

Multi-Agent Reinforcement Learning for Context-Aware Real-Time Error Suppression

Arthur Strauss^{1,2}, Aniket Chatterjee³, Lukas Voss^{1,4}, Tom Dvir², and Hui Khoon Ng¹

¹ Centre for Quantum Technologies, National University of Singapore

² Quantum Machines Ltd., Tel Aviv, Israel

³ Oxford University, United Kingdom

⁴ Ulm University, Germany

Abstract. Quantum processors are limited by coherent errors that depend on the circuit context in which each gate operates, a dependence that conventional calibration protocols do not capture because they characterize gates in isolation. We present CARTES (Context-Aware Real-Time Error Suppression), a framework that applies model-free multi-agent reinforcement learning (MARL) to adapt pulse-level gate parameters to the surrounding circuit context during execution.

Our work features two main contributions:

1. In simulations of a six-qubit superconducting chain that is subject to microwave crosstalk and static ZZ coupling, three RL agents calibrate three parallel two-qubit gates. The agents raise the local average gate fidelity of every gate above 0.9979 and recover 87% to 95% of the gap between the context-agnostic baseline and the numerical contextual optimum.
2. Experimentally, we integrate the MARL workflow directly into the Quantum Orchestration Platform, which reduces communication and compilation latency while it preserves a flexible parametrization of custom calibrations. This integration enables contextual calibration within minutes of wall-clock time, versus the many hours typical of traditional closed-loop methods.

The method requires no extra controller memory and no extra latency, since a compact classical neural network infers the corrections in real time during circuit execution. Because the framework embeds learning into the control stack, it provides context-dependent error suppression at the cost of standard circuit execution, with direct applications in drift compensation and quantum error correction.

Keywords: Error Suppression · Reinforcement Learning · Quantum Control.

Extended Abstract

1 Introduction

As quantum processors approach the error rates required for fault-tolerant computing, further gains depend on the suppression of coherent, circuit-dependent

noise such as classical crosstalk and gate bleedthrough [10]. Conventional calibration relies on context-agnostic characterization, including randomized benchmarking [12, 6, 1], which averages gate performance over randomized circuits and therefore misses errors that arise from the structure of the target computation. Context-aware characterization methods [3, 9] estimate gate fidelity within its circuit environment, yet they diagnose contextual errors rather than correct them. Existing context-dependent calibration approaches, in turn, address static hardware properties such as coupling strengths or frequency collisions [22, 15, 20, 17, 11], leaving circuit-dependent calibration itself largely unexplored.

In this work, we introduce **CARTES**, Context-Aware Real-Time Error Suppression, a framework that calibrates quantum gates directly against their surrounding circuit context. CARTES combines model-free reinforcement learning [18, 19], direct fidelity estimation (DFE) [7], and real-time parameter inference [14, 13] to learn pulse-level corrections without an explicit device model. The corrections are computed during circuit execution on the Quantum Orchestration Platform (QOP) [5] and its core technology known as the Pulse Processing Unit (PPU), so that neither waveform memory nor compilation overhead grows with the number of contexts. We validate CARTES in simulation, where multiple agents calibrate parallel entangling gates on a six-qubit chain under microwave crosstalk, and in experiment, where the workflow runs on hardware at the Israeli Quantum Computing Center (IQCC) with the learning agent embedded in the control stack.

2 Methods and preliminary results

2.1 AI embedded in the quantum control stack

CARTES reformulates gate calibration as an online learning problem. Conventional software stacks associate every native gate with a single pulse implementation, which offline calibration determines in advance; this abstraction simplifies compilation, but it forces coherent circuit-dependent errors to be treated at the circuit level, where dynamical decoupling, randomized compiling, Pauli twirling, or gate insertion trade error suppression against circuit depth, compiled circuit instances, or repeated waveform uploads.

CARTES explores a complementary paradigm. It exploits pulse-level degrees of freedom that modern control systems already expose and adapts the physical implementation of native gates during execution, while the intended logical operation is preserved. For every gate occurrence, pulse parameters must be inferred from the surrounding circuit context such that the resulting gate fidelity is maximal; the objective is therefore a context-conditioned calibration policy that generalizes across a family of circuit configurations, not a single optimal pulse. This execution model relies on three ingredients: (i) runtime pulse parametrization, in which pulse templates keep amplitudes, phases, and durations adjustable after compilation; (ii) embedded real-time execution, in which pulse updates and parameter streaming run directly on the PPU; and (iii) online

data-driven learning, in which pulse parameters are optimized from experimental observations alone.

