Case Study: Agentic Calibration of Superconducting Qubits

Beatriz Yankelevich[1,2], Eva Zhang[3], JonLuca DeCaro[3], Jeffrey A. Grover[1], and William D. Oliver[1,2,4]

[1] *Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[2] *Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[3] *OpenAI, San Francisco, CA, USA*
[4] *Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

September 4, 2026

Contents

1 Introduction

Quantum computing is an emerging computational paradigm that encodes and processes information according to the laws of quantum mechanics. Phenomena such as superposition, interference, and entanglement allow quantum algorithms to manipulate information in ways that differ fundamentally from classical computation. For certain problems, this enables dramatic speedups over the best known classical methods, most famously Shor's algorithm for integer factorization, while also offering a natural path toward simulating complex quantum systems in physics, chemistry, and materials science.

Superconducting circuits are a leading platform for quantum computing. Often described as "artificial atoms," superconducting qubits are integrated circuits whose discrete energy levels encode quantum information. At MIT, the Engineering Quantum Systems Group (EQuS) researches superconducting quantum systems across the stack, from new fabrication techniques and qubit designs to many-qubit experiments that explore quantum error correction and emergent quantum phenomena.

Designing and executing a superconducting-qubit experiment employs a diverse skill set that draws from various scientific and engineering disciplines. The work can be roughly grouped into five categories: analytical and numerical modeling of quantum systems, chip design, chip fabrication, cryogenic and microwave engineering, and qubit calibration and measurement. Although AI can

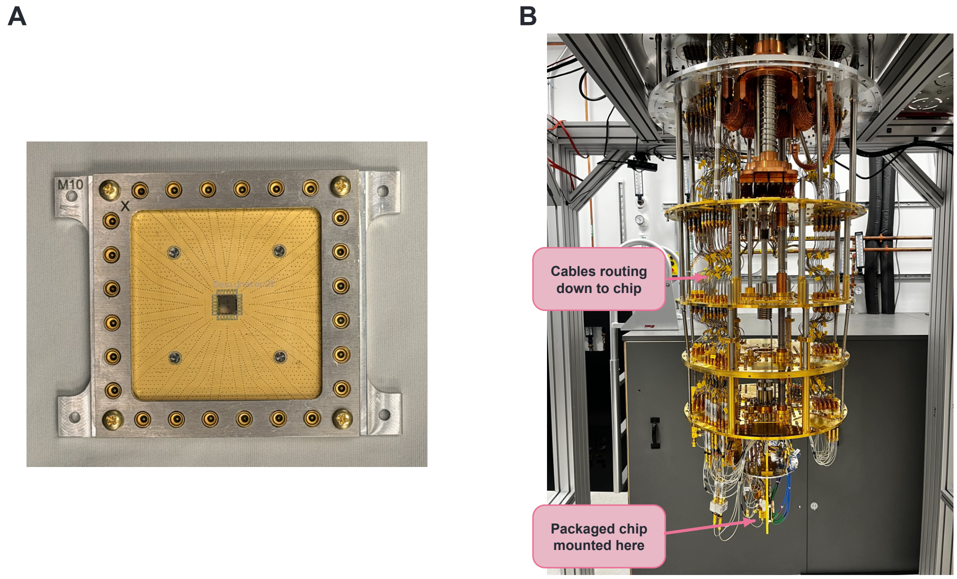

Figure 1: **Experimental setup**. **(a)** A packaged 5×5 mm chip. A printed circuit board routes signal lines from the chip in the center to connectors at the edge of the package, where cables connect. **(b)** An open dilution refrigerator. The chip package is mounted on the bottom flange. Cables connect to the package and route up through the top of the dilution refrigerator, over to a separate, room-temperature electronics rack. During operation, nested thermal shields and an outer vacuum can enclose the assembly.

accelerate nearly every stage, its most compelling role may be in qubit measurement, where it participates directly in the experiment[1–4]. Here, we describe how EQuS researchers enabled a general-purpose agent powered by GPT-5.6 Sol, on Ultra reasoning level, to conduct superconducting-qubit measurements end to end—controlling live laboratory hardware, interpreting the resulting data, and adapting subsequent measurements through the group's existing orchestration software.

