

Empowering Quantum Error Traceability with MoE for Automatic Calibration

1st Tingting Li
Zhejiang University
Hangzhou, China
litt2020@zju.edu.cn

2nd Ziming Zhao
Zhejiang University
Hangzhou, China
zhaoziming@zju.edu.cn

3rd Liqiang Lu^{*}
Zhejiang University
Hangzhou, China
liqianglu@zju.edu.cn

4th Siwei Tan
Zhejiang University
Hangzhou, China
siweitan@zju.edu.cn

5th Jianwei Yin^{*}
Zhejiang University
Hangzhou, China
zjuyjw@cs.zju.edu.cn

Abstract—Quantum computing offers the potential for exponential speedups over classical computing in tackling complex tasks, such as large-number factorization and chemical molecular simulation. However, quantum noise remains a significant challenge, hindering the reliability and scalability of quantum systems. Therefore, effective characterization and calibration of quantum noise are critical to advancing these systems. Quantum calibration is a process that heavily relies on expert knowledge, and there currently is a range of research focused on automatic calibration. However, traditional calibration methods often need an effective error traceback mechanism, leading to repeated calibration attempts without identifying root causes. To address the issue of error traceback in calibration failures, this paper proposes an automatic calibration error traceback algorithm facilitated by a Mixture of Experts (MoE) system inspired by the current large language model technologies. Our approach enables traceability of quantum calibration errors, allowing for the rapid identification and correction of deviations from the calibration state. Extensive experimental results demonstrate that the MoE-based automatic calibration method significantly outperforms traditional error traceability and calibration efficiency techniques. Notably, our approach improved the average visibility of 77 qubits by 25.5%, surpassing the outcomes of fixed calibration processes. This work presents a promising path toward more reliable and scalable quantum computing systems.

Index Terms—Quantum computing, quantum calibration, calibration error traceability, mixture of Experts (MoE).

I. INTRODUCTION

Quantum computing [1] promises exponential speedups in complex tasks such as large-number factorization [2] and chemical molecular simulation [3]. In December 2023, IBM released a “Condor” quantum chip with 1,121 superconducting qubits [4], showcasing the potential of quantum computing to solve problems that are intractable for classical computers. However, quantum noise remains a significant barrier to the reliability and scalability of quantum systems [5]. To achieve high-fidelity quantum computation, frequent and complex quantum calibrations are required, *e.g.*, IBM Quantum performs calibrations hourly and daily [6]. Effective noise characterization and calibration of quantum noise are critical to advancing these systems. As shown in Figure 1, the calibration process of a quantum computer consists of interdependent operations that can be represented as a Directed Acyclic Graph (DAG) representation [7], which involves multiple parameter

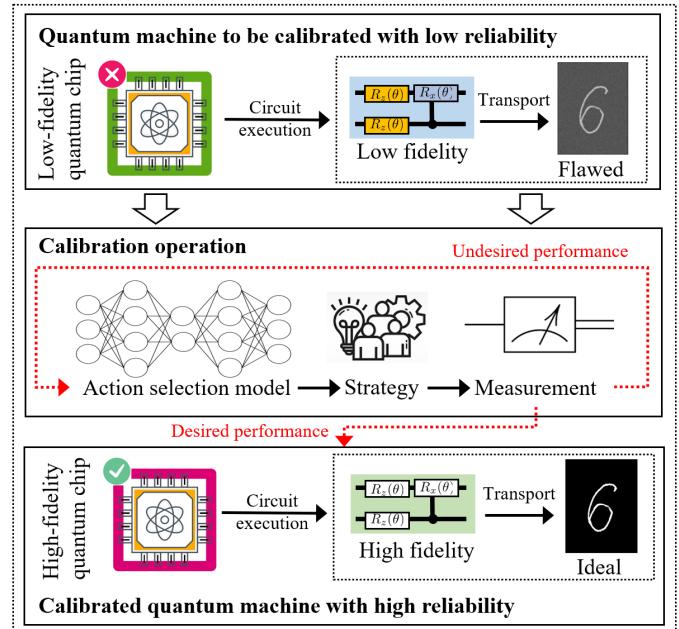


Fig. 1: Illustrative explanation of automatic calibration.

settings, such as qubit frequency, coupler frequency, and so on. Due to the susceptibility of quantum devices to various noise sources [8], such as hardware operation defects and environmental coupling [9], [10], [11], quantum calibration heavily relies on expert knowledge. Recent advancements include fast Bayesian optimization methods for efficient parameter space search [12], real-time calibration techniques for dynamic parameter adjustment [13], and high-speed processor calibration strategies that reduce overhead and enhance stability [14]. However, current calibration protocols are often based on fixed calibration processes that are designed using expert knowledge. When a calibration error is encountered, it is usually simply repeated to execute the calibration operation [7], which lack an effective error traceback mechanism [7]. Therefore, it is crucial to develop an adaptive and intelligent calibration system, inferring the possible causes of calibration errors, and being able to perform error backtracking and adjust calibration control parameters autonomously.