Model-free reinforcement learning (RL) matches this setting, because it accommodates unknown, nonlinear, and time-varying error processes and relies solely on experimental feedback [18, 2, 8, 14]. In CARTES, RL does not optimize a single pulse calibration offline, but it learns a runtime calibration policy that continuously adapts pulse implementations to the executed circuit. To scale to many gates, CARTES adopts a distributed multi-agent (MARL) formulation: each native two-qubit gate is assigned an independent agent, which observes the local circuit context c and samples pulse-level corrections a from its policy $\pi_\theta(a|c)$. Each agent is optimized with proximal policy optimization (PPO) [16] from a local reward that subsystem-level fidelity estimation provides, so the cost of reward extraction depends on the subsystem of the target gate rather than on the processor size. Figure 2 summarizes the optimization loop.

In practice, the bottleneck of closed-loop calibration is often not quantum execution but the repeated compilation and upload of new circuit implementations: superconducting circuits execute within microseconds, whereas compilation and waveform transfer require seconds [19, 4]. CARTES eliminates this overhead, because optimization is restricted to pulse parameters that remain adjustable at runtime; both policy sampling and dynamic pulse synthesis are embedded within the PPU, and only lightweight parameters or observations cross a low-latency interface to the host. Figure 1 summarizes the resulting paradigm, in which real-time pulse synthesis, embedded machine learning, and distributed reinforcement learning combine into compilation-free context-aware error suppression.

2.2 Compilation-free contextual calibration

To validate the proposed framework, we calibrate three native CNOT gates that execute in parallel on a six-qubit chain, as depicted in Figure 3. Each entangling gate is assigned an independent PPO agent, which generates 13 pulse-level corrections conditioned on its local circuit context. The context is the preceding layer of continuously parametrized single-qubit rotations $RX(\phi_i)$, whose microwave drives leak onto neighboring qubits with amplitudes of up to 9% of the drive angle; in addition, a static ZZ coupling of $2\pi \times 3.5$ kHz acts on the two spectator pairs between the gates. Each agent receives a DFE reward on its own gate subsystem, which the environment amplifies through repeated application of the noisy gate layer (randomized depth $n \in \{1, \dots, 17\}$ per policy update) so that small coherent errors remain resolvable at a fixed measurement budget.

The top panel of Figure 4 shows the learning dynamics. Although the agents train over a continuous family of circuit contexts, the collective reward exceeds 0.9 after 3.2×10^7 measurement shots. Training uses a resolution curriculum, which raises the number of sampled Pauli fiducials required for channel DFE [7] from 400 to 3000 as the policies converge and thereby lowers the total measurement cost by 18% at equal final fidelity. The policies are converged after approximately 3.4×10^8 shots, which corresponds to below one hour of quantum execution time at a repetition period of 10 μ s per shot (assuming active reset).

Because recompilation and waveform uploads are eliminated, this training over varying circuit configurations costs the same runtime as repeated shots of one fixed circuit.

Generalization is evaluated on 1000 held-out circuit contexts, which are drawn from the same distribution but never seen in training. The bottom panel of Figure 4 compares, per gate, the context-agnostic baseline, the learned policy, and the numerical contextual optimum. The learned policies improve every gate: the local average gate fidelity rises from 0.99553 to 0.99924 for gate (q0,q1), from 0.97501 to 0.99810 for gate (q2,q3), and from 0.98336 to 0.99796 for gate (q4,q5), which recovers 87%, 95%, and 90% of the respective gap between baseline and contextual optimum. On the circuit level, the output-state fidelity of the full six-qubit layer rises from 0.96053 to 0.99637 and approximately factorizes into the three local fidelities; the policy outperforms the baseline on 97.1% of the contexts after 7.5×10^7 shots and on 99.7% at convergence. The contextual optimum itself stays below unity, because the static ZZ coupling is not fully correctable within the chosen gate parametrization.

