates are planned to be performed on our Q5 chip. The fiducial preparation and measurement sequences and germ sequences will be created using pyGSTi [19], a toolkit for characterizing the QPU performance by executing various quantum protocols.

C. Quantum Volume (QV)

To measure the performance, we have to consider at least three key attributes: scale, quality and speed. The QV is designed to include these three points plus the connectivity of qubits and the error rate of gates and measurements. Further, the dephasing time is also considered together [14]. From random $\mathrm{SU}(4)$ unitaries we take randomly permuted gates to form quantum circuits with k qubits and k layers, and run the circuits for many trials. If the measured results for the largest k are above a certain threshold, then we compute QV as $QV = 2^k$. Generally, the higher the gate quality the bigger the QV value. The random QV circuits include two-qubit gates applied to two arbitrary qubits. In the architecture of superconducting circuits, however, two-qubit gates exist only between the nearest-neighboring qubits, resulting in restrictions to the desired gates applied to long-distance qubits. To overcome this limitation, SWAP gates can be introduced to move qubit states between connected qubits and with that to achieve the configuration that the desired gate applies to. A swap gate is represented by 3 CZ gates with 6 Hadamard gates. Thus, the transpiled circuit running on a QPU includes additional layers. In this sense, QV also measures — to some extent in software — the qubit mapping, routing, allocation and movement.

QV cannot capture the overall scale of QPU and also limits the scale of its classical computation. To overcome these shortcomings, layer circuit benchmarking combined with the RB technique is proposed [20]. Instead of random Clifford gates being interleaved by the desired gates, layer circuits with only two-qubit gates are interleaved by layer circuits with only random single-qubit gates. The full layer with two-qubit gates needs to be first divided into disjoint layers without overlapping qubits. Using the RB sequences, one can then calculate the so-called layer fidelity. Layer fidelity is the product of the fidelity of each disjoint layer. The fidelity of each disjoint layer is the product of the fidelity of the

RB process. The layer fidelity is easy to be scaled and will therefore be used for our future benchmarking.

D. Application Benchmark: VQE

From QV we can get a sense of overall performance, but we cannot predict how good an application's results will be. Here we choose VQE to examine the application-specific performance of Q5.

Given a Hamiltonian $\mathcal{H}$ one wants to find the state $|\psi\rangle$ that minimizes the energy $E = \langle\psi|\mathcal{H}|\psi\rangle$. In general, the degrees of freedom of the Hamiltonian increase exponentially. One solution is to use variational principles [21], where $|\psi(\theta)\rangle$ is parameterized with a finite set of parameters θ, to minimize

$$\min_{\theta} \frac{\langle\psi(\theta)|\mathcal{H}|\psi(\theta)\rangle}{\langle\psi(\theta)|\psi(\theta)\rangle}. \tag{1}$$

VQE is a hybrid quantum-classical variational algorithm that combines quantum measurements with classical optimization and is particularly useful for finding the lowest energy of a given material/molecule. The VQE scheme consists of a quantum part and a classical part. The task of the quantum computer is to prepare the initial state $|\psi(0)\rangle$, evolves it under the parametric circuit $U(\theta)$ and measures the expectation values of the required Hamiltonian terms. The task of the classical computer is to construct the energy value based on the expectation values provided by the quantum computer and use the optimizer to suggest a new set of parameters θ for the next quantum task. Measurements of Hamiltonian terms demand only shallow circuits. After each run, the quantum circuit resets, and the short coherence time is enough to get the job done.

IV. Handling errors

A. Calibration Process

After the installation of a new quantum computer, it requires a full initial calibration. The whole calibration procedure is complex and we only explain the central steps in the following. We first use spectroscopy experiments to determine the rough qubit readout frequencies via resonances of the readout resonators. We then calibrate the π pulse by performing Rabi experiments and determine the qubit frequency through Ramsey experiments. To avoid leakage to higher energy levels, we calibrate and use DRAG pulses. The flux biases of tuneable couplers are scanned in order to obtain idle points which negligible coupling between adjacent qubits. With these updated parameters, we can turn to the readout to further optimise it. After calibrating individual qubits, we turn to the two-qubit CZ gates between adjacent qubits. The calibration procedure involves multiple optimization steps to ensure proper implementation of a CZ gate [22]. To investigate the quality of single- and two-qubit gates, we perform RB experiments as well as GHZ state benchmarks. In addition, the T_1 and T_2 coherence times are determined by suitable experiments. When the calibration sequence completes successfully, we generate a set of calibration parameters.