Addressing error traceback in calibration processes, especially in the face of the intricate quantum noise, is a formidable task. In this paper, we design an automatic error traceback

*Corresponding author.

algorithm that leverages the unique capabilities of a Mixture of Experts (MoE) system [15]. The MoE's distinctiveness lies in its dynamic allocation of expert networks, a feature that optimizes performance across varying noise patterns. This dynamic allocation enhances both accuracy and efficiency in identifying and correcting calibration issues, thereby enabling rapid error traceability and significantly advancing the potential of quantum computing. This paper aims to investigate the application of the MoE calibration system for the extraction of characteristic data features inherent to the quantum calibration process. In summary, this paper makes three key contributions:

- We utilize the MoE architecture to identify erroneous calibration operations for error traceback and then employ the integration policies to analyze the calibration data features, reselecting parameters to execute the error calibration operation.
- We design a series of adjustment expert policies and tailored operations for different actions that integrate measurement information.
- We conduct experiments on physical chips, and our results demonstrate that the proposed method can achieve efficient automatic calibration.

II. BACKGROUND AND RELATED WORK

A. Quantum Calibration

Quantum calibration is vital for enhancing quantum device performance, particularly in the Noisy Intermediate-Scale Quantum (NISQ) era where noise sensitivity demands precise control [16], [5], [17]. Calibration accuracy, including qubit frequency, amplitude, and phase adjustments, significantly affects quantum operation stability [18]. There are recent research works to optimize quantum calibration [18], *e.g.*, method based on Bayesian optimization [12], FPGA-based automatic calibration protocol [19], or focus on high efficiency [13] and low overhead [14]. Despite the calibration system introducing expert knowledge, when calibration failures or unexpected deviations occurred, traditional methods often repeating execute calibration operations [7], this trial-and-error method lacks a robust mechanism for addressing calibration failures and lacks a calibration backtracking mechanism, limiting the efficiency and effectiveness of the calibration process. In this context, there is a critical need to develop calibration methodologies that deeply integrate expert knowledge while incorporating mechanisms for backtracking and recalibration in the event of calibration failures or discrepancies in performance.

B. Mixture of Experts (MoE)

The Mixture of Experts (MoE) model is a powerful machine learning paradigm used in various domains, including generative AI [20], [21] and bioinformatics [22], [23]. At its core, MoE is an ensemble learning technique that combines expert models to enhance accuracy and robustness compared to individual models [24]. In automatic calibration systems for quantum devices, the MoE model presents a promising approach due to its ability to handle complex, non-linear behaviors inherent in quantum systems [7], [25].

However, integrating MoE into quantum calibration faces several technical challenges: (i) *Lack MoE deployment experience to quantum calibration*: There is no existing precedent for applying MoE to quantum calibration, necessitating the development of innovative methodologies tailored to the unique characteristics of quantum systems. (ii) *Abstracting the complex quantum calibration system*: Formulating the MoE system presents another technical hurdle, as defining the mixture components and expert models requires careful consideration of the specific goals and requirements of quantum calibration. (iii) *Calibration parameters determination*: Updating the parameters of quantum calibration with a large number of complicated parameter relations presents a formidable challenge, as the optimization process must navigate a high-dimensional parameter space while satisfying various constraints and objectives. This paper delves into these challenges in detail and proposes potential strategies and methodologies for addressing them, paving the way for future research in this emerging field.

III. PROBLEM SPACE AND FORMULATION

System Model and Goals. We define a system as the tuple $S_{\mathcal{M}} = (\mathcal{Q}, \mathcal{O})$, where \mathcal{Q} denotes the quantum computer to be calibrated and \mathcal{O} represents the set of operations that can be executed. Specifically, for Machine \mathcal{Q} , we can analyze current device reliability based on quantum state measurements, denoted as performance \mathcal{P} . Particularly, the calibration performance of quantum machines encompasses parameters such as signal-to-noise ratio (SNR), sigma values, state error, separation error, visibility, and separation. If the performance of quantum machines does not meet our desired requirements (*i.e.*, threshold), we need to perform appropriate calibration operations to improve device reliability. Subsequently, we introduce candidate calibration operations that can be performed.

Calibration Operations. For quantum device calibration, several operations ensure accurate and reliable performance. This procedure begins with *readout frequency determination*, establishing the qubit's resonance frequency for interaction with the resonant cavity. Next, *readout power relationship*, involves scanning the $|S_{21}|$ 2D planar graph to correlate readout frequency with power levels, identifying the optimal readout power spectrum. The *low-speed Z bias relationship* is then determined by the application of a sustained frequency bias, utilizing the $|S_{21}|$ measurement to assess the qubit's response to slow Z variations. The *high-speed Z bias relationship* is calibrated using a Digital-to-Analog Converter (DAC) board that delivers precise frequency modulation to the qubit. Following this, a *qubit resonance frequency spectrum scan* is executed, employing a π pulse to transition the qubit from the $|0\rangle$ to the $|1\rangle$ state, thus pinpointing the qubit's resonance frequency. *Rabi oscillation* is then leveraged to identify the precise π pulse characteristics necessary for qubit excitation. Finally, *Ramsey experiment* is conducted to measure the qubit's coherence time, crucial for mitigating decoherence effects and enhancing quantum operational reliability. More details on the calibration operation will be elaborated in § IV-C.