As a representative example of the experimental workflow, we demonstrate in Fig. 5 the calibration of a controlled-Z (CZ) gate implemented on a flux-tunable superconducting quantum processor. The calibration is performed by adjusting in real time three control parameters of a flux pulse—its amplitude and two additional virtual phases. The total wall-clock duration of the procedure is approximately 10 minutes, which is primarily dictated by current experimental constraints, including the latency associated with ethernet-based communication (this particular experiment did not yet employ the dedicated, low-latency link expected to reduce the communication delay to the microsecond scale) and the use of a thermalization-based reset protocol, with a characteristic reset time of approximately 150 μ s per shot (where total number of shots for training was around 350 000).

These results illustrate the central idea behind CARTES: native gate implementations become adaptive runtime objects, which are generated directly from the circuit context, instead of entries in an exponentially growing library of static pulse calibrations. The same architecture extends to larger processors, richer context representations, and cooperative multi-agent strategies in which neighboring adaptive gates jointly compensate coherent interactions across the device. Further, the model-free character of the approach qualifies it as a generic calibration subroutine for circuit contexts that require feedback, such as quantum error correction and fault-tolerant circuits [21].

References

- [1] Frank Arute et al. “Quantum supremacy using a programmable superconducting processor”. en. In: *Nature* 574.7779 (Oct. 2019). Number: 7779, pp. 505–510. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1666-5. URL: <https://www.nature.com/articles/s41586-019-1666-5> (visited on 01/28/2021).