However, an initial one-time calibration is not sufficient for a proper operation. Quantum systems and control hardware drift over time, which degrades the gate's accuracy. This fact gives rise to the need for regular recalibration of certain parameters. One can choose between various recalibration procedures, ranging from full recalibration to only adjusting a single parameter. The employed recalibration is a compromise between partial down-time due to recalibration and best performance. Generally, the previous calibration is still reasonably good such that one starts the recalibration from the previous set of calibration parameters. To maintain stable performance, regular recalibration is necessary. Currently, we do the full recalibration at our center twice a week, to keep a stable performance of the system. The calibration process of our Q5 system is shown in Fig. 2.

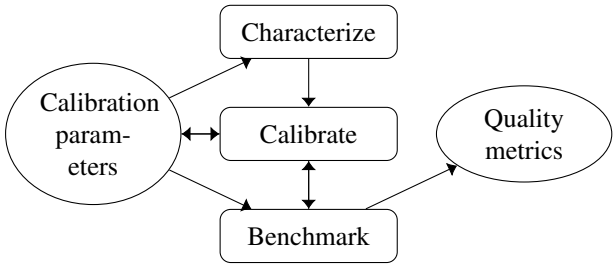


Fig. 2. The scheme of the calibration process performed for our Q5 chip.

In a superconducting circuit quantum computer, each qubit is different from each other. Accurate quantum gate implementation and manipulation requires accurate individual characterization of qubitsand quantum operations. To obtain high-quality gates, complex microwave pulses need to be designed and optimized, which is even more challenging if crosstalk and noise needs to be considered. This means that as the number of qubits increases, the complexity of calibrating all required control parameters for the qubits and couplers will become enormous. Under such situations the calibration process will take a long time, which means a large amount of time will be spent on calibration rather than quantum jobs. Further, it will no longer be feasibable as a manual process as common with today's laboratory-born systems. Therefore, setting up an automated and efficient calibration procedure becomes increasingly important. In the automated calibration, the chosen sequence of calibration experiments should be carefully redesigned to enable efficient scheduling and optimization of time-consuming experiments. In particular, when two-qubit gates do not involve the same qubit, it is expected that their calibration processes can be carried out in parallel.

Our automated process is currently run through cron jobs and are still in experimental testing. Most calibration steps are included in a python script, which will be triggered by a cronjob. The script also contains error handling if an experiment fails, e.g., simply retrying the failed experiment or changing the experimental parameters slightly. When a complete recalibration begins, the "Maintenance" flag will be set true. On the Q5 machine, the Hellinger fidelity (i.e. the

fidelity of two counts of distributions) of the GHZ state is currently set as the trigger. The 5-qubit GHZ state is checked every few hours. If the 5-qubit Hellinger fidelity falls down below a threshold (e.g. 0.5), a check of the 2-qubit GHZ states is triggered and a recalibration process for the qubits of the smallest Hellinger is initiated. The 5-qubit GHZ will then be checked again. If larger than the threshold, the recalibration stops. Otherwise, a further process is triggered for the qubits of the second smallest Hellinger, then a further process for the third smallest Hellinger, and so on. For simplicity, we can also start up from the 3-qubit GHZ state.

The controlling software of room-temperature electronics and calibration software are put into a docker container and installed on a node of the Linux cluster. It has 2 sockets, each containing a 36-core Xeon Platinum 8360Y CPU. The Ethernet connecting to the electronics is of bandwidth 10Gbits/s. The calibration software (in Python) is developed and provided by the quantum computer vendor.