Superconducting-qubit experiments are unusual among experimental fields in that, once a chip is fabricated, packaged, and cooled, the experiment is controlled almost entirely via software. The chip sits under vacuum inside a dilution refrigerator—a specialized cryogenic device that cools the chip to millikelvin temperatures. Computer programs coordinate sequences of pulses generated by room-temperature instruments. These pulses travel through cables into the refrigerator, interact with the quantum circuits on the chip, and return as signals that are amplified, digitized, and analyzed.

Each measurement is a miniature experiment that characterizes a property of the quantum system or calibrates an interaction. Measurements form an interdependent sequence: each step yields physical parameters—such as resonance frequencies, pulse amplitudes, and decay times—that often define the next step. Calibrating a novel multi-qubit experiment can require months and thousands of preliminary measurements before the target experiment can be performed. Only a handful of these measurements make it into the publication. Certain stretches of this workflow can be automated, when the analysis is unambiguous and behavior is predictable. But this is often not the case when dealing with new chip designs and exploring novel physical phenomena. The challenge is compounded by the fact that superconducting processors are imperfect physical systems that cannot be totally isolated from the environment. Parameters drift over time, and chips occasionally exhibit stochastic behavior, requiring adaptable and intelligent calibration.

Qubit calibration therefore presents an exciting challenge for agentic AI systems. Current AI models already possess extensive knowledge of the physics theory, software engineering skills, and specific experimental infrastructure used in these experiments. The question is whether agents can bring that knowledge to bear through the flexible, empirical style of reasoning needed to work with real-world experimental systems. EQuS researchers have seen promising early results with agentic calibration. Here, we present the workflow and results from a relatively simple superconducting-qubit chip, where an agent successfully performed a set of standard measurements on a never-before-calibrated chip.

2 How superconducting qubits work

The simplest superconducting qubit is a circuit comprising a capacitor and a Josephson junction, a superconducting circuit element that functions as a nonlinear inductor [5]. Together, these elements form a weakly nonlinear oscillator, which, in the quantum regime, has a series of energy levels with disparate transition energies—usually referred to by their resonant frequency. The two lowest levels are typically used as the computational states—"0" and "1," analogous to bits in a classical computer—but the higher states remain relevant. Higher levels can facilitate operations between qubits or cause unwanted errors.

We measure the qubit state by coupling it to a resonator, another circuit element that has a single resonant mode. The interaction between the qubit and the resonator is designed such that the resonant frequency of the resonator shifts depending on the state of the qubit. Thus, we can read out the qubit state by probing the resonator transition. These elements form the basic building blocks of a superconducting-qubit chip; much more complexity is added to perform novel, many-qubit experiments.

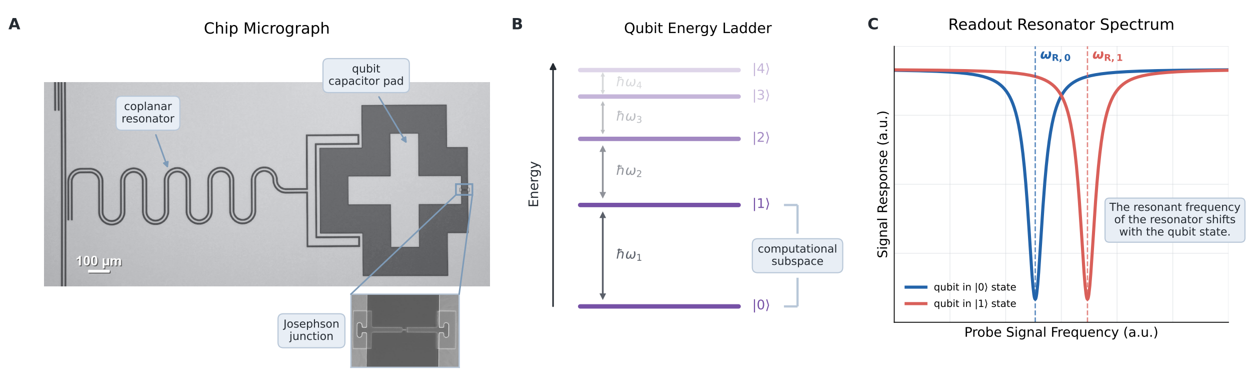

Figure 2: **(a)** On-chip qubit and coupled readout resonator. **(b)** Qubit energy spectrum—the lowest two levels are used as the computational basis. Energy transitions are typically referred to in units of frequency ω ($E = \hbar\omega$, where $\hbar$ is the reduced Planck constant.) A "ket" $|x\rangle$ denotes a quantum state. **(c)** Qubit-state-dependent resonator response. Sweeping the probe frequency across a resonant mode produces a Lorentzian response—a "dip" centered on the resonance frequency. The readout resonator resonant frequency shifts depending on the state of the qubit, enabling us to infer the qubit state.