CARTES: Context-Aware Real-Time Error Suppression

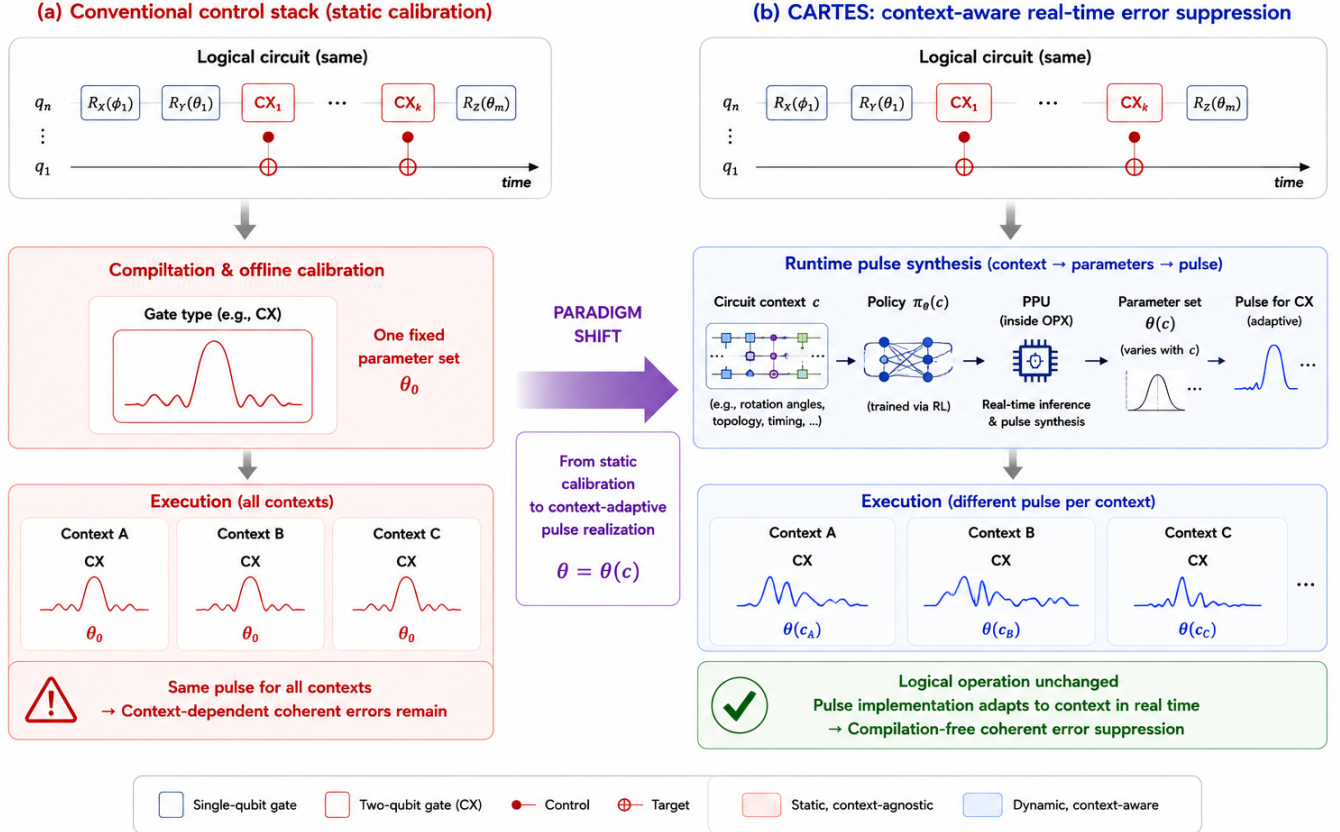


Fig. 1. CARTES conceptual paradigm shift: the intended logical gate is disentangled from its pulse-level implementation. With real-time classical processing and pulse-level control, the traditional gate model is extended so that a gate definition carries a pulse set, which is adjusted dynamically to compensate circuit-dependent noise, for example as a function of gate parameters, layer structure, or wall-clock time of execution.

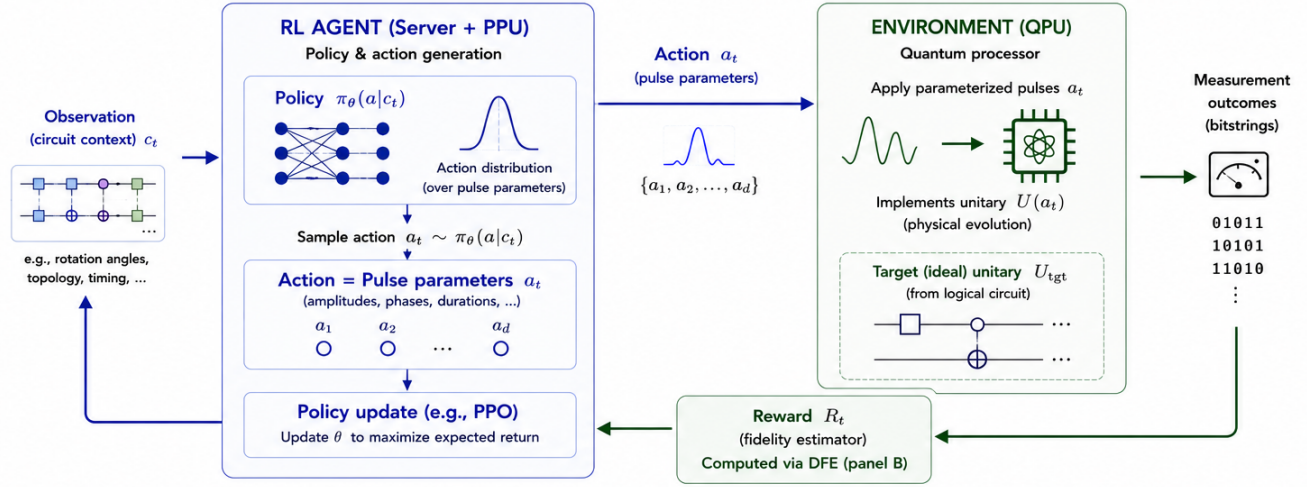
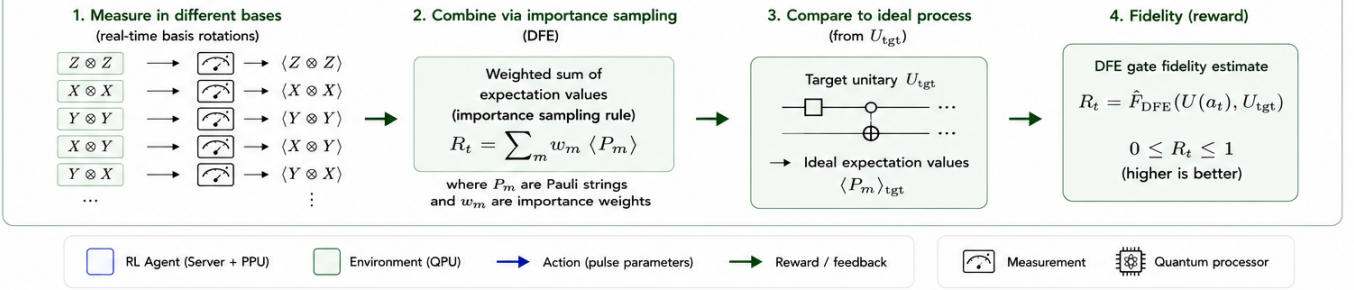
Figure 2. CARTES RL workflow: from context to fidelity reward**A. Closed-loop RL workflow****B. Reward: fidelity estimation via DFE**