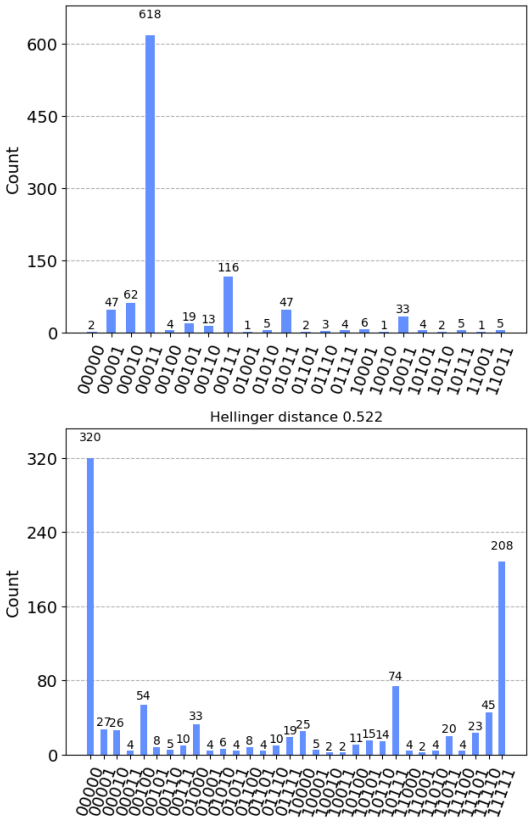


Fig. 3. Top: The measurements of the state $|00011\rangle$ by 2 X gates on qubits q_1 and q_2, respectively. Bottom: The measurements of the GHZ state $\frac{1}{\sqrt{2}}(|00000\rangle + |11111\rangle)$. For each of experiments 1000 shots are performed. The Hellinger fidelity of the GHZ state is about 0.52 in this experiment.

B. Error Mitigation

Errors and gate infidelities can be reduced through a systemic calibration process. However, even in calibrated QPUs, there are still different small errors. We give two specific examples of readout errors in Fig. 3. Two types of states $|00011\rangle$ (product state) and $\frac{1}{\sqrt{2}}(|00000\rangle + |11111\rangle)$ (entangled state) are generated and measured with 1000 shots each. From the results, we can see that the major error occurs in the first-order process. For state $|00011\rangle$, errors mainly cause states that are 1 bit different from state $|00011\rangle$. The same goes for the GHZ state. The GHZ state is entangled, but after measurement it collapses to either $|00000\rangle$ or $|11111\rangle$. Similar as the product state, readout errors expand mainly to states 1 bit different from $|00000\rangle$ and $|11111\rangle$. As we can see from these examples, these errors are a considerable problem for the current system. An increase in the number of qubits or quantum bases/states will increase the likelihood and complexity of errors, ultimately leading to loss of information and spoiling the results.

It is important to understand what these errors are and how to characterize and measure them, and then design better algorithms by overcoming them, or use error mitigation techniques to reduce them and reach better benchmark results. The RB method to obtain the gate fidelity discussed in Section III is actually an averaged error characterization. Error mitigation techniques have been proposed recently by adding mitigation gates to the circuit or performing post-processing after measurement, such as zero-noise extrapolation [23], probabilistic error cancellation [24], dynamic decoupling [25], readout error mitigation [26], etc.

V. Experimental Results

The computing performance of the superconducting quantum computer has been carefully tested at our center. In this section, we present experimental results on Q5 through quantum volume and varational quantum eigensolver and how the performance can be improved by error mitigation.

T1 (μs)	T2 (μs)	Single	CZ (%)	Readout (%)
30.11	9.23	99.86	97.9	91.6

TABLE I

A. Quality Metrics

Through the characterization protocols in Section II, we obtain the quality metrics of Q5, as shown in Table I. The average coherence times are $T_1 \approx 30.11\mu$s and $T_2 \approx 9.23\mu$s, when all five qubits are characterized individually. Single-qubit gate and CZ gate fidelity are determined using standard and interleaved RB methods, showing averaged CZ fidelity of 97.9% and averaged single-qubit gate fidelity of 99.86%. The Q5's average readout fidelity is 91.6% which is limited by crosstalk and the absence of a travelling-wave parametric amplifier. The system under investigation here is a research system and specifications of a production can be found elsewhere [27].