3 Measurement workflow

To understand how a measurement is actually performed, consider the resonator spectroscopy measurement. This is typically the first measurement performed on a new chip, with the objective of finding the resonator frequency. The control electronics generate a pulse and send it to the resonator. The returning voltage signal is then digitized: a single point is represented as a complex number that encodes the amplitude and phase of the signal. This process is repeated many times so that the averaged result has a higher signal-to-noise ratio. Then, the frequency of this pulse is stepped, with the aim to sweep through the resonator frequency. The resulting raw data is an array of complex numbers. In a successful resonator spectroscopy measurement, the signal amplitude has a Lorentzian profile, with a dip at the resonant frequency (see Figure 2c). Another key parameter is the power of the pulse. Past a certain power, the pulse is strong enough to also excite the coupled qubit, and the resonator will shift in response. Sweeping the pulse power reveals this transition. The goal of this measurement is to extract the resonator frequency and the optimal pulse power, which are then persisted to calibration settings. Subsequent measurements use these values to define the qubit readout pulse. This workflow is driven by EQuS's in-house orchestration software that coordinates the various room-temperature instruments. Measurements take anywhere from a few seconds to tens of minutes, depending on the length of the pulse sequence, number of averages, and sweep lengths. Not every measurement has a predefined analysis model or target parameter; some are exploratory or diagnostic.