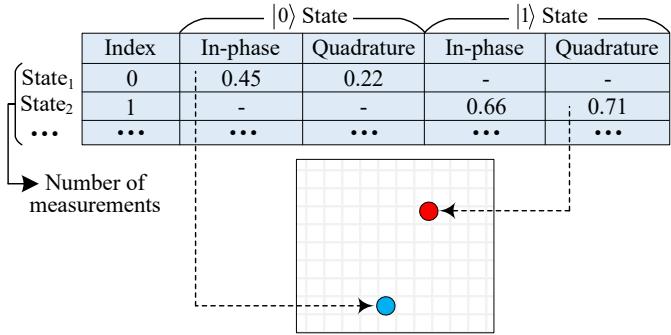


Fig. 2: Illustrative explanation of quantum state reliability characterization data.

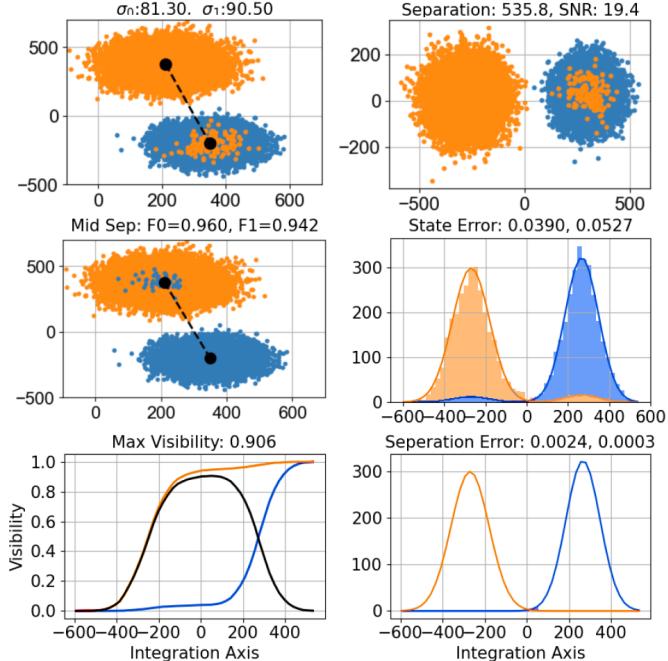


Fig. 3: Reliability assessment metrics for calibration.

IV. DESIGN DETAILS

Firstly, we explore the utilization of an MoE model to autonomously identify and revert to the essential calibration procedures that demand refinement. Subsequently, we delve into meticulously designed calibration strategies aimed at refining the parameters of the aforementioned calibration actions. This systematic approach enhances the quality of calibration results, thereby optimizing the operational efficiency of quantum computing systems.

A. Quantum State Reliability Characterization

As shown in Figure 2, for the reliability characterization, we store the quantum state measured data and further construct the IQ plane (In-phase and Quadrature phase components of the signal) [26], [27]. The qubits are prepared to either the $|0\rangle$ state or the $|1\rangle$ state and measured T times respectively, and the coordinate points of their distribution on the IQ plane are recorded, with blue color indicating the $|0\rangle$ state and red color indicating the $|1\rangle$ state. By plotting the distribution of these points on the IQ plane, and further by clustering analysis,

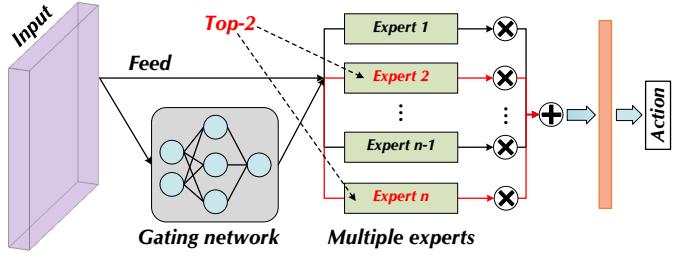


Fig. 4: The selection model of erroneous calibration operation.

the $|0\rangle$ and $|1\rangle$ states are projected onto the circle center. The line connecting the centers of the two states is used as the x -axis for projection, showing a Gaussian distribution. Further, the two states are integrated and the difference is calculated, and the maximum difference is the maximum distinguishable line, as shown in Figure 3. We select 9 calibration metrics to characterize calibration performance of each qubit in detail, including Signal-to-Noise Ratio (SNR), separation, visibility, σ_0 and σ_1 , separation error se_0 and se_1 , state preparation error pe_0 and pe_1 .

B. Calibration Action Selection Model

We leverage the Mixture of Experts (MoE) model [15] for action selection, enabling us to address the requirements of multiple performance metrics and thereby integrate the outcomes of various experts to determine the calibration actions.

The Mixture-of-Experts Layer. The MoE layer is composed of a set of n individual “expert networks,” denoted as E_1, \dots, E_n , along with a “gating network” G that generates a sparse vector of size n . For a visual summary of the MoE structure, refer to Figure 4. Each expert within the MoE is a neural network equipped with its own set of parameters. Essentially, the requirement for these experts is that they must be able to accept inputs of a consistent size and produce outputs of the same dimensions.