B. QV

We use Qiskit [28] functionality to perform and analyse the QV experiments [29]. The results are shown in Fig. 4. Due

to the limitation of $T2$ time, the largest number of layers we can measure is $k = 3$. For each k, $n = 100$ trials of circuits are generated and transpiled in the set of basis gates $[r, cz]$ under the layout that the hardware qubit $q0$ is always on. In the $k = 3$ circuits the required SWAP gate to fulfil the connectivity constraints is controlled by Qiskit and transformed to the corresponding CZ and Hadamard gates. Then, we run the jobs in the Q5 backend and do the measurements with 1024 shots for each circuit. The experimental results together with the ideal output of these random circuits simulated by the statevector simulator are added into the QV fitter. The QV fitter gives us the mean heavy output probability (HOP) and the confidence level. When HOP $> 2/3$ with 97% certainty the QV verification is successful.

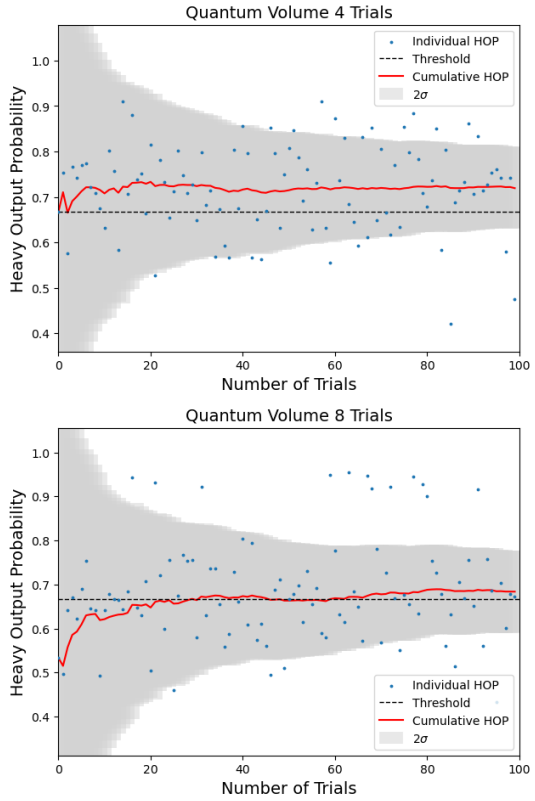


Fig. 4. QV4 (top) and QV8 (bottom) verification. A total of 100 trials are run for each. Blue dots mean HOP for each trial. Black dashed line means the QV success threshold at $2/3$. Red line means cumulative HOP. From the gray area we expect that more trials will reduce the error bars and get closer to the value $2/3$. For QV4 150 trials will make the QV experiment successful.

C. VQE

As a specific application in Q5, we use VQE to determine the potential curve of the hydrogen molecule for different interatomic distances. Molecular orbitals are encoded in qubits. Under the Bravyi-Kitaev transformation the molecular hydrogen is represented as a spin Hamiltonian. We consider the simplified effective Hamiltonian on only 2 qubits [30]

$$\mathcal{H} = g_0 I + g_1 Z_0 + g_2 Z_1 + g_3 Z_0 Z_1 + g_4 X_0 X_1 + g_5 Y_0 Y_1, \quad (2)$$

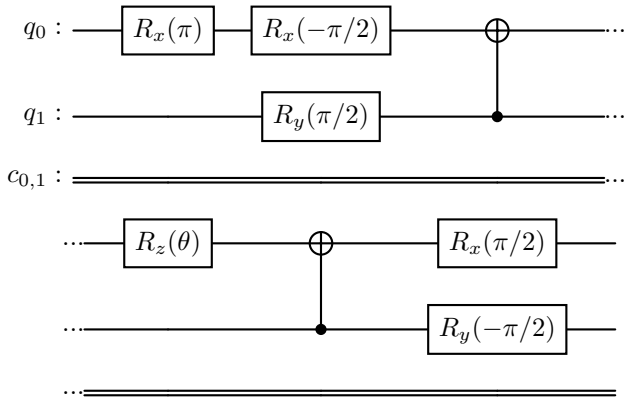


Fig. 5. The parameterized VQE circuit with a parameter θ. There are two two-qubit gates and six single-rotation gates. The Ansatz is the same as the one used in Reference [30].

