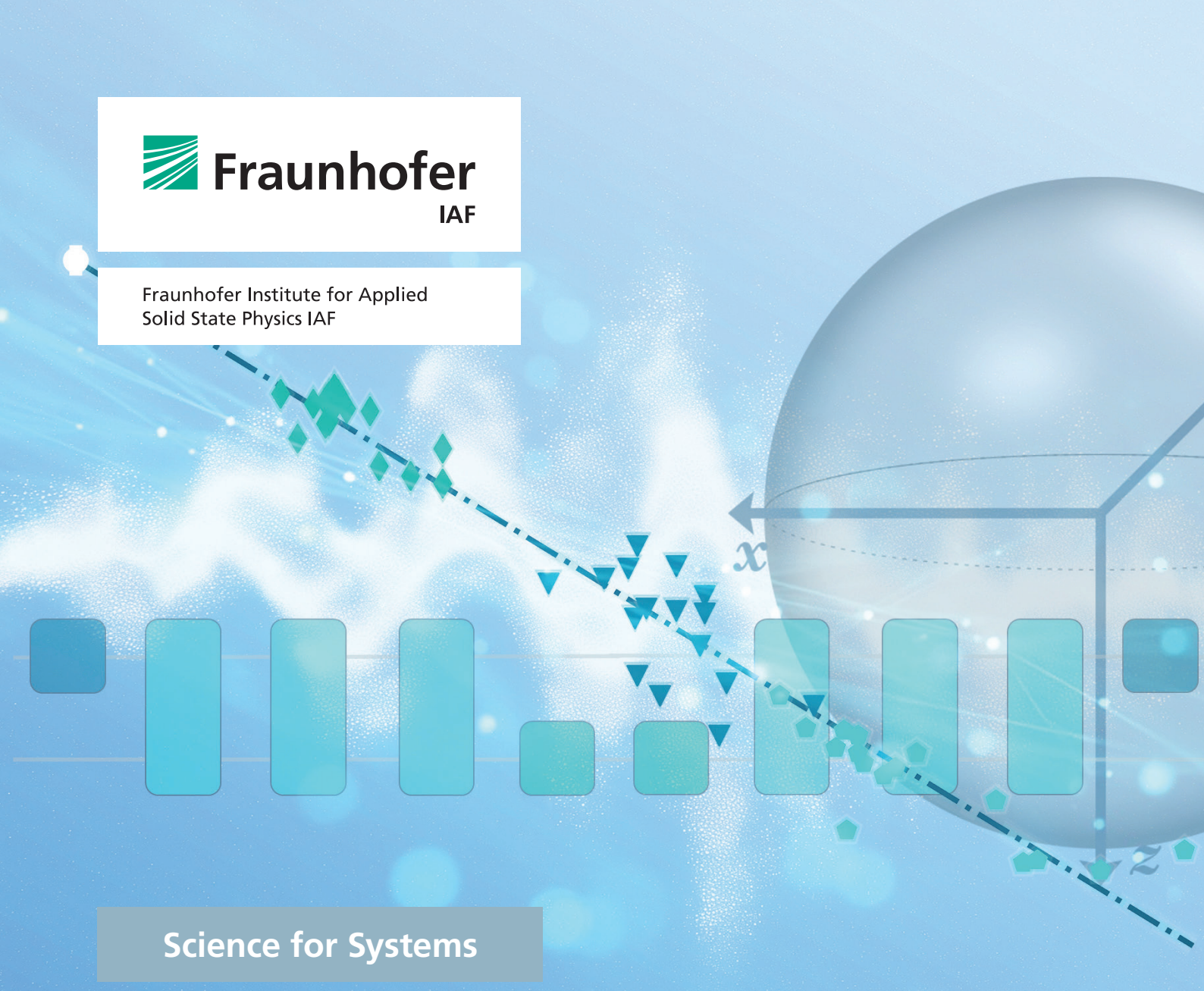


Science for Systems

Advanced Techniques for Error Reduction, Mitigation and Their Applications

An Approach Towards Fault-Tolerant Quantum Computing

Kathrin Friederike König

Fraunhofer Verlag

Fraunhofer Institute for
Applied Solid State Physics IAF

Science for Systems

No. 69

Kathrin Friederike König

Advanced Techniques for Error Reduction, Mitigation and Their Applications

An Approach Towards Fault-Tolerant Quantum Computing

Fraunhofer Verlag

Contact:
Fraunhofer Institute for
Applied Solid State Physics IAF
Tullastrasse 72
79108 Freiburg
Germany
Phone +49 761 5159-0
info@iaf.fraunhofer.de
www.iaf.fraunhofer.de

Cover illustration: © Kathrin Friederike König & Moritz Ringwald

Bibliographic information of the German National Library:
The German National Library has listed this publication in its Deutsche Nationalbibliografie; detailed bibliographic data is available on the internet at www.dnb.de.

ISBN (print edition): 978-3-8396-2199-8
DOI (free Open Access version): https://doi.org/10.24406/publica-8354

Zugl.: Freiburg, Univ., Diss., 2026

Print and finishing: Fraunhofer-Druckerei, Stuttgart

The book was printed with chlorine- and acid-free paper.



This work is licensed under a Creative Commons Attribution 4.0 International Public License: https://creativecommons.org/licenses/by/4.0/legalcode

© Fraunhofer Verlag, 2026
Nobelstrasse 12
70569 Stuttgart
Germany
verlag@fraunhofer.de
www.verlag.fraunhofer.de

is a constituent entity of the Fraunhofer-Gesellschaft, and as such has no separate legal status.

Fraunhofer-Gesellschaft zur Förderung
der angewandten Forschung e.V.
Hansastrasse 27 c
80686 München
Germany
www.fraunhofer.de

Dissertation zur Erlangung des Doktorgrades der
Technischen Fakultät der Albert-Ludwigs-Universität
Freiburg im Breisgau

Towards Fault-Tolerant Quantum Computing: Advanced Techniques for Error Reduction, Mitigation and their Applications

vorgelegt von
Kathrin Friederike König

2026

universität freiburg

Dekan
Prof. Dr. Frank Balle

Erstgutachter
Prof. Dr. Dr. Oliver Ambacher

Zweitgutachter
PD. Dr. Thomas Wellens

Datum der Disputation
01. April 2026

"... small errors will accumulate and cause the computation to go off track." - Rolf Landauer (1995, Is Quantum Mechanics Useful? in Philosophical Transactions)

Declaration

Affirmation in lieu of oath in accordance with §7, section 1, sentence 3, no. 7 of the doctoral degree regulations of the Albert-Ludwigs-Universität for the Faculty of Engineering.

The submitted doctoral thesis with the topic "*Towards Fault-Tolerant Quantum Computing: Advanced Techniques for Error Reduction, Mitigation and their Applications*" is an academic achievement solely made on my own merits.

I used only the listed sources and resources and that I did not get any undue help from a third party. I have identified any direct or indirect quotations taken from other published works as such.

The doctoral thesis or parts was not yet submitted to a University in Germany or abroad as part of an examination or qualification performance.

I confirm the accuracy of the above statements.

I am aware of the significance of an affirmation in lieu of oath and the criminal consequences of an untrue or incomplete affirmation in lieu of oath.

I, Kathrin Friederike König, hereby affirm in lieu of oath that I stated the whole truth in all conscience and that I did not withhold any facts in this matter.

Freiburg, January 20, 2026

Signature: _____

Kathrin Friederike König

Abstract

Quantum computing holds the potential to address problems that are currently considered intractable for classical processors by leveraging quantum mechanical effects for simulation, optimization, and chemistry. Despite recent technological progress, quantum hardware remains in the noisy intermediate-scale regime, where coherence limits, gate imperfections, and readout errors restrict circuit depth and algorithmic reliability. Although first demonstrations of small-scale Quantum Error Correction exist, logical processors with low logical error rates are still distant. In the near term, even error-corrected qubits are expected to operate with residual error, which means that improving calibration and mitigating errors at the physical and circuit level will remain essential for obtaining meaningful results. Established calibration routines and mitigation methods such as Zero-Noise-Extrapolation (ZNE) are valuable, but need to be further improved and more rigorously validated with respect to their underlying assumptions, stability and quantitative benefits.

This thesis addresses error reduction techniques on transmon qubits through calibration, gate analysis, and circuit-level mitigation. First, a systematic comparison of two single-qubit amplitude calibration techniques is carried out (cross-point and ping-pong calibration), followed by an analysis of two implementations of a two-qubit gate, based on the cross-resonance interaction. The results show that cross-point calibration offers faster convergence, higher precision and a more favorable signal to noise ratio than ping-pong calibration. Hamiltonian tomography confirms the validity of a reduced model for the cross-resonance interaction, and algorithmic tests reveal that direct and echoed CNOT gates offer complementary advantages depending on the application.

Second, the thesis develops and evaluates the method of Inverted-Circuit Zero-Noise-Extrapolation (IC-ZNE). Its main idea consists of measuring the effective error strength of a circuit through inverted circuits. This quantity is then used for extrapolation towards zero noise rather than relying on simple, approximate noise scaling assumptions as in standard ZNE. This approach constitutes a novel and original method. IC-ZNE consistently produces results closer to ideal reference values than standard ZNE. The method achieves up to 77% error reduction in observable error on the tested circuits compared to standard ZNE and simple post processing filters can yield up to 39% reduction in selected cases compared to standard ZNE. Further refinements introduce noise-adaptive circuit scaling strategies that enhance mitigation outcomes without significant additional cost. The performance of these techniques is examined across several circuits and in the context of variational quantum algorithms. Applications to the Quantum Approximate Optimization Algorithm and the Variational Quantum Eigensolver demonstrate that these techniques extend the useful operating regime of current devices, although their effectiveness decreases for very deep circuits or unfavorable variational ansätze.

Taken together, the calibration and mitigation techniques investigated and developed in this thesis demonstrate practical routes toward enhancing the reliability of near-term quantum processors. While they cannot replace full fault-tolerant quantum computation,

they offer useful and complementary capabilities that support progress toward scalable quantum technologies.

Zusammenfassung

Quantencomputer haben das Potenzial, Probleme zu lösen, die für klassische Computer als unzugänglich gelten, indem sie quantenmechanische Effekte beispielsweise für Simulationen in der Chemie und für Optimierungen nutzen. Trotz jüngster technologischer Fortschritte befindet sich die Quantenhardware weiterhin im „Noisy Intermediate Scale"-Regime, in dem Kohärenzlimits, Gatterfehler und Auslesefehler die Schaltkreistiefe und die Zuverlässigkeit von Algorithmen einschränken. Obwohl erste Demonstrationen der Quantenfehlerkorrektur in kleinem Maßstab vorliegen, sind logische Prozessoren mit niedrigen logischen Fehlerraten noch in weiter Ferne. Kurzfristig wird daher selbst bei fehlerkorrigierten Qubits mit Restfehlern zu rechnen sein, sodass Verbesserungen in der Kalibrierung und Fehlerminderung auf physikalischer und Schaltkreisebene weiterhin entscheidend bleiben. Etablierte Kalibrationsverfahren und Fehlerreduktionsmethoden wie die Zero-Noise-Extrapolation (ZNE) sind vielversprechend, ihre Annahmen, Stabilität und quantitativen Vorteile wurden jedoch auf aktueller Hardware noch nicht ausreichend systematisch untersucht.

Diese Arbeit befasst sich mit Techniken zur Fehlerreduktion bei Transmon-Qubits durch Kalibrierung, Gatteranalyse und Schaltkreismitigation. Zunächst wird ein systematischer Vergleich zweier Kalibrationstechniken für Ein-Qubit-Gatter (Cross-Point- und Ping-Pong-Kalibrierung) durchgeführt, gefolgt von einer Analyse zweier Implementierungen eines Zwei-Qubit-Gatters auf Basis der Cross-Resonance-Wechselwirkung. Die Ergebnisse zeigen, dass die Cross-Point-Kalibrierung eine schnellere Konvergenz, höhere Präzision und ein günstigeres Signal-Rausch-Verhältnis bietet. Die Hamilton-Tomographie bestätigt die Validität eines reduzierten Modells für die Cross-Resonance-Wechselwirkung, und algorithmische Tests zeigen, dass direkte und „echoed" CNOT-Implementierungen je nach Anwendung komplementäre Vorteile besitzen.

Anschließend entwickelt und evaluiert die Dissertation die Inverted-Circuit Zero-Noise-Extrapolation (IC-ZNE). Die Grundidee besteht darin, die effektive Fehlerstärke eines Schaltkreises mithilfe invertierte Schaltkreise zu messen. Diese Größe wird anschließend zur Extrapolation in Richtung Nullrauschen verwendet, anstatt sich wie bei der Standard-ZNE auf einfache, approximative Annahmen zur Rauschskalierung zu stützen. Dieser Ansatz stellt eine neuartige Methode dar und liefert durchweg Ergebnisse, die näher am idealen Referenzwert liegen als die Standard-ZNE. Die IC-ZNE erreicht auf den getesteten Schaltkreisen eine Fehlerreduktion von bis zu 77% im Vergleich zur Standard-ZNE, während einfache Nachbearbeitungsfilter in ausgewählten Fällen eine Verbesserung von bis zu 39% ermöglichen. Weitere Verfeinerungen führen zu fehleradaptiven Skalierungsstrategien, die die Mitigationsleistung ohne erheblichen Mehraufwand verbessern. Die Leistungsfähigkeit dieser Methoden wird über verschiedene Schaltkreise hinweg und im Kontext variationeller Quantenalgorithmen untersucht. Anwendungen auf den Quantum Approximate Optimization Algorithm und den Variational Quantum Eigensolver zeigen, dass diese Techniken den nutzbaren Bereich aktueller Geräte erweitern, ihre Wirksamkeit jedoch bei sehr tiefen Schaltungen oder ungünstigen variationellen Ansätzen abnimmt.

Insgesamt zeigen die in dieser Arbeit untersuchten und entwickelten Kalibrations- und Mitigationsmethoden praktische Wege zur Verbesserung der Zuverlässigkeit von Quantenprozessoren der nahen Zukunft. Sie können zwar keine vollständig fehlertolerante Quantenberechnung ersetzen, bieten jedoch nützliche und komplementäre Fähigkeiten, die den Fortschritt hin zu skalierbaren Quantentechnologien unterstützen.

Contents

Abstract . . . . . **III**

Zusammenfassung . . . . . **V**

1 Quantum computing in the noisy intermediate-scale era **1**

2 Quantum noise, hardware control, and error mitigation **5**
2.1 General setting and notation 5
2.2 Noise and errors in quantum computation 7
2.3 Error reduction at the superconducting quantum hardware and pulse level 9
2.3.1 Amplitude calibration techniques 11
2.3.2 The Controlled-Not gate and its associated Hamiltonian 17
2.4 Algorithm- and software-level error mitigation techniques 20
2.4.1 Zero-Noise Extrapolation 21
2.4.2 Randomized Compiling 23
2.5 Variational quantum algorithms 25
2.5.1 Quantum Approximate Optimization Algorithm 25
2.5.2 Variational Quantum Eigensolver 27

3 Calibration, optimization and analysis of quantum gates at the pulse level 29
3.1 Comparison of ping-pong and cross-point calibration for single-qubit gates 29
3.2 Experimental analysis of the two-qubit CR Hamiltonian in transmon qubits 36
3.3 Comparison between the implementation of direct and echoed CNOT gates 41
3.4 Summary of calibration and gate analysis at the pulse level 45

4 Inverted-Circuit Zero-Noise Extrapolation and further refined error mitigation techniques **49**
4.1 Definition of error strength 50
4.2 Measuring error strength 52
4.3 Structure of Inverted-Circuit Zero-Noise Extrapolation 56
4.4 Error scaling behavior in Zero-Noise Extrapolation 58
4.5 Comparison of sZNE and IC-ZNE using real backend data 61
4.6 Influence of error strength on the performance of sZNE and IC-ZNE . . . 63
4.7 Influence of error randomization on (Inverted-Circuit) Zero-Noise Extrapolation 67
4.7 Influence of error randomization on (Inverted-Circuit) Zero-Noise Extrapolation 67
4.8 Zero-Noise Extrapolation with Adaptive Scaling Factors 69
4.8.1 Equally spaced scaling factors versus exponentially spaced scaling factors 70
4.8.2 Impact of circuit error estimation and the choice of scaling factors 76

4.9 Filtering methods . . . 78
4.9.1 ϵ_0-filtering . . . 79
4.9.2 Local filtering . . . 80
4.9.3 Global filtering . . . 81
4.9.4 Local and global ellipse filtering . . . 81
4.9.5 Combined ϵ_0- and local (ellipse) filtering . . . 82
4.9.6 Hardware evaluation of filtering methods . . . 83
4.10 Analysis of folding schemes . . . 88
4.11 Summary of Inverted-Circuit ZNE and mitigation refinements . . . 91

5 Quantum error mitigation for variational algorithms . . . 93
5.1 Error Mitigation for the Quantum Approximate Optimization Algorithm 94
5.2 Error Mitigation for the Variational Quantum Eigensolver . . . 97
5.3 Summary of error mitigation for variational algorithms . . . 102

6 Conclusion and Outlook . . . 105

Bibliography . . . i

Appendix A Experimental details, algorithmic implementations, and theoretical derivations . . . xiii
A.1 Error rates of the *ibmq_ehningen* quantum device . . . xiii
A.2 Noise models for quantum hardware and depolarizing noise . . . xiii
A.2.1 Quantum hardware Noise Model . . . xiii
A.2.2 Depolarizing Noise Model . . . xvi
A.3 Box plot definition . . . xvii
A.4 Quantum algorithms and circuits implemented in this work . . . xvii
A.4.1 Grover's search algorithm . . . xviii
A.4.2 Harrow-Hassidim-Lloyd algorithm . . . xx
A.4.3 Hadamard-Ladder Circuit . . . xxii
A.4.4 Random circuits . . . xxiii
A.5 QAOA implementations and problem definitions . . . xxiii
A.5.1 Portfolio Optimization . . . xxiii
A.5.2 Maximum Cut . . . xxvi
A.5.3 Traveling Salesman Problem . . . xxvi
A.6 VQE implementations and problem definitions . . . xxix
A.7 Derivation of the adjoint quantum channel property . . . xxx
A.8 Details of the data-filtering procedure . . . xxxiii

Appendix B Glossary . . . xxxv
B.1 Acronyms . . . xxxv
B.2 List of symbols . . . xxxvi
B.3 List of figures . . . xl

Appendix C List of publications . . . xliii

Acknowledgments . . . xlv

1 Quantum computing in the noisy intermediate-scale era

Quantum computing was first proposed by Richard Feynman in 1982 [1]. He argued that classical computers cannot efficiently simulate general quantum systems because the associated Hilbert space grows exponentially with system size. According to Feynman, a device that follows the laws of quantum mechanics could simulate a broad range of quantum phenomena more efficiently. This includes electrons in atoms and molecules, spin systems and chemical reactions. The field became much more active after the introduction of Shor's factoring algorithm in 1994 [2], which provided the first convincing example of an exponential quantum advantage. Since then, many physical platforms have been developed with the aim of building a scalable quantum computer.

In recent years the field has seen rapid progress, and several quantum computing platforms have matured considerably. Each technology provides specific strengths and faces individual challenges. Figure 1.1 illustrates this development by comparing coherence times across leading platforms. Despite significant achievements such as early indications of quantum utility [3] and the development of systems with hundreds of qubits [4], the field is still far from the stage of fault tolerance through Quantum Error Correction (QEC) [5–7]. A central criterion for assessing and comparing platforms is their sensitivity to operational errors, which limits the achievable circuit depth and the reliability of computations.

Current hardware approaches include solid-state platforms such as superconducting qubits [9] and nitrogen vacancy centers in diamond [10, 11] as well as atomic systems

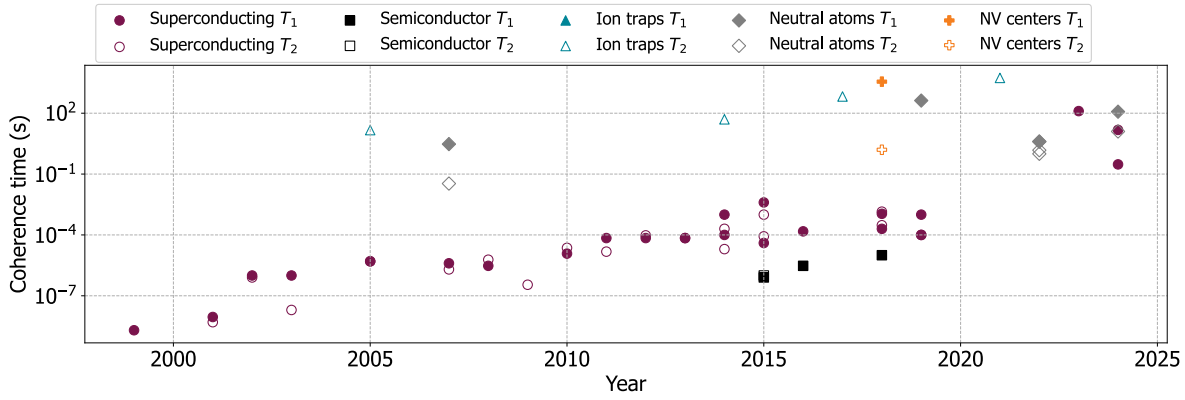


Figure 1.1: *Coherence times T_1 and T_2 for a range of quantum computing platforms, based on data compiled in Ref. [8]. The plot illustrates the substantial variation in achievable coherence times (energy-relaxation time T_1 and dephasing time T_2) across technologies such as superconducting qubits, trapped ions, neutral atoms, NV centres, and semiconductor spin qubits. These differences strongly influence the attainable circuit depth and overall performance of each platform.*

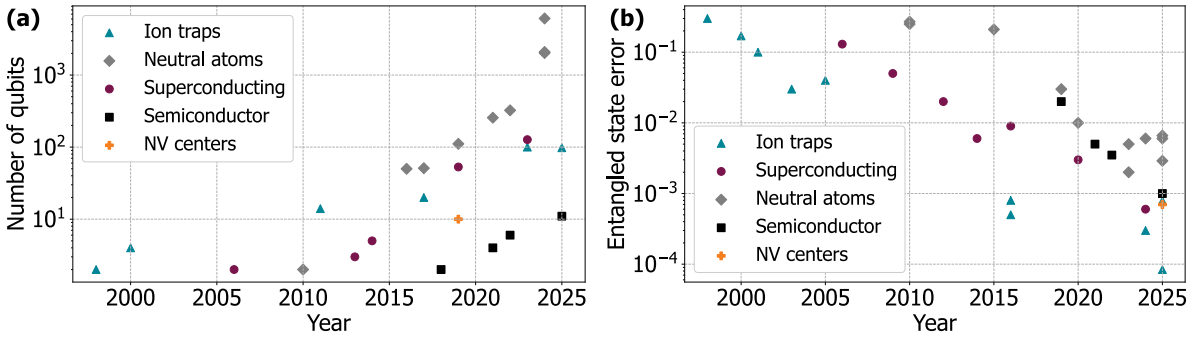


Figure 1.2: *(a) Comparison of qubit numbers and (b) two-qubit (entangling) gate errors for a range of quantum computing platforms, using data from Ref. [8]. (a) The number of available qubits provides an indication of scalability, while in (b) the entangling gate error is a key metric for algorithmic feasibility and proximity to fault-tolerance thresholds. Together, these plots highlight the technological maturity and limitations of each platform.*

such as trapped ions [12] and neutral atoms [13]. Other solid-state approaches involve semiconductor spin qubits in SiGe [14] and GaAs [15]. Among these technologies, superconducting qubits have become the most widely used platform. They are based on Josephson junctions [16] that are typically fabricated from aluminum or niobium. Several implementations exist, for example flux qubits, phase qubits and charge qubits [17].

All quantum hardware platforms pursue the same central goals. These include increasing coherence times (characterized by the relaxation time T_1 and the dephasing time T_2), reducing gate errors with a focus on two-qubit gates, since they typically exhibit higher error rates than single-qubit gates, and increasing the number of qubits to support more complex algorithms. Longer coherence times allow deeper circuits and therefore more complex computation. Lower gate errors reduce the accumulated noise in a quantum circuit and can often be improved through more sophisticated control and calibration. These developments across coherence, gate errors and qubit counts are summarized in Figures 1.1 and 1.2 using data from Ref. [8]. Over the years researchers have achieved significant progress in all of these aspects.

Superconducting qubits are widely used and accessible via well-established cloud toolchains. This thesis focuses on transmons [18], a charge-insensitive regime of the charge qubit, as they combine fast, high-fidelity control with strong tooling support and device availability. While these platforms allow for extensive calibration and mitigation studies, coherence, gate errors and other noise sources still limit achievable circuit depth. This motivates the techniques developed here.

Although quantum computing platforms have steadily improved over the years, their error rates remain above fault-tolerance thresholds [19, 20], and the qubit overhead required for full fault-tolerant error correction is still beyond current practical capabilities [21, 22], even though, for example, *IBM* claims that error correction will be available by 2029 [23]. The first logical qubits will not be perfect either, so additional techniques will likely remain relevant. We therefore remain in the Noisy Intermediate-Scale Quantum (NISQ) era [24], where hardware noise continues to limit algorithmic performance. As a result, some form of error mitigation will be essential for practical progress and for enabling meaningful computations on today's and future devices. One aspect of this effort involves gate-calibration techniques [25–30], since precise calibration directly reduces operational

errors and improves the reliability of computational results. Quantum Error Mitigation (QEM) [31] complements these approaches by reducing the impact of noise without requiring the large qubit counts demanded by full fault-tolerant error correction.

Quantum gate calibration covers a broad set of techniques that differ in their scope, their underlying assumptions, and their level of hardware integration. On one side of this spectrum are model-based calibration methods, which rely on Hamiltonian modeling and optimal-control tools to determine suitable pulse parameters [26, 32]. On the other side are model-free or closed-loop calibration approaches [25, 33]. These approaches adjust pulse parameters based on direct measurements of the qubit response and do not require an explicit model of the device. As quantum processors grow in size, the need for automated and scalable calibration frameworks becomes more important [26,29,34]. Such frameworks coordinate many interdependent calibration routines and aim to maintain stable device performance over time.

In superconducting quantum processors, particularly in *IBM*'s quantum devices, a widely used and well-established method is the ping-pong calibration technique [25]. This method calibrates single-qubit rotation amplitudes by repeatedly applying specific gate sequences, measuring the resulting qubit population, and adjusting the control amplitude in small steps. The procedure is repeated until the deviation from the target rotation becomes sufficiently small. Although the method is robust in practice, it often requires several iterations to reach the desired accuracy. Recently, a new calibration technique based on the cross-point protocol was introduced [35]. At the outset of this work, no systematic comparison between these two approaches was available, and no analysis of their respective strengths and weaknesses had been presented.