Let us denote by $G(x)$ and $E_i(x)$ the output of the gating network and the output of the i -th expert network for a given input x . The output y of the MoE module can be written as:

$$y = \sum_{i=1}^n G_i(x)E_i(x) \quad (1)$$

We leverage the sparsity of the output of $G(x)$ to conserve computational resources by avoiding the computation of $E_i(x)$ whenever $G_i(x)$ is zero. If the number of experts becomes excessively large, we can reduce complexity by implementing a two-tiered hierarchical MoE structure. In this structure, a primary gating network selects a sparse, weighted aggregation of “experts,” each of which consists of a secondary MoE with its own gating network. Our approach resembles other models of conditional computation. An MoE with experts as simple weight matrices parallels the parameterized weight matrix concept from prior research [28]. Additionally, an MoE with experts comprising a single hidden layer resembles the block-wise dropout technique described in [29], where layers with dropped-out units are flanked by fully activated layers.

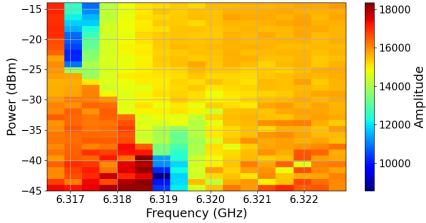


Fig. 5: Qubit readout power & qubit readout frequency.

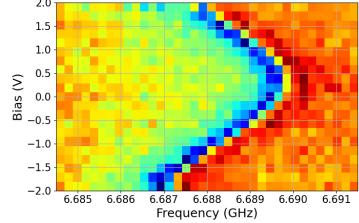


Fig. 6: Low-speed Z control & qubit readout frequency.

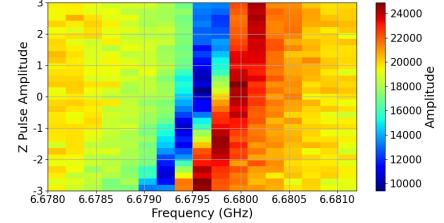


Fig. 7: High-speed Z control & qubit readout frequency.

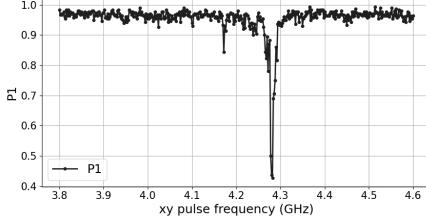


Fig. 8: Qubit resonance frequency of π pulse curve.

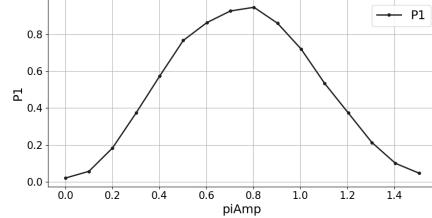


Fig. 9: Qubit resonance amplitude of π pulse curve.

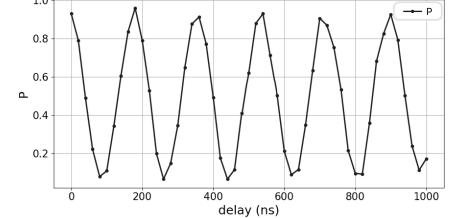


Fig. 10: Qubit decoherence characterization.

Gating Network. A straightforward selection for a non-sparse gating mechanism [30] entails the multiplication of the input by a trainable weight matrix W_g , followed by the application of the Softmax function: $G_\sigma(x) = \text{Softmax}(x \cdot W_g)$. For the noisy Top-K Gating, we can add two components to the Softmax gating network, *i.e.*, sparsity and noise. Before the application of the Softmax function, we introduce tunable Gaussian noise to the system and then retain only the top k values, designating the remainder to $-\infty$. This assignment results in the corresponding gate values becoming zero. The implementation of sparsity is a strategic move aimed at computational efficiency, as previously discussed. The noise term helps with load balancing, the amount of noise per component is controlled by a second trainable weight W_{noise} :

$$G(x) = \text{Softmax}(\text{KeepTopK}(H(x), k)) \quad (2)$$

$$H(x)_i = (x \cdot W_g)_i + \text{StandardNormal()} \cdot \text{Softplus}((x \cdot W_{\text{noise}})_i) \quad (3)$$

$$\text{KeepTopK}(v, k)_i = \begin{cases} v_i & \text{if } v_i \text{ is in the top } k \text{ elements} \\ -\infty & \text{otherwise} \end{cases} \quad (4)$$

For training the gating network, we train the gating network by simple back-propagation, along with the rest of the model. If we choose $k > 1$, the gate values for the top k experts have nonzero derivatives with respect to the weights of the gating network. This type of occasionally-sensitive behavior is described in [31] with respect to noisy rectifiers. Gradients also backpropagate through the gating network to its inputs. Ultimately, the trained model is utilized for inference to obtain the selected calibrated actions.