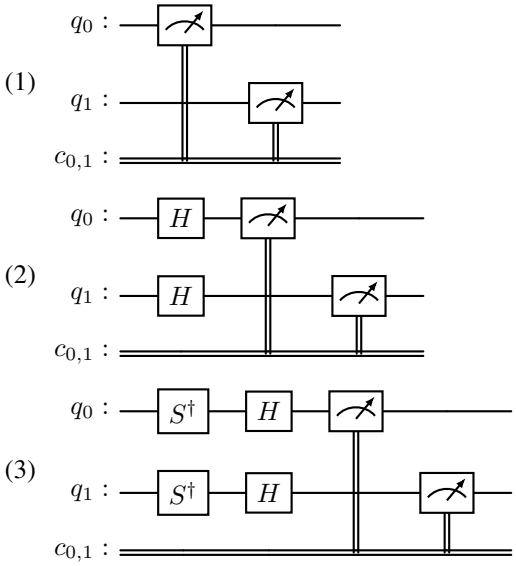


Fig. 6. The measurements in the Z basis for the Hamiltonian parts (1) $Z_0 \otimes Z_1$; (2) $X_0 \otimes X_1$; and (3) $Y_0 \otimes Y_1$.

where g_i are coefficients from the bond length of the molecule and I, X, Y and Z are the identity and Pauli operators. The parameterized VQE Ansatz for the Hamiltonian Eq. (eq:h2) is designed same as the one in Reference [30], see Fig. 5.

Measuring the Hamiltonian Eq. (2) is equivalent to

$$\begin{aligned}\langle \mathcal{H} \rangle = g_0 I + g_1 \langle Z_0 \rangle + g_2 \langle Z_1 \rangle \\ + g_3 \langle Z_0 Z_1 \rangle + g_4 \langle X_0 X_1 \rangle + g_5 \langle Y_0 Y_1 \rangle.\end{aligned} \quad (3)$$

Since the measurement is performed on the Z basis, the terms Z_0, Z_1 and $Z_0 Z_1$ can be directly measured. As for the terms $X_0 X_1$ and $Y_0 Y_1$ it is necessary to first do rotations and then measure, more precisely, $\langle X_0 X_1 \rangle = \langle H Z_0 H \otimes H Z_1 H \rangle$ and $\langle Y_0 Y_1 \rangle = \langle H S Z_0 S^\dagger H \otimes H S Z_1 S^\dagger H \rangle$, where $S = Z^{1/2}$, as shown in Fig. 6. In our experiments, we do not measure $\mathcal{H}$ directly, but measure Z_0, Z_1, $X_0 \otimes X_1$, $Y_0 \otimes Y_1$ and $Z_0 \otimes Z_1$ respectively. Their expectation values can be constructed from the probability of the shots in the states $|00\rangle$, $|01\rangle$, $|10\rangle$

and $|11\rangle$. Eq. (3) finally gives the expectation value of H in the state of $U(\theta)|\psi(0)\rangle$. Then in the classical optimizer we choose the Powell method to predict the next θs minimizing the energies.

We map the Qiskit qubits to our backend qubits $[q0, q1]$. For each part of the Hamiltonian and a selected θ, we perform 1000 shots. The experiments and the ideal simulations were carried out for different atomic distances, as shown in Fig. 7. Considering the low readout fidelity and the absence of any error mitigation, the experimental results are reasonable, albeit with a relative error of $\approx 20\%$.