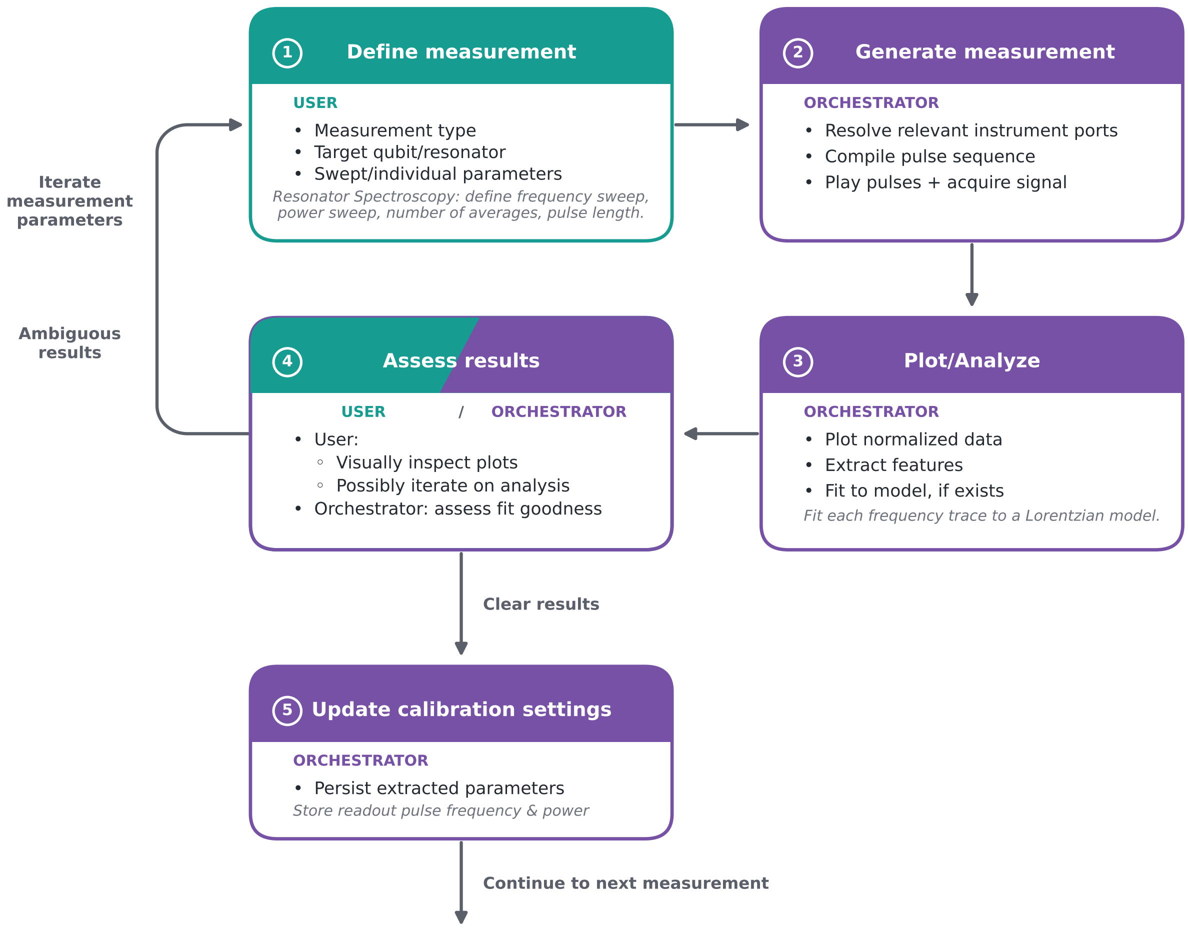

Figure 3: **Measurement workflow.** EQuS's in-house orchestration software drives the process, while an agent can assume the user's role.

4 Agentic calibration results

EQuS researchers instructed an agent to follow the same measurement workflow as human researchers, using the orchestration software through a Jupyter notebook. The agent ran through the Codex app with no specialized harness beyond a simple EQuS in-house Jupyter MCP. It had access to live measurement parameters, programs, plots, raw data, logs, and the measurement database. The agent could also analyze and plot data in the notebook and maintain laboratory notes in Markdown files. The main challenge was identifying the context agents needed to work effectively. After several months of iteration, the researchers converged on a combination of the experimental setup details, chip designs, and measurement-specific skills, as well as access to the orchestration software's source code. The skill for a particular measurement includes information about the template code for execution and analysis, the crucial calibrations that must be completed before the measurement can be performed, tips to choose good measurement and analysis parameters, common physical reasons that the measurement might fail, characteristics of a successful or failed measurement/analysis, and example plots of both successful and failed measurements.

The chip measured here consists of six uncoupled qubits, four with fixed frequencies and two with tunable frequencies, each with a coupled readout resonator. The qubits are nominally designed identically, but fabrication imperfections and environmental interactions cause each one to behave differently. This chip had never been measured before, so the agent had to discover the parameters, given the design targets. The agent successfully discovered all six resonators and appropriate initial powers for readout pulses.