Complementary to calibration, which reduces error rates at the level of physical qubits, and to error suppression techniques such as dynamical decoupling [36, 37], QEM focuses on reducing bias in measured observables by improving noisy results through (classical) post-processing. Major approaches include inversion (probabilistic error cancellation [38], PEC), projection (symmetry verification [39, 40] and subspace expansion [41–43]), purification (virtual distillation [44]), extrapolation (Zero-Noise Extrapolation [38, 45, 46], ZNE), and noise tailoring (randomized compiling [47]). Each method trades bias for variance and incurs sampling overhead, additional depth, or both. Since most mitigation approaches assume relatively low noise levels, they only perform well when error rates are moderate. If noise levels are too high, these mitigation techniques cease to function reliably. ZNE is particularly widely used because it is hardware-agnostic, requires no ancillas nor detailed noise models, integrates via simple circuit-level noise scaling (e.g., gate folding or pulse stretching), and offers a flexible cost–accuracy trade-off. It composes well with readout error mitigation and other mitigation methods (e.g. Probabilistic Error Amplification [3, 48], PEA), making it useful on current quantum devices.

Despite the practical appeal of QEM, especially ZNE, several caveats remained insufficiently examined at the outset of this work. ZNE assumes a controllable, approximately monotonic noise-scaling mechanism and sufficiently stationary noise during data collection. These assumptions are rarely validated on current devices. Folding increases effective circuit depth and can push coherence to be the limiting factor, yet principled ways to curb overhead while preserving monotonic scaling (e.g., selective or local folding, schedule-aware folding) are not well characterized. The impact of non-Markovian, coherent, or drift-dominated errors on extrapolation bias and variance is poorly under-

stood, and combined strategies to re-Markovianize noise (randomized compiling, Pauli twirling) alongside ZNE need systematic analysis. Finally, post-processing pipelines, variance-aware weighting, outlier handling, and filtering for time-correlated runs, lack standardized, validated methodology.

Structure of the thesis

The challenges and open questions described above are addressed in this thesis as follows. Chapter 2 introduces the fundamental concepts on which this work is based. General settings and notation are presented, followed by an overview of noise mechanisms in quantum computing. The superconducting hardware used for the experimental evaluation is briefly characterized. Relevant error-reduction strategies, including single-qubit gate calibration, are described. Furthermore, a two-qubit gate is introduced, and two distinct implementations of this gate are outlined. Mitigation techniques such as ZNE and randomized compiling are discussed. Finally, two algorithms that serve as application cases for these mitigation techniques are introduced.
An analysis and comparison of two single-qubit gate calibration methods is presented in Chapter 3, together with an examination of the cross-resonance Hamiltonian and a comparison of two implementations of the two-qubit gate.
The focus of Chapter 4 is the development of techniques for discrete ZNE, including a new and original method: Inverted-Circuit Zero-Noise Extrapolation (IC-ZNE). The theory for measuring the error strength of a circuit, its implementation, its scaling behavior, and a comparison between standard ZNE and IC-ZNE are investigated and discussed. Further refinements to improve extrapolation accuracy through adaptation to circuit error rates and post-processing, also referred to as filtering techniques, are introduced and analyzed.
In Chapter 5, applications of ZNE and IC-ZNE using variational quantum algorithms are presented briefly and compared.
The thesis closes with conclusions and an outlook in Chapter 6. The Appendix A contains supplementary technical material, including quantum hardware specifications, noise-model descriptions, and detailed algorithmic implementations. A list of abbreviations, symbol definitions, and figures is provided in Appendix B. Publications associated with this thesis are listed in Appendix C.

2 Quantum noise, hardware control, and error mitigation

Building on the general introduction to quantum computing platforms presented in the previous chapter, this chapter focuses on the theoretical foundations relevant to the operation and characterization of superconducting qubit systems. The aim is to establish the formal framework and notation used throughout this work, providing the basis for the description of noise processes, the calibration procedures discussed in Chapter 3, and introduce state-of-the-art error-mitigation techniques which we will further develop in Chapter 4. The emphasis is on concepts directly applicable to transmon-based architectures, including their gate implementation, control mechanisms, and error modeling. Finally, in Chapter 5, two representative algorithms are examined to illustrate how error-mitigation methods can improve their performance.

This chapter is organized as follows. In Section 2.1, we first introduce the general theoretical framework and notation used throughout this work. Section 2.2 provides a concise overview of the main error sources in quantum computation and their mathematical descriptions. We then describe the experimental hardware employed in this work, including its connectivity layout, calibration procedures for single-qubit gates, the implementation and Hamiltonian of the fundamental two-qubit gate in Section 2.3. Following this, Section 2.4 outlines a well-established error-mitigation technique based on identity insertions and presents randomized compiling as an additional approach to suppressing the impact of noise. Finally, Section 2.5 introduces the theoretical background of two variational algorithms that will be used later in this work.

2.1 General setting and notation

We consider a register of q qubits with Hilbert space $\mathcal{H} = \mathcal{H}_2^{\otimes q}$. Each qubit is modeled as a two-level subsystem with logical states $|0\rangle$ and $|1\rangle$. Unless stated otherwise, the computational basis $\{|0\rangle, |1\rangle\}^{\otimes q}$ is used.

States and operators. Pure quantum states are represented by vectors $|\psi\rangle \in \mathcal{H}$, while mixed states are described by density matrices ρ, which are positive semidefinite operators with unit trace. Observables correspond to Hermitian operators A, and their ideal expectation values are computed as

$$\langle A \rangle_{\text{ideal}} = \langle \psi | A | \psi \rangle, \tag{2.1}$$

where

$$|\psi\rangle = U|0\rangle \tag{2.2}$$

and U is the targeted ideal circuit unitary. In practice, the implemented operation is affected by noise and described by a quantum channel $\mathcal{E}_U$, giving the measured noisy expectation value

$$\langle A\rangle = \mathrm{Tr}[A\rho], \tag{2.3}$$

with

$$\rho = \mathcal{E}_U(|0\rangle\langle 0|). \tag{2.4}$$

Quantum operations. An ideal noiseless operation on q qubits is represented by a unitary operator $U \in U(2^q)$. A noisy implementation is modeled by a completely positive trace-preserving (CPTP) map [16], also called a quantum channel,

$$\mathcal{E}(\rho) = \sum_k E_k \rho E_k^\dagger, \tag{2.5}$$

where $\{E_k\}$ are Kraus operators satisfying $\sum_k E_k^\dagger E_k = \mathbb{I}$ [16, 49]. For a unitary operation U, the ideal channel is $\mathcal{U}(\rho) = U\rho U^\dagger$. The deviation between $\mathcal{E}$ and $\mathcal{U}$ represents the error associated with that operation.

Quantum gates. Quantum computations are realized through a sequence of logical operations called *quantum gates*. Any unitary operation on q qubits can be decomposed into single-qubit and controlled-NOT (CNOT) gates, forming a universal gate set. The primary single-qubit gates used in this work are based on the Pauli matrices and the Hadamard gate. Their matrix representations are given by

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{2.6}$$

Two-qubit entangling operations are typically realized by the CNOT gate

$$\mathrm{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \tag{2.7}$$

The symbols used to represent quantum gates in circuit diagrams are shown in Figure 2.1. These include the previously mentioned gates, as well as the measurement symbol. These

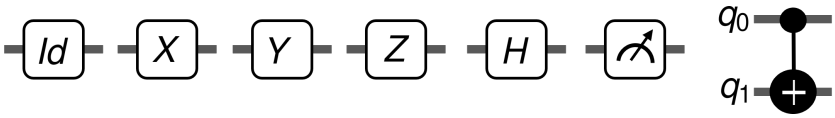


Figure 2.1: *Gate symbols in quantum computing. From left to right:* $\mathbb{I}, X, Y, Z, H$*, measurement, and CNOT* ($q_0 \to q_1$)*. CNOT: a filled dot on* q_0 *(control) connected by a vertical line to* $\oplus$ *on* q_1 *(target); applies* X *to* q_1 *if* $q_0 = 1$*.*

conventions are used for all circuit schematics throughout this work. For the CNOT (on the right), the first qubit represents the control and the second the target.

Observables and measurements. Ideal projective measurements in the computational basis are described by the projectors $P_i = |i\rangle\langle i|$. For a pure state $|\Psi\rangle$ the probability p_i to obtain the measurement outcome i is

$$p_i = \langle\Psi|P_i|\Psi\rangle = |\langle i|\Psi\rangle|^2. \tag{2.8}$$

For a mixed state ρ the probability is

$$p_i = \mathrm{Tr}[\rho P_i] = \langle i|\rho|i\rangle. \tag{2.9}$$

The expectation value of an observable A is estimated from the measurement statistics of repeated circuit executions, with respect to the basis in which A is diagonal. Unless explicitly stated otherwise, all expectation values in this work refer to measurements in the computational Z basis, therefore to observables that are diagonal in this basis. If A has the spectral decomposition $A = \sum_i a_i P_i$, then its expectation value is

$$\langle A\rangle = \sum_i a_i p_i = \sum_i a_i \langle P_i\rangle. \tag{2.10}$$

If A is not diagonal in Z, we apply a basis change to A's eigenbasis before performing a Z-basis measurement. Pauli-X and Pauli-Y measurements are implemented by applying the appropriate single-qubit rotations followed by a Z-basis measurement.

2.2 Noise and errors in quantum computation

Errors in quantum computing arise from imperfections in both the quantum hardware and the associated classical control electronics. These imperfections manifest themselves as unwanted interactions with the environment, crosstalk between qubits, control inaccuracies, and measurement errors. Together, they lead to deviations between the ideal and implemented quantum operations. Such effects are collectively described as noise, modeled by quantum channels acting on the system state, except for measurement errors.

A central assumption in most quantum error models is the Markov approximation, in which each noise process acts independently and does not depend on the system's past evolution [16]. Under this assumption, the system dynamics can be described by a quantum operation, that is, a CPTP map (see Equation (2.5)). In contrast, non-Markovian noise involves memory effects, where the environment retains information

about previous system states. Such correlations can arise from slow parameter drift, spectral crowding, or shared control lines.

The effect of noise and quantum operations on an open quantum system is described by a quantum channel, a CPTP map that transforms the density matrix ρ, see Equation (2.5). Under the Markov approximation and in the absence of crosstalk, the overall noisy evolution $\mathcal{E}_U$ for a circuit composed of sequential operations $U = U_n \cdots U_2 U_1$ can be decomposed into a product of independent quantum channels, $\mathcal{E}_U = \mathcal{E}_{U_n} \circ \cdots \circ \mathcal{E}_{U_2} \circ \mathcal{E}_{U_1}$. If the noise is also time-independent (without drift), the same channel $\mathcal{E}_{U_i}$ describes each instance of a given gate U_i. A key complication is crosstalk: when U_i is intended as a local operation (e.g., a one- or two-qubit gate), unwanted coupling can cause $\mathcal{E}_{U_i}$ to affect additional (neighboring) qubits. Moreover, if a circuit layer consists of multiple gates acting simultaneously on different qubits, $U_i = U_{i,m} \cdots U_{i,2} U_{i,1}$, then the decomposition $\mathcal{E}_{U_i} = \mathcal{E}_{U_{i,m}} \circ \cdots \circ \mathcal{E}_{U_{i,2}} \circ \mathcal{E}_{U_{i,1}}$ is generally valid only in the absence of crosstalk. Otherwise, the entire layer must be treated as a whole [50].

Errors can be broadly classified as coherent or incoherent [31]. Coherent errors are systematic and unitary in nature, such as over-rotations, calibration drifts, or unwanted phase accumulations. They correspond to small, consistent unitary deviations. Coherent errors have been shown to accumulate faster than stochastic Pauli-type errors and can therefore be converted into Pauli errors through suitable randomization techniques [31]. Incoherent errors arise from random processes such as energy relaxation or dephasing and are typically modeled by non-unitary channels acting probabilistically on the system state. In practice, both types coexist and contribute differently to experimental error metrics.

The noise acting on qubits is often approximated by simple one-qubit quantum channels that capture the dominant physical error processes. In this work, noise is mainly modeled by the depolarizing channel, which represents random applications of the Pauli operators X, Y, and Z. It is a simple Markovian model that describes the replacement of the ideal quantum state by a mixture of these three operations. The action of the depolarizing channel on a density matrix ρ is given by

$$\mathcal{E}_{\text{dep}}(\rho) = (1-p)\rho + \tfrac{p}{3}(X\rho X + Y\rho Y + Z\rho Z), \tag{2.11}$$

where p is the depolarizing probability. A generalized form of the depolarizing channel for a quantum system of dimension d is given by

$$\mathcal{E}_{\text{dep}}(\rho) = p\,\tfrac{\mathbb{I}}{d} + (1-p)\rho, \tag{2.12}$$

where ρ denotes the state being considered, p the probability that the system is depolarized, $d = 2^q$ with q the number of qubits, and $\mathbb{I}$ is the identity operator [16, p. 378]. With probability p, the state is replaced by the completely mixed state, $\frac{\mathbb{I}}{d}$. This model provides a symmetric approximation of stochastic errors and is frequently used to describe gate imperfections in superconducting qubit systems.

Other types of noise such as bit-flip, phase-flip, amplitude damping, and phase damping channels are often used to describe specific error mechanisms [16]. Amplitude damping is associated with energy relaxation characterized by the time constant T_1, and phase damping with dephasing characterized by the time constant T_2. Although these channels

are not directly used in the analysis presented here, they represent important physical processes contributing to qubit decoherence.

In addition to gate-related noise, State-Preparation And Measurement (SPAM) errors significantly affect experimental results [31]. Unlike gate errors, measurement errors are not described by a quantum channel acting on the state, but instead by a classical probabilistic model or a generalized measurement formalism. These errors are often modeled by a classical confusion matrix A. This matrix maps the ideal outcome probabilities p_{ideal} to the observed ones through $p_{\text{noisy}} = Ap_{\text{ideal}}$ [51–53]. More formally, readout errors can be described within the Positive Operator-Valued Measure (POVM) formalism [16]. A POVM consists of measurement operators M_i that satisfy $M_i \geq 0$ and $\sum_i M_i = \mathbb{I}$. For a quantum state ρ the probability to observe outcome i is $p_i = Tr[\rho M_i]$. In the confusion matrix model, the POVM elements M_i corresponding to noisy readout, are assumed to be linear combinations of the ideal computational basis projectors $P_j = |j\rangle\langle j|$. That is, one writes $M_i = \sum_j A_{i|j} P_j$, which yields the same probability relation $p_{\text{noisy}} = Ap_{\text{ideal}}$. SPAM errors are typically characterized independently from gate errors and corrected using measurement calibration or post-processing techniques [51, 53–56].

2.3 Error reduction at the superconducting quantum hardware and pulse level

Following the introduction of general settings, notation, and error channels, we turn to the physical implementation of qubits. Early proposals and demonstrations in trapped-ion systems (e.g., Cirac and Zoller [12]; Monroe and Wineland [57]) initiated a comprehensive exploration of platforms, encompassing semiconductor spins, quantum dots, ultracold atoms, nitrogen-vacancy centers, and photonic qubits [58]. In this work we focus on superconducting transmission line shunted plasma oscillation (transmon) qubits [18], realized as planar millimeter-scale circuits with Josephson junctions. Transmons are weakly anharmonic artificial atoms that achieve charge-noise insensitivity while retaining sufficient nonlinearity for selective control [18]. Their fabrication leverages thin-film deposition, lithography, and junction oxidation processes common to semiconductor manufacturing, enabling rapid design iteration [59, 60]. Operating in the microwave regime with frequencies of $\sim 4 - 7$ GHz benefits from mature control and readout instrumentation. However, transmons require millikelvin environments (10–20 mK), which imposes constraints on the wiring and thermal loads of dilution refrigerators.

The *IBM Falcon* (revision 5.11) chip used in this work is a fixed-frequency, superconducting transmon processor with heavy-hex connectivity [61]. This connectivity slightly reduces qubit connectivity but also minimizes qubit frequency collisions and spectator errors [62, 63]. Each qubit is a capacitively shunted Josephson junction coupled to its neighbors via bus resonators. Single-qubit control is performed with dedicated microwave drive lines. Readout uses individual dispersively coupled resonators frequency-multiplexed onto shared feedlines and amplified at base temperature (typically ≈ 10 mK) using a cryogenic low-noise amplifier. *IBM*'s heavy-hex layout limits each qubit connection to 1–3 nearest-neighbor connections, fewer than a square lattice, reducing frequency collisions and long-range crosstalk [63]. Two-qubit entanglement gates are implemented with the cross-resonance (CR) interaction on connected qubit pairs [64–66]. The architecture

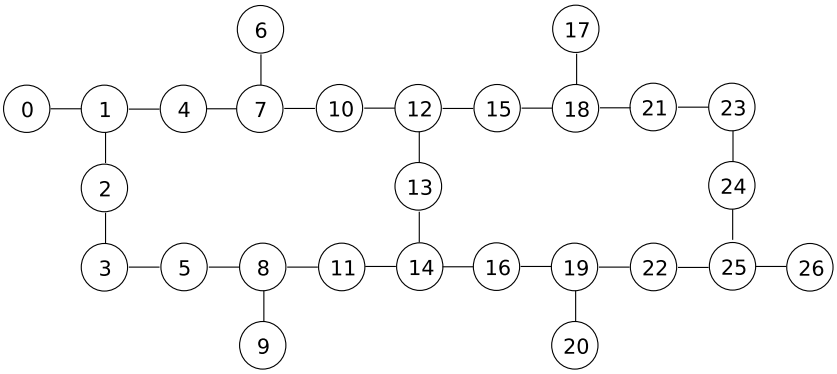


Figure 2.2: *Architecture of the ibmq_ehningen device. The graph shows the connectivity (lines) between the 27 transmon qubits (circles) on IBM's Falcon r5.11 processor.*



Figure 2.3: *Additional gate symbols available on the ibmq_ehningen backend. From left to right: the rotation gate $R_Z(\phi)$ with phase ϕ, and the square-root-of-X gate SX.*

of the *ibmq_ehningen* device is shown in Figure 2.2. The hardware contains 27 qubits (circles) with their connections (lines). The native gates available and implemented in this hardware are the identity, square-root-X (SX or $\sqrt{X}$), X, controlled-X (CX or CNOT), qubit reset operations, and the virtual Z-rotation (R_z). The R_z gate is not implemented through a physical microwave pulse. Instead, it is realized by updating the classical control frame, meaning that the phase reference for subsequent X and Y rotations is shifted by the desired angle [67]. As a consequence, an $R_z(\phi)$ gate has ideally zero duration and contributes negligible extra error [67]. These gates form a universal gate set for superconducting qubits and serve as the building blocks for all compiled circuits on the *ibmq_ehningen* device. All error and gate data can be found in Section A.1. The R_Z and SX gate are described through

$$R_Z(\phi) = \begin{pmatrix} e^{-i\frac{\phi}{2}} & 0 \\ 0 & e^{i\frac{\phi}{2}} \end{pmatrix}, \quad SX = \sqrt{X} = \frac{1}{2}\begin{pmatrix} 1+i & 1-i \\ 1-i & 1+i \end{pmatrix}. \tag{2.13}$$

The symbols used to represent those quantum gates in circuit diagrams are shown in Figure 2.3.

The following sections describe two amplitude calibration techniques for single-qubit control [25, 35] in Subsection 2.3.1. They also discuss the Hamiltonian formulation of the CNOT gate for transmon qubits and introduce two pulse schemes: an echoed CR scheme that yields an effective $Z \otimes X$ interaction [64–66], and a direct CR scheme that offers reduced execution time and in some cases lower error rates, discussed in Subsection 2.3.2. Together, these tools form the foundation for minimizing operational errors and advancing error mitigation and correction in scalable quantum architectures.

2.3.1 Amplitude calibration techniques

Quantum gates are the fundamental building blocks of quantum algorithms. Achieving low-error gate implementations is essential for reliable quantum computation [16]. Precise control and calibration are crucial for mitigating errors that inherently challenge the stability and fidelity of quantum operations. In superconducting transmon systems, gate performance is limited by decoherence and control imperfections, making systematic calibration and accurate Hamiltonian modeling essential [18, 59, 60]. Each qubit q_i is characterized by its transition frequency $\omega_{0,1}$, anharmonicity α_i, and relaxation and dephasing times $T_{1,i}, T_{2,i}$. Control pulses acting on qubit i are parameterized by the amplitude A_i, duration τ_i, and detuning Δ_i. Calibration procedures aim to determine the optimal set A_i, τ_i, Δ_i minimizing the average gate infidelity.

This section outlines the calibration hierarchy used to achieve high-fidelity single-qubit control. We begin with a rough calibration to determine approximate qubit and pulse parameters, followed by two fine amplitude calibration techniques: the ping-pong method, which iteratively compensates amplitude-dependent rotation errors, and the cross-point amplitude calibration, which employs an alternative ansatz for error compensation and offers a more hardware-efficient protocol with reduced error rates. The performance of these two fine amplitude calibration schemes is compared in Section 3.1.

2.3.1.1 Rough calibration

This section provides a brief overview of the rough amplitude calibration process used in *IBM* superconducting qubit systems. This procedure yields a first approximation of the parameters needed to operate quantum logic gates, specifically the qubit transition frequency ω_{01} and the approximate amplitude for a π-rotation (a Pauli-X gate). Small deviations due to residual errors may remain. The first step in characterizing a transmon qubit is to determine its transition frequency ω_{01}, corresponding to the energy difference between the ground state $|0\rangle$ and the first excited state $|1\rangle$. This is achieved through a spectroscopy experiment. A microwave drive pulse with a Gaussian envelope is applied to the qubit. The driving Hamiltonian in the frame rotating at the drive frequency ω_d is given by [60]

$$H_{drive}(t)/\hbar = \frac{\Omega}{2}(I(t)\sigma_x + Q(t)\sigma_y), \tag{2.14}$$

where $I(t)$ and $Q(t)$ denote the dimensionless in-phase and quadrature envelope functions of the microwave drive, σ_x the Pauli matrix X, and Ω is the Rabi frequency proportional to the drive amplitude. In the rotating frame the qubit Hamiltonian is

$$H_{\text{qubit}}/\hbar = \frac{1}{2}(\omega_{01} - \omega_d)\sigma_z, \tag{2.15}$$

so the spectroscopy response is maximized when the drive frequency ω_d is resonant with the qubit transition frequency ω_{01}. The experiment involves sweeping the drive frequency ω_d across a range encompassing the expected design frequency of the qubit. For each frequency, the excited-state population P_1 is measured after applying the microwave signal.

The resulting spectrum typically exhibits a Lorentzian lineshape [59, 60]

$$P_1(\omega) = A_{\text{off}} + \frac{A_{\text{amp}}}{\pi} \frac{\frac{1}{2}\Gamma}{(\omega - \omega_{01})^2 + (\frac{1}{2}\Gamma)^2}, \tag{2.16}$$

where A_{amp} is the signal amplitude, Γ the full width at half maximum of the resonance, and A_{off} is a constant offset. A nonlinear least-squares fit of the measured data to Equation (2.16) yields the rough transition frequency ω_{01}. An example of such a spectroscopy measurement is shown in Figure 2.4 (a). Since spectroscopy yields only a rough estimate limited by line broadening, a follow-up Ramsey interferometry is usually performed to fine-tune the transition frequency [59].

Once the resonance frequency ω_{01} is known, the next step is to calibrate the amplitude of the microwave pulse required to perform a π-rotation. This is accomplished by observing Rabi oscillations. A resonant microwave pulse ($\omega_d = \omega_{01}$) with a fixed duration and a variable amplitude A is applied to the qubit initialized in $|0\rangle$. The excited-state probability oscillates as a function of the pulse amplitude (or, equivalently, its pulse duration)

$$\langle Z \rangle(A) = \frac{A_{\text{osc}}}{2} \cos(2\pi f_{\text{Rabi}} A + \phi) + B. \tag{2.17}$$

Here A_{osc} is the oscillation amplitude, f_{Rabi} is the Rabi frequency, which defines the oscillation period, ϕ is the phase offset, and B is a constant offset. The amplitude A_π that corresponds to a π-pulse is found at the maximum of the oscillation where $\langle Z \rangle$ is minimal (see Figure 2.4 (b)). This occurs when the argument of the cosine is π, giving

$$A_\pi = \frac{\pi - \phi}{2\pi f_{\text{Rabi}}}. \tag{2.18}$$

Fitting the measured data to Equation (2.17) provides the parameters f_{Rabi} and ϕ, allowing for the extraction of A_π. This value serves as the initial, rough estimate for the amplitude of an X-gate. An example of Rabi oscillations is shown in Figure 2.4 (b).

Beyond the rough frequency and amplitude calibration, several other characterizations are essential for operating a transmon qubit. The Gaussian pulse is often modified to further suppress leakage to the noncomputational $|2\rangle$ level due to the weak anharmonicity of the transmon. This correction is achieved using the Derivative Removal by Adiabatic Gate (DRAG) technique [68,69], which adds a quadrature component proportional to the time derivative of the main pulse to reduce leakage and phase errors. Further calibrations include state-discrimination calibration (to distinguish $|0\rangle$ and $|1\rangle$), measurement of the energy-relaxation time T_1 [59], and determination of the coherence time T_2 using, for example, a Hahn-echo sequence [59]. Although not the primary focus of this work, these measurements provide the essential characterization needed to assess device performance and stability, forming the foundation for subsequent fine-calibration procedures like error-amplification sequences for pulse duration and amplitude refinement.