Fig. 2. Online context-aware workflow for error suppression. The circuit context, which corresponds to the observation, is the representation input to the agent; the agent network generates a probability distribution over pulse-level corrections, the policy. The reward signal R is extracted from the experiment through post-processing of measurement outcomes, which builds a fidelity estimator.

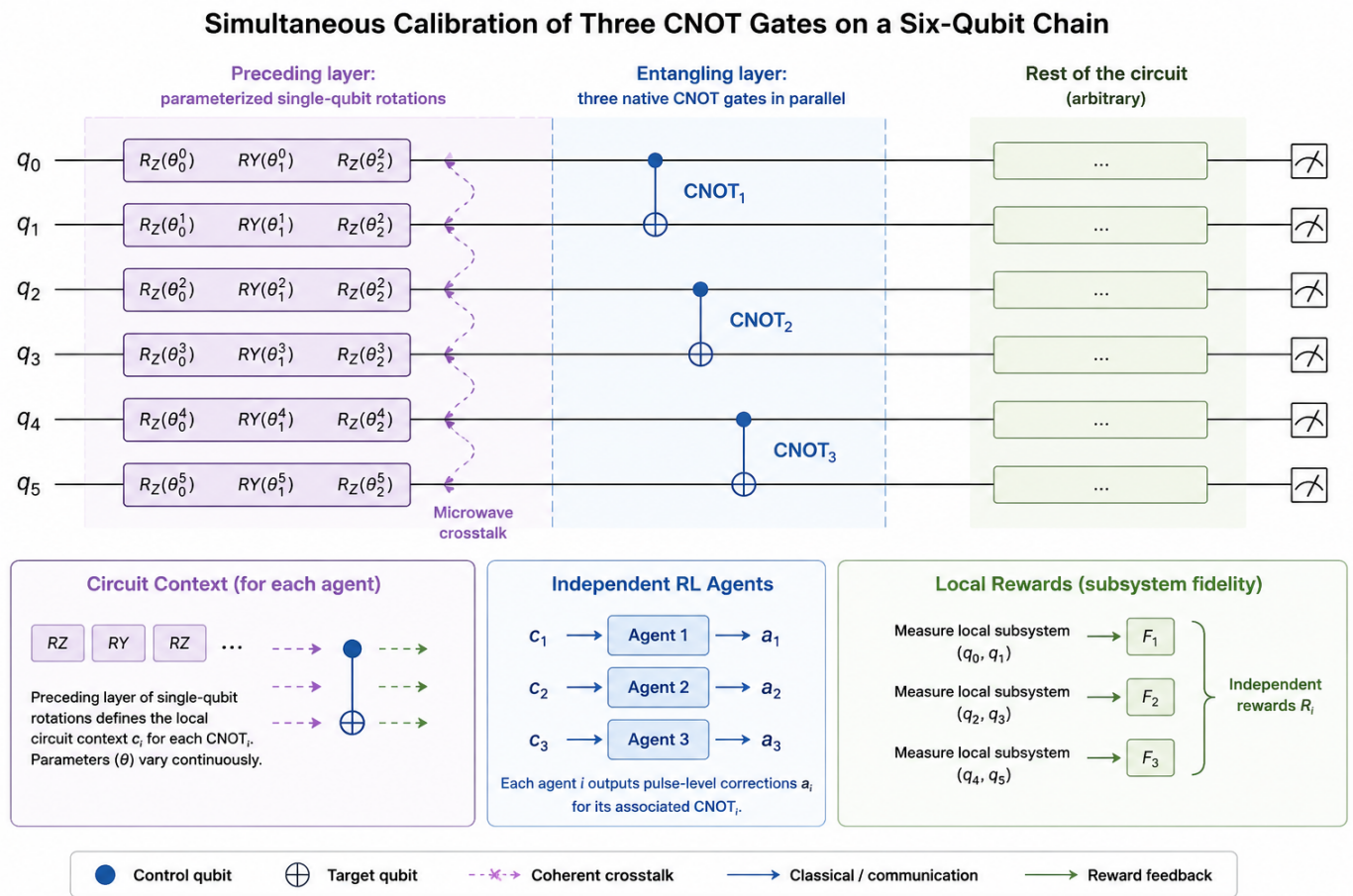


Fig. 3. Use case: compensation of microwave crosstalk on a six-qubit chain. The CNOT gates, which a traditional stack would treat as fixed, carry dynamic pulse-level adjustments that compensate the noise induced by the rotation angles in the preceding single-qubit layer.

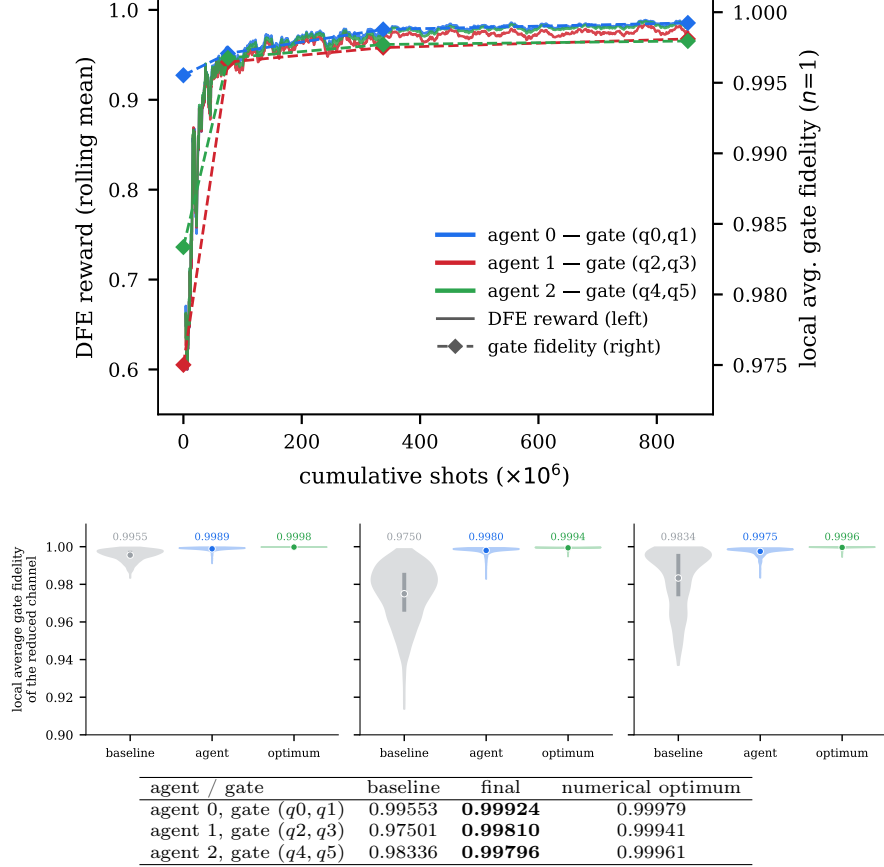


Fig. 4. Simulation results for the joint calibration of three parallel CNOT gates. Top: training dynamics versus the cumulative number of measurement shots. Solid curves show the rolling-mean DFE reward of each agent, which is measured at a randomized amplification depth $n \in \{1, \dots, 17\}$ per policy update; dashed curves show the local average gate fidelity of the corresponding gate per cycle ($n=1$), and each dashed series starts at its context-agnostic baseline. Bottom: validation on 1000 held-out circuit contexts, which compares the context-agnostic baseline, the learned policy, and the numerical contextual optimum for each gate; the table lists the corresponding mean fidelities. The optimum stays below unity, because the simulations include static ZZ crosstalk that the gate parametrization cannot fully correct; the residual distance between policy and optimum is a learning gap.