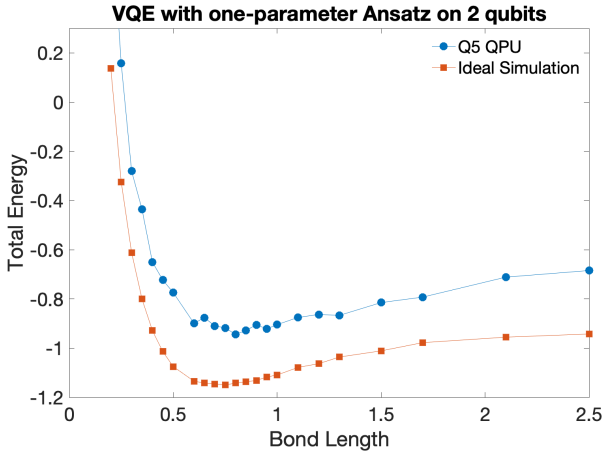


Fig. 7. Energy of molecular hydrogen H_2 as a function of bond lengths determined by VQE. Red line corresponds to the ideal simulation results by Qiskit. Blue line corresponds to the experimental results on Q5. The relative error between the real and the ideal is roughly 20%.

Using Q5's quality metric set we generate a fake backend and perform noise simulations by the means of Qiskit's Noise Model. Comparing to our experiments on Q5, noise model simulations with the assigned parameters from the quality metrics produced better results than the real quantum computer. This means that noise simulations using current quality metric parameters do not fully reflect the dissipative effects. Some impacts on/of other qubits and some projection errors on the measured qubits are may not be properly taken into account.

D. Readout Error Mitigation

In superconducting quantum processors, readout errors can be appear due to, for example, correlated readout errors or low signal-to-noise-ratio. However, some readout errors can be mitigated with suitable techniques. As a demonstration, we use the readout error mitigation technique provided by Qiskit [28] to improve GHZ state measurements and verify its performance. The local readout mitigator uses a series of circuits with X gates to measure states from $|00000\rangle$ to $|11111\rangle$ and get probabilities that can be used to compensate the effects from the errors. The mitigated results of the GHZ state are greatly improved, as shown in Fig. 8.

VI. Conclusions

In summary, we have investigated the QPU characterization, calibration, benchmarking, and error mitigation needed to

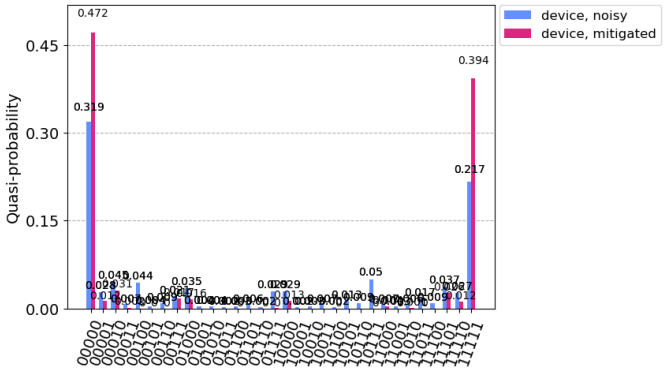


Fig. 8. The readout error mitigation technique performed on the GHZ state measurement on Q5. A significant improvement is obtained: the Hellinger fidelity is increased roughly from 0.5 to 0.8.

maintain stable and high-performance quantum computing in HPC centers. We have characterized the performance of our in-house R&D QPU chip Q5 and set up the calibration process and measurement of the quality metrics. We have addressed the benchmarking of algorithms and error mitigations as well. As the number of qubits grows and chip architectures become more complex, automated calibration and related optimization factors will become increasingly important, especially for HPC centers.

In HPC centers, quantum computing services are production-oriented, which requires maintenance operations to be not only reliable but also efficient. Obviously, as the systems scale up (the number of qubits continues to increase), the calibration scheduling problem becomes increasingly challenging. The calibration process should ensure high fidelity of qubits and gates as well as timely return to production. At this point, the trade-off between higher fidelity, more calibration effort and a limited time budget becomes important. The challenge of calibration scheduling problems deserves further research.

To deal with the trade-offs we can try different strategies. Suppose we have a list of qubits and gates, a list of well-defined calibration routines and their predicted duration times, priorities, expected fidelities (results), and the time budget. Some constraints should be set. First, only a subset of qubits will be calibrated at a time. This subset can be random, coarse-grained or user-defined so that time scales and system sizes are significantly reduced and there are always qubits available for service. Second, a time threshold will be set. If the predicted calibration time exceeds this threshold the qubits will be labelled as "unusable". Third, a fidelity threshold will be set. If qubit fidelity falls below this threshold, another round of calibration procedures is required. Based on these scheduling rules, we aim to maximise the number of qubits with higher fidelity within a given limited time budget. We hope this strategy can help us understand the complexity of scheduling calibration jobs and find an optimal sequence of calibration routines.