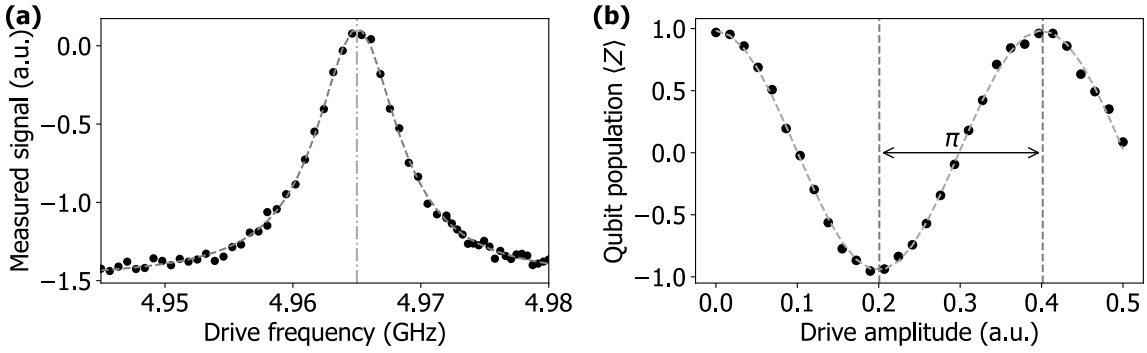


Figure 2.4: *Example data from the rough calibration procedure. (a) Qubit spectroscopy showing a Lorentzian response of the excited-state probability as a function of drive frequency. The vertical dashed line marks the extracted qubit transition frequency ω_{01}. (b) Amplitude-sweep Rabi experiment showing oscillations of the measured qubit expectation value $\langle Z \rangle$ as the drive amplitude is varied. The dotted vertical lines represent a cosine fit used to determine the π-pulse amplitude A_π. The drive amplitude and measured signal are given in arbitrary units, as they correspond to scaled control parameters that depend on the specific hardware configuration and is shown here for illustration.*

2.3.1.2 Ping-Pong Amplitude calibration for single-qubit control

Fine calibration is required to correct small coherent over- or under-rotations that remain after the rough amplitude calibration (e.g., for the π-pulse). For small coherent rotation errors δ, the gate infidelity scales quadratically as $\frac{\delta^2}{6}$ [25]. Here, δ denotes the small gate deviation of the implemented rotation angle from its ideal value, i.e., $\theta_{\text{actual}} = \theta_{\text{ideal}} + \delta$. In the fine calibration procedure, the small over- or under-rotations of the qubit are first amplified so that they can be measured and compensated. The method used is known as ping-pong calibration, named after the characteristic oscillating pattern observed in the measurement results. The basic idea is to apply the same gate repeatedly, forming an error-amplifying sequence. If each gate has a small rotation error δ from the ideal angle, applying the gate n times results in an accumulated error of $n \cdot \delta$. By fitting this amplified response, the correct drive amplitude can be determined with high precision. The pulse amplitude is then updated accordingly. Increasing the number of repetitions improves the sensitivity of the calibration but is ultimately limited by the coherence time of the qubit and SPAM errors.

The pulse sequence for calibrating an X_π or a $X_{\pi/2}$ pulse is $X_{\pi/2} - (X_{\pi,\pi/2})^{2n}$, both following the scheme of Ref. [25]. The initial rough amplitude estimate for the $X_{\pi/2}$ pulse is taken as half of the X_π pulse amplitude with the same pulse duration and shape, since the rotation angle scales approximately linearly with the drive amplitude.

The measured population of the excited state $P_{|1\rangle}(n)$ oscillates as a function of the repetition number n. The data points are fitted to

$$P_{|1\rangle}(n) = a\cos(\beta\, n + \varphi_0) + b, \tag{2.19}$$

where n is the number of gate repetitions, a the oscillation amplitude, β the effective rotation angle per gate repetition, φ_0 a phase offset, and b an offset [70]. In contrast to formulations that make the even–odd alternation arising from repeated $\pi/2$ rotations explicit through a factor of $(-1)^n$, as in Ref. [25], this approach incorporates such effects

into the fitted parameters β and φ_0. This allows the same model to be used for both X_π and $X_{\pi/2}$ pulses. The rotation angle error per gate is then given by

$$\delta = \beta - \theta_0, \tag{2.20}$$

where $\theta_0 = \pi$ for the X_π gate and $\theta_0 = \pi/2$ for the $X_{\pi/2}$ gate. The pulse amplitude is updated as

$$A_{\text{new}} = \frac{A_{\text{old}}\,\theta_0}{\beta}. \tag{2.21}$$

Note that on *IBM* hardware, the SX gate is implemented as a pure rotation $R_X(\pi/2)$. Alternative conventions such as $\sqrt{-iX}$ differ only by a global phase, which is irrelevant for this calibration.

As the residual error becomes smaller, a larger number of repetitions is required to achieve a meaningful update. The ping-pong calibration can therefore be repeated iteratively until a predefined threshold δ is reached, enabling sub-percent precision in gate amplitudes [25].

However, as emphasized in Ref. [25], this approach assumes that the dominant error is a coherent, gate-local over- or under-rotation about the intended rotatation axis. Detuning and axis misalignment, crosstalk, or slow drift can bias the inferred angle error. Moreover, increasing the number of repetitions amplifies T_1 and T_2 decay, which reduces the oscillation contrast and can make the cosine fits less reliable or even ambiguous, for example by producing weakly defined minima in the fit landscape. SPAM errors are only partially mitigated via an offset parameter and can further distort the fit in this regime.

In addition, for weakly anharmonic qubits, repeated strong drives can induce leakage to $|2\rangle$, especially if the DRAG correction is not well tuned. This breaks the effective two-level cosine model and can cause leakage effects to be misinterpreted as amplitude errors [68]. In practice, frequency and DRAG calibrations should therefore precede ping-pong calibration, and the repetition range should be chosen to balance sensitivity against decoherence and leakage [25, 68].

In Figure 2.5 an exemplary calibration is shown. On the left, the rotations in a Bloch sphere are shown for the first three points of the sequence. The measured population plotted against the number of repetitions is shown on the right.

Overall, the ping-pong calibration provides a powerful way to amplify and correct small coherent rotation errors, achieving sub-percent precision in gate amplitudes [25], but it is sensitive to detuning, drift, SPAM-induced offsets, and leakage without proper DRAG, which can all degrade the fidelity and interpretability of the fit [25, 68]. In the following section, we turn to a second calibration method that could overcome some of the limitations of the ping-pong calibration method that we introduced in this section.

2.3.1.3 Cross-point amplitude calibration for single-qubit control

Like the ping-pong method, the cross-point amplitude calibration uses an error-amplifying pulse sequence, but it replaces the trigonometric fitting function with a simpler zero-crossing and linear-fit approach [35, 71].

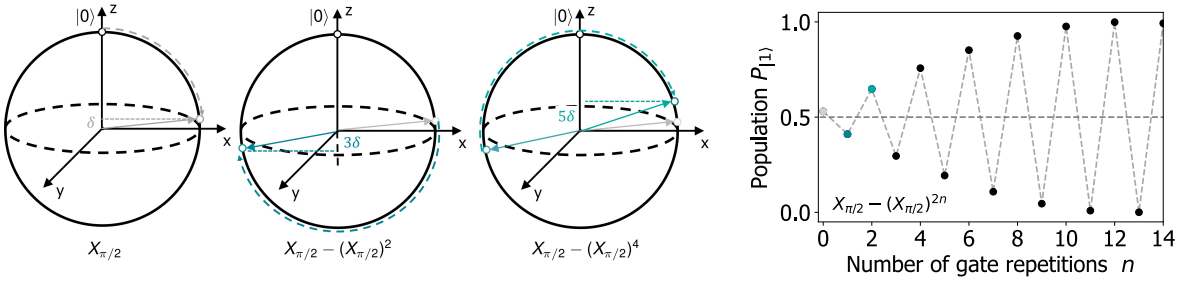


Figure 2.5: *Schematic of the amplitude-calibration (ping-pong) error-amplification protocol (adapted from [25]). (a) Bloch-sphere depiction of the first three steps of the sequence. (b) Ground-state population $P_{|1\rangle}$ as a function of the number of gate repetitions n. A small initial offset δ after the first gate is coherently amplified under repetition, providing a sensitive indicator of amplitude miscalibration.*

The cross-point calibration determines the pulse amplitude A that produces the target rotation $\theta(A) = \theta_0$ by executing two mirrored sequences and identifying the value at which their measurement outcomes are equal, the so-called 'cross-point'. The symmetric construction is expected to mitigate systematic effects such as SPAM errors or slow readout drifts while maintaining first-order sensitivity to small coherent amplitude errors.

Let $\theta(A)$ denote the rotation angle induced by a single Gaussian-shaped pulse with amplitude A on the qubit's $|0\rangle \leftrightarrow |1\rangle$ transition. Near the correct amplitude θ_0 the rotation can be written as $\theta(A) = \theta_0 + \delta$, where $|\delta| \ll 1$ represents a small coherent over- or under-rotation error. Two mirrored error-amplifying sequences, $S_+(A, n)$ and $S_-(A, n)$, are constructed such that they respond with opposite sign to the same small error. Here n is the number of gate repetitions. Each sequence consists of n identical pulses with amplitude A (or opposite phase $-A$), followed by a single $X_{\pi/2}$ pulse. This final pulse converts the residual accumulated rotation into a measurable population difference.

The repetition of identical pulses amplifies the coherent error approximately linearly with n. After n applications, the total rotation error is close to $n\,\delta$. The measured population $P_{|1\rangle}$ can therefore be expanded to first order around the ideal operating point. For small δ one obtains

$$P_{|1\rangle,\pm}(A, n) \approx \tfrac{1}{2} \pm c(n)\,[\theta(A) - \theta_0] + b, \tag{2.22}$$

where $c(n) > 0$ is a proportionality factor that captures the sensitivity of the sequence to a small rotation error and increases approximately linearly with the number of repetitions n. The term b accounts for slowly varying offsets such as contributions from SPAM errors or readout drift. The sign change reflects the mirroring of the two sequences.

The cross-point follows from the intersection condition

$$P_{|1\rangle,+}(A, n) = P_{|1\rangle,-}(A, n) \quad \Longrightarrow \quad \theta(A) = \theta_0, \tag{2.23}$$

which defines the amplitude at which the target rotation θ_0 is realized. This zero-crossing condition is, to first order, independent of the offset b, which makes the calibration relatively robust against slow drifts and small SPAM asymmetries.

In experiments, the amplitudes A are swept over a small symmetric range $A \in [A_0 - \Delta_A,\ A_0 + \Delta_A]$ around a rough estimate A_0, where Δ_A is a deliberately applied amplitude offset that allows the cross-point to be determined through linear fitting. The measured data $P_{|1\rangle,\pm}(A, n)$ are then fitted to extract the intersection point Δ_A^*, which corresponds to the applied offset value at which both mirrored sequences yield equal populations. The new calibrated amplitude is obtained as

$$A_{\text{new}} = A_0 + \Delta_A^*. \tag{2.24}$$

Close to the solution, the change of the rotation angle can be linearized as $\theta(A) - \theta_0 \approx \beta(A - A^*)$, where β is a proportionality constant that quantifies the local sensitivity of the rotation to the pulse amplitude. Substituting this into Eq. (2.22) gives the linear model used for the experimental fit

$$P_{|1\rangle,\pm}(A, n) \approx a_\pm\,(A - A^*) + b, \quad \text{with} \quad a_\pm = \pm\, c(n)\,\beta. \tag{2.25}$$

Equivalently, when expressed in terms of the scanned offset Δ_A, the fitting model becomes

$$P_{|1\rangle,\pm}(\Delta_A, n) = a_\pm\,(\Delta_A - \Delta_A^*) + b, \tag{2.26}$$

where Δ_A^* is the fitted cross-point on the Δ_A axis. This symmetric form ensures that the two mirrored curves intersect at the offset corresponding to the correct amplitude.

As in the ping-pong calibration, the cross-point method can be applied to both the X_π and $X_{\pi/2}$ pulses. For the X_π-pulse calibration the pulse sequence is defined as

$$(X_{\pi+\Delta})^n - X_{\pi/2}. \tag{2.27}$$

The corresponding mirrored sequence replaces all X_π pulses by phase-inverted pulses $-X_{\pi+\Delta}$, which are implemented experimentally through virtual Z gates (or $R_Z(\pi)$-rotations) before and after the block of $X_{\pi+\Delta}$ pulses. Similarly, for the $X_{\pi/2}$ (or SX) calibration the sequence

$$(X_{\pi/2+\Delta})^n - X_{\pi/2} \tag{2.28}$$

is used, where the rough amplitude estimate for the $X_{\pi/2}$ pulses is taken as half the amplitude of the calibrated X_π pulse with identical duration and shape.

An example of the calibration process is illustrated in Figure 2.6. On the left, the mirrored sequences are represented as Bloch-sphere trajectories, while the right panel shows the measured excited-state population as a function of the applied amplitude offset Δ_A. The intersection of the two mirrored curves marks the correction value Δ_A^* for the initial amplitude A_0.

The cross-point calibration is a conceptually simple and experimentally efficient alternative to trigonometric fitting–based methods such as the ping-pong calibration. The expected advantages of this method are the partial cancellation of SPAM-related offsets through the symmetric sequence construction, the linear relation between the measured signal and small rotation errors, and the use of a straightforward linear or zero-crossing fit instead of nonlinear regression. Unlike ping-pong calibration, where the

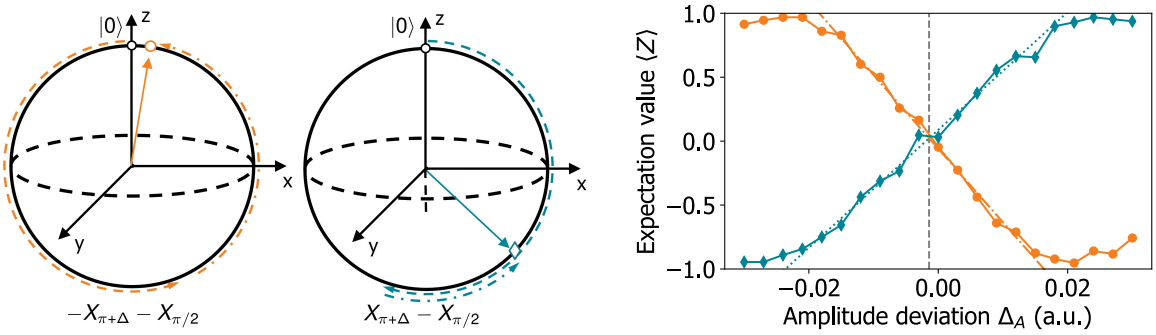


Figure 2.6: *Schematic of the amplitude-calibration (cross-point) error-amplification protocol. (a) Bloch-sphere depiction of the mirrored error-amplifying sequences. (b) Expectation value $\langle Z \rangle$ (equivalently the excited-state population $P_{|1\rangle} = (1 - \langle Z \rangle)/2$) as a function of the applied amplitude offset Δ_A. The orange points indicate the negative sequence S_- and the green diamonds indicate the positive sequence S_+. The dash-dotted orange and dotted green lines indicate the linear fits. The intersection (vertical dashed line) corresponds to the correction Δ_A^* that yields the desired rotation.*

amplitude is updated empirically until a δ threshold is reached, the optimal amplitude for cross-point can be directly extracted from the experiment. Consequently, the method is relatively robust against slow drifts and typically converges fast. However, the linear approximation might limit its dynamic range. For large initial miscalibrations or in the presence of strong decoherence, the mirrored responses may deviate from perfect antisymmetry, which reduces the accuracy of the extracted cross-point. Determining which method, cross-point or ping-pong, delivers better performance under realistic experimental conditions is still unresolved and will be analyzed in Section 3.1.

2.3.2 The Controlled-Not gate and its associated Hamiltonian

Two-qubit gates create entanglement between qubits and are essential for implementing quantum algorithms. In fixed-frequency superconducting transmon devices, such as the *IBM Eagle* architecture used in this work, the CNOT gate is realized via the CR gate. The CR gate uses a microwave drive applied to the control qubit at (or near) the resonance frequency of the target qubit [72]. Through the fixed coupling between neighboring qubits (typically mediated by a resonator), this drive induces an effective conditional interaction on the target. The concept behind this gate implementation is simple, as it requires the same type of electronic control as single-qubit gates.

The effective interaction generated by the CR drive is usually modeled by the reduced Hamiltonian [65]

$$H_{\mathrm{CR}} = a_{IX}\, I\otimes X + a_{IY}\, I\otimes Y + a_{IZ}\, I\otimes Z + a_{ZX}\, Z\otimes X + a_{ZY}\, Z\otimes Y + a_{ZZ}\, Z\otimes Z, \quad (2.29)$$

where X, Y, and Z are single-qubit Pauli matrices and I denotes the identity. Operators of the form $A \otimes B$ represent two-qubit Pauli operators, the first acting on the control, and the second on the target qubit, and a_{ab} is the corresponding interaction strength. The desired entangling interaction is $Z \otimes X$, the remaining terms represent single-qubit

drives ($I \otimes X$, $I \otimes Y$, $I \otimes Z$) acting on the target qubit and residual couplings ($Z \otimes Y$, $Z \otimes Z$) that arise from finite detuning, anharmonicity, and coupling. It is assumed that the control qubit remains in its computational state, without resonant $|0\rangle \leftrightarrow |1\rangle$ transitions, due to the drive being off-resonant with the control's transition frequency.

The idealized CR interaction is described by the Hamiltonian

$$H_{CR} = a_{ZX} Z \otimes X, \tag{2.30}$$

where Z and X are Pauli operators, the first acting on the control, and the second on the target qubit, and a_{ab} is the corresponding interaction strength. Under this interaction, the target undergoes a rotation about X conditioned on the control's Z state, yielding the unitary

$$CR(\phi) = e^{-i(\phi/2) Z \otimes X}, \tag{2.31}$$

with the rotation angle ϕ set by the drive amplitude, phase, and duration [73].

Additional terms that appear in Equation (2.29) must be mitigated. A widely used strategy is the echoed CR (ECR) gate [65], which divides the CR drive into two halves with opposite sign and inserts a π pulse on the control to refocus errors. The decomposition, that is locally equivalent to a CNOT is

$$CNOT = CR(-\pi/4)(X \otimes I)\; CR(\pi/4)\, U_1 \otimes U_2, \tag{2.32}$$

where U_1 and U_2 are single-qubit unitary operations. A cancellation pulse on the target qubit is added to suppress unwanted single-qubit drive terms, so that a_{IX} and a_{IY} are compensated, while the echoed structure strongly suppresses a_{IZ} and reduces sensitivity to the static a_{ZZ} term.

A key calibration is the drive phase, which sets the orientation of the conditional target drive in the $X - Y$ plane. A relative phase Φ between the $Z \otimes Y$ and $Z \otimes X$ components induces a rotation in the $Z \otimes Y$-$Z \otimes X$ subspace. This mixing can be expressed as

$$a_{ZX}(\Phi) = a_{ZX} cos(\Phi) + a_{ZY} sin(\Phi)\,,\; a_{ZY}(\Phi) = -a_{ZX} sin(\Phi) + a_{ZY} cos(\Phi). \tag{2.33}$$

By calibrating Φ, such that $a_{ZY}(\Phi) \approx 0$, the interaction is aligned along $Z \otimes X$. If instead the interaction is aligned along $Z \otimes Y$ (e.g. $\Phi = 90°$), a local Z rotation on the target maps $Y \to X$.

The main advantages of this echoed construction are its robustness and insensitivity to small calibration drifts. It is therefore well suited for large-scale experiments where long-term stability is required. The disadvantages are that the echo increases the effective gate length and introduces additional single-qubit operations, which limit the achievable gate speed and expose the system more strongly to decoherence.

An alternative approach is the direct or echo-free CR gate [62]. In this scheme, a single CR pulse implements the entangling operation without the use of intermediate inversion pulses. The cancellation of parasitic terms is instead realized by active calibration methods. A continuous microwave tone applied to the target qubit (a rotary drive) can be used to cancel residual $I \otimes X$ and $I \otimes Y$ terms from the cross-resonance interaction,

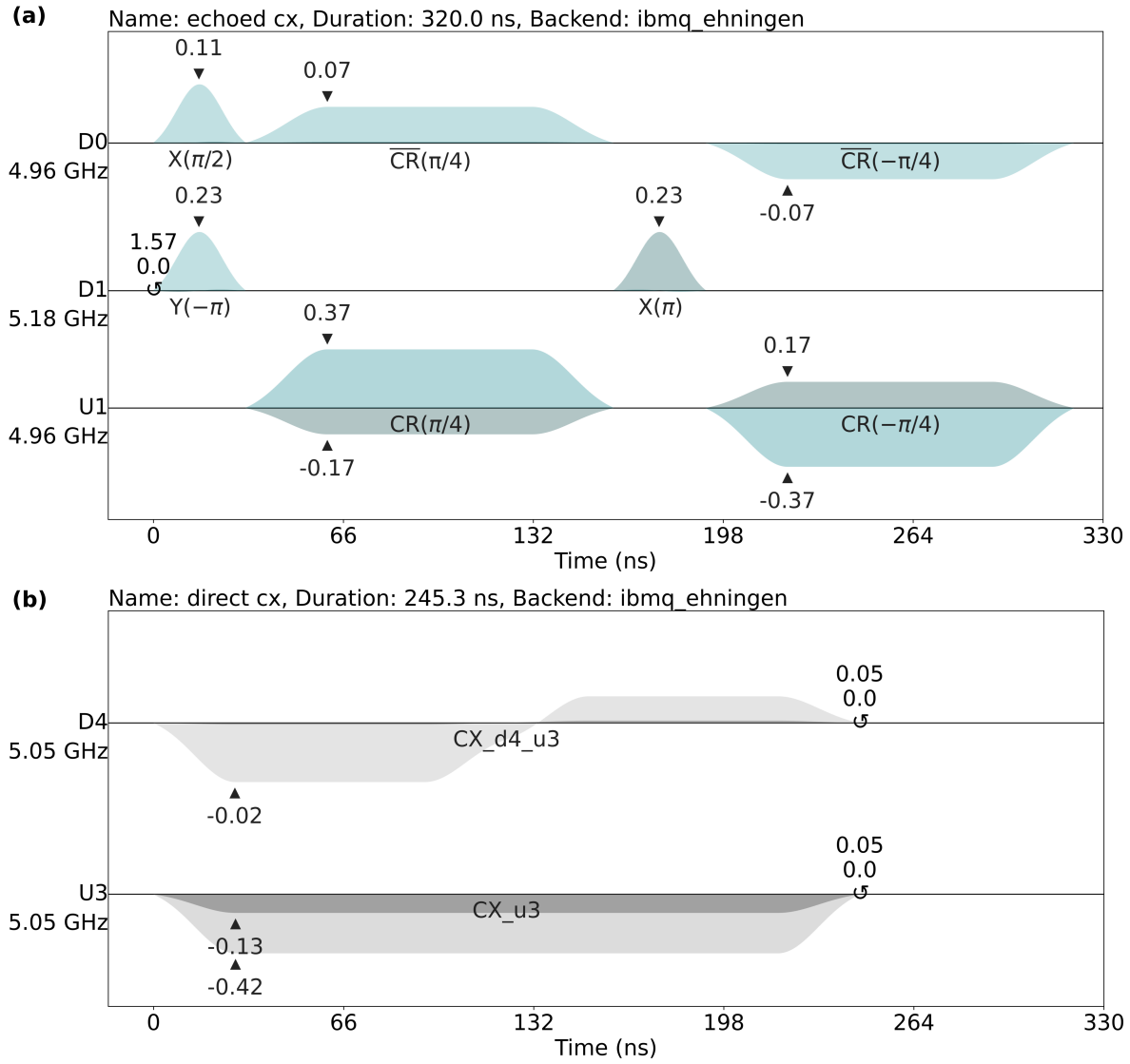


Figure 2.7: *Comparison of a echoed and direct CNOT gate. Pulse-level implementation of (a) an echoed CNOT (320 ns) between qubits 1 and 0 and (b) a direct CNOT (245.3 ns) between qubits 1 and 4 on ibmq_ehningen. The CR drive is applied on the control qubit's channel (U1/U3), while single-qubit operations and auxiliary tones are applied on the drive channels (D0/D1/D4). In the direct CNOT, the additional pulse on the target qubit (`CX_d4_u3`) serves both as a cancellation tone to remove unwanted terms and as a rotary drive to stabilize the target frame during the gate.*

while also compensating for other coherent errors [62, 73]. If tuned carefully, this allows for a shorter and potentially higher-fidelity two-qubit gate.

The direct implementation reduces gate time and thus the exposure to decoherence. However, it is significantly more sensitive to calibration errors and drift [62]. Moreover, residual $Z \otimes Z$ interactions and spectator effects are less effectively suppressed, and the method tends to show a faster degradation of fidelity when the pulse duration is reduced aggressively. The overhead of frequent recalibration could therefore be higher compared to the echoed case.

The two different implementations of the CNOT gate are depicted in Figure 2.7 at the pulse level with their respective pulse envelopes, i.e. the control signals sent to

the qubits. Figure 2.7 (a) represents an echoed CNOT gate between qubits 1 and 0 on *ibmq_ehningen* with a duration of 320 ns, and Figure 2.7 (b) a direct CNOT gate between qubits 1 and 4 with a duration of 245.3 ns.
The structure of the pulse sequences [74] are as follows: The horizontal axis indicates the pulse duration. Each row corresponds to a hardware channel, with the frequency of the applied microwave pulse indicated. The drive channels (D0, D1, D4) carry signals that directly manipulate the qubit states. The control channels (U1 and U3) use the same physical lines, though they are used to represent pulses for multi-qubit interactions (namely, the CR drive). They are also specially marked to distinguish them from single-qubit drives. For the drive channels, the pulse frequency is set to the resonance frequency of the addressed qubit (i.e., resonant single-qubit control), whereas for the control channels, the control qubit is driven at the frequency of the target qubit (U1: qubit 1 is driven at the frequency of qubit 0; U3: qubit 1 is driven at the frequency of qubit 4).
The colored shapes are *GaussianSquare* pulses, i.e. square pulse with Gaussian-shaped rising and falling edges. The small black arrows denote calibrated pulse amplitude, and circular arrows indicate virtual Z-rotations (phase or frame shifts). The labels identify the applied pulses, with rotation angles in brackets (e.g. $X(\pi/2)$ corresponds to a rotation around the x-axis). Pulses may be positive (above the axis) or negative (below the axis), corresponding to a 180° phase shift and thus driving the qubit in the opposite direction. The lighter color refers to the real (in-phase) and the darker to the imaginary (quadrature-phase) components of the waveform.

In the echoed CNOT (Figure 2.7 (a)), the CR drive is split into two halves of opposite amplitudes with an intermediate π-pulse on the control qubit, which cancels unwanted single-qubit terms. In contrast, the direct CNOT (Figure 2.7 (b)) employs a single continuous CR pulse on the control qubit (U3). Simultaneously, an additional pulse is applied to the target qubit (D4), labelled `CX_d4_u3`. This auxiliary drive acts as a cancellation tone to compensate the unwanted $I \otimes X$ and $I \otimes Y$ rotations induced by the CR interaction, while also functioning as a rotary drive that stabilizes the target qubit's frame throughout the gate. Together, these simultaneous drives implement the entangling operation without the need for an echo sequence.

2.4 Algorithm- and software-level error mitigation techniques

The high error rates of quantum gates in current Noisy Intermediate-Scale Quantum (NISQ) devices [75] make error mitigation essential. In the long term, these errors are expected to be addressed through Quantum Error Correction (QEC) [75, 76]. However, QEC requires a large overhead, often exceeding 1,000 physical qubits per logical qubit [77], which remains impractical given the limited qubit counts of present-day hardware.

Quantum Error Mitigation (QEM) offers a near-term alternative by reducing the noise-induced bias in measured expectation values without using full QEC. It combines repeated circuit executions with classical post-processing to estimate noise-free observables, while the noise level of individual circuit runs remains unchanged or may even be deliberately

increased. These techniques complement ongoing hardware improvements such as enhanced gate fidelities and reduced decoherence.