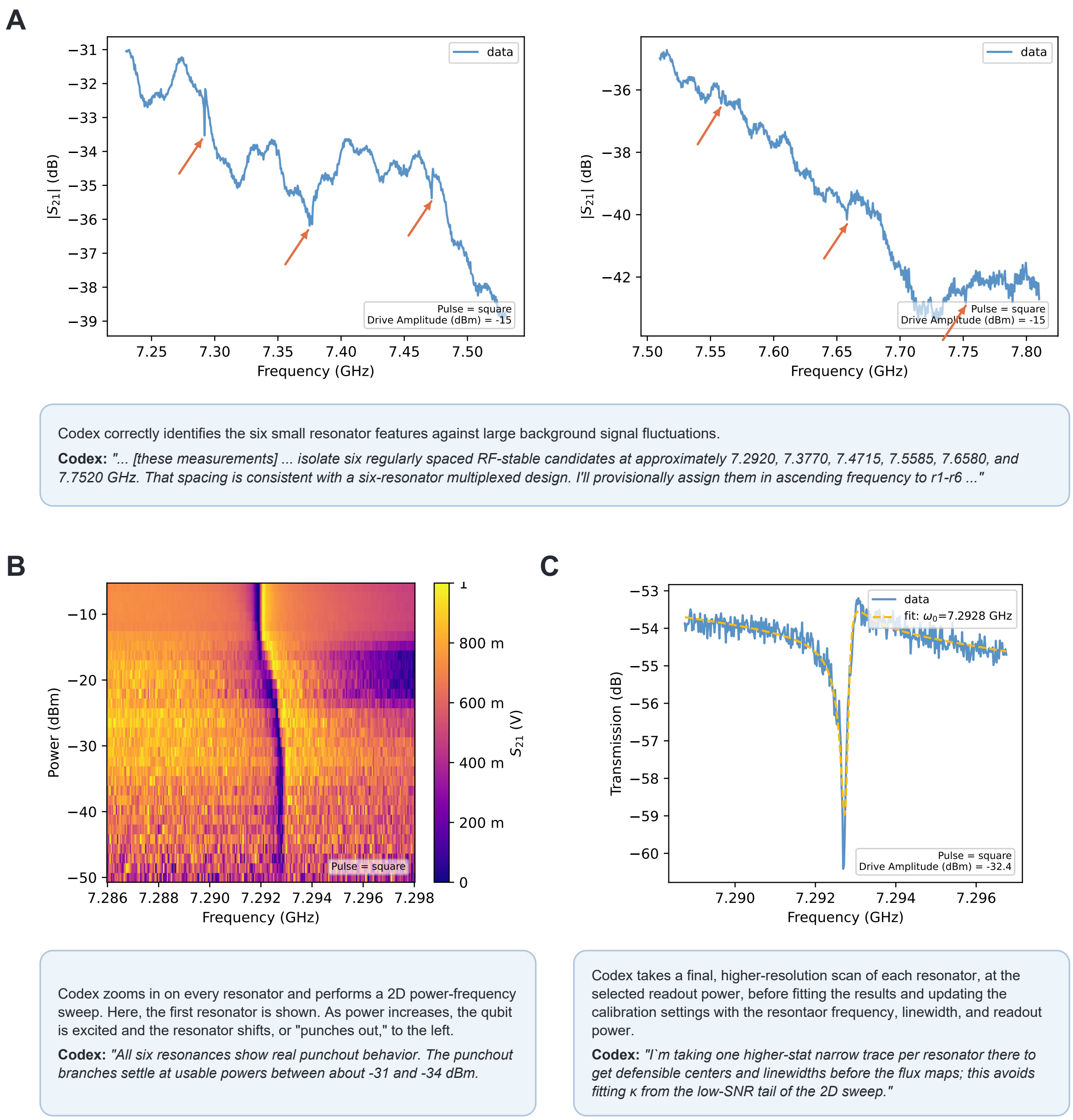

Figure 4: **Resonator calibration,** performed by an agent

After completing resonator spectroscopy, EQuS researchers instructed the agent to perform the remaining target measurements on the fixed-frequency qubits. This standard set includes prerequisite calibrations and measurements of circuit parameters, transition energies, and coherence times, which quantify how long a qubit preserves its state. The figure below shows the final results for the first qubit, obtained fully autonomously by the agent. Across the 40 target measurements for the four fixed-frequency qubits, researchers intervened to improve only four.