C. Adjust Strategy

The calibration procedure for bringing up individual qubits on a quantum chip involves several key steps. It begins with determining the readout frequency by measuring the $|S_{21}|$

amplitude-frequency curve through flux bias, identifying the qubit readout frequency based on the design frequency. Next, the relationship between readout frequency and power is analyzed via a 2D scan of the $|S_{21}|$ graph to establish the effective power range, avoiding excessive power that could decouple the qubits from their resonant cavities. Then, a long-term constant frequency bias is applied to measure the qubit's readout frequency and low-speed Z variations, fitting the data to understand their relationship. Following this, a high-speed Z bias is controlled through a DAC, testing its effects after initially setting a low-speed bias. A π pulse is then used to drive the qubit from the $|0\rangle$ to the $|1\rangle$ state, allowing for the identification of the resonance frequency where the probability of the $|1\rangle$ state is maximized. The amplitude and duration of the π pulse are measured by applying a fixed frequency microwave signal and adjusting its amplitude, with the maximum probability of the $|1\rangle$ state indicating the effective pulse strength. Finally, the Ramsey experiment employs two $\pi/2$ pulses separated by a time interval to measure decoherence (T_2) and refine the qubit frequency, addressing decoherence challenges and enhancing the fidelity of quantum operations for improved performance in quantum applications.

D. Calibration Workflow

The overall calibration workflow is shown in Algorithm 1. Specifically, the calibration process for a quantum machine consists of several steps. It begins with initialization, setting the machine parameters, followed by pre-operations that determine the readout frequency and establish a temporary calibration notation. Next, quantum state characterization assesses the reliability of the quantum state before each iteration. The MoE model is then used to select the most suitable action based on this characterization, which is translated into a concrete adjustment strategy for the calibration operation. After performing the operation, the machine's performance is evaluated to determine if it meets the desired threshold. If

TABLE I: Average calibration performance evaluation of 77 qubit tests.

Calib.	Separation \uparrow	SNR \uparrow	Sigma_0 \downarrow	Sigma_1 \downarrow	Sep_err_0 \downarrow	Sep_err_1 \downarrow	State_err_0 \downarrow	State_err_1 \downarrow	Visibility \uparrow
Base	522.443	14.294	155.211	148.693	0.062	0.033	0.049	0.260	0.597
Ours	734.705	33.105	104.134	115.178	0.015	0.010	0.025	0.098	0.852

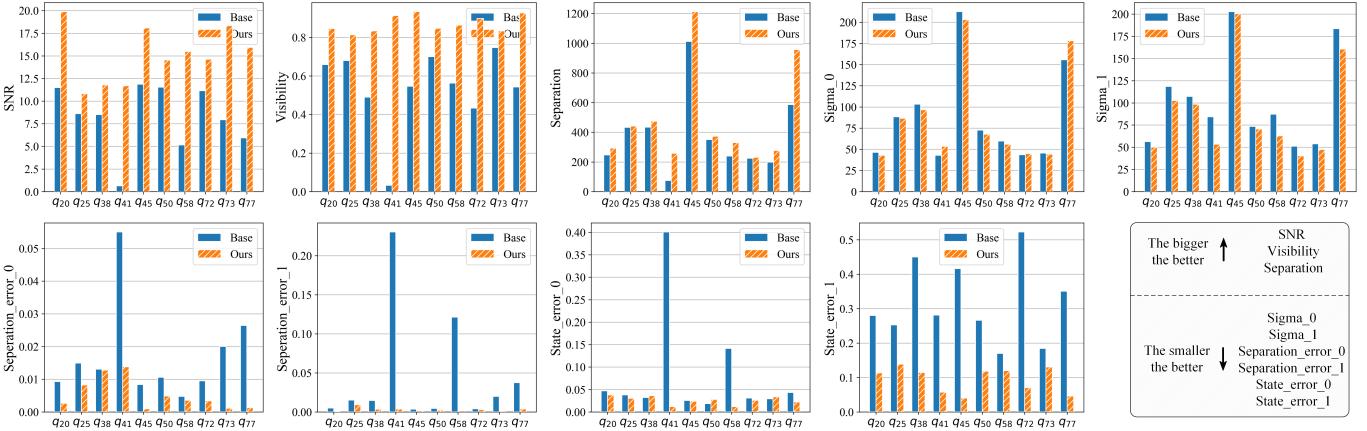


Fig. 11: Select 10 test qubits for detailed comparisons for 9 calibration metrics.

Algorithm 1 The calibration pipeline

Input: The quantum machine Q_{init} to be calibrated, the trained MoE model \mathcal{M} , the adjustment strategy \mathcal{S} , and the desired performance threshold \mathcal{T}

Output: Calibrated quantum machine Q_{calib} , and calibration characterization performance \mathcal{P}_{calib}

- 1: Initialize quantum machine parameters
- 2: # Perform pre-operations
- 3: Readout frequency determination
- 4: Calculate initial performance \mathcal{P}_{init}
- 5: Temporary calibration notation $\mathcal{Q}' = Q_{init}$, $\mathcal{P}' = \mathcal{P}_{init}$
- 6: **while** $\mathcal{P}' < \mathcal{T}$ **do**
- 7: # Quantum state reliability characterization (§ IV-A)
- 8: Calculated to obtain *input*
- 9: # Action selection (§ IV-B)
- 10: $A_{sel} = \mathcal{M}(input)$
- 11: # Adjust strategy (§ IV-C)
- 12: $Operation = \mathcal{S}(A_{sel})$
- 13: Perform calibration operation $\mathcal{Q}' = (Operation, \mathcal{Q}')$
- 14: Calculate current calibration performance \mathcal{P}'
- 15: **end while**
- 16: Update calibration notation $Q_{calib} = \mathcal{Q}'$, $\mathcal{P}_{calib} = \mathcal{P}'$
- 17: **return** Calibrated machine Q_{calib} , \mathcal{P}_{calib}

not, the process iterates, involving new state characterization and action selection until the threshold is reached. Finally, the calibration notation is updated to reflect the calibrated state, and the algorithm returns the calibrated quantum machine along with its performance characterization, completing the process.