Several error-mitigation strategies requiring fewer resources than QEC have been proposed. Among them, Zero-Noise Extrapolation (ZNE) estimates results in the limit of vanishing noise [45, 46, 75], and measurement-error mitigation corrects readout errors through calibration-matrix inversion [51, 53]. Randomized compiling converts coherent errors into stochastic Pauli errors [31, 47]. Other approaches include probabilistic error cancellation [38] and virtual distillation [44, 78].

These methods differ in their resource demands, assumptions about the noise model, and achievable suppression levels. In this work, the focus is on ZNE and randomized compiling, which are experimentally accessible on current superconducting-qubit platforms. The following sections provide an overview of these techniques and their implementation: Subsection 2.4.1 introduces ZNE with its folding schemes, and Subsection 2.4.2 describes the randomized compiling protocol used in this work.

2.4.1 Zero-Noise Extrapolation

Zero-Noise Extrapolation [38, 45] is an error-mitigation technique that estimates the ideal expectation value of an observable A by deliberately increasing the noise in a quantum circuit and extrapolating the measured results to the zero-noise limit. The central idea is to evaluate the same circuit under several controlled noise-scaling conditions, described by scaling factors $\lambda_i \geq 1$, which effectively amplify the strength of the native noise processes while leaving the logical computation unchanged. For each scaling factor λ_i, the corresponding expectation value $\langle A \rangle_{\lambda_i}$ is measured, and these data points are fitted with a function that models the dependence of $\langle A \rangle_\lambda$ on λ. Extrapolating this function to the noiseless limit $\lambda \to 0$ yields an estimate of the ideal expectation value $\langle A \rangle_{\text{ideal}} \approxeq \langle A \rangle_{\lambda \to 0}$. ZNE protocols differ in how the noise is amplified (for example, through gate folding [46, 75] or pulse stretching [38, 79]), in the number of noise-scaled circuits that are executed, and in the choice of extrapolation function, such as linear, polynomial, or exponential fits.

In this work, a noise-model agnostic approach for scaling the noise strength is used, in which unitary operations U are replaced by $U(U^\dagger U)^n$ [46, 75], corresponding to n repetitions of the operation. The additional operation $U^\dagger U$ does not modify the logical action of the circuit, since $U^\dagger U = \mathbb{I}$ corresponds to the identity in the absence of noise. The unitary folding techniques can be applied either locally or globally: in local folding, individual gates (e.g., CNOT gates) are folded, whereas in global folding the entire circuit U is repeated as a unitary block. While global folding amplifies all noise sources uniformly, it can become impractical for circuits with high intrinsic noise levels, as the accumulated errors increase rapidly with the circuit depth. The effects of these two schemes are examined in Section 4.10.
Errors occurring in a noisy implementation of U, $\mathcal{E}_U$, are amplified by a factor

$$\lambda = 2n + 1. \tag{2.34}$$

Assuming a depolarizing noise channel Λ_p with error probability p, which commutes with all unitaries, the noisy implementations of U and $U^\dagger$ can be written as

$$\mathcal{E}_U = \Lambda_p \, \mathcal{U}, \tag{2.35}$$
$$\mathcal{E}_{U^\dagger} = \Lambda_p \, \mathcal{U}^\dagger, \tag{2.36}$$

where $\mathcal{U}(\rho) = U\rho U^\dagger$ and $\mathcal{U}^\dagger(\rho) = U^\dagger \rho U$ denote the unitary quantum channels corresponding to U and $U^\dagger$, respectively. The depolarizing channel is defined as

$$\Lambda_p(\rho) = (1-p)\rho + p\rho_*, \tag{2.37}$$

where $\rho_* = \mathbb{I}/2^q$ denotes the maximally mixed state of a q-qubit system and $\mathbb{I}$ is the identity operator.

Under these assumptions, the noisy implementation of the folded sequence $U(U^\dagger U)^n$ again results in a depolarizing error, now with a scaled error probability λp to first order in p [46, 75]. This scaling arises from the fact that depolarizing noise commutes with unitary operations, $\Lambda_p \, \mathcal{U} = \mathcal{U}\Lambda_p$ and $\Lambda_p \, \mathcal{U}^\dagger = \mathcal{U}^\dagger \Lambda_p$, which yields

$$\begin{aligned} \mathcal{E}_U \left(\mathcal{E}_{U^\dagger}\mathcal{E}_U\right)^n &= \Lambda_p \, \mathcal{U} \left(\Lambda_p \, \mathcal{U}^\dagger \Lambda_p \, \mathcal{U}\right)^n \\ &= \Lambda_p^{2n+1} \, \mathcal{U} \left(\mathcal{U}^\dagger \mathcal{U}\right)^n \\ &= \Lambda_{1-(1-p)^{2n+1}} \, \mathcal{U}. \end{aligned} \tag{2.38}$$

We thus recover the same structure as in Equation (2.35), but with the depolarizing probability rescaled as $p \to 1-(1-p)^{2n+1} \approx \lambda p$ ($\lambda p \ll 1$), where $\lambda = 2n+1$. For other error models, e.g. in the presence of coherent or gate-dependent errors, such a simple scaling behavior does not necessarily hold. In particular, for coherent errors, the noise channels $\mathcal{E}_U$ and $\mathcal{E}_{U^\dagger}$ may partially or completely cancel, such that $\mathcal{E}_{U^\dagger}\mathcal{E}_U = \mathbb{I}$ and no effective noise amplification occurs. This effect will be discussed in Section 4.4.

With this folding-based method, only odd-valued amplification factors are possible, that is $\lambda \in 1, 3, 5, \ldots$. To achieve non-integer scaling factors, the insertion of identity operations can be distributed randomly across the circuit [46, 75]. In this case, for a target amplification factor, each gate in the circuit, for example a CNOT gate, is independently folded with a certain probability. For each gate n_i, the local amplification factor becomes

$$\lambda_i = 2n_i + 1. \tag{2.39}$$

The global, or effective, amplification factor is then given by the average over all gates,

$$\lambda = \frac{1}{N}\sum_{i=1}^{N} \lambda_i, \tag{2.40}$$

where N is the total number of gates in the circuit. By controlling the probability that a given gate is folded, arbitrary effective scaling factors can be realized. For example, folding 25% of all gates results in an average amplification factor of $\lambda = 1.5$. Averaging over many random folding realizations of the same circuit yields an overall

noise amplification equivalent to deterministic folding, while enabling continuous and fractional values of λ and leading to a more uniform distribution of noise across the circuit. This stochastic approach is particularly useful for fine-grained noise-scaling experiments and for smooth extrapolation in ZNE protocols, which will be used in Section 4.8.

Due to its simplicity and moderate sampling overhead, the ZNE method based on identity insertions is frequently used to mitigate gate errors. We will now describe a common, simple implementation. It will subsequently be referred to as standard ZNE (sZNE) in Chapter 4. We focus on the two-qubit gate (CNOT) errors, which are the dominant source of errors in the *IBM* superconducting device used in this work (i.e., one or two orders of magnitude larger than single-qubit gate errors) [64, 65, 80–82]. We use noise scaling factors $\lambda = 1$ (the original circuit), $\lambda = 3$ (every CNOT gate replaced by 3 CNOT gates) and $\lambda = 5$ (5 CNOT gates) and measure the expectation value $\langle A \rangle$ for each value of λ.

Within the depolarizing noise model, we obtain

$$\langle A \rangle = (1-p)^{\lambda} \langle A \rangle_{\text{ideal}} + [1-(1-p)^{\lambda}] \langle A \rangle_{*}, \tag{2.41}$$

where $\langle A \rangle_{*} = \text{Tr}[A\rho_{*}]$ with ρ_{*} as the completely mixed state and p the probability of an error occurring. With the small noise approximation, using the first-order Taylor expansion, we get

$$\langle A \rangle \approxeq (1-\lambda p) \langle A \rangle_{\text{ideal}} + p\lambda \langle A \rangle_{*}, \tag{2.42}$$

for $\lambda p \ll 1$.

Motivated by Equation (2.41) the measured data points of sZNE are fitted by an exponential function

$$f(\lambda) = a_1 e^{-a_2 \lambda} + a_3, \tag{2.43}$$

which is expected to yield more accurate results than a linear fit [81]. The fit parameters a_1 and a_3 are restricted to the interval $[A_{min}, A_{max}]$ given by the minimum and maximum eigenvalue of A.

While Equation (2.41) is derived under the depolarizing noise model, in practice the scaling of the error strength may differ, as will be discussed in Chapter 4.

2.4.2 Randomized Compiling

Randomized compiling [47] is a method in quantum computing that aims to mitigate the influence of noise and operational errors which occur during the execution of quantum gates. The essential idea is to insert random unitary operations, usually taken from the Pauli group, which preserve the logical computation while averaging out systematic error biases. By averaging over a large ensemble of randomized circuits, this approach can increase the overall fidelity of a quantum computation while requiring only a small amount of additional resources [48, 83–87].

In the following we introduce the concept of Pauli twirling [47], where only Pauli gates are inserted. The principle of Pauli twirling relies on the fact that the CNOT gate is a

two-qubit entangling Clifford operation [88]. As a result, every Pauli operator is mapped onto another Pauli operator when it is conjugated by a CNOT gate

$$\text{CNOT}\ \mathcal{P}_i\ \text{CNOT} = \mathcal{P}_{\pi(i)}, \tag{2.44}$$

where $\mathcal{P}_i$ with $i = 1, 2, \ldots, 4^q$ denotes multi-qubit Pauli operators, that is, tensor products of the single-qubit Pauli operators I, X, Y or Z acting on each qubit, and π represents the corresponding permutation. Since $\text{CNOT}^\dagger = \text{CNOT}$ and $\mathcal{P}_i^2 = \mathbb{I}$, one can write

$$\mathcal{P}_i\ \text{CNOT}\ \mathcal{P}_{\pi(i)} = \text{CNOT}. \tag{2.45}$$

Applying two Pauli operators $\mathcal{P}_i$ and $\mathcal{P}_{\pi(i)}$ before and after a CNOT gate with a randomly chosen index i therefore preserves the logical action of the CNOT in the absence of noise.

In the presence of noise, the error channel Λ of a noisy CNOT gate $\mathcal{E}_{\text{CNOT}} = \Lambda \circ \text{CNOT}$ is transformed, on average, into a Pauli channel $\Lambda_{\mathcal{P}}$ of the form

$$\Lambda_{\mathcal{P}}(\rho) = \sum_{i=1}^{4^q} p_i \mathcal{P}_i \rho \mathcal{P}_i, \quad p_i > 0, \quad \sum_{i=1}^{4^q} p_i = 1. \tag{2.46}$$

In this expression, the Pauli operator $\mathcal{P}_i$ is applied with probability p_i and produces an error whenever $\mathcal{P}_i \neq \mathbb{I}$.

Twirling can also be extended to neighboring qubits that are subject to crosstalk effects [89, 90], as discussed in Ref. [91]. Twirling randomizes the coherent interactions between qubits, transforming them into stochastic, uncorrelated Pauli-type errors that do not accumulate coherently throughout the circuit, thereby reducing crosstalk [92]. The twirling process is shown in Figure 2.8. Applying the Pauli twirling gates $\mathcal{P}_i$ and $\mathcal{P}_{\pi(i)}$ before and after each CNOT gate is shown in the middle panel of Figure 2.8. Combining these gates with adjacent single-qubit operations avoids an increase in circuit depth and execution time (see Figure 2.8, right).

If the noisy CNOT gate originally suffers from a coherent error, meaning its error channel Λ is unitary, Pauli twirling converts this error into an incoherent one. As mentioned in Subsection 2.4.1, the scaling behaviour of the noise arising from repeated CNOT gates is expected to be more predictable in this case.

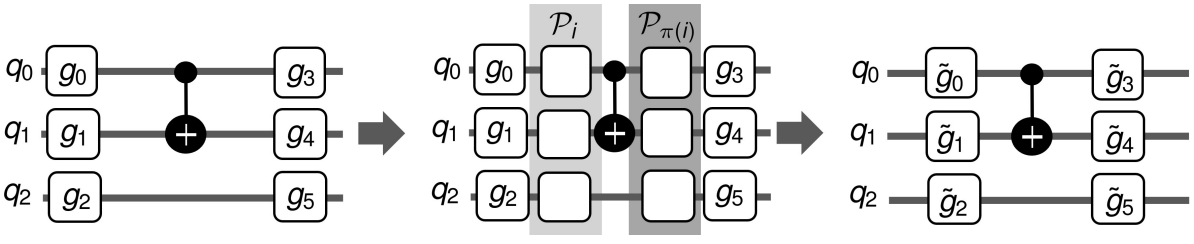


Figure 2.8: *Illustration of the Pauli twirling process. On the left, a circuit with single-qubit gates (g_j) and a single CNOT gate is shown. In the middle, random Pauli gates ($\mathcal{P}_i$, $\mathcal{P}_{\pi(i)}$) are inserted before and after each CNOT gate, as well as on adjacent qubits that may experience crosstalk, without altering the logical action of the circuit. On the right, consecutive single-qubit gates are merged into combined gates ($\tilde{g}_j$) to minimize the circuit depth.*

Therefore, randomized compiling provides a practical and theoretically grounded approach to suppressing coherent errors, effectively transforming them into stochastic noise that

can be more easily characterized and mitigated. The Pauli twirling scheme described here will be implemented in all experimental circuits presented in Chapter 4, unless otherwise stated. Including it ensures that the measured noise behavior is dominated by stochastic rather than coherent components. This enables a consistent analysis of the error-mitigation techniques discussed in this work.

2.5 Variational quantum algorithms

Quantum algorithms promise exponential speedups over classical methods for tasks such as integer factorization [2], the simulation of quantum systems [93] and the solution of linear systems of equations [94]. However, the presence of noise, limited qubit numbers and restricted connectivity continues to prevent practical implementations of these algorithms on NISQ devices [95].

To make use of current quantum hardware, variational quantum algorithms (VQAs) have become the most widely adopted approach. They are designed specifically to operate within the constraints of NISQ systems by combining parametrized quantum circuits with classical optimization. The quantum circuits prepare states with parameters that are updated iteratively on a classical computer. This keeps the circuit depth shallow and mitigates the effects of noise. VQAs therefore have the potential to provide a practical pathway toward demonstrating quantum advantage with imperfect hardware.

In this section two prominently used VQAs are presented. These algorithms are applied in Section 3.3 to compare two CNOT implementations and in Chapter 5 to evaluate the improvements obtained through error mitigation techniques. The algorithms discussed here are the Quantum Approximate Optimization Algorithm (QAOA) in Subsection 2.5.1 and the Variational Quantum Eigensolver (VQE) in Subsection 2.5.2.

2.5.1 Quantum Approximate Optimization Algorithm

The QAOA is a hybrid quantum-classical approach designed to approximate the optimal solution of combinatorial optimization problems [96]. It combines a parameterized quantum circuit with a classical optimization routine, creating an iterative feedback loop between parameterized quantum state preparation and classical parameter adjustment.

The quantum circuit in QAOA is composed of three main components: state preparation, the cost Hamiltonian (or phase separator), and the mixer Hamiltonian. The overall structure and details of these components depend on the specific problem formulation and the chosen ansatz.

State Preparation
The first step in the QAOA process is to prepare a reference state $|\Psi_0\rangle$, which is straightforward to implement on the quantum device. In the standard formulation, $|\Psi_0\rangle = |+\rangle^{\otimes n}$, which represents a uniform superposition over computational basis states. For constrained or problem-tailored variants, $|\Psi_0\rangle$ may instead be chosen to lie within the feasible subspace, for example by preparing the ground state of the chosen mixer [97, 98].

Cost Hamiltonian (Phase Separator)
The cost function of the optimization problem is encoded into the cost Hamiltonian

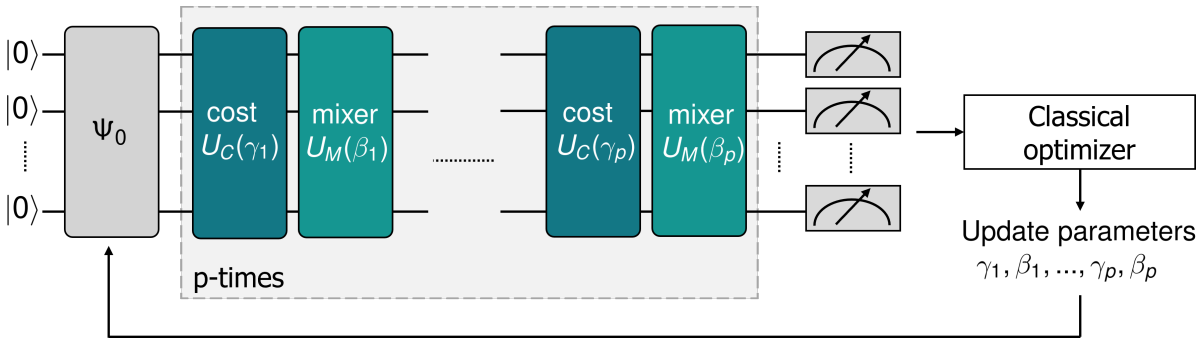


Figure 2.9: *Structure of QAOA. Each qubit is initialized in the state $|0\rangle$ and prepared in the initial state $|\Psi_0\rangle$. The cost and mixer unitaries $U_C(\gamma_i)$ and $U_M(\beta_i)$ are then applied p times. After measurement, the resulting expectation value is evaluated on a classical computer to update the parameters γ_i and β_i. This hybrid optimization cycle is repeated until convergence or until a predefined stopping criterion is reached.*

H_C, also referred to as the phase separator. This operator applies phase shifts to the quantum state according to the problem-specific cost function and depends on a tunable parameter γ. In the circuit representation, this operation is implemented as the unitary $U_C(\gamma_i) = e^{-i\gamma_i H_C}$.

Mixer Hamiltonian

The mixer Hamiltonian H_M enables transitions between different computational basis states, allowing the algorithm to explore the search space and satisfy problem constraints. It introduces an additional tunable parameter, denoted β. In the circuit representation, it is denoted as $U_M(\beta_i) = e^{-i\beta_i H_M}$.

The parameters (γ, β) are optimized by a classical optimizer, which minimizes the expectation value of the problem Hamiltonian. For a circuit with QAOA depth p, the cost and mixer Hamiltonians are applied alternately p times, each with their own pair of parameters (γ_i, β_i). The total number of variational parameters is therefore $2p$, forming the sequence $\gamma_1, \beta_1, ..., \gamma_p, \beta_p$. Larger values of p allow the QAOA circuit to better approximate the optimal solution of the combinatorial problem, while smaller depths provide coarser approximations with reduced circuit complexity [96].

The optimization process proceeds iteratively. In each iteration, the quantum circuit prepares a parameterized quantum state $|\Psi(\beta, \gamma)\rangle$, and the corresponding expectation value of the energy is computed as [71]

$$E(\beta, \gamma) = \langle \Psi(\beta, \gamma) | H | \Psi(\beta, \gamma) \rangle, \tag{2.47}$$

where H is a Hamiltonian (usually $H = H_C$,i.e., the cost Hamiltonian). A classical computer updates the parameters β and γ to minimize Equation (2.47),

$$E_{\min} = \min_{\beta,\gamma} E(\beta, \gamma). \tag{2.48}$$

Each such iteration of state preparation, measurement, and parameter update is referred to as a run. The hybrid optimization cycle is repeated until convergence or until a predefined stopping criterion is met. Figure 2.9 illustrates the general structure of the QAOA algorithm.

2.5.2 Variational Quantum Eigensolver

VQE is a hybrid quantum algorithm widely used for quantum-chemistry calculations on a quantum computer to estimate molecular ground-state energies [99]. In VQE, a parametrized quantum circuit prepares trial states on a quantum device, while a classical optimizer updates the circuit parameters to minimize the measured energy expectation value. Due to its variational structure, VQE can be implemented using relatively shallow circuits when an appropriate ansatz is chosen. This makes VQE well-suited to near-term quantum hardware [54].

For a given molecule, the electronic structure problem is mapped to an effective qubit Hamiltonian via a fermion-to-qubit transformation, such as the Jordan-Wigner or Bravyi-Kitaev mappings [100–102]. The Hamiltonian operator H is thus expressed as a weighted sum of Pauli operators P_i

$$H = \sum_i h_i P_i, \tag{2.49}$$

where the coefficients h_i encode the information about the molecular system under consideration.

The VQE algorithm minimizes the expectation value of the Hamiltonian with respect to a parametrized quantum state. The energy function is

$$E(\vec{\theta}) = \langle \Psi(\vec{\theta}|H|\Psi(\vec{\theta})\rangle, \tag{2.50}$$

where the trial state $|\Psi(\vec{\theta})\rangle$ is generated by a parametrized unitary ansatz $U(\vec{\theta})$ that acts on an initial reference state $|\Psi_0\rangle$ [95]. The vector $\vec{\theta}$ contains all variational parameters of the ansatz. Typical ansätze include hardware-efficient layers or chemically motivated unitary coupled-cluster [103–105]. Because each Pauli operator P_i can be measured independently on the quantum device, the total energy can be reconstructed by summing the measured expectation values weighted by the corresponding coefficients h_i. In practice, commuting Pauli terms are often grouped to reduce the number of measurements, while error mitigation techniques can be employed to improve the accuracy of the estimated energy [54, 79, 106, 107].

By evaluating the ground-state energy for different internuclear distances R, VQE can be used to construct molecular potential energy curves, from which the equilibrium distance R_0 is obtained as the location of the energy minimum.

The algorithm proceeds in iterative steps. First the quantum device prepares the state $|\Psi(\vec{\theta})\rangle$. Second the expectation values of the Pauli operators P_i are measured and used to evaluate $E(\vec{\theta})$. Third a classical optimizer updates the parameters $\vec{\theta}$ to reduce the energy. The procedure is repeated until convergence. The workflow is shown in Figure 2.10.

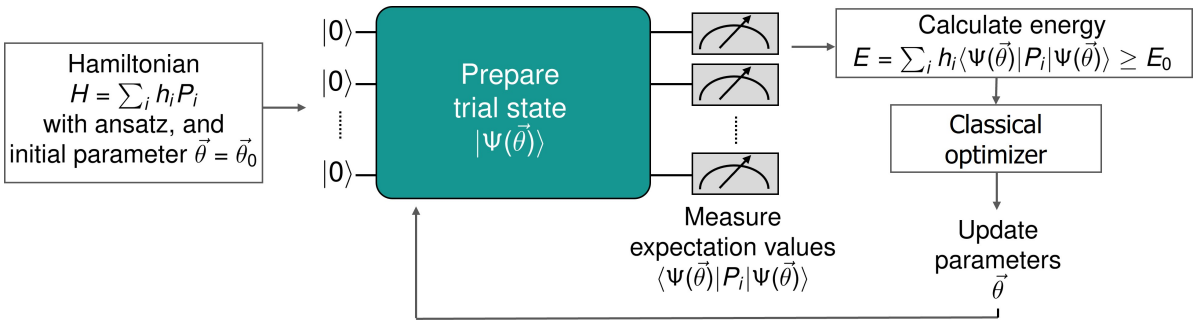


Figure 2.10: *Structure of VQE. Each qubit is initialized in the state $|0\rangle$ and the trial state $|\Psi(\vec{\theta})\rangle$ is prepared, starting with an ansatz and initial parameters $\vec{\theta}_0$. Expectation values of Pauli terms are measured and combined to compute $E(\vec{\theta})$ on a classical computer. A classical optimizer updates the parameters $\vec{\theta}$. This hybrid optimization loop repeats until convergence or until a predefined stopping criterion is met.*

The Rayleigh–Ritz variational principle [108, 109] provides the theoretical foundation of the method. For any normalized state $|\Psi\rangle$ expressed in the eigenbasis $|\Psi_i\rangle$ of the Hamiltonian with eigenvalues b_i one obtains

$$\langle\Psi|H|\Psi\rangle = \sum_i b_i |\langle\Psi_i|\Psi\rangle|^2 \geq b_{\min}, \tag{2.51}$$

where $b_{\min}$ is the smallest eigenvalue of H. Therefore

$$E(\vec{\theta}) \geq E_0, \tag{2.52}$$

where E_0 is the exact ground state energy. The estimated energy thus approaches the true ground state energy from above in the noiseless and infinite-shot limit [99].

3 Calibration, optimization and analysis of quantum gates at the pulse level

The performance of a quantum processor depends on how precisely its control pulses implement the intended unitary operations. Small errors in amplitude, detuning, or phase accumulate to cause gate infidelities, resulting in coherent over- or under-rotations, undesired Stark shifts, or leakage out of the computational subspace. Therefore, achieving reliable, high-fidelity gates requires continuous calibration and a clear mapping from experimental control parameters to the effective qubit Hamiltonian.

This chapter validates, and compares pulse-level calibrations for transmon qubits, moving from single-qubit to two-qubit operations and linking gate-level metrics to algorithmic performance. First, we examine single-qubit amplitude calibration techniques using two error-amplifying sequences in Section 3.1. We detail their sensitivity to state preparation and measurement (SPAM) errors and fit instability, and demonstrate how robust fitting and reduced shot counts can still yield stable results. We quantify the trade-offs between the number of shots and the number of rounds for each method.

Next, we examine the echoed Controlled-NOT (CNOT) gate, a two-qubit gate that enables entanglement and conditional logic operations, in Section 3.2. We examine the general form of a two-qubit Hamiltonian in the Pauli basis, identify the relevant interaction terms, and determine whether a reduced Hamiltonian model is sufficient. We validate the accuracy of the model and identify the dominant relevant terms through a combination of simulation and experimentation.

Finally, in Section 3.3 we compare two CNOT gate implementations. We characterize performance trade-offs, such as gate time, error rates, and crosstalk. Then we evaluate their effect on algorithmic output using the Quantum Approximate Optimization Algorithm (QAOA) as a benchmark.

3.1 Comparison of ping-pong and cross-point amplitude calibration for single-qubit gates

Part of the results in this chapter originated during the Qiskit Advocate Mentorship Program (QAMP) and were later extended in collaboration with Naoki Kanazawa (IBM Research–Japan).

In order to perform useful quantum computations, it is essential to use well-calibrated control pulses that minimize systematic errors and maintain high gate fidelity. Small

deviations in the drive amplitude lead to coherent over- or under-rotations, which accumulate as algorithmic error or manifest themselves as instability in benchmarking results. In Subsection 2.3.1, two amplitude-calibration techniques based on error-amplifying pulse sequences were introduced: the ping-pong and cross-point methods. Both rely on repeated gate application to amplify small coherent errors, but they differ in how the calibration parameter is extracted from the measured data.