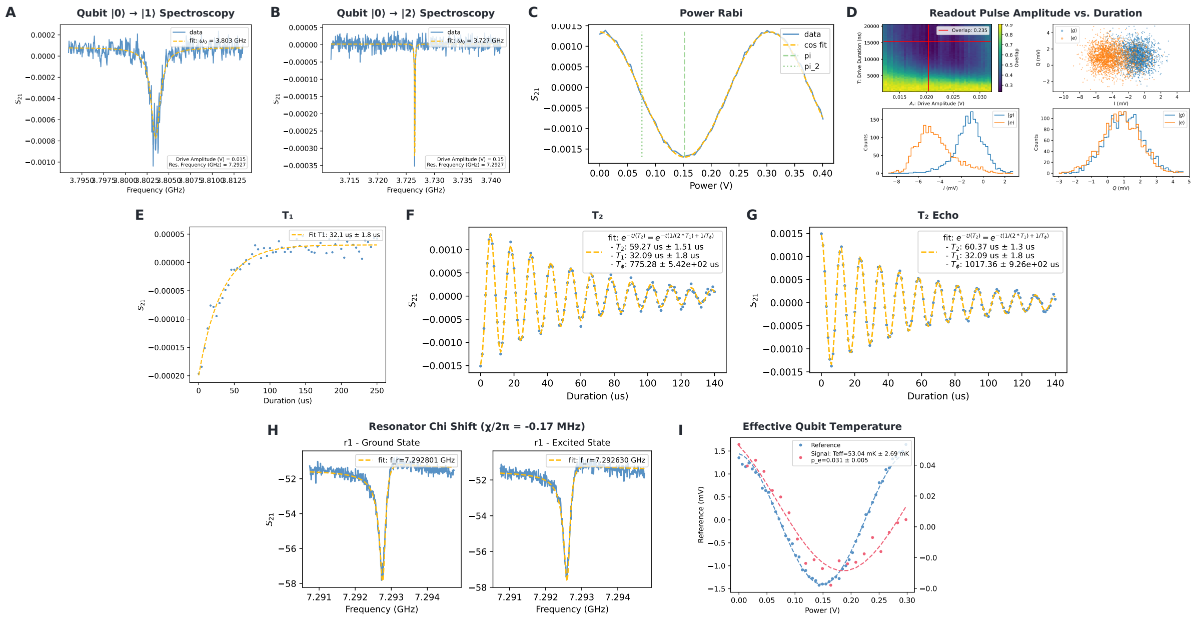

Figure 5: **Completed calibration of qubit 1**. **(a)** Qubit spectroscopy measurement to find the transition energy from the ground state to the first excited state of the qubit. **(b)** Qubit spectroscopy measurement to find the transition energy to the second excited state. **(c)** A Rabi measurement calibrates the pulse power needed to perform an X gate (or "π pulse") that takes the qubit from the $|0\rangle$ to the $|1\rangle$ state. **(d)** Optimizing the readout pulse parameters. **(e, f, g)** Coherence measurements that quantify how long the qubit can keep its state before it decays. **(h)** Precisely measuring the qubit-state-dependent resonator shift (χ). **(i)** Measuring the residual population in the excited state to calculate the effective temperature of the qubit.

The agent had more issues with the frequency-tunable qubit. Tunable qubits are controlled by an applied current, conventionally expressed as magnetic flux. The qubit frequency varies approximately sinusoidally with flux (see Figure 6a). The tunable qubit measurements had a poorer signal-to-noise ratio farther from the maximum frequency, for reasons to do with the chip design. These measurements illustrate how noisy measurements can challenge the agent.

The qubit spectroscopy measurement is intended to find the qubit frequency. A pulse is sent to the qubit, followed by a readout pulse to the resonator. The frequency of the qubit pulse is swept. When it overlaps with the qubit frequency and excites the qubit, the resonator shifts in response, which is reflected in the signal received back from the readout pulse. Repeating this sweep at successive flux values traces the qubit resonance as a moving peak or dip. Although the agent initially located the qubit spectrum, it required significant instruction from the researcher to converge on a satisfactory result. The qubit signal here is faint over much of the flux range, and it takes an experienced eye to visually identify it. The agent resorted to performing additional analysis on individual traces, which seemed to moderately improve its performance.