V. EVALUATION

A. Experiment Setup

We collected a calibration dataset from a superconducting computer with 121 qubits with 11×11 grid, conducting data

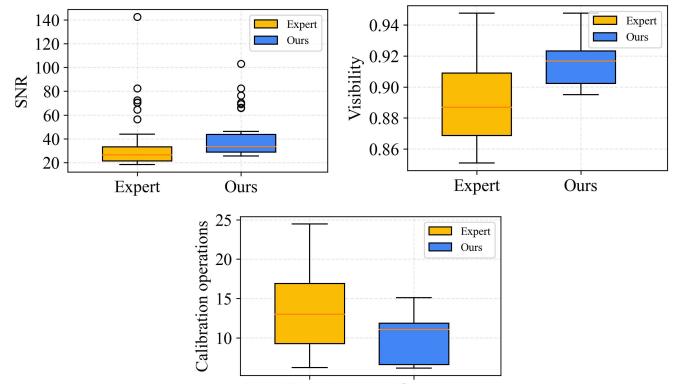


Fig. 12: The calibration overhead and performance comparison between our method and expert-led multi-iteration calibration.

collection and experimental testing on 77 of these qubits. The data is divided by qubit and segmented according to the IQraw operation. The input for the MoE consists of the distribution histogram of the quantum states $|0\rangle$ and $|1\rangle$, fitted by the IQraw double Gaussian, with a size of [2, 50]. We focus on actions that modify the parameters in IQraw, which include 7 actions detailed in Section IV-C. The qubits are prepared to either the $|0\rangle$ state or the $|1\rangle$ state and measured T ($T = 3000$) times respectively, and the coordinate points of their distribution on the IQ plane are recorded. We select 77 qubits for the full dataset of test average performance and 10 qubits at random for detailed analysis, respectively.

B. Calibration Effect Evaluation

To evaluate the performance of our method compared to the base calibration without error backtracking (denoted as “Base”, *i.e.*, performing the base calibration process just one time), we tested 77 qubits and characterized their performance, and Table I shows the average performance. For each calibration

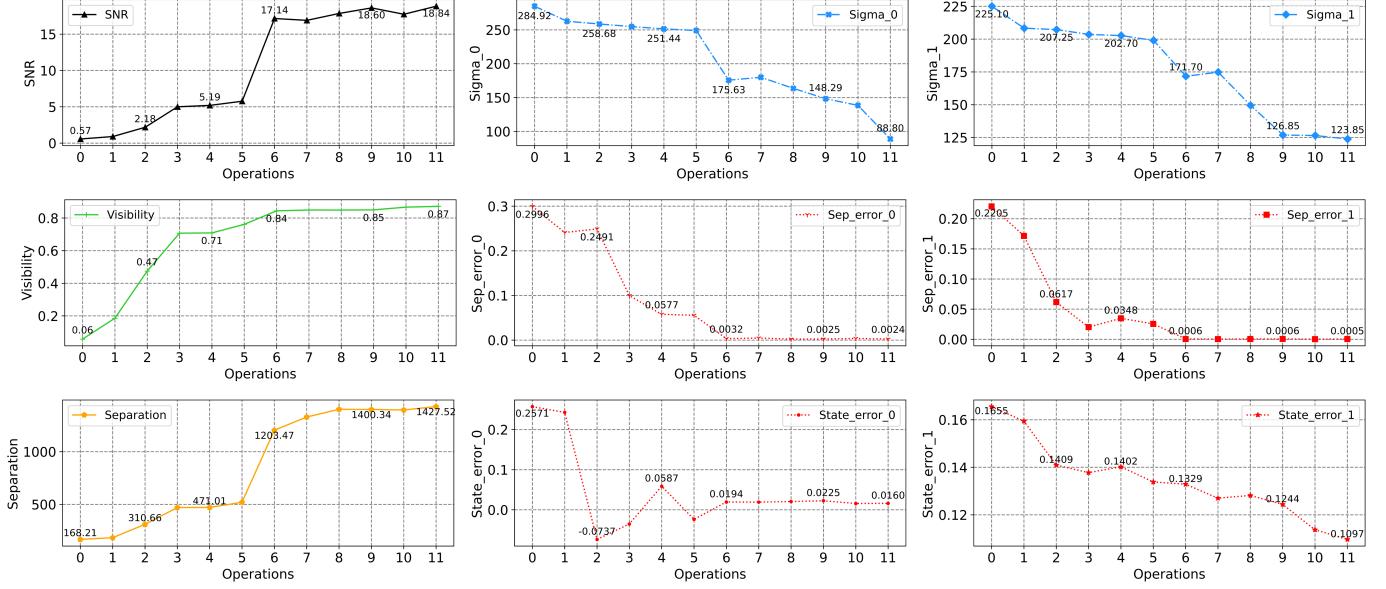


Fig. 13: The case study of performance improvements in overall calibration processes of qubit q_{22} .