Furthermore, such scheduling problem can be formulated

as a different problem such as a knapsack-like problem, etc. Scheduler mechanism for HPC resource management may also help solve this problem. By deploying a 20-qubit QPU, we will be able to provide long-term large-scale characterization data sets for calibration routines. Based on this, we can extract more features of calibration jobs and further leverage machine learning techniques to build more accurate prediction models.

ACKNOWLEGMENT

The authors thank Dr. Mario Hernandez Vera (LRZ) for helpful discussions and for providing the nuclei's repulsive energy of hydrogen molecules and Dr. Muhammad Farooqi (LRZ) for helpful discussions on HPC resource scheduling and management. The authors also thank the Systems & Labs team of LRZ's Quantum Computing and Technologies Department for maintaining the stable performance of quantum systems. The work is supported by the BMBF projects DAQC (13N15689) and Q-Exa (13N16063) and the Munich Quantum Valley (MQV) funded through the state of Bavaria's High-tech Agenda.

REFERENCES

[1] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, aug 2018.

[2] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, "Charge-insensitive qubit design derived from the cooper pair box," *Phys. Rev. A*, vol. 76, p. 042319, Oct 2007.

[3] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver, "A quantum engineer's guide to superconducting qubits," *Applied Physics Reviews*, vol. 6, p. 021318, jun 2019.

[4] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. L. Campbell, T. P. Orlando, S. Gustavsson, and W. D. Oliver, "Tunable coupling scheme for implementing high-fidelity two-qubit gates," *Phys. Rev. Appl.*, vol. 10, p. 054062, Nov 2018.

[5] J. M. Gambetta, A. D. Córcoles, S. T. Merkel, B. R. Johnson, J. A. Smolin, J. M. Chow, C. A. Ryan, C. Rigetti, S. Poletto, T. A. Ohki, M. B. Ketchen, and M. Steffen, "Characterization of addressability by simultaneous randomized benchmarking," *Phys. Rev. Lett.*, vol. 109, p. 240504, Dec 2012.

[6] E. Magesan, J. M. Gambetta, B. R. Johnson, C. A. Ryan, J. M. Chow, S. T. Merkel, M. P. da Silva, G. A. Keefe, M. B. Rothwell, T. A. Ohki, M. B. Ketchen, and M. Steffen, "Efficient measurement of quantum gate error by interleaved randomized benchmarking," *Phys. Rev. Lett.*, vol. 109, p. 080505, Aug 2012.

[7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[8] P. W. Shor, "Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer," *SIAM Journal on Computing*, vol. 26, p. 1484–1509, Oct. 1997.

[9] L. K. Grover, "A fast quantum mechanical algorithm for database search," in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96, (New York, NY, USA), p. 212–219, Association for Computing Machinery, 1996.

[10] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014.

[11] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor," *Nature Communications*, vol. 5, July 2014.

[12] C. Dankert, R. Cleve, J. Emerson, and E. Livine, "Exact and approximate unitary 2-designs and their application to fidelity estimation," *Phys. Rev. A*, vol. 80, p. 012304, Jul 2009.

[13] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz, "Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit," 2013.

[14] A. W. Cross, L. S. Bishop, S. Sheldon, P. D. Nation, and J. M. Gambetta, "Validating quantum computers using randomized model circuits," *Phys. Rev. A*, vol. 100, p. 032328, Sep 2019.

[15] R. Blume-Kohout, J. K. Gamble, E. Nielsen, K. Rudinger, J. Mizrahi, K. Fortier, and P. Maunz, "Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography," *Nature Communications*, vol. 8, no. 1, p. 14485, 2017.

[16] E. Nielsen, J. K. Gamble, K. Rudinger, T. Scholten, K. Young, and R. Blume-Kohout, "Gate Set Tomography," *Quantum*, vol. 5, p. 557, Oct. 2021.