In this section, a comparison is made between the performance of two calibration methods on single-qubit X_π and $X_{\pi/2}$ gates. First, we present representative datasets for ping-pong and cross-point calibration. Using these datasets, we demonstrate issues such as poor data quality affecting the trigonometric fit and nonlinear terms affecting the linear fit. These instances highlight the method's sensitivity to fitting noise, initial parameter guesses, and other errors. Subsequent experiments systematically vary the number of measurement shots and the number of calibration rounds to quantify their effect on the stability and accuracy of both approaches.

The data are fitted with the functions described in Subsection 2.3.1, see Equations (2.19) and (2.26). Both calibration methods start from the same rough amplitude estimate. For the ping–pong calibration, Python's `curve_fit` function is applied with suitable initial guesses and parameter bounds. For the cross–point calibration, an additional Savitzky–Golay filter [110] is applied to smoothen the experimental data before performing a linear fit. This preprocessing is justified since a local linear dependence is assumed in the cross-point model, and the Savitzky–Golay filter acts as a one-dimensional smoothing kernel. All experiments and analyses were implemented using the *Qiskit Experiments* library [70], which was extended by the author and Naoki Kanazawa to incorporate the cross-point calibration method.

Both calibration methods were evaluated on *ibmq_ehningen* using qubit 0 with 256 shots per data point. For the cross–point calibration, 21 Δ_A steps were used in the range $[-0.03, 0.03]$, while the ping–pong calibration used 14 gate repetitions. The cross–point method therefore requires more distinct amplitude settings, but it may need fewer shots per point to reach the same accuracy. This could result in similar total resource costs for both approaches, which will be examined later. Figure 3.1 presents the results from the first calibration round following the initial rough X_π-pulse amplitude estimate.

Although both calibrations use 256 shots per data point, the ping-pong sequence, see Figure 3.1 (a), produces an irregular oscillatory signal, which complicates a stable cosine fit. This can arise from undersampling of the oscillation period or from systematic effects such as SPAM offsets, residual detuning, and small deviations from a linear relationship between pulse amplitude and rotation angle. These distortions modify the sinusoidal shape and can cause the nonlinear fit to fail.

In contrast, the cross-point calibration, see Figure 3.1 (b), yields two monotonic, mirrored responses whose intersection can be determined through a simple linear fit. As discussed in Subsubsection 2.3.1.3, for small amplitude errors the signal can be linearized, which justifies the local linear fit. However, when errors becomes large, the linear approximation no longer holds and the full periodic dependence reappears, producing sinusoidal tails at the edges of the sweep. To maintain the validity of the linear approximation, only the data points within a limited region around $\Delta_A = 0$ are included in the fit. This region can be determined automatically by a statistical criterion such as a χ^2 test or by comparing the residuals of linear fits over variable windows.

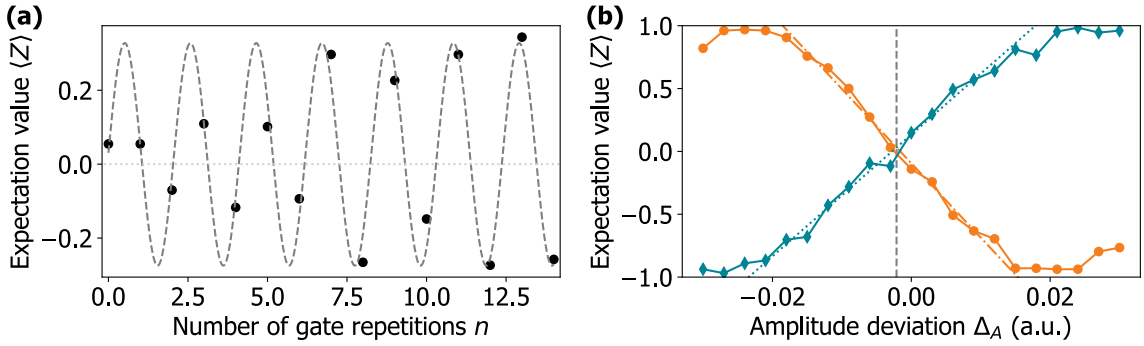


Figure 3.1: *Ping-pong and cross-point calibration experiment for a X_π-pulse. (a) Ping-pong experiment. The measured population (black circles) oscillates as a function of the number of gate repetitions n. The expected sinusoidal dependence appears to be distorted, but the trigonometric fit (dotted gray line) still converges to a consistent period that follows the data points. Some measurements deviate from the expected trend, illustrating the sensitivity of this method to noise (SPAM offsets, and residual detuning). (b) Cross-point experiment. The central region (near $\Delta_A = 0$) shows two nearly linear, oppositely sloped traces used to determine the cross-point Δ_A^* (gray dashed line). At larger $|\Delta_A|$, the traces regain a periodic dependence as the accumulated rotation $n\theta(A)$ approaches $\mathcal{O}(1)$ radians and wraps modulo 2π.*

Table 3.1: *Fitting parameters for ping-pong and cross-point calibration according to the models in Subsection 2.3.1, Equations (2.19) and (2.26) for the datasets shown in Figure 3.1.*

ping-pong	a	β	φ_0	b
	0.301	3.040	-1.557	0.027
cross-point	a_+	a_-	b	Δ_A^*
	-59.1	48.4	0.0314	-0.00213

Despite the less regular appearance of the ping-pong calibration data, both methods result in nearly identical amplitude values, yielding $A_{\text{new},\pi,p} = 0.1986$ for the ping-pong calibration and $A_{\text{new},\pi,c} = 0.1983$ for the cross-point calibration. The corresponding fit parameters according to the models in Subsection 2.3.1 are summarized in Table 3.1. The initial amplitude A_{old}, obtained from the rough calibration, was used for both experiments, and the updated amplitudes were calculated using Equations (2.21) and (2.24).

After having analyzed each calibration method with a single example, we now examine their behavior through experiments performed with different numbers of measurement shots. Figure 3.2 shows experiments with (a) 108, (b) 204, and (c) 300 shots per calibration round, each consisting of 50 independent repetitions. Both calibration methods were executed in two consecutive rounds: ping–pong (PP_1, PP_2) and cross–point (CP_1, CP_2). In the first round, PP_1 and CP_1 start from the amplitude obtained from the rough calibration, while in the second round, PP_2 and CP_2 use the amplitude extracted in the respective first round. For reference, the initial rough calibration results are labeled as "Rough." All experiments were carried out on the *ibmq_ehningen* backend.

We quantify the signal quality by the signal-to-noise ratio (SNR). The SNR is defined as the absolute mean μ of the data divided by its standard deviation σ

$$SNR = \frac{|\mu|}{\sigma}, \tag{3.1}$$

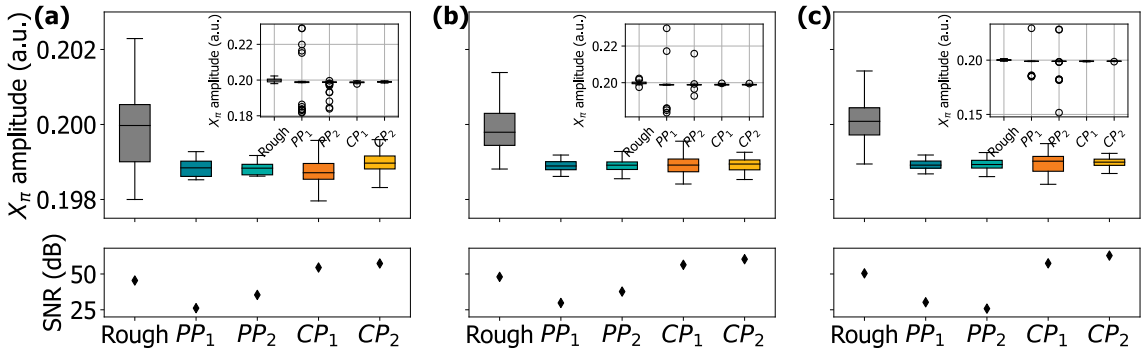


Figure 3.2: *Box plot of the extracted X_π-pulse amplitudes for different calibration methods: rough ("Rough"), ping–pong (PP_1, PP_2), and cross–point (CP_1, CP_2). Panels (a–c) correspond to datasets with 108, 204, and 300 shots per calibration round, respectively. Insets show the full distribution including outliers. The ping–pong calibration exhibits a higher number and magnitude of outliers compared to the cross–point calibration. The scatter plots below each box plot display the corresponding signal-to-noise ratios (SNR) in dB. Increasing the number of shots reduces the number of outliers, while the SNR values remain comparable across all experiments. The cross–point calibration consistently shows the highest SNR, although when outliers are excluded, its within-round variability appears slightly larger than ping-pong calibration. For higher shot counts, the rough calibration data becomes more stable and shows reduced variance.*

evaluated for 50 experiments. Table 3.2 shows μ and σ. A large SNR indicates that the average signal is well above the noise level, while a small SNR implies that random fluctuations dominate. For comparison between datasets, the SNR is expressed in decibel

$$SNR_{\mathrm{dB}} = 20\log_{10}(SNR). \tag{3.2}$$

The results are summarized in Figure 3.2 and Table 3.2. For an explanation of how box plots are composed, see Appendix A.3. As the number of measurement shots increases, the rough calibration results become more stable and show reduced variation (i.e., larger SNR). Cross–point calibration yields the lowest standard deviation among all methods and shot numbers, indicating higher repeatability of the extracted amplitudes. In contrast, the ping-pong calibration shows a larger variance and more pronounced outliers, particularly at low shot numbers, consistent with the higher sensitivity of the nonlinear trigonometric fit to noise and initial parameter guesses. Although increasing the number of shots reduces the statistical fluctuations in the measured amplitude values, this reduction is comparatively small. The remaining variations in the data are dominated by effects that are not improved by additional repetitions, such as small residual imperfections in the pulse amplitudes that remain. As a result, the SNR changes only marginally when the number of measurement shots is increased, and the main visible effect is a slight reduction in statistical outliers rather than a clear improvement in SNR.

Cross-point calibration appears to be more precise and robust, with a higher SNR, lower standard deviation, and fewer outliers. Ping-pong calibration, on the other hand, is more sensitive to noise and initial guesses.

Table 3.2: *Mean and standard deviation of the X_π-pulse amplitudes extracted from the box plots in Figure 3.2 with different numbers of measurement shots. Values are given as mean ± standard deviation.*

Method	108 shots	204 shots	300 shots
Rough	0.1999 ± 0.0011	0.19989 ± 0.00079	0.20009 ± 0.00060
PP_1	0.1987 ± 0.0097	0.1988 ± 0.0064	0.1982 ± 0.0060
PP_2	0.1977 ± 0.0033	0.1991 ± 0.0026	0.20 ± 0.010
CP_1	0.19875 ± 0.00037	0.19891 ± 0.00030	0.19897 ± 0.00026
CP_2	0.19895 ± 0.00027	0.19893 ± 0.00019	0.19898 ± 0.00014

SPAM primarily reduces contrast and introduces offsets in measured populations. This increases variance and can bias fits. Ping-pong calibration, with its sinusoidal fit, is more sensitive to SPAM-induced offsets or contrast changes and to local non-linearities, which explains the larger variance and outliers at low shot numbers. Cross-point calibration, with its intersection based fit, is less sensitive to global offsets or contrast and appears more robust, reflected in higher SNR and lower standard deviation. However, the current data do not separate SPAM from control or drive noise or drift. To estimate SPAM, a measurement calibration should be added with a report of the assignment error, and optionally a readout error mitigation should be applied.

To assess whether the observed distribution of extracted amplitudes is dominated by shot noise, we analyze the scaling of the standard deviation σ with the number of shots N. In the shot-noise limit, the standard deviation should follow $\sigma \propto 1/\sqrt{N}$, corresponding to a variance $\sigma^2 \propto 1/N$. Deviations from this scaling indicate additional, N-independent noise sources (e.g., control amplitude fluctuations, drift, or fit instability). We therefore fit the empirical variance to the model

$$\sigma^2(N) = \frac{a}{N} + b, \tag{3.3}$$

where a captures the shot-noise contribution and b represents a variance floor from technical noise that does not improve with more shots. The fitted parameters are summarized in Table 3.3. For the rough amplitude calibration $\sigma\sqrt{N}$ is nearly constant and the fitted floor $b \approx 0$, consistent with shot-noise limit behavior over all different shots. For PP_1, $\sigma\sqrt{N}$ is similar across all N, b is slightly negative due to fit uncertainty. This is broadly consistent with shot-noise-dominated behavior, though the non-monotonic pattern suggests some technical noise. For PP_2 the mean is much larger at 300 shots producing an unphysical negative a and a large positive floor. This indicates nonstationary behavior (drift, fit failure, or outliers) at 300 shots. The simple $a/N + b$ model does not capture this regime. For CP_1, $\sigma\sqrt{N}$ increases modestly with N and the fit yields a small positive floor ($b > 0$; 46% of the variance at $N = 300$). This suggests a combination of shot noise and an additional N-independent noise source (e.g., slow drift or control amplitude fluctuations). For CP_2, $\sigma\sqrt{N}$ is roughly constant and $b \approx 0$, indicating shot-noise-limited performance across the tested shot counts. Cross-point calibration is the most repeatable, approaching the shot-noise limit and exhibiting the smallest standard deviations. CP_1 shows evidence of a small technical noise floor. Ping-pong calibration is more sensitive to noise and fit stability. PP_2 displays a clear instability at higher shots, consistent with the nonlinear fit's sensitivity to initial parameter guesses or

Table 3.3: *Fit parameters a and b from the variance model $\sigma^2(N) = a/N + b$ for all calibration methods. A value of $b \approx 0$ indicates shot-noise-limited behavior, positive b implies an additional technical noise floor.*

Method	$\sigma \cdot \sqrt{N}$			a	b
	$N = 108$	$N = 204$	$N = 300$		
Rough	0.0114	0.0113	0.0104	$1.4 \cdot 10^{-4}$	$-9 \cdot 10^{-8}$
PP_1	0.1008	0.0914	0.1039	$1.0 \cdot 10^{-2}$	$-3 \cdot 10^{-6}$
PP_2	0.0343	0.0371	0.1730	$-1.2 \cdot 10^{-2}$	$1.1 \cdot 10^{-4}$
CP_1	0.00385	0.00429	0.00450	$1.2 \cdot 10^{-5}$	$3.1 \cdot 10^{-8}$
CP_2	0.00281	0.00271	0.00243	$8.9 \cdot 10^{-6}$	$-9 \cdot 10^{-9}$

model mismatch. Excluding the PP_2 outliers at 300 shots or making the fit more robust may cause its behavior to revert toward shot-noise-like scaling.

However, it should be noted that the cross-point calibration requires longer execution times per round ($\approx 15 - 17$ s) than the ping-pong calibration ($\approx 12 - 14$ s). However, the cross-point method consistently produced smaller amplitude errors and lower variability across repeated experiments. Mean relative amplitude errors were on the order of 10^{-3} for cross-point and $10^{-2} - 10^{-1}$ for ping-pong, indicating higher precision and stability for the cross-point approach. Both calibration schemes showed reduced errors in the second round compared with the first. For reliable, high-precision amplitude calibration, cross-point is the better choice under the current conditions.

In Figure 3.3, the calibration of the $X_{\pi/2}$ amplitude is shown for two different numbers of shots, 100 and 4000, with ten experiments each. Both calibration methods start from the same rough calibration data, and in this case the same rough data were also used for all ten experiments to ensure better comparability. Similar to the X_π calibration, we observe that with more shots the number of outliers decreases, which shows that a part of the variability at low shot numbers comes from limited statistics and unstable fitting. The width of the boxes is similar for both shot numbers. For both shot numbers, the cross-point calibration starts with higher amplitudes in the first round and remains only slightly higher than the ping-pong calibration in the second round. This indicates that both methods converge to similar results when more data are available, but cross-point tends to start from a slightly overestimated amplitude.

In summary, increasing the number of measurement shots reduces statistical noise and the number of outliers for all calibration methods, which is consistent with the expected shot-noise scaling. Cross-point calibration demonstrates the most stable and repeatable performance, approaching the shot-noise limit with only a slight increase in noise. Ping-pong calibration remains more sensitive to fit instabilities and outliers, especially at low shot numbers. However, both methods yield consistent amplitudes when sufficient statistics are collected. While the cross-point method requires more distinct amplitude settings per calibration, the present data indicate that it achieves significantly lower variance and a higher SNR with the same total number of shots. This suggests that cross-point calibration could be performed with fewer shots per point for comparable accuracy, resulting in similar or even reduced total resource cost.

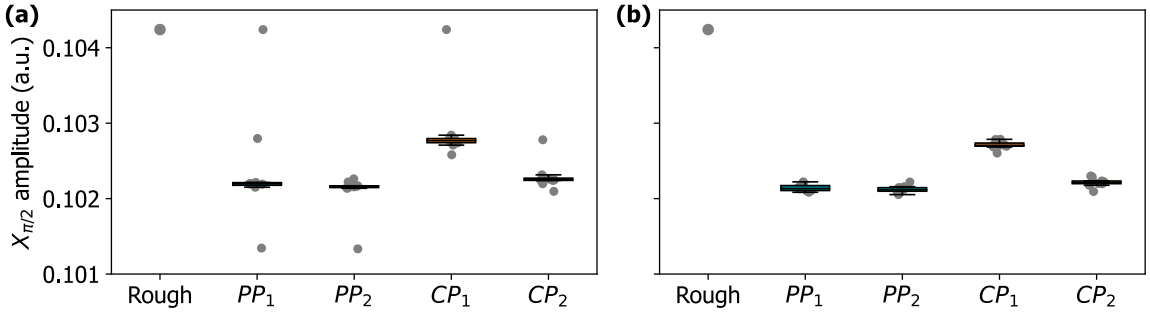


Figure 3.3: *$X_{\pi/2}$ amplitude for different calibration methods and shot numbers with identical rough values across experiments. Panel (a) shows 100 shots and panel (b) shows 4000 shots, each over 10 experiments. Rough is plotted as a single gray point since all values are the same. For PP_1, PP_2, CP_1, and CP_2, colored box plots summarize the distribution across experiments and jittered points indicate individual runs. Increasing the number of shots markedly reduces the variability while central tendencies remain consistent across methods.*

After examining the influence of the number of shots on calibration precision, we will now turn to the effect of repeated calibration rounds. Increasing the number of rounds enables us to investigate the convergence behavior and long-term stability of the ping-pong and cross-point methods.

In Figure 3.4 we assess the effect of increasing the number of calibration rounds on the extracted X_π pulse amplitudes. We fix the shot number to 256 and perform 12 independent repetitions per calibration method and round on *ibmq_ehningen*. The data includes the common rough calibration 'Rough' and data for ping-pong (PP_1-PP_4), and cross-point (CP_1-CP_4) calibration. Each calibration round uses the amplitude obtained from the preceding round, starting from the rough calibration and updating the value sequentially through all four rounds.

Two clear trends emerge in the multi-round calibration experiment. First, ping-pong calibration exhibits an increasing number of outliers with each additional round. Several data points reach the specified amplitude bounds $[0, 0.4]$, indicating fitting failures or optimizer constraints. The standard deviation increases from 0.0001 to 0.1 for PP_1–PP_4, while it remains below 0.0001 for all cross-point rounds. This behavior is consistent with the nonlinear trigonometric fit's higher sensitivity to noise, initial parameter guesses, and slow drift accumulated between rounds. Second, cross-point calibration achieves its lowest variability at round two (CP_2). Further rounds do not improve precision and slightly increase variability, suggesting diminishing returns once the intersection-based estimate has converged. Overall, the cross-point protocol is more repeatable under these conditions. Additional rounds primarily expose instability in the ping-pong protocol rather than delivering increased accuracy. These results indicate that while additional calibration rounds do not significantly improve accuracy, they can help identify the stability limits of the respective fitting procedures.

The calibration experiments compared two amplitude-calibration techniques, ping-pong and cross-point, which are both based on error-amplifying pulse sequences. While both methods converge to consistent pulse amplitudes, the cross-point calibration shows lower variance, higher signal-to-noise ratio, and increased robustness against fitting instabilities and outliers. These improvements arise from the symmetric, linear intersection fit used in

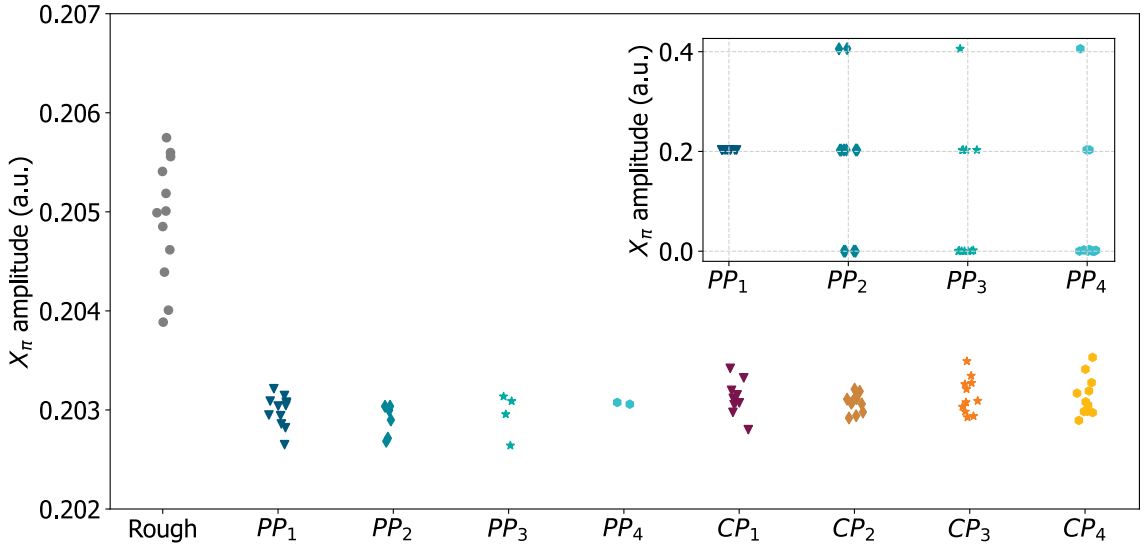


Figure 3.4: *Extracted X_π pulse amplitudes over four calibration rounds for each method, including the rough initial calibration. Ping-pong calibration shows greater variability and an increasing number of outliers with the round index. These include solutions that saturate at the amplitude bounds [0, 0.4], which are characteristic of fit or optimizer failures. Cross-point calibration yields the lowest variability at round two (CP_2), but additional rounds provide limited benefit and can slightly increase variability. Under a fixed shot budget, the cross-point protocol shows higher repeatability, whereas extra rounds amplify instability in the ping-pong calibration.*

cross-point calibration, which reduces sensitivity to SPAM offsets and initial parameter guesses, whereas the nonlinear trigonometric fit in ping-pong calibration amplifies such effects. An additional advantage of the cross-point method is that it provides a direct estimate of the optimal pulse amplitude from the intersection of the fitted curves. In contrast, the ping-pong calibration follows an empirical update rule and must typically be repeated until the amplitude deviation falls below a chosen threshold, making it practically more time-consuming. Under equal shot budgets, the cross-point protocol achieves comparable or better precision with simpler data analysis and faster convergence. It can therefore serve as an efficient alternative for the ping-pong calibration. Further investigation is required to assess the quantitative impact of SPAM errors and to confirm the observed trends on a broader dataset.

3.2 Experimental analysis of the two-qubit cross-resonance Hamiltonian in transmon qubits

In quantum computing, the ability to implement controlled operations between qubits is essential for executing quantum algorithms. One of the most important two-qubit gates is the CNOT gate, which enables entanglement and conditional logic operations. This section focuses on the experimental characterization of CNOT gates in *IBM*'s superconducting quantum devices based on fixed-frequency transmon qubits. The theoretical description of the underlying interaction is summarized in Subsection 2.3.2.

IBM's superconducting devices implement CNOT gates using the cross-resonance (CR) interaction. This mechanism allows for an effective two-qubit coupling by applying a resonant microwave drive to one qubit, the control qubit, at the transition frequency of a neighboring qubit, the target qubit. The drive induces a controlled rotation on the target qubit conditioned on the state of the control qubit. This approach avoids tunable couplers and variable qubit frequencies and is compatible with scalable fixed-frequency architectures.

The underlying dynamics of the CR interaction can be expressed in the Pauli basis. Each term has the form $P_i = \sigma_a \otimes \sigma_b$, where $\sigma_a, \sigma_b \in \{I, X, Y, Z\}$ are single-qubit Pauli matrices, and I is the identity. The first operator acts on the control qubit, and the second on the target qubit. A full discussion of the effective CR Hamiltonian is given in Subsection 2.3.2. Building on that, we now analyze the time evolution of a two-qubit system under the action of the CR Hamiltonian H.

We initialize the system in either the $|00\rangle$ or $|10\rangle$ state and evolve it under the Hamiltonian H for a time t. Since the goal is to isolate the action of the effective Hamiltonian on the target qubit, it is sufficient to fix the target in $|0\rangle$ while varying the control state. The corresponding time-evolved states are given by

$$|\Psi^{(0)}(t)\rangle = e^{-iHt}|00\rangle, \qquad |\Psi^{(1)}(t)\rangle = e^{-iHt}|10\rangle. \tag{3.4}$$

In order to analyze the system dynamics, we measure the expectation values

$$\langle\Psi^{(k)}(t)|P_i|\Psi^{(k)}(t)\rangle, \quad \text{for } k = 0, 1, \tag{3.5}$$

where each term takes the form $P_i = \sigma_a \otimes \sigma_b$. This allows us to reconstruct the time-dependent action of the effective Hamiltonian.

The most general form of a two-qubit Hamiltonian in the Pauli basis is

$$H_{\text{gen}} = \sum_{a,b\in\{I,X,Y,Z\}} a_{ab}\, \sigma_a \otimes \sigma_b, \tag{3.6}$$

where each coefficient a_{ab} denotes the interaction strength associated with the corresponding Pauli term. In implementations of the CR gate, many of these terms are strongly suppressed and do not contribute measurably to the observed dynamics. In particular, terms which flip the state of the control qubit, such as $X \otimes I$ and $Y \otimes I$ (or, more generally, all terms where $\sigma_a = X$ or Y), are typically suppressed, since the control qubit ideally remains in a computational basis state during the gate application. The $Z \otimes I$ term merely shifts the energy of the control qubit. Since the corresponding phase factor is canceled due to the echo, it has no influence on the intended CNOT operation.

Under these assumptions, the effective Hamiltonian used in standard CR modeling retains the six dominant terms listed in Equation (2.29), corresponding to single-qubit rotations on the target $(I \otimes X, I \otimes Y, I \otimes Z)$ and the conditional two-qubit interactions $(Z \otimes X, Z \otimes Y, Z \otimes Z)$ [65].