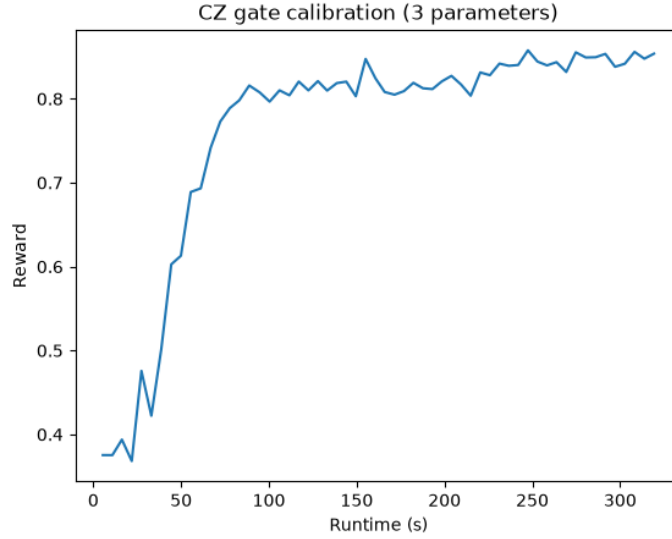


Fig. 5. Experimental real-time calibration on Israeli Quantum Computing Center for a CZ gate on flux-tunable transmon qubit. Latency is constrained by thermal reset, but is much lower than standard closed loop calibrations.

- [2] Yuval Baum et al. “Experimental Deep Reinforcement Learning for Error-Robust Gate-Set Design on a Superconducting Quantum Computer”. In: *PRX Quantum* 2.4 (Nov. 2021), p. 040324. DOI: 10.1103/PRXQuantum.2.040324. URL: <https://link.aps.org/doi/10.1103/PRXQuantum.2.040324> (visited on 02/09/2023).
- [3] Dripto M. Debroy et al. “Context-aware fidelity estimation”. In: *Physical Review Research* 5.4 (Dec. 2023), p. 043202. DOI: 10.1103/PhysRevResearch.5.043202. URL: <https://link.aps.org/doi/10.1103/PhysRevResearch.5.043202> (visited on 04/10/2024).
- [4] Leon Ding et al. *High-Fidelity, Frequency-Flexible Two-Qubit Fluxonium Gates with a Transmon Coupler*. arXiv:2304.06087 [quant-ph]. Apr. 2023. URL: <http://arxiv.org/abs/2304.06087> (visited on 04/20/2023).
- [5] Lior Ella et al. *Quantum-classical processing and benchmarking at the pulse-level*. arXiv:2303.03816 [quant-ph]. Mar. 2023. DOI: 10.48550/arXiv.2303.03816. URL: <http://arxiv.org/abs/2303.03816> (visited on 09/07/2023).
- [6] Alexander Erhard et al. “Characterizing large-scale quantum computers via cycle benchmarking”. en. In: *Nature Communications* 10.1 (Nov. 2019), p. 5347. ISSN: 2041-1723. DOI: 10.1038/s41467-019-13068-7. URL: <https://www.nature.com/articles/s41467-019-13068-7> (visited on 11/18/2025).