[17] T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout, "Measuring the capabilities of quantum computers," *Nature Physics*, vol. 18, no. 1, pp. 75–79, 2022.

[18] G. O. Samach, A. Greene, J. Borregaard, M. Christandl, J. Barreto, D. K. Kim, C. M. McNally, A. Melville, B. M. Niedzielski, Y. Sung, D. Rosenberg, M. E. Schwartz, J. L. Yoder, T. P. Orlando, J. I.-J. Wang, S. Gustavsson, M. Kjaergaard, and W. D. Oliver, "Lindblad tomography of a superconducting quantum processor," *Phys. Rev. Appl.*, vol. 18, p. 064056, Dec 2022.

[19] E. Nielsen, K. Rudinger, T. Proctor, A. Russo, K. Young, and R. Blume-Kohout, "Probing quantum processor performance with pygsti," *Quantum Science and Technology*, vol. 5, p. 044002, jul 2020.

[20] D. C. McKay, I. Hincks, E. J. Pritchett, M. Carroll, L. C. G. Govia, and S. T. Merkel, "Benchmarking quantum processor performance at scale," 2023.

[21] D. J. Griffiths and D. F. Schroeter, *Introduction to quantum mechanics*. Cambridge ; New York, NY: Cambridge University Press, third edition ed., 2018.

[22] F. Marxer, A. Vepsäläinen, S. W. Jolin, J. Tuorila, A. Landra, C. Ockeloen-Korppi, W. Liu, O. Ahonen, A. Auer, L. Belzane, V. Bergholm, C. F. Chan, K. W. Chan, T. Hiltunen, J. Hotari, E. Hyyppä, J. Ikonen, D. Janzso, M. Koistinen, J. Kotilahti, T. Li, J. Luus, M. Papic, M. Partanen, J. Räbinä, J. Rosti, M. Savytskyi, M. Seppälä, V. Sevriuk, E. Takala, B. Tarasinski, M. J. Thapa, F. Tosto, N. Vorobeva, L. Yu, K. Y. Tan, J. Hassel, M. Möttönen, and J. Heinsoo, "Long-distance transmon coupler with cz-gate fidelity above 99.8%," *PRX Quantum*, vol. 4, p. 010314, Feb 2023.

[23] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, "Error mitigation extends the computational reach of a noisy quantum processor," *Nature*, vol. 567, p. 491–495, Mar. 2019.

[24] E. van den Berg, Z. K. Minev, A. Kandala, and K. Temme, "Probabilistic error cancellation with sparse pauli–lindblad models on noisy quantum processors," *Nature Physics*, vol. 19, p. 1116–1121, May 2023.

[25] P. Das, S. Tannu, S. Dangwal, and M. Qureshi, "Adapt: Mitigating idling errors in qubits via adaptive dynamical decoupling," in *MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO '21, ACM, Oct. 2021.

[26] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, "Mitigating measurement errors in multiqubit experiments," *Physical Review A*, vol. 103, Apr. 2021.

[27] "https://meetiqm.com/resources/press-releases/finland-launches-a-20-qubit-quantum-computer."

[28] Qiskit contributors, "Qiskit: An open-source framework for quantum computing," 2023.

[29] P. Jurcevic, A. Javadi-Abhari, L. S. Bishop, I. Lauer, D. F. Bogorin, M. Brink, L. Capelluto, O. Günlük, T. Itoko, N. Kanazawa, A. Kandala, G. A. Keefe, K. Krsulich, W. Landers, E. P. Lewandowski, D. T. McClure, G. Nannicini, A. Narasgond, H. M. Nayfeh, E. Pritchett, M. B. Rothwell, S. Srinivasan, N. Sundaresan, C. Wang, K. X. Wei, C. J. Wood, J.-B. Yau, E. J. Zhang, O. E. Dial, J. M. Chow, and J. M. Gambetta, "Demonstration of quantum volume 64 on a superconducting quantum computing system," *Quantum Science and Technology*, vol. 6, p. 025020, Mar. 2021.

[30] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis, "Scalable quantum simulation of molecular energies," *Phys. Rev. X*, vol. 6, p. 031007, Jul 2016.