To validate whether the commonly used six-term Hamiltonian model [65] accurately describes the hardware, we perform a Hamiltonian tomography on the *ibmq_ehningen* backend and extract coefficients for all 15 non-trivial two-qubit Pauli terms from device

measurements. Backend parameters, including pulse durations, amplitudes, and single-qubit correction terms (e.g., virtual Z-rotations), are first obtained. On *ibmq_ehningen*, the calibrated CR duration for the CNOT between qubit 0 (control) and qubit 1 (target) is approximately 128 ns. The calibrated CNOT gate consists of a sequence of pulses with an echoed CR component and single-qubit rotations (see Subsection 2.3.2). In our experiment, we omit the echo and potential single-qubit corrections to observe the pure CR interaction. We vary the CR pulse duration from 0 ns to approximately 1.78 μs in equally spaced steps, while preparing the control qubit in either the ground ($c = |0\rangle$) or excited state ($c = |1\rangle$).

Each configuration is executed with 32 000 shots, and measurement error mitigation is performed via full calibration matrix inversion [38] using 10 000 calibration shots. To access all 15 Pauli terms, single-qubit basis rotations are applied prior to measurement, enabling tomography of the target state in all Pauli bases.

The experimental data are then fit to a simulated unitary evolution under a general two-qubit Hamiltonian using a constrained nonlinear least-squares routine (`curve_fit` in Python). The simulation generates expectation values for each Pauli term at each CR duration, which are compared against the measured values. This fit yields the effective Hamiltonian parameters a_{ab} directly from hardware data, enabling a quantitative comparison between the experimentally realized interaction and the idealized CR model.

The experimental data together with the fitted expectation values for all 15 possible two-qubit Pauli terms are shown in Figure 3.5. In the following, we describe the contributions of each Pauli term and give a detailed discussion of their origin.

First of all, the Pauli expectation values $I \otimes X$, $X \otimes X$, $X \otimes Y$, $Z \otimes X$, $Y \otimes X$, $Y \otimes Y$, $Z \otimes X$ do not show any dynamics and remain stable at zero throughout the experiment. Secondly, the Pauli terms $X \otimes I$, $X \otimes Z$, $Y \otimes I$ or $Y \otimes Z$ show small oscillations. They have a negligible amplitude and are likely due to imperfections in the control electronics or other small error influences (e.g. imperfect readout error mitigation). The terms that show a significant oscillation are $I \otimes Y$, $I \otimes Z$, $Z \otimes Y$ and $Z \otimes Z$. These oscillations originate from a direct drive of the target qubit during the CR pulse. These are exactly the expected oscillations when the target qubit rotates around the x-axis (i.e., Hamiltonian $I \otimes X$ and $Z \otimes X$ components).

The $Z \otimes I$ term remains stable at $+1$ when the control qubit is in state $|1\rangle$ and at -1 when the control qubit is in state $|0\rangle$, indicating that the control qubit itself is stable. When the control qubit is in state $|1\rangle$, a small offset is observed, which is likely due to imperfect measurement mitigation. This offset can also be observed in the other subgraphs containing oscillatory terms.

Figure 3.6 shows the interaction strengths of the fit parameters of the full Hamiltonian (gray). The difference between the full fit, with all Pauli terms, and the reduced fit, with the six major terms used in the literature, is also plotted as green hatched bars (note the different scale on the right-hand side). The biggest differences lie in the $X \otimes X$, $Y \otimes X$ and $Z \otimes I$ terms. The $X \otimes X$ and $Y \otimes X$ terms are set to zero in the reduced fit, so the $Z \otimes I$ term must compensate for this difference. To assess the quality of the fit, we

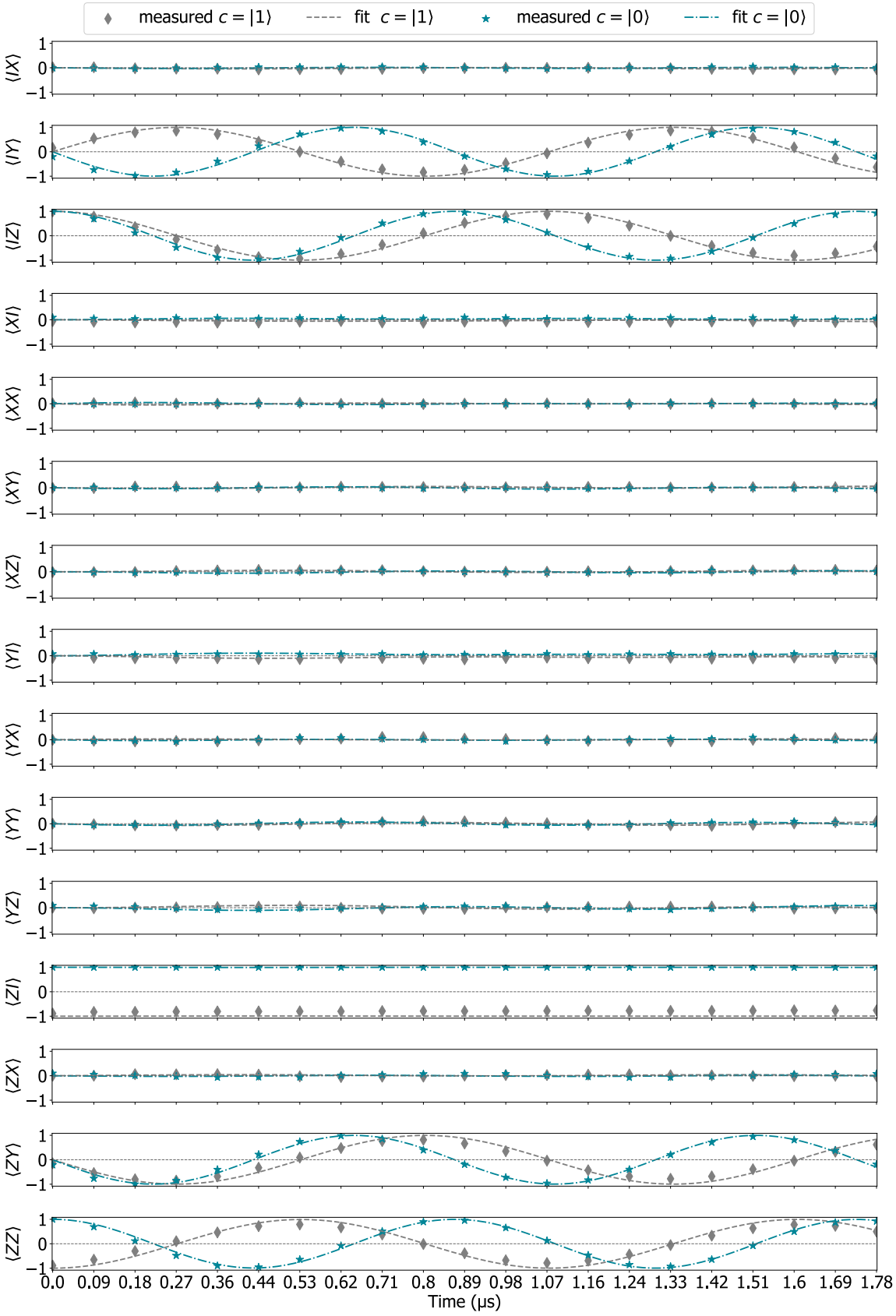


Figure 3.5: *Expectation values of all 15 two-qubit Pauli bases under CR gate evolution over varying pulse duration. Experimental data shown for initial control states $|0\rangle$ and $|1\rangle$. Dotted lines represent a single Hamiltonian model fit incorporating all Pauli terms.*

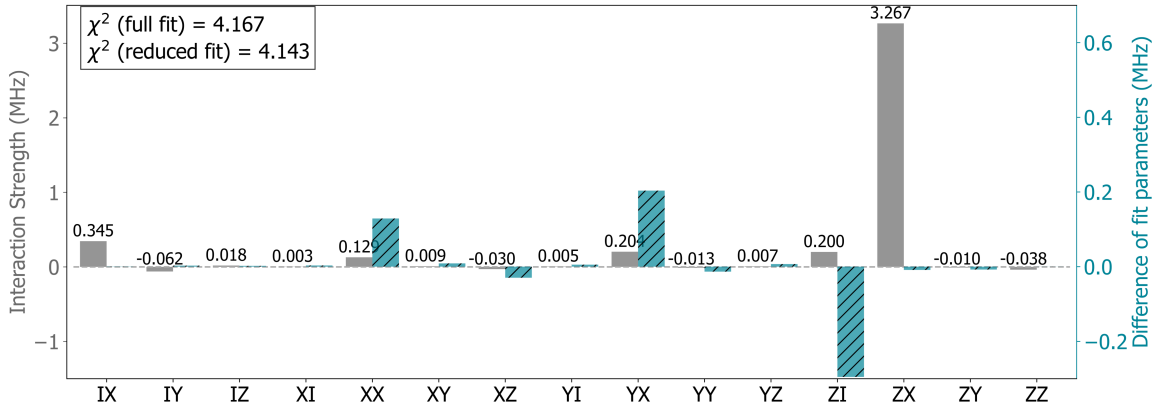


Figure 3.6: *The gray bars represent the fitted coefficients for all 15 possible Pauli terms. The green hatched bars indicate the change in the fitted coefficients when the fit is restricted to the six principal terms $I \otimes X$, $I \otimes Y$, $I \otimes Z$, $Z \otimes X$, $Z \otimes Y$, $Z \otimes Z$, with all remaining coefficients constrained to zero (note the different scale of the corresponding y-axis on the right-hand side). In the restricted fit, the $X \otimes X$, $Y \otimes X$, and $Z \otimes I$ terms show the largest deviations from their full-fit values. The corresponding $\mathcal{X}^2$-values for the full and reduced fits are approximately equal (see inset), indicating that the reduced Hamiltonian model describes the data with comparable accuracy despite having fewer parameters. This supports the validity of modeling the Hamiltonian with only these six dominant terms.*

calculated the $\mathcal{X}^2$ value

$$\mathcal{X}^2 = \sum_{i=1}^{m} \frac{(a_{ab,i} - a_{ab,i_{model}})^2}{a_{ab,i_{model}}}. \tag{3.7}$$

The difference between the full and reduced fit ($\mathcal{X}^2$(full fit) = 4.167 and $\mathcal{X}^2$(reduced fit) = 4.143) is minimal, indicating that the reduced Hamiltonian model accurately describes the data despite having fewer parameters. Therefore, the reduced model is also valid for the device under investigation.

As described in Subsection 2.3.2, another CNOT schedule, known as the direct CNOT, can be implemented (see Figure 2.7). In the subsequent sections, a comparative analysis will be conducted on these pulse schemes, addressing their respective advantages and disadvantages, as well as their applications in backends. This is due to the fact that not every CNOT gate can be efficiently realized by a direct version, which will be explained below.

3.3 Comparison between the implementation of direct and echoed CNOT gates

The results in this chapter have been partially published in:

Yanjun Ji, K. F. Koenig, Ilia Polian,
Optimizing quantum algorithms on bipotent architectures
Physical Review A **108**, 022610 (2023).

Yanjun Ji, K. F. Koenig, Ilia Polian,
Improving the performance of digitized counterdiabatic quantum optimization via algorithm-oriented qubit mapping
Physical Review A **110**, 032421 (2024).

The current generation of quantum hardware and software is undergoing continuous refinement. One development on the software and calibration level is the introduction of an alternative implementation of the CNOT gate, the so-called direct CNOT. In contrast to the conventional CNOT based on the echoed cross-resonance gate (Section 3.2), the direct version achieves a shorter duration, which reduces decoherence-induced errors. However, any miscalibration or residual Hamiltonian terms can contribute to coherent errors that are less efficiently suppressed than in the echoed version. Here, we compare both implementations, the direct and the echoed CNOT, first looking at the architecture, then at the gate times and error probabilities, the characteristics of the qubits and finally compare the two implementations of the CNOT gate within the context of the Quantum Approximate Optimization Algorithm (QAOA) introduced in Subsection 2.5.1.

While the reduced gate time of the direct CNOT is expected to lower decoherence-related errors, its practical realization introduces additional challenges, as outlined in Subsection 2.3.2. The direct CNOT requires more extensive calibration than the echoed variant, and is more sensitive to crosstalk, parameter drift, and frequency crowding. The absence of the echo sequence further implies that certain Hamiltonian error terms cannot be compensated as effectively as in the echoed case. Consequently, if these error contributions exceed tolerable thresholds on a given backend, a direct CNOT cannot be reliably implemented.

As a result, devices often employ a hybrid architecture, where some couplings are executed as direct CNOTs and others as echoed CNOTs. This is the case, for example, for *ibmq_ehningen*, illustrated in Figure 3.7. In this representation, direct CNOT gates are marked as solid gray connections, echoed CNOT gates in blue with double lines, and qubits are annotated according to which implementation is available: gray with dashed outlines (only direct), blue with solid outlines (only echoed), and green with dash-dotted outlines (both options). This data can be obtained from the backend by evaluating the gate length of the CNOT gate and analyzing the corresponding pulse schedules (see Figure 2.7).

Error rates and gate times for direct and echoed CNOT gates on the quantum device *ibmq_ehningen* are shown in Figure 3.8 (a) and (b). CNOT pairs with a direct CNOT implementation are depicted with blue bars and hatching, echoed CNOT gates in gray. Two of the CNOT gate pairs ([15, 18], [18, 15]) are faulty, have a high error rate and are depicted as white bars with a dashed outline. The analysis of the CNOT gates reveals

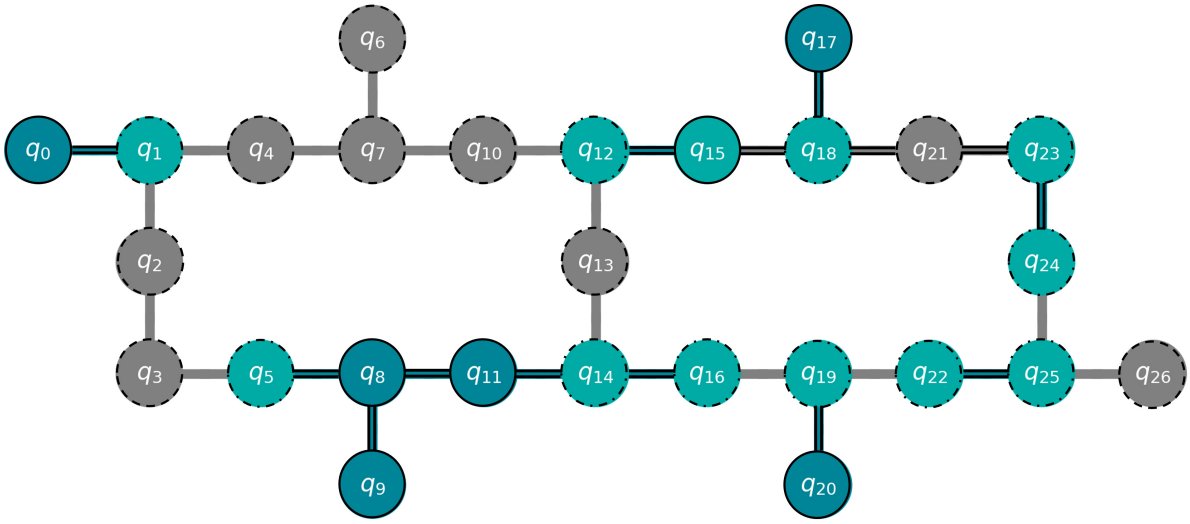


Figure 3.7: *Architecture of ibmq_ehningen indicating two different implementations of the CNOT gate. The backend supports two types of CNOT gates: direct CNOTs (shown as thick lines) and echoed CNOTs (shown as double lines). Qubits are depicted as circles, and lines between qubits indicate connections where a CNOT gate can be applied. When only one type of CNOT gate can be applied, the qubit takes the same color as that gate type: gray with a dashed outline for direct CNOTs, and blue with a solid outline for echoed CNOTs. If both gate types are available, the qubit is shown in green with a dash–dot outline. The architecture is considered hybrid because it contains two different implementations of the CNOT gate.*

significant performance differences between the two implementations. As illustrated in Figure 3.8 (a), the direct CNOT gate exhibits a median error probability (dashed and dash-dotted horizontal line) that is 1.3 times lower than that of the echoed CNOT gate, which corresponds to a reduction in median error rate about 38%. The faulty CNOT gates were excluded from the calculation of the median error probability, as they are non-functional. Furthermore, the gate times, illustrated in Figure 3.8 (b), indicate that the direct CNOT gate operates 1.4 times faster than its echoed counterpart, reflecting a significant reduction in median gate time about 40%. In summary, direct CNOT gates exhibit both lower error rates and shorter durations, which should enhance the fidelity of a wide range of quantum computations by allowing deeper circuits before decoherence dominates and enhance scalability. Even though the median metrics seem to be improved for direct CNOT gates, homogeneities across couplers (like the faulty CNOT gates) must be considered when mapping algorithms to hardware. While direct CNOT gates are faster and less error-prone, the calibration instability and backend-dependence mean that in practice the advantage might not be universal. This comparison provides the foundation for assessing how the choice of CNOT implementation impacts QAOA performance.

Since the advantages of direct CNOT gates are partly offset by calibration complexity and sensitivity to noise, it is important to also consider other qubit and gate parameters. Beyond two-qubit performance, coherence times, single-qubit gate fidelities, and readout errors determine the practical usefulness of different qubit categories. These characteristics are summarized in Figure 3.9.

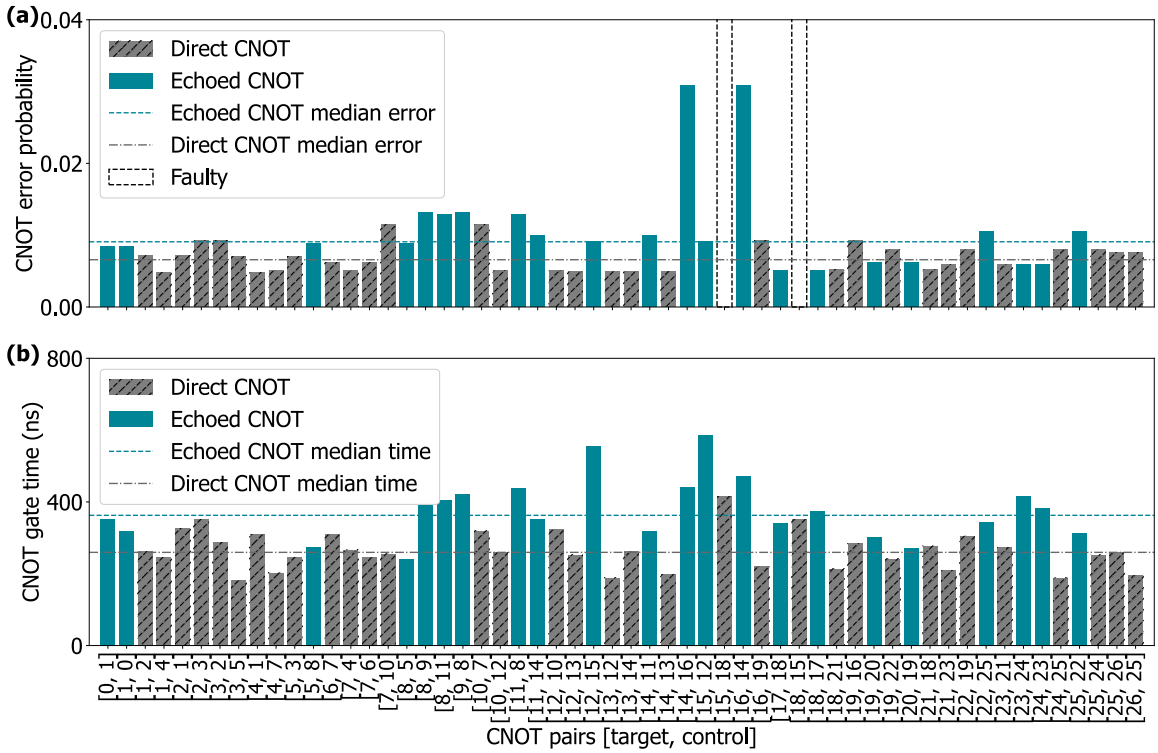


Figure 3.8: *Error rates and gate times of direct and echoed CNOT gates on ibmq_ehningen. (a) The error probability of the CNOT gate is depicted for direct (blue hatched bars) and echoed (gray bars) implementations. The median error rates are indicated by horizontal dash-dotted and dashed lines. Notably, the median error rate of the direct CNOT gate is 1.3 times lower than that of the echoed CNOT gate. The two bars outlined with dashed lines represent CNOT gates with a reported fault, exhibiting an error rate of 1. (b) The gate times for the two different CNOT gate implementations are presented similarly. The median gate time for the direct CNOT gate is 1.4 times shorter compared to that of the echoed CNOT gate.*

The coherence times of the 27 qubits are shown in Figure 3.9 (a) and (b). Qubits connected to both CNOT types (hereafter referred to as Q-hybrid) exhibit higher median values for both T_1 and T_2 compared to qubits restricted to only echoed CNOTs (Q-echoed).

Single-qubit gate errors and readout errors are summarized in Figure 3.9 (c) and (d). While the median single-qubit error rates are comparable across all three categories, the median readout error is noticeably larger for Q-echoed. By contrast, Q-direct and Q-hybrid qubits display better performance in this respect.

The enhanced coherence of Q-hybrid qubits may reflect a hardware allocation strategy, where qubits supporting both CNOT types occupy more favorable positions on the device. In contrast, the systematically higher readout errors of Q-echoed qubits suggest additional hardware constraints, such as frequency crowding or less optimized readout resonators. Qubits located at the edges of the chip may be subject to less favorable conditions, such as poorer isolation and greater external interference. In contrast, interior qubits could benefit from better control wiring and shielding. The positive effects (e.g., reduced latency) and negative effects (e.g., crosstalk, increased routing length, or frequency collisions) of these conditions could contribute to the enhanced coherence of hybrid qubits. However, the absence of a clear and consistent advantage

across all parameters indicates that hardware-level single-qubit and two-qubit calibration data alone is insufficient to determine which qubit category performs best in practice. Consequently, benchmarking within the context of algorithm execution is necessary to assess the effective usefulness of the different qubit combinations.

Here, QAOA is used as a representative variational algorithm, chosen because its performance is strongly influenced by two-qubit gate quality and qubit coherence times. To evaluate the two CNOT implementations and their combinations, we considered two example problems: MaxCut and Portfolio Optimization (PortOpt). A short overview of QAOA is provided in Subsection 2.5.1, while implementation details are given in Appendix A.5.

In Max-Cut, the goal is to partition the vertices of a graph such that the number of edges between partitions is maximized. In PortOpt, the objective is to select a portfolio of assets that balances high expected return with minimal risk. QAOA addresses these problems using a parameterized quantum circuit with alternating cost and mixer layers. The depth of the circuit is determined by the parameter p, with higher p generally enabling better approximations at the expense of longer circuits and greater susceptibility to noise.

We evaluate three performance metrics: the Approximation Ratio (AR), the Success Probability (SP), and the circuit duration. AR quantifies how close the QAOA solution is to the exact optimum after averaging over all samples (i.e. measurement shots), while SP captures the probability of sampling the optimal solution. Shorter circuit duration is beneficial because it reduces exposure to decoherence.

On *ibmq_ehningen*, the maximum linear connectivity available for QAOA circuits was limited to five qubits, which we used for all experiments. Each circuit was executed 50 000 times. The variational parameters of the QAOA circuit were optimized using a simulator-based workflow (see Appendix A.5). Qubit combinations with the highest predicted fidelity were selected using *MAPOMATIC* [111]. For echoed- and direct-CNOT gate circuits, *MAPOMATIC* chooses the best highest-fidelity linear arrangement of qubits that supports the required two-qubit interactions on all neighboring pairs. For circuits with varied CNOT implementations, it restricts the search to qubits with below-average single- and two-qubit error rates and chooses the arrangement with the shortest schedule duration. This ensures that each CNOT type is evaluated under comparably optimal hardware conditions.

The results are shown in Figure 3.10. Direct CNOT circuits consistently achieve shorter duration, reflecting their intrinsically faster gate times. For PortOpt (see Figure 3.10 (a)), direct and echoed gates yield comparable AR, with the hybrid circuit performing less favorably. For MaxCut (see Figure 3.10 (b)), direct gates generally improve AR compared to echoed, while the hybrid circuit also performs competitively. In terms of SP of MaxCut, direct gates tend to outperform echoed ones, although the relative advantage depends on the value of p. In PortOpt, the echoed variant for higher p outperforms the varied and direct CNOTs. For Maxcut it can be beneficial to use the mixed approach with both kinds of CNOT gates.

These findings suggest that the performance trade-offs between direct and echoed CNOTs are problem-dependent: direct gates benefit from reduced gate times and errors, while echoed gates may retain an advantage in calibration stability for certain circuit structures.