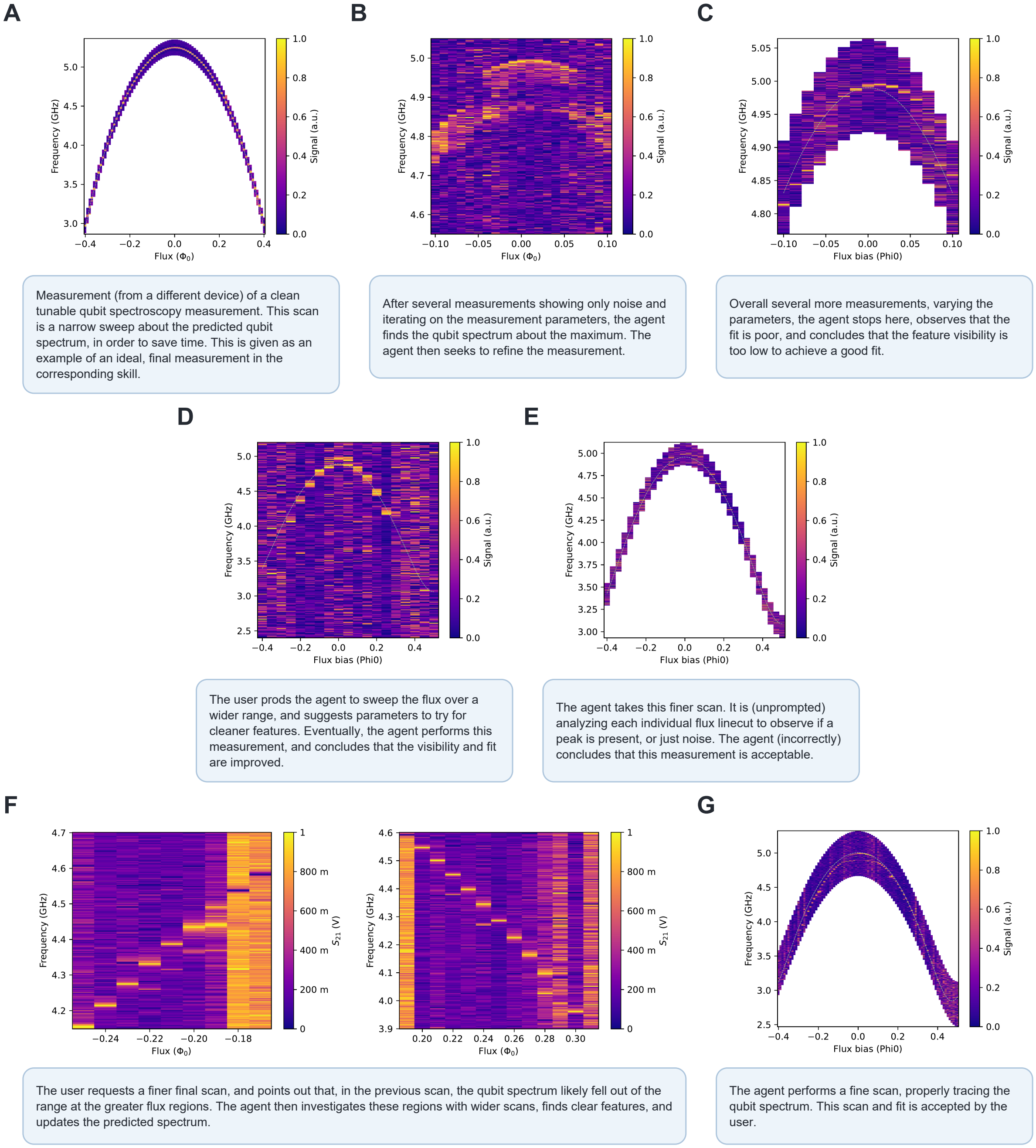

Figure 6: **Selected tunable qubit spectroscopy measurements**, performed by the agent.

The final measurement, accepted by the researcher (Figure 6g), contains unexpected features: an unknown mode crossing near 4.8 GHz, where the second qubit transition also appears, and an unexplained asymmetry that is likely physical in origin. These features exemplify the ambiguity inherent in experiments. Experience is needed to judge whether such effects will matter to the experimental goal and whether their causes can even be identified—a potential pitfall for agentic calibration.

The researcher then instructed the agent to perform the rest of the calibration measurements at the zero flux point, which it did successfully, then use these to seed an automated loop controlled by the orchestrator. This loop steps over a flux range and takes the coherence measurements at each point, programmatically varying the measurement parameters and skipping over failed points. This was successful; the agent monitored the loop over the twelve hours it ran overnight (taking 200 measurements), and then the agent manually investigated a few points where the measurements failed.

5 Implications for superconducting qubit research

The chip presented here is a simple design meant to benchmark the fabrication process or noise environment of the experimental setup. The measurements demonstrated are standard in the field, and very simple compared to those that would be developed for a novel, multi-qubit experiment. An experienced researcher can characterize a set of fixed-frequency qubits in approximately one day, whereas characterizing a set of tunable qubits takes approximately one week. The EQuS group regularly produces a large quantity of these chips and is now routinely using agents to help characterize them.

The agents perform very well when the measurements have a high signal-to-noise ratio, and the qubits physically perform as expected. EQuS researchers have observed that agent performance has improved as models become more advanced - particularly with regards to visual analysis of measurements. Furthermore, elaborating the potential physical failure modes in the measurement-specific skills has proven to be an important strategy to improve agentic measurement performance.

Anecdotally, agents take longer than experienced researchers, even when they successfully converge on the correct measurement parameters. Often, agents appear to lack experimental 'intuition' - pursuing an incorrect line of investigation or not realizing an obvious physical reason why a measurement has failed. Because measurements can take several minutes and run serially, physical acquisition itself is the rate-limiting step; agent swarms therefore cannot accelerate the process through parallel, brute-force exploration. Model fine-tuning in order to confer experimental intuition may be necessary to produce agents that pursue the most efficient path of exploration. The bulk of published literature contains only a tiny fraction of the data acquired in the course of an experiment and mostly consists of the final successful results.

Nevertheless, agents save researcher time by measuring overnight or while researchers work on other tasks. Because multiple chips can be measured simultaneously in one refrigerator—limited primarily by cabling and available control-electronics ports—agents can also increase characterization throughput. Often, many chips of mostly identical design have to be characterized in order to find the best one with which to pursue the full experiment; this can be extremely time consuming.