performance metric, our results surpass those of the Base experiment. Specifically, the ‘Separation’ metric has shown an improvement of 212.26 units. The ‘SNR’ has experienced an average enhancement of 18.8 dB. Notably, the ‘Visibility’ metric has demonstrated an average improvement of 25.5%. Furthermore, the ‘Separation Error’ and ‘State Preparation Error’ have average decreased by 3.5% and 9.3%, respectively.

For a more intuitive demonstration, we randomly select 10 qubits for detailed performance comparison, the results as shown in Figure 11. Most of the qubits in our method show significant improvement compared to Base, and among the ensemble of 10 qubits, there is a notable variation in performance trends. For example, the error of q_{41} is significantly larger than that of other qubits, and after fine error analysis and recalibration, the visibility performance of our method is improved to 0.9. We analyze that the possible reasons are as follows. We analyze that it may be since the qubit nature of q_{41} is different from that of other qubits, and its weak sensitivity to the calibration parameters (such as readout power), so that the conventional fixed calibration process with expert rules does not apply to it. These results state that our method can show good optimization of the calibration performance for some abnormal qubits.

C. Calibration Operation Overhead

We compared our method against the default calibration process (Base) and multiple manual adjustments by experts in Figure 12. Conducted on 77 qubits, the boxplots illustrate three calibration performance metrics: SNR, visibility, and calibration operations. The results demonstrate that our method outperforms expert manual adjustments, achieving higher SNR and visibility with fewer calibration iterations.

D. Case Study

We conducted a case study on qubit calibration, involving 11 calibration steps and analyzing 9 performance metrics (Figure 13). Key findings show that calibration improves signal isolation (SNR), enhances state distinguishability (Separation), and boosts coherence (Visibility). Error rates for ground and excited states (Σ_0 , Σ_1) and separation errors (sep_error_0 , sep_error_1) decrease, indicating refined state discrimination. State errors ($state_error_0$, $state_error_1$) also decline, demonstrating effective error correction. Overall, calibration enhances readout performance, reduces errors, and supports high-precision qubit control and reliable state manipulation.

VI. CONCLUSION

In this paper, we present the first error-traceable calibration workflow optimization algorithm that integrates the MoE system to enhance error analysis and calibration flow backtracking in quantum automated calibration. Our algorithm identifies calibration actions needing backtracking and adjusts their parameters using a formulated updating strategy. This approach offers greater flexibility than fixed calibration processes, enabling the identification of suitable methods for abnormal qubits. Additionally, it significantly reduces the number of calibration iterations compared to manual expert-led calibration, effectively decreasing manual calibration time.

VII. ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (No.62472374) and the National Key Research and Development Program of China (No. 2023YFF0905200). This work was also funded by Zhejiang Pioneer (Jianbing) Project (No. 2023C01036).