- [7] Steven T. Flammia and Yi-Kai Liu. “Direct Fidelity Estimation from Few Pauli Measurements”. In: *Physical Review Letters* 106.23 (June 2011), p. 230501. DOI: 10.1103/PhysRevLett.106.230501. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.106.230501> (visited on 09/07/2022).
- [8] Thomas Fösel et al. “Reinforcement Learning with Neural Networks for Quantum Feedback”. In: *Physical Review X* 8.3 (Sept. 2018), p. 031084. DOI: 10.1103/PhysRevX.8.031084. URL: <https://link.aps.org/doi/10.1103/PhysRevX.8.031084> (visited on 02/09/2023).
- [9] Jonathan A. Gross et al. *Characterizing Coherent Errors using Matrix-Element Amplification*. arXiv:2404.12550 [quant-ph]. Apr. 2024. DOI: 10.48550/arXiv.2404.12550. URL: <http://arxiv.org/abs/2404.12550> (visited on 05/20/2024).
- [10] J. Kelly et al. “Optimal quantum control using randomized benchmarking”. In: *Physical Review Letters* 112.24 (June 2014). arXiv:1403.0035 [cond-mat, physics:quant-ph], p. 240504. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.112.240504. URL: <http://arxiv.org/abs/1403.0035> (visited on 02/09/2023).
- [11] Paul V. Klimov et al. “Optimizing quantum gates towards the scale of logical qubits”. en. In: *Nature Communications* 15.1 (Mar. 2024), p. 2442. ISSN: 2041-1723. DOI: 10.1038/s41467-024-46623-y. URL: <https://www.nature.com/articles/s41467-024-46623-y> (visited on 05/20/2024).
- [12] Easwar Magesan et al. “Robust randomized benchmarking of quantum processes”. In: *Physical Review Letters* 106.18 (May 2011). arXiv: 1009.3639, p. 180504. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.106.180504. URL: <http://arxiv.org/abs/1009.3639> (visited on 08/18/2021).
- [13] Malthe A. Marciniak et al. *Millisecond-Scale Calibration and Benchmarking of Superconducting Qubits*. arXiv:2602.11912 [quant-ph]. Feb. 2026. DOI: 10.48550/arXiv.2602.11912. URL: <http://arxiv.org/abs/2602.11912> (visited on 03/20/2026).
- [14] Kevin Reuer et al. “Realizing a deep reinforcement learning agent for real-time quantum feedback”. en. In: *Nature Communications* 14.1 (Nov. 2023), p. 7138. ISSN: 2041-1723. DOI: 10.1038/s41467-023-42901-3. URL: <https://www.nature.com/articles/s41467-023-42901-3> (visited on 05/20/2024).
- [15] Kenneth Rudinger et al. “Experimental Characterization of Crosstalk Errors with Simultaneous Gate Set Tomography”. In: *PRX Quantum* 2.4 (Nov. 2021), p. 040338. DOI: 10.1103/PRXQuantum.2.040338. URL: <https://link.aps.org/doi/10.1103/PRXQuantum.2.040338> (visited on 05/21/2023).
- [16] John Schulman et al. *Proximal Policy Optimization Algorithms*. arXiv:1707.06347 [cs]. Aug. 2017. URL: <http://arxiv.org/abs/1707.06347> (visited on 08/31/2022).

- [17] Alireza Seif et al. *Suppressing Correlated Noise in Quantum Computers via Context-Aware Compiling*. arXiv:2403.06852 [quant-ph] version: 1. Mar. 2024. URL: <http://arxiv.org/abs/2403.06852> (visited on 03/17/2024).
- [18] V. V. Sivak et al. “Model-Free Quantum Control with Reinforcement Learning”. In: *Physical Review X* 12.1 (Mar. 2022), p. 011059. DOI: 10.1103/PhysRevX.12.011059. URL: <https://link.aps.org/doi/10.1103/PhysRevX.12.011059> (visited on 07/04/2022).
- [19] V. V. Sivak et al. *Real-time quantum error correction beyond break-even*. arXiv:2211.09116 [quant-ph]. Nov. 2022. DOI: 10.48550/arXiv.2211.09116. URL: <http://arxiv.org/abs/2211.09116> (visited on 02/09/2023).
- [20] Vinay Tripathi et al. “Suppression of Crosstalk in Superconducting Qubits Using Dynamical Decoupling”. In: *Physical Review Applied* 18.2 (Aug. 2022), p. 024068. DOI: 10.1103/PhysRevApplied.18.024068. URL: <https://link.aps.org/doi/10.1103/PhysRevApplied.18.024068> (visited on 02/09/2023).
- [21] Thomas Wagner et al. “Learning Logical Pauli Noise in Quantum Error Correction”. In: *Physical Review Letters* 130.20 (May 2023), p. 200601. DOI: 10.1103/PhysRevLett.130.200601. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.130.200601> (visited on 05/20/2024).
- [22] Gregory A. L. White et al. *Unifying non-Markovian characterisation with an efficient and self-consistent framework*. arXiv:2312.08454 [quant-ph]. Dec. 2023. DOI: 10.48550/arXiv.2312.08454. URL: <http://arxiv.org/abs/2312.08454> (visited on 05/20/2024).