Furthermore, agents can greatly expedite the experimental process by accelerating associated coding tasks and performing them in tandem. Once a chip has multiple qubits that are coupled to each other, the calibration becomes substantially more complex. The target experiment usually comprises a sequence of simultaneous operations on multiple qubits. The qubits can interact with each other in both desirable and undesirable ways, each of which must be accounted for. A substantial amount of the effort in developing a new calibration process for a novel experiment is in writing the software for multi-qubit control and new analysis. Coding agents already reduce that time significantly, but having agents that can iterate on the code and validate it with real measurements is a major improvement. Furthermore, our experiments can be simulated to a high degree of accuracy. We often perform simulations to plan an experiment, but then discover unpredicted behavior, often caused by a physical factor that was overlooked in the original model. Agents will likely excel at fast iteration between simulation and measurement, to guide the experimental process.

Future work will evaluate how agents operate the measurement process over a longer time scale. Experiments are unpredictable. The flux periodicity and offset of a tunable qubit drift over time, for a variety of environmental factors. The coherence times of qubits regularly vary as well. A microscopic defect on the chip can spontaneously manifest as a mode that interacts with the qubit. Such instabilities may mean that a measurement that worked one day fails the next. Diagnosing such failures may require recalibrating the qubit and revisiting measurements completed dozens of steps earlier. Thus, experimental reasoning must accommodate a drifting system, hidden physical variables, and ambiguous failures.

6 Contributions

MIT authors developed the agentic measurement infrastructure and performed the experiments. OAI authors provided advice on the agentic infrastructure and comments on the text.

References

[1] Shiheng Li, Jacob M. Miller, Phoebe J. Lee, Gustav Andersson, Christopher R. Conner, Yash J. Joshi, Bayan Karimi, Amber M. King, Howard L. Malc, Harsh Mishra, Hong Qiao, Minseok Ryu, Xuntao Wu, Siyuan Xing, Haoxiong Yan, Jian Shi, and Andrew N. Cleland. Large language model-assisted superconducting qubit experiments. *arXiv preprint arXiv:2603.08801*, March 2026. doi: 10.48550/arXiv.2603.08801. URL https://arxiv.org/abs/2603.08801.

[2] Shuxiang Cao, Zijian Zhang, Mohammed Alghadeer, Simone D. Fasciati, Michele Piscitelli, Mustafa Bakr, Peter Leek, and Alán Aspuru-Guzik. Automating quantum computing laboratory experiments with an agent-based AI framework. *Patterns*, 6(10):101372, October 2025. doi: 10.1016/j.patter.2025.101372. URL https://doi.org/10.1016/j.patter.2025.101372.

[3] Shuxiang Cao, Zijian Zhang, Abhishek Agarwal, Grace Bratrud, Niyaz R. Beysengulov, Daniel C. Cole, Alejandro Gómez Frieiro, Elena O. Glen, Hao Hsu, Gang Huang, Raymond Jow, Greshma Shaji, Tom Lubowe, Ligeng Zhu, Luis Mantilla Calderón, Nicola Pancotti, Joel Pendleton, Brandon Severin, Charles Etienne Staub, Sara Sussman, Antti Vepsäläinen, Neel Rajeshbhai Vora, Yilun Xu, Varinia Bernales, Daniel Bowring, Elica Kyoseva, Ivan Rungger, Giulia Semeghini, Sam Stanwyck, Timothy Costa, Alán Aspuru-Guzik, and Krysta Svore. QCalEval: Benchmarking Vision-Language Models for Quantum Calibration Plot Understanding. *arXiv preprint arXiv:2604.25884*, April 2026. doi: 10.48550/arXiv.2604.25884. URL https://arxiv.org/abs/2604.25884.

[4] Huikai Xu, Jiaxiu Han, Shigang Ou, Cheng Ye, Zisong Shen, Jing Gao, Yijia Wang, Tianrui Che, Yu Song, Weiyang Liu, Lei Wang, Lin-Feng Zhang, Pan Zhang, and Hai-Feng Yu. Vibe calibration: Autonomous bring-up of a 112-Qubit superconducting quantum processor by a skill-orchestrating language agent. *arXiv preprint arXiv:2606.22376*, June 2026. doi: 10.48550/arXiv.2606.22376. URL https://arxiv.org/abs/2606.22376.

[5] Philip Krantz, Morten Kjaergaard, Fei Yan, Terry P. Orlando, Simon Gustavsson, and William D. Oliver. A quantum engineer's guide to superconducting qubits. *Applied Physics Reviews*, 6(2):021318, June 2019. doi: 10.1063/1.5089550. URL https://doi.org/10.1063/1.5089550.