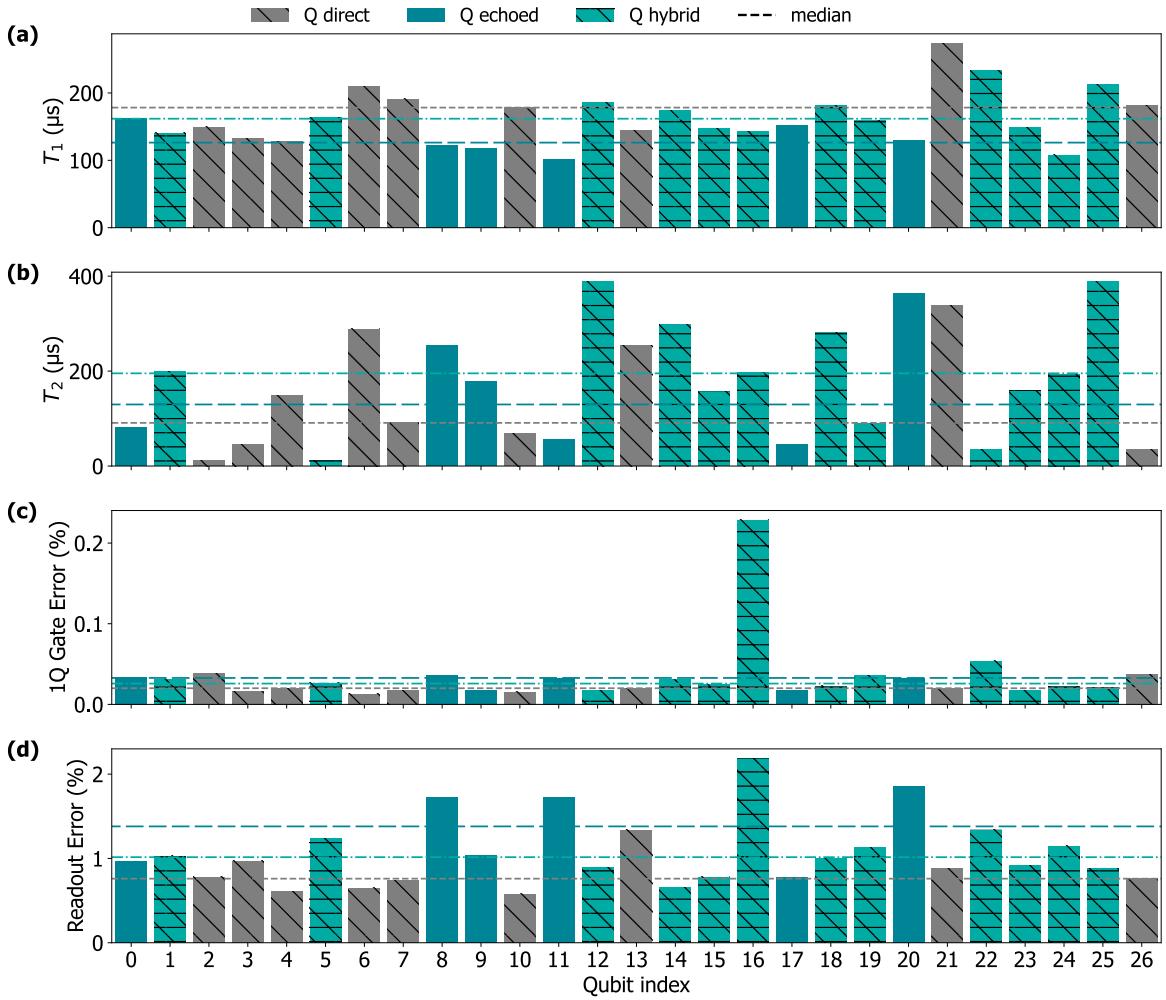


Figure 3.9: *Qubit characteristics of ibmq_ehningen. (a) Relaxation time T_1 and (b) dephasing time T_2. (c) Single-qubit gate error and (d) readout error. Qubits are grouped by the type of CNOT connectivity: Q-hybrid (green, cross-hatched) support both echoed and direct CNOTs, Q-echoed (blue) support only echoed CNOTs, and Q-direct (gray, diagonally hatched) support only direct CNOTs. Hybrid qubits show systematically higher median coherence times (T_1, T_2) than echoed, while direct and hybrid qubits achieve lower readout error compared to echoed qubits.*

Also the usage of both implementations can be beneficial, depending on the exact CNOT fidelities, as well as single-qubit fidelities, T_1, T_2 times and the readout fidelity.

3.4 Summary of calibration and gate analysis at the pulse level

This chapter investigated the calibration and characterization of quantum gates at the pulse level, spanning single-qubit amplitude calibration, Hamiltonian analysis of the cross-resonance (CR) interaction, and the implementation of two-qubit CNOT gates. It compares ping-pong and cross-point calibration in experiment, which to our knowledge constitutes the first direct experimental comparison of these two implemented amplitude-calibration techniques, validates the sufficiency of a reduced six-term CR model, and

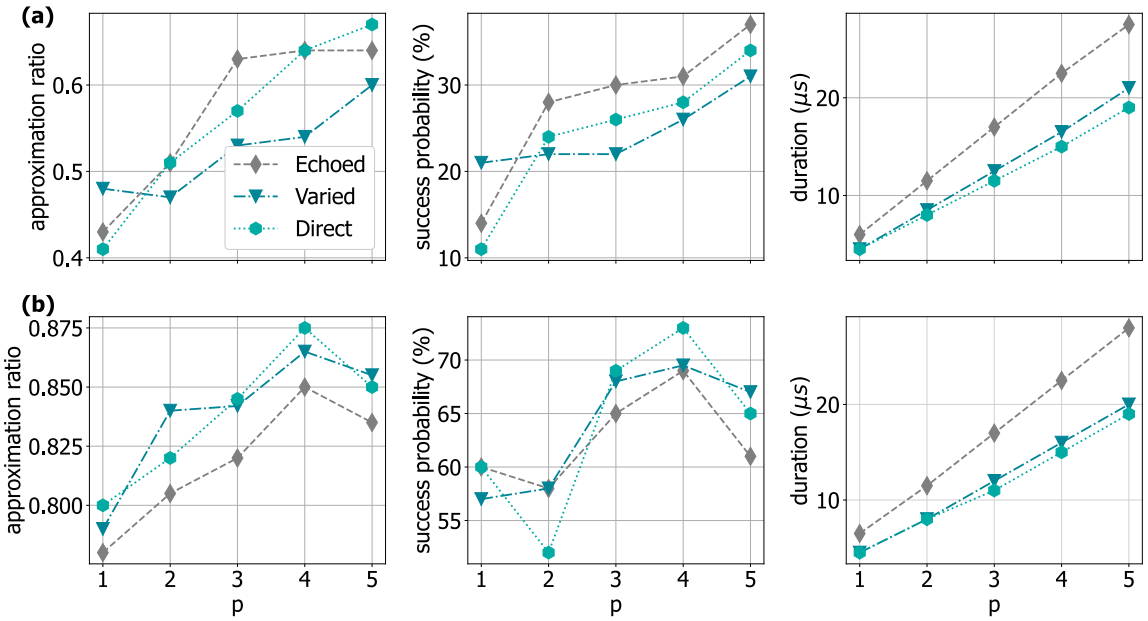


Figure 3.10: *QAOA results with 5 qubits on ibmq_ehningen using direct, echoed, and mixed CNOT implementations. (a) Portfolio Optimization and (b) MaxCut. Shown are the AR, SP, and circuit duration across different CNOT types. Direct gates achieve consistently shorter durations, and often higher AR and SP, while echoed gates remain competitive in PortOpt. Mixed implementations perform comparably for MaxCut but worse for PortOpt. Qubit assignments and circuit details are provided in Appendix A.5.*

evaluates how direct and echoed CNOT implementations, used individually or together, affect algorithmic performance.

Cross-point amplitude calibration was found to be more precise and more robust than ping-pong calibration. While both methods converge to similar pulse amplitudes when sufficient statistics are available, cross-point calibration consistently exhibits smaller variance, higher signal-to-noise ratio (SNR), and markedly fewer outliers. Across 108–300 measurement shots, cross–point calibration achieved standard deviations at the $10^{-4} - 10^{-3}$ level (e.g., 0.00026-0.00037), whereas ping-pong calibration showed substantially broader distributions ($\approxeq 10^{-2} - 10^{-1}$) and greater sensitivity to noise and fit instability, particularly at low shot counts and in later calibration rounds. Analysis of the shot-count dependence indicates that cross-point calibration approaches shot-noise-limited behavior by the second calibration round, while ping-pong calibration exhibits nonstationary behavior consistent with fit instability, drift, or outliers. Cross-point calibration converges by the second round, with additional rounds providing little benefit, whereas ping-pong calibration tends to accumulate errors. Although a cross-point round requires slightly longer execution time ($\approx 15 - 17$ s versus $\approx 12 - 14$ s for ping–pong calibration), it yields mean relative amplitude errors at the $\approx 10^{-3}$ level compared to $10^{-2} - 10^{-1}$ for ping–pong calibration. Under equal shot budgets, cross-point calibration therefore achieves superior precision and can be operated with fewer shots per calibration.

Hamiltonian tomography confirmed that a reduced six-term cross-resonance model ($I \otimes X$, $I \otimes Y$, $I \otimes Z$, $Z \otimes X$, $Z \otimes Y$, $Z \otimes Z$) describes the measured dynamics with accuracy comparable to a full fifteen-term Pauli decomposition. The full and reduced fits yield

nearly identical quality ($\mathcal{X}^2$(full fit) = 4.167, $\mathcal{X}^2$(reduced fit) = 4.143), indicating that the simplified model captures the dominant interaction terms on the investigated device and is sufficient for modeling and pulse-level design.

A comparison of direct and echoed CNOT implementations shows that direct CNOTs achieve approximately 1.4-times shorter median gate durations and 1.3-times lower median error rates than echoed CNOTs, at the cost of increased sensitivity to calibration stability and crosstalk. Algorithm-level benchmarks using the Quantum Approximate Optimization Algorithm demonstrate that the relative performance of direct and echoed CNOT gates is problem-dependent: each implementation can yield improved results depending on the circuit structure and depth, and hybrid use of both gate types can be advantageous when accounting for two-qubit fidelities, coherence times, readout errors, and qubit connectivity.

Overall, the results highlight that precise single-qubit calibration, validated Hamiltonian modeling, and an informed choice of two-qubit gate implementations are essential for achieving good algorithmic performance on present-day superconducting hardware. While these techniques improve gate quality and can enhance algorithmic performance, residual errors may persist and cannot be eliminated by calibration or by selecting particular gate implementations alone. As a result, improved gate fidelities do not necessarily guarantee unbiased expectation values or reliable circuit-level results. This observation motivates the use of error mitigation techniques that operate at the level of circuit execution rather than control calibration and aim to reduce the impact of remaining noise. In the subsequent chapter, we therefore turn to Zero-Noise Extrapolation and the development of extended mitigation strategies based on this approach.

4 Inverted-Circuit Zero-Noise Extrapolation and further refined error mitigation techniques

The results in this chapter have been partially published in:

K. F. Koenig, F. Reinecke, W. Hahn, T. Wellens,
Inverted-circuit zero-noise extrapolation for quantum-gate error mitigation
Physical Review A **110**, 042625 (2024).

K. F. Koenig, F. Reinecke, T. Wellens,
Adaptive folding and noise filtering for robust quantum error mitigation
arXiv:2505.04463 (2025).

Zero-Noise Extrapolation (ZNE), as outlined in the theoretical chapter (Subsection 2.4.1), is a widely used technique for mitigating quantum errors by extrapolating noisy results obtained at various noise strengths to the zero-noise limit [38, 46, 75]. However, standard ZNE (sZNE) is limited in accuracy due to assumptions about how noise scales and the reliability of extrapolation methods. In particular, it assumes that repeating a gate λ-times amplifies its error by a factor of λ. As we will demonstrate in this section, and as was mentioned in Subsection 2.4.1, this only holds under specific conditions, such as for small errors commuting with the ideal gate. For instance, under small depolarizing noise each gate independently adds a small error, so the total error probability scales approximately linearly with λ. By contrast, for coherent errors such as systematic over-rotations, repeating the same gate can cause constructive or destructive interference of errors, leading to non-linear scaling or even partial cancellation.

For ZNE to yield meaningful results, it is crucial to strike a balance: the extrapolation must be precise enough to capture the actual error behavior, yet practical enough to be implementable on current quantum hardware with limited resources and noise constraints. Achieving this balance is key to ensuring both reliability and usability of error mitigation in practice.

To address the limitations of sZNE, we introduce a set of refined error mitigation strategies centered around the Inverted-Circuit Zero-Noise Extrapolation (IC-ZNE), which we introduce in this chapter. Rather than assuming a particular scaling behavior, we describe a direct method to quantify the error strength ϵ of a quantum circuit U executed on a noisy quantum device. We begin by formally defining the error strength in Section 4.1, followed by techniques to measure it in Section 4.2.

Building on this foundation, we describe the structure of the IC-ZNE method in Section 4.3, which enables a direct measurement of error amplification rather than mere estimation. To make the later comparison meaningful, we first revisit the assumptions

underlying sZNE in Section 4.4. We evaluate the presumed error scaling of sZNE based on prior experiments and simulations employing various error models, highlighting where these assumptions hold and where they break down. We also demonstrate how IC-ZNE addresses some of the gaps left by sZNE, thereby extending the scope of general ZNE methods. In Section 4.5, we present a comparative analysis of sZNE and IC-ZNE through individual experiments on quantum hardware. Finally, in Section 4.6, we assess the optimal performance of both IC-ZNE and sZNE under varying error strengths using a simulator.

After having investigated how IC-ZNE can enrich the ZNE mitigation toolbox and the situations in which it fills gaps left by established methods, we will examine how IC-ZNE and sZNE can be further improved. In Section 4.7, we investigate the effect of error randomization, implemented through Pauli twirling, on the performance of sZNE and IC-ZNE, emphasizing its role in reducing bias and improving robustness. Additionally, in Section 4.8, we introduce Adaptive Scaling Factors (ASF), which may be applied as a standalone method for ZNE, reducing overhead relative to IC-ZNE. ASF dynamically adjusts noise amplification to improve the stability and accuracy of extrapolated results based on the measured or estimated error strength. We will compare equally and exponentially spaced scaling factors, analyzing their respective advantages in error mitigation, and assess the potential improvements in performance when comparing sZNE and IC-ZNE.
So far, the techniques rely on extrapolating results through different complementary methods. In Section 4.9, we investigate a post-processing technique called filtering, which can be applied to all the methods introduced. The aim is to suppress statistical noise and to improve extrapolation results by removing unwanted fluctuations in the data.
Section 4.10 examines the combination of global and local circuit folding with ASF, as the interplay of these strategies can potentially have a greater impact than in other settings. Previous studies have suggested that global folding can outperform local folding under time-correlated noise. As ASF aims to adaptively adjust scaling factors, the effect of different folding strategies can become more subtle. By comparing these strategies for both ZNE and IC-ZNE, we aim to identify the most effective approach.

4.1 Definition of error strength

In the context of quantum error mitigation and specifically within the ZNE framework, accurately quantifying the error strength ϵ of an imperfect circuit execution is crucial for effective extrapolation. We define the error strength in terms of the state fidelity F between the ideally prepared pure state $|\Psi\rangle$ and the density matrix ρ produced by executing the corresponding circuit on a noisy device, i.e.

$$\epsilon = 1 - F, \tag{4.1}$$

where F is given by

$$F(|\Psi\rangle, \rho) = \langle\Psi|\rho|\Psi\rangle. \tag{4.2}$$

Based on the definition of ϵ, it is restricted to the interval $[0,\ 1]$.

In a quantum computation, we aim to implement a unitary circuit U such that the ideal final state is $U|0\rangle = |\Psi\rangle$, assuming the system is initially prepared with all qubits in state $|0\rangle$. In experimental settings, the implemented process is instead given by a noisy quantum channel $\mathcal{E}_U$, leading to a potentially mixed final state

$$\rho = \mathcal{E}_{\mathcal{U}}(|0\rangle\langle 0|). \tag{4.3}$$

To derive our definition of the fidelity, we turn to the concept of Uhlmann fidelity, which is a quantity to compare quantum states. The Uhlmann fidelity [112] is defined between two density matrices σ and ρ

$$F_U(\sigma, \rho) = \left(\mathrm{Tr}\left[\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}\right]\right)^2. \tag{4.4}$$

Here, for any positive semidefinite operator X (in particular ρ and σ) with spectral decomposition $X = \sum_i x_i|x_i\rangle\langle x_i|$, we define $\sqrt{X} = \sum_i \sqrt{x_i}|x_i\rangle\langle x_i|$. The Uhlmann fidelity F_U captures the degree of similarity between the two density matrices and takes values between 0 and 1. A value of $F_U = 1$ indicates that the two states are identical and thus indistinguishable, while $F_U = 0$ corresponds to orthogonal states that are maximally distinguishable. When calculating the fidelity between a pure state Ψ and a noisy state ρ, the corresponding expression of the Uhlmann fidelity simplifies to

$$F(|\Psi\rangle, \rho) = \langle\Psi|\rho|\Psi\rangle, \tag{4.5}$$

thus arriving at Equation (4.2). Intuitively, Equation (4.5) corresponds to the overlap between the two states.

With this we can decompose ρ in terms of the fidelity and some additional noisy part σ

$$\rho = (1-\epsilon)|\Psi\rangle\langle\Psi| + \epsilon\sigma. \tag{4.6}$$

From Equations (4.1) and (4.2), it follows that $\langle\Psi|\sigma|\Psi\rangle = 0$. As an example, let us now derive ϵ and σ for a concrete noise channel, namely the depolarizing noise channel $\mathcal{E}_{\mathrm{dep}}(\rho)$ see Equation (2.12). For the depolarizing channel applied to a pure state $|\Psi\rangle\langle\Psi|$, we obtain

$$\begin{aligned}
\mathcal{E}_{dep}(|\Psi\rangle\langle\Psi|) &= (1-p)|\Psi\rangle\langle\Psi| + p\frac{\mathbb{I}}{d} \\
&\quad \text{with} \quad \mathbb{I} = |\Psi\rangle\langle\Psi| + (\mathbb{I} - |\Psi\rangle\langle\Psi|) \\
&= (1-p)|\Psi\rangle\langle\Psi| + \frac{p}{d}\Big[|\Psi\rangle\langle\Psi| + (\mathbb{I} - |\Psi\rangle\langle\Psi|)\Big] \\
&= \left[(1-p) + \frac{p}{d}\right]|\Psi\rangle\langle\Psi| + \frac{p}{d}(\mathbb{I} - |\Psi\rangle\langle\Psi|) \\
&\quad \text{with} \quad 1 - p + \frac{p}{d} = 1 - p\left(\frac{d-1}{d}\right)
\end{aligned}$$

$$= \underbrace{1 - p\left(\frac{d-1}{d}\right)}_{1-\epsilon} |\Psi\rangle\langle\Psi| + \underbrace{\frac{p}{d}\Big(\mathbb{I} - |\Psi\rangle\langle\Psi|\Big)}_{\epsilon\sigma}. \tag{4.7}$$

By comparing Equation (4.6) and Equation (4.7), σ can be expressed in the case of depolarizing noise as

$$\sigma = (\mathbb{I} - |\Psi\rangle\langle\Psi|)/(2^q - 1), \tag{4.8}$$

where q is the number of qubits in the quantum circuit. Note that σ is independent of ϵ in this model. It depends only on the Hilbert-space dimension and $|\Psi\rangle$. The noise strength of the depolarizing channel then results as $\epsilon = p(1 - \frac{1}{2^q})$. In general, however, the noise channel is, a priori, not known, such that both σ and ϵ in Equation (4.6) are also unknown, as they depend on the unknown target state, $|\Psi\rangle$, and the unknown depolarizing probability p.
Returning to the decomposition in Equation (4.6), the expectation value of an observable A can generally be written as

$$\langle A\rangle = (1-\epsilon)\underbrace{\langle\psi|A|\psi\rangle}_{=\langle A\rangle_{\text{ideal}}} + \epsilon \ \mathrm{Tr}[A\sigma]. \tag{4.9}$$

For $\epsilon \to 0$ the expectation value converges to $\langle A\rangle_{\text{ideal}}$. In a ZNE scheme, $\langle A\rangle$ is measured for different values of ϵ. However, different realizations of the noisy circuit lead to different σ's. Therefore, two realizations with the same error strength ϵ might lead to different values of $\langle A\rangle$. Nevertheless, fitting many data points obtained from different realizations should produce smooth behavior

$$\langle A\rangle(\epsilon) = (1-\epsilon)\langle A\rangle_{\text{ideal}} + \epsilon \ \mathrm{Tr}[A\,\bar{\sigma}(\epsilon)], \tag{4.10}$$

where $\bar{\sigma}(\epsilon)$ corresponds to the average of σ over all realizations with noise strength ϵ.

With this definition established, we are now in a position to examine the measurement and utilization of the error strength ϵ in realistic quantum circuits. This precise characterization is compulsory for the development of more reliable and adaptable error mitigation strategies.

4.2 Measuring error strength

Having established the formal definition of the error strength, this section will proceed to derive the methodology for its measurement.
To measure the error strength of a quantum circuit U, we execute U followed by its inverse $U^\dagger$ and measure the probability P_0 of all qubits returning to the initial state $|0\rangle$. In the noiseless case, $P_0 = |\langle 0|U^\dagger U|0\rangle|^2 = |\langle 0|0\rangle|^2 = 1$. However, in the noisy case,

$$P_0 = \langle 0|\rho'|0\rangle \leq 1, \quad \text{where} \quad \rho' = (\mathcal{E}_{U^\dagger}\mathcal{E}_U)(|0\rangle\langle 0|), \tag{4.11}$$

with $\mathcal{E}_U$ and $\mathcal{E}_{U^\dagger}$ representing the noise channels associated with U and $U^\dagger$, respectively (see Equation (2.5) for the definition of a quantum channel).

Substituting ρ' into the definition of P_0 from Equation (4.11) and applying the definition of the adjoint map with respect to the Hilbert-Schmidt inner product [16, p. 76], we obtain

$$\begin{aligned} P_0 &= \left\langle 0 \middle| \left(\mathcal{E}_{U^\dagger}\mathcal{E}_U\right)\left(|0\rangle\langle 0|\right) \middle| 0 \right\rangle \\ &= \mathrm{Tr}\Big[|0\rangle\langle 0| \ \left(\mathcal{E}_{U^\dagger}\mathcal{E}_U\right)\left(|0\rangle\langle 0|\right)\Big] \\ &= \mathrm{Tr}\Big[\mathcal{E}_{U^\dagger}^\dagger\left(|0\rangle\langle 0|\right) \ \mathcal{E}_U\left(|0\rangle\langle 0|\right)\Big] \\ &= \mathrm{Tr}\left[\tilde{\rho}\rho\right], \end{aligned} \tag{4.12}$$

with

$$\tilde{\rho} = \mathcal{E}_{U^\dagger}^\dagger\left(|0\rangle\langle 0|\right). \tag{4.13}$$

In the case of a depolarizing noise channel, which is both self-adjoint and commutes with unitary operations, one can show that $\mathcal{E}_{U^\dagger}^\dagger = \mathcal{E}_U$ and therefore $\rho = \tilde{\rho}$ (see Appendix A.7). In other words, applying either the operator $\mathcal{E}_U$ or the adjoint channel of the noisy implementation of the inverse circuit $U^\dagger$, $\mathcal{E}_{U^\dagger}^\dagger$, to the initial state $|0\rangle\langle 0|$ leads to identical outputs. For non-depolarizing noise channels, however (such as dephasing, amplitude damping or correlated noise), the two resulting states can differ, i.e. $\rho \neq \tilde{\rho}$. Nevertheless, we will assume in the following that both channels $\mathcal{E}_U$ and $\mathcal{E}_{U^\dagger}^\dagger$ degrade the fidelity in a symmetric manner, justifying the approximation $\langle\Psi|\tilde{\rho}|\Psi\rangle \approx \langle\Psi|\rho|\Psi\rangle$ when estimating the noise strength. This assumption can be further motivated by considering the operational equivalence of forward and reverse noise in symmetric environments and will be justified as we show in the following. It is possible to decompose every circuit into single-qubit gates, denoted by S, and CNOT gates, denoted by C [16, pp. 191–194]. Combined, they can be used to implement an arbitrary unitary operation on the state space of n qubits. For example, consider the circuit

$$U = S_3 C_2 S_2 C_1 S_1, \tag{4.14}$$

where the indices indicate the order of gates in the circuit decomposition (i.e., S_1 is the first single-qubit gate, C_1 the first CNOT, and so on). Since the CNOT gate is self-adjoint, we have

$$U^\dagger = S_1^\dagger C_1 S_2^\dagger C_2 S_3^\dagger. \tag{4.15}$$

To model realistic execution, we assume that single-qubit gates are essentially error-free and that randomized compiling (see Subsection 2.4.2) is applied to the CNOTs. In this setting, the noise affecting each CNOT can be represented as a replacement rule $C \to \Lambda_{\mathcal{P}}\mathcal{C}$, where $\mathcal{C}$ denotes the ideal CNOT channel and $\Lambda_{\mathcal{P}}$ is a stochastic Pauli channel arising from the randomized compiling procedure [47]. Applying this replacement to every CNOT in the decomposition of U gives one noise block $\Lambda_{\mathcal{P}_i}$ per CNOT. Thus,

$$\mathcal{E}_{\mathcal{U}} = S_3 \Lambda_{\mathcal{P}_2} \mathcal{C}_2 S_2 \Lambda_{\mathcal{P}_1} \mathcal{C}_1 S_1. \tag{4.16}$$

Similarly, for the inverse circuit,

$$\mathcal{E}_{\mathcal{U}^\dagger} = S_1^\dagger \Lambda_{\mathcal{P}_1} \mathcal{C}_1 S_2^\dagger \Lambda_{\mathcal{P}_2} \mathcal{C}_2 S_3^\dagger. \tag{4.17}$$

Note that CNOT gates are self-adjoint and that the noise channels are not inverted when constructing $\mathcal{E}_{U^\dagger}$, since the physical implementation of $\mathcal{C}$ and $\mathcal{C}^\dagger = \mathcal{C}$ is identical. Since Pauli channels are self-adjoint, we obtain for

$$\mathcal{E}_{U^\dagger}^\dagger = S_3 \mathcal{C}_2 \Lambda_{\mathcal{P}_2} S_2 \mathcal{C}_1 \Lambda_{\mathcal{P}_1} S_1. \tag{4.18}$$

It is similar to Equation (4.16), with the exception that the noise channels are applied prior to, rather than subsequent to, the ideal CNOT gates. Let us consider the Pauli operators, which form a basis of Hermitian operators that do not all commute with the CNOT gate. This commutation relationship leads to another Pauli operator, $P_{\pi(i)}$ as shown in Equation (2.44), so that

$$P_i C = C P_{\pi(i)}. \tag{4.19}$$

Therefore, $\mathcal{E}_{\mathcal{U}^\dagger}$ and $\mathcal{E}_{U^\dagger}^\dagger$ may, in principle, differ for Pauli channels where $p_i \neq p_{\pi(i)}$ for at least one i. However, we expect that the effect on the fidelity will be negligible, since the latter mainly originates from the Pauli terms proportional to the identity operator P_1 (without error), which commutes with CNOT.

By explicitly decomposing circuits into single-qubit and CNOT gates with randomized compiling noise models, we clarified how the noisy implementations of U, $U^\dagger$, and their adjoint channels differ in detail, but also why their effect on fidelities can be treated as approximately symmetric, thereby justifying the approximation $\langle\Psi|\tilde{\rho}|\Psi\rangle = \langle\Psi|\rho|\Psi\rangle$.

In the effort to derive an expression for P_0 in terms of the error strength, we decompose $\tilde{\rho}$ analogously to ρ in Equation (4.6)

$$\tilde{\rho} = (1-\epsilon)|\psi\rangle\langle\psi| + \epsilon\tilde{\sigma}, \tag{4.20}$$

where $\langle\psi|\tilde{\sigma}|\psi\rangle = 0$ and , as justified above, the error strength ϵ is the same as for ρ in Equation (4.6). Using Equations (4.6) and (4.20), we obtain

$$\begin{aligned} P_0 &= \mathrm{Tr}[((1-\epsilon)|\psi\rangle\langle\psi| + \epsilon\sigma)((1-\epsilon)|\psi\rangle\langle\psi| + \epsilon\tilde{\sigma}))] \\ &= (1-\epsilon)^2 \mathrm{Tr}[|\psi\rangle\langle\psi|\psi\rangle\langle\psi|] + \epsilon(1-\epsilon)(\mathrm{Tr}[|\psi\rangle\langle\psi|\sigma] + \mathrm{Tr}[|\psi\rangle\langle\psi|\tilde{\sigma}]) + \epsilon^2 \mathrm{Tr}[\sigma\tilde{\sigma}] \\ &= (1-\epsilon)^2 + \epsilon^2 \underbrace{\mathrm{Tr}[\sigma\tilde{\sigma}]}_{=:a}, \end{aligned} \tag{4.21}$$

where we used that $\langle\psi|\psi\rangle = 1$, $\mathrm{Tr}\Big[|\psi\rangle\langle\psi|\sigma\Big] = 0$, and $\mathrm{Tr}\Big[|\psi\rangle\langle\psi|\tilde{\sigma}\Big] = 0$. The quantity a is defined as the trace of the product of the operators σ and $\tilde{\sigma}$. The parameter a, is generally unknown but expected to be small (as shown below). This corresponds to scenarios where the purities of σ and $\tilde{\sigma}$ are low, which is also anticipated in the presence of predominantly incoherent errors.