REFERENCES

- [1] Tingting Li and Ziming Zhao. Moirai: Optimizing quantum serverless function orchestration via device allocation and circuit deployment. In *2024 IEEE International Conference on Web Services (ICWS)*, pages 707–717. IEEE, 2024.
- [2] Hyundo Jung, Hyunjin Kim, Woojin Lee, Jinwoo Jeon, Yohan Choi, Taehyeong Park, and Chulwoo Kim. A quantum-inspired probabilistic prime factorization based on virtually connected boltzmann machine and probabilistic annealing. *Scientific reports*, 13(1):16186, 2023.
- [3] Yudong Cao, Jonathan Romero, Jonathan P Olson, Matthias Degroote, Peter D Johnson, Mária Kieferová, Ian D Kivlichan, Tim Menke, Borja Peropadre, Nicolas PD Sawaya, et al. Quantum chemistry in the age of quantum computing. *Chemical reviews*, 119(19):10856–10915, 2019.
- [4] IBM. IBMQ Quantum. <https://quantum-computing.ibm.com/>, 2022.
- [5] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2018.
- [6] IBM Daily Calibration on Quantum Machines. <https://quantum-computing.ibm.com/>, Accessed: Jun. 5, 2023.
- [7] Julian Kelly, Peter O’Malley, Matthew Neeley, Hartmut Neven, and John M Martinis. Physical qubit calibration on a directed acyclic graph. *arXiv preprint arXiv:1803.03226*, 2018.
- [8] Tingting Li, Ziming Zhao, and Jianwei Yin. Minerva: Enhancing quantum network performance for high-fidelity multimedia transmission. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 3704–3712, 2024.
- [9] Tirthak Patel, Abhay Potharaju, Baolin Li, Rohan Basu Roy, and Devesh Tiwari. Experimental evaluation of nisq quantum computers: Error measurement, characterization, and implications. In *International Conference for High Performance Computing, Networking, Storage and Analysis (SC)*, 2020.
- [10] H. P. Breuer and F. Petruccione. *The theory of open quantum systems*. Oxford University Press, Great Clarendon Street, 2002.
- [11] Konstantinos Georgopoulos and et al. Modeling and simulating the noisy behavior of near-term quantum computers. *PRA*, 104(6):062432, 2021.
- [12] Peng Qian, Shahid Qamar, Xiao Xiao, Yanwu Gu, Xudan Chai, Zhen Zhao, Nicolo Forcellini, and Dong E Liu. Fast quantum calibration using bayesian optimization with state parameter estimator for non-markovian environment. *arXiv preprint arXiv:2205.12929*, 2022.
- [13] Yiding Liu, Zedong Li, Alan Robertson, Xin Fu, and Shuaiwen Leon Song. Enabling efficient real-time calibration on cloud quantum machines. *IEEE Transactions on Quantum Engineering*, 2023.
- [14] Max Werninghaus, Daniel J Egger, and Stefan Filipp. High-speed calibration and characterization of superconducting quantum processors without qubit reset. *PRX Quantum*, 2(2):020324, 2021.
- [15] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc V. Le, Geoffrey E. Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In *ICLR (Poster)*. OpenReview.net, 2017.
- [16] Tingting Li, Liqiang Lu, et al. Qust: Optimizing quantum neural network against spatial and temporal noise biases. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2024.
- [17] Tingting Li, Ziming Zhao, and Jianwei Yin. Task-driven quantum device fingerprint identification via modeling qnn outcome shift induced by quantum noise. In *Companion Proceedings of the ACM on Web Conference 2024*, pages 557–560, 2024.
- [18] Nicolas Wittler, Federico Roy, Kevin Pack, Max Werninghaus, Anurag Saha Roy, Daniel J Egger, Stefan Filipp, Frank K Wilhelm, and Shai Machnes. Integrated tool set for control, calibration, and characterization of quantum devices applied to superconducting qubits. *Physical review applied*, 15(3):034080, 2021.
- [19] Yilun Xu, Gang Huang, Jan Balewski, Alexis Morvan, Kasra Nowrouzi, David I Santiago, Ravi K Naik, Brad Mitchell, and Irfan Siddiqi. Automatic qubit characterization and gate optimization with qubic. *ACM Transactions on Quantum Computing*, 4(1):1–12, 2022.
- [20] Timothy R McIntosh, Teo Susnjak, Tong Liu, Paul Watters, and Malka N Halgamuge. From google gemini to openai q*(q-star): A survey of reshaping the generative artificial intelligence (ai) research landscape. *arXiv preprint arXiv:2312.10868*, 2023.
- [21] Dongbo Xi, Zhen Chen, Peng Yan, Yingter Zhang, Yongchun Zhu, Fuzhen Zhuang, and Yu Chen. Modeling the sequential dependence among audience multi-step conversions with multi-task learning in targeted display advertising. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3745–3755, 2021.
- [22] Kodai Minoura, Ko Abe, Hyunha Nam, Hiroyoshi Nishikawa, and Teppi Shimamura. A mixture-of-experts deep generative model for integrated analysis of single-cell multimics data. *Cell reports methods*, 1(5), 2021.
- [23] Yue Cao, Thomas Andrew Geddes, Jean Yee Hwa Yang, and Pengyi Yang. Ensemble deep learning in bioinformatics. *Nature Machine Intelligence*, 2(9):500–508, 2020.
- [24] Nan Du, Yanping Huang, Andrew M Dai, Simon Tong, Dmitry Lepikhin, Yuanzhong Xu, Maxim Krikun, Yanqi Zhou, Adams Wei Yu, Orhan Firat, et al. Glam: Efficient scaling of language models with mixture-of-experts. In *International Conference on Machine Learning*, pages 5547–5569. PMLR, 2022.
- [25] Paul V Klimov, Julian Kelly, John M Martinis, and Hartmut Neven. The snake optimizer for learning quantum processor control parameters. *arXiv preprint arXiv:2006.04594*, 2020.
- [26] David C McKay, Christopher J Wood, Sarah Sheldon, Jerry M Chow, and Jay M Gambetta. Efficient Z gates for quantum computing. *Physical Review A*, 96(2):022330, 2017.
- [27] Simon Gustavsson, Olger Zwier, Jonas Bylander, Fei Yan, Fumiki Yoshihara, Yasunobu Nakamura, Terry P Orlando, and William D Oliver. Improving quantum gate fidelities by using a qubit to measure microwave pulse distortions. *Physical review letters*, 110(4):040502, 2013.
- [28] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.
- [29] Emmanuel Bengio, Pierre-Luc Bacon, Joelle Pineau, and Doina Precup. Conditional computation in neural networks for faster models. *arXiv preprint arXiv:1511.06297*, 2015.
- [30] Michael I Jordan and Robert A Jacobs. Hierarchical mixtures of experts and the em algorithm. *Neural computation*, 6(2):181–214, 1994.
- [31] Yoshua Bengio, Nicholas Léonard, and Aaron Courville. Estimating or propagating gradients through stochastic neurons for conditional computation. *arXiv preprint arXiv:1308.3432*, 2013.