If the gate errors are represented by Pauli channels, which can be implemented for CNOT gates through randomized compiling (refer to Subsection 2.4.2), then σ and $\tilde{\sigma}$ result from a combination of many states due to the many potential errors at each

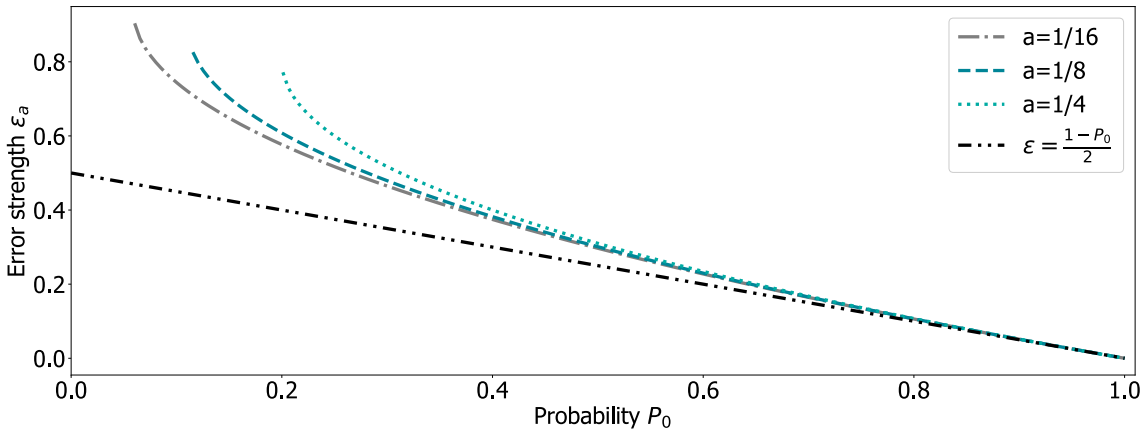


Figure 4.1: *The noise strength ϵ_a, as defined in Equation (4.23), is shown as a function of the probability P_0 of measuring each qubit in state $|0\rangle$ for different values of a as indicated in the legend. For low noise levels, the particular value of a does not significantly impact the value of the error strength. For comparison, the black dash–dot–dot line represents the low noise approximation in Equation (4.24).*

gate. Consequently, σ and $\tilde{\sigma}$ correspond to highly mixed states, i.e., $\mathrm{Tr}[\sigma^2] \ll 1$ and $\mathrm{Tr}[\tilde{\sigma}^2] \ll 1$. Using the inequality $0 \leq \mathrm{Tr}[(\sigma - \tilde{\sigma})^2] = \mathrm{Tr}[\sigma^2] + \mathrm{Tr}[\tilde{\sigma}^2] - 2\mathrm{Tr}[\sigma\tilde{\sigma}]$, we obtain the bound

$$a = \mathrm{Tr}[\sigma\tilde{\sigma}] \leq \frac{1}{2}\left(\mathrm{Tr}[\sigma^2] + \mathrm{Tr}[\tilde{\sigma}^2]\right) \ll 1. \tag{4.22}$$

For a depolarizing error, for instance, we obtain $\tilde{\sigma} = \sigma$, with σ given by Equation (4.8) above. This yields $a = Tr[\sigma\tilde{\sigma}] = \frac{1}{2^q - 1}$.

Solving Equation (4.21) for the error strength ϵ yields solutions expressed in terms of a and P_0, where P_0 denotes the probability of measuring all qubits in the $|0\rangle$ state. When $P_0 > a$, the physical constraint $0 \leq \epsilon \leq 1$ leads to a unique solution

$$\epsilon_a = \frac{1 - \sqrt{P_0 - a(1 - P_0)}}{1 + a}. \tag{4.23}$$

The subscript emphasizes the functional dependence on a. The condition $P_0 > a$ guarantees a real-valued solution by ensuring that the argument of the square root remains non-negative.

A second mathematical solution exists for $\frac{a}{1+a} < P_0 \leq a$, while no real solutions exist when $P_0 < \frac{a}{1+a}$. This bifurcation arises from the quadratic nature of Equation (4.21). However, only Equation (4.23) is physically valid, as it correctly approaches the noiseless limit $\epsilon \to 0$ when $P_0 \to 1$. The alternative solution $\epsilon = [1 + \sqrt{P_0 - a(1 - P_0)}]/(1 + a)$ violates this fundamental requirement and may yield unphysical values with $\epsilon > 1$. The absence of solutions for $P_0 < \frac{a}{1+a}$ indicates that the measured probability lies outside the compatible regime with the underlying model.

The values of ϵ_a as a function of P_0, shown in Figure 4.1, converge in the low noise regime to

$$\epsilon_a \to \frac{1 - P_0}{2} \text{ for } P_0 \to 1, \tag{4.24}$$

independent of the value of a. Intuitively, the factor of two in the denominator arises from the fact that the error associated with $\mathcal{E}_U$ is doubled when adding $\mathcal{E}_{U\dagger}$. It is important to note that Equation (4.24) does not hold for coherent errors, where the assumption $a \ll 1$ is violated. In such cases, we have confirmed that $a \propto \frac{1}{\epsilon}$ for $\epsilon \to 0$.

In case of incoherent errors, we account for deviations from the asymptotic behavior described in Equation (4.24) by choosing the value of a as follows: $a = \frac{1}{2^q}$ which is slightly larger than the value obtained for depolarizing errors. If $P_0 < \frac{1}{2^q}$, we set $a = P_0$ to prevent situations where there is no solution or multiple solutions, as previously mentioned. Finally, we arrive at the following estimate for the noise strength ϵ based on the measured probability P_0 of finding all q qubits in the state $|0\rangle$:

$$\epsilon = \begin{cases} \frac{1-\sqrt{P_0 - \frac{1-P_0}{2^q}}}{1+\frac{1}{2^q}} & P_0 > \frac{1}{2^q} \\ \\ \frac{1-P_0}{1+P_0} & P_0 \leq \frac{1}{2^q}. \end{cases} \tag{4.25}$$

4.3 Structure of Inverted-Circuit Zero-Noise Extrapolation

IC-ZNE builds on the concepts introduced in the preceding chapters. In summary, IC-ZNE allows for the accurate determination of the scaling of error strength, which deviates from the scaling assumption made in sZNE. This is achieved without requiring any knowledge of the underlying noise or complex error characterization.

IC-ZNE differs structurally from sZNE (see Subsection 2.4.1) in several key aspects. sZNE amplifies noise by repeating individual gates that scale the overall error strength, relying on the assumption that errors increase linearly or in a predictable manner with circuit depth. This assumption is only valid for specific types of noise, such as small, stochastic, and commuting errors (e.g., depolarizing channels). In contrast, IC-ZNE circumvents this assumption and measures the error strength directly by applying the original quantum circuit followed by its inverse, such that the ideal unitary evolution cancels while the noise aggregates. This approach allows for a more interpretable noise amplification without depending on the gate-level scaling assumptions.

While IC-ZNE may introduce an implementation overhead due to the need for an additional circuit, it offers improved robustness against various types of errors, particularly when combined with error randomization techniques such as Pauli twirling. Measuring the effective error strength transforms the impact of noise processes into a directly measurable scalar quantity, which is guaranteed to vanish in the absence of noise. As a result, extrapolation is performed on ϵ rather than on an assumed error amplification factor λ that may fluctuate and not grow linearly. Consequently, IC-ZNE is more robust

to noise fluctuations. However, it should be noted that some noise contributions may not be captured by the IC-ZNE method and are therefore not included in the extrapolation. For instance, systematic over- or under-rotations that are applied identically by the inverse circuit would ideally cancel out and thus remain undetected. While techniques such as Pauli twirling can mitigate this effect, they are not always employed in practice. Overall, the structural differences between IC-ZNE and ZNE reflect a shift from local, assumption-driven noise scaling to a more holistic and model-independent approach to error amplification and mitigation.

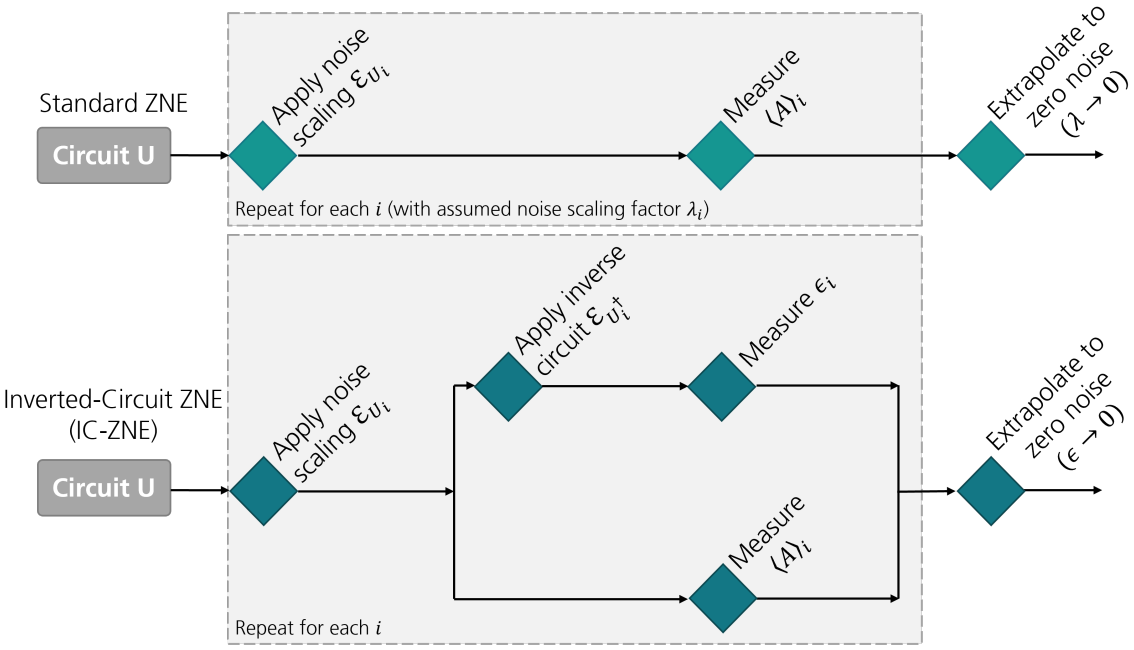


Figure 4.2: *Comparison of the Zero-Noise Extrapolation (ZNE) workflow (upper part) with the Inverted-Circuit ZNE (IC-ZNE) approach (lower part). Standard ZNE is applied to a set of circuits U_i, that are logically equivalent to U, but differ in their noise amplification factors. The IC-ZNE method incorporates two additional steps: applying the inverse of the circuit and measuring the error strength, denoted as ϵ_i. This process is executed for each amplification factor, yielding distinct error strength measurements for all circuits. In practice, different strategies can be used to obtain scaled circuits U_i, such as applying amplification to the entire circuit (global folding), or, as in most of this work, selectively to more error-prone gates, such as two-qubit gates (local folding).*

A comparison between the sZNE workflow and the IC-ZNE approach is shown in Figure 4.2. The ZNE procedure consists of three steps: noise scaling is first applied to the circuit, the expectation value is then measured, and this process is repeated for all amplification factors to enable an extrapolation to the noise-free limit. The IC-ZNE workflow introduces two additional steps. After noise scaling, the inverse circuit is applied to estimate the error strength ϵ_i. This measurement is carried out for each amplification factor, yielding a distinct error strength for every circuit. A discussion of the computational overhead is provided in Section 4.6.

Different strategies can be used to obtain noise scaling, such as global folding or local folding (see Subsection 2.4.1). In the case of global folding, the amplified circuit is given by $U_i = U(U^\dagger U)^{n_i}$, where $n_i \in \{0, n_1, .., n_{max}\}$ denotes the number of folding repetitions. For local folding, only a subset of gates, typically CNOT gates, are folded.

If U_c represents such a gate, the corresponding amplified circuit is obtained by replacing each occurrence of the gate U_c in U with $U_c(U_c^\dagger U_c)^{n_i}$. In IC-ZNE, after applying either global or local folding, the inverse of the amplified circuit is appended in order to measure the error strength ϵ for each amplification factor n_i. It is important that this error strength measurement directly follows the execution of the folded circuit, since otherwise time-dependent fluctuations could distort the result. Unless stated otherwise, all results presented in this work employ local folding, with folding applied exclusively to CNOT gates.

In summary, IC-ZNE offers a promising alternative to sZNE, especially due to its robustness against different sources of error. By directly measuring the error strength ϵ using inverted circuits, rather than relying on assumptions about its scaling behavior, IC-ZNE provides a more reliable framework for error mitigation in quantum computing. The measured error strength can be used to examine assumptions made in ZNE (see Subsection 2.4.1). In the following section, we therefore analyze the scaling behavior of the measured error strength as a function of the noise scaling factor λ. By comparing $\epsilon(\lambda)$ with the exponential scaling implied by sZNE, deviations from the assumed behavior are identified for different circuits, noise models, and real quantum hardware.

4.4 Error scaling behavior in Zero-Noise Extrapolation

In the context of ZNE, it is assumed that errors scale exponentially (see Subsection 2.4.1). However, this simple scaling behavior is predicated on the assumption that only depolarizing errors are present. Nevertheless, additional errors may emerge in real quantum devices (see Section 2.2), potentially causing deviations from the anticipated scaling behavior, since noise is not limited to depolarizing noise. This chapter examines different scaling behaviors in various contexts, including real backends, simulations with depolarizing error models, and backend error models.

To demonstrate how the behavior diverges from the assumed scaling, we illustrate the dependence of the normalized error strength ϵ/ϵ_0, obtained by inverted circuits, as a function of the noise scaling factor λ, in Figure 4.3. Here ϵ_0 denotes the error strength of the unscaled circuit ($\lambda = 1$). The three test circuits used here are the Grover circuit, Harrow-Hassidim-Lloyd (HHL) circuit, and Hadamard-Ladder (H-Ladder) circuit. Detailed descriptions are provided in Appendix A.4.

The dependence of the expectation value $\langle A \rangle$ on the noise scaling factor λ in sZNE is commonly approximated by an exponential fit,

$$\langle A \rangle = a_1 e^{-a_2 \lambda} + a_3, \tag{4.26}$$

(see also Subsection 2.4.1). The ideal, noise-free value corresponds to $\lambda = 0$, giving $\langle A \rangle_{\text{ideal}} = a_1 + a_3$. Substituting this into Equation (4.26) allows the expression to be rewritten as

$$\langle A \rangle = e^{-a_2 \lambda} \langle A \rangle_{\text{ideal}} + \left(1 - e^{-a_2 \lambda}\right) \langle A \rangle_{\text{noisy}}, \tag{4.27}$$

where $\langle A \rangle_{\text{noisy}} = a_3$ is the expectation value reached in the strong-noise limit $\lambda \to \infty$.

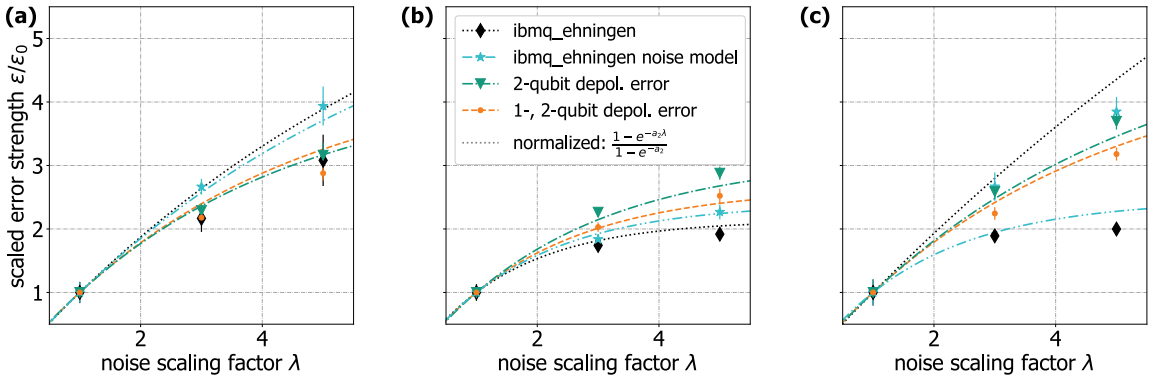


Figure 4.3: *Scaled error strength ϵ/ϵ_0 as a function of the noise scaling factor λ for representative runs of (a) Grover, (b) HHL, and (c) H-Ladder circuit. The error strength is normalized such that the mean value $\langle\epsilon\rangle/\epsilon_0 = 1$ for $\lambda = 1$. Black diamonds show data obtained from the real quantum device (ibmq_ehnignen). Simulated results are shown for a 1- and 2-qubit depolarizing noise model (orange circles), a 2-qubit-only depolarizing model (green triangles), and the backend-specific noise model (blue stars). The expected exponential scaling behavior assumed in sZNE is indicated by the corresponding lines (dotted, dashed, dash-dotted, and dash-dot-dotted). Across all circuits, deviations from exponential scaling are observed, particularly for the real hardware data, demonstrating the limitations of standard noise models.*

Equation (4.27) can be interpreted as stating that, at finite λ, the effective quantum state is a mixture of the ideal state $\rho_{\text{ideal}} = |\psi\rangle\langle\psi|$ and the noisy state ρ_{noisy} obtained in the strong-noise limit

$$\rho(\lambda) = e^{-a_2\lambda}|\psi\rangle\langle\psi| + \left(1 - e^{-a_2\lambda}\right)\rho_{\text{noisy}}. \tag{4.28}$$

In the case of depolarizing noise, this reduces to $\rho_{\text{noisy}} = \rho_*$ (see Subsection 2.4.1).

Combining Equation (4.28) with the definitions of the fidelity and ρ in Equations (4.2) and (4.6), one obtains the relation between the noise scaling factor λ and the error strength ϵ

$$\epsilon = \left(1 - \langle\psi|\rho_{\text{noisy}}|\psi\rangle\right)\left(1 - e^{-a_2\lambda}\right), \tag{4.29}$$

or, normalized to the error strength at $\lambda = 1$,

$$\frac{\epsilon}{\epsilon_0} = \frac{1 - e^{-a_2\lambda}}{1 - e^{-a_2}}. \tag{4.30}$$

Equation (4.30) shows that the error strength ϵ increases monotonically with the amplification factor λ, starting from $\epsilon = 0$ at $\lambda = 0$ and saturating for large λ at a maximum value determined by the overlap $\langle\psi|\rho_{\text{noisy}}|\psi\rangle$. In other words, λ controls how quickly the state is driven from the ideal state towards the fully noisy limit, with a_2 setting the characteristic scale of this transition.

We analyze the scaled error strength ϵ/ϵ_0 as a function of the noise scaling factor λ for representative runs of the Grover, HHL, and H-Ladder circuits. The complete experiments will be discussed in the following section. The results, shown in Figure 4.3,

Table 4.1: *Initial error strengths ϵ_0 and exponential decay rates a_2 obtained by fitting Equation (4.26) to Grover, HHL, and H-Ladder datasets under various noise models.*

	Grover		HHL		H-Ladder	
	ϵ_0	a_2	ϵ_0	a_2	ϵ_0	a_2
ibmq_ehningen	0.15	0.135	0.42	0.630	0.45	0.071
ibmq_ehningen noise model	0.091	0.162	0.32	0.537	0.059	0.522
2-qubit depol. error	0.22	0.262	0.29	0.383	0.17	0.205
1-, 2-qubit depol. error	0.20	0.244	0.25	0.473	0.14	0.235

include data obtained on quantum hardware (*ibmq_ehningen*), as well as simulations using the backend's full noise model, and two different depolarizing noise models. Both depolarizing models are parameterized by an average gate infidelity of $r = 0.02$. The first model includes single- and two-qubit depolarizing channels (1-, 2-qubit depol. error), while the second applies depolarizing noise only to two-qubit gates (2-qubit depol. error). These parameters are chosen to reflect typical noise levels of the *IBM* quantum hardware used in this work. A detailed description of these models is provided in Appendices A.2 and A.1. The corresponding exponential decay rates a_2 are obtained by fitting the measured expectation values $\langle A \rangle$, with the ZNE model in Equation (4.26). The initial error strength ϵ_0 and resulting a_2 values are listed in Table 4.1.

The expected exponential scaling behavior matches well for the two-qubit depolarizing error model (green triangles, dash-dotted line in Figure 4.3) and for the combined one- and two-qubit depolarizing error model (orange circles, dashed line). In contrast, the results from some of the remaining cases exhibit significant but partial deviations from this idealized behavior. Deviations from the expected behavior are observed in: the Grover circuit (run on *ibmq_ehningen*, see Figure 4.3 (a)), the H-Ladder circuit (run on *ibmq_ehningen*, see Figure 4.3 (c)), and the *ibmq_ehningen* noise model (see Figure 4.3 (c). These results show that the underlying assumptions do not accurately reflect reality. Additionally, even with the backend-specific noise model, deviations from the expected exponential scaling persist, particularly for the H-Ladder circuit. Furthermore, the scaled error strength for the H-Ladder circuit data from the backend exhibits the greatest deviation. One possible explanation is the presence of correlated noise and idle-time decoherence among the utilized qubits. Another explanation could be that, due to crosstalk effects similar to those reported in Ref. [91], the H-Ladder circuit exhibits higher deviations. Although the specific qubits used in this work differ, similar interactions may occur and could account for the increased deviations observed for the H-Ladder circuit.

These discrepancies highlight that real device noise cannot be adequately described by a simple depolarizing model or its associated exponential scaling. This reinforces the need for error mitigation methods such as IC-ZNE that do not rely on oversimplified assumptions but instead use empirically measured error strengths. Having established that the ZNE assumptions about error scaling do not hold in general, we now turn to a comparison of real backend data obtained with ZNE and IC-ZNE.

4.5 Comparison of sZNE and IC-ZNE using real backend data

Building on the theoretical introduction to IC-ZNE, we now compare sZNE and IC-ZNE using experimental data collected on *IBM*'s quantum device *ibmq_ehningen*. As test cases, we consider three quantum circuits: Grover's algorithm, the HHL algorithm, and the H-Ladder circuit. Detailed descriptions of their structures are provided in Appendix A.4. While Grover and HHL represent full algorithms, H-Ladder is not an algorithm but rather a parametrizable circuit that can be extended in depth.

Noise scaling factors $\lambda \in \{1, 3, 5\}$ were applied to each circuit, and for each λ, 16 randomly twirled versions of the original circuit were executed 625 times each, totaling 10,000 measurement repetitions per scaling factor. Twirling facilitates averaging over coherent noise effects (see Subsection 2.4.2). In the case of IC-ZNE, each twirled circuit was additionally measured to estimate the corresponding error strength ϵ, requiring another 625 repetitions per circuit.

The results are shown in Figures 4.4 to 4.6. In each figure, subfigure (a) shows the sZNE approach, where the expectation values $\langle A \rangle$ are plotted for the chosen values of λ and extrapolated using an exponential fit. Subfigures (b) show the IC-ZNE results, where the same expectation values are plotted as a function of the measured error strengths ϵ, and a first-order linear fit is used for extrapolation.

The spread of data points at fixed scaling factor is primarily due to variations introduced by the choice of twirling gates, statistical fluctuations, and backend noise. For the extrapolation, an exponential fit in λ is used for sZNE, while a linear fit in ϵ is employed for IC-ZNE. The quoted uncertainties of the extrapolated zero-noise values are calculated from the covariance matrix of the fit parameters, ensuring a consistent statistical treatment across both methods.

Table 4.2: *Comparison of extrapolated expectation values obtained with sZNE and IC-ZNE for the three benchmark circuits, shown in Figures 4.4 to 4.6, along with the corresponding ideal values. Reported uncertainties correspond to one standard deviation.*

	$\langle A \rangle_{\text{ideal}}$	$\langle A \rangle_{\text{sZNE}}$	$\langle A \rangle_{\text{IC-ZNE}}$
Grover	1.000	0.970 ± 0.021	0.994 ± 0.009
HHL	0.625	0.594 ± 0.025	0.633 ± 0.017
H-Ladder	0.500	0.459 ± 0.017	0.510 ± 0.018

Across all three benchmarking circuits, IC-ZNE systematically yields extrapolated values that are closer to the respective ideal results and, in most cases, reduces statistical uncertainty compared to sZNE (see Table 4.2). For the Grover circuit, IC-ZNE nearly reproduces the exact expectation value, demonstrating both improved accuracy, with an 81% reduction in deviation, and a more than twofold reduction in variance. The HHL circuit, which involves a larger number of CNOT gates, exhibits stronger noise effects overall. Nevertheless, IC-ZNE achieves a 75% improvement over sZNE. The confidence interval of the IC-ZNE result already includes the exact value, whereas the sZNE interval lies entirely below it. Hence, only IC-ZNE achieves statistical consistency with the ideal outcome.

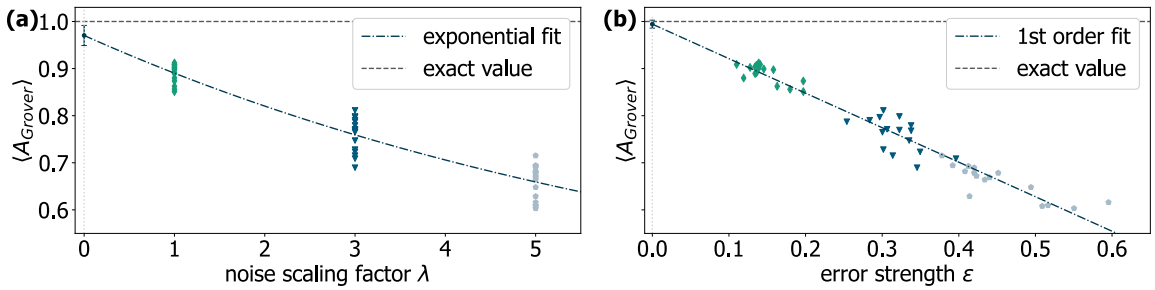


Figure 4.4: *sZNE (a) and IC-ZNE (b) for the Grover circuit executed on IBM's quantum device ibmq_ehningen. For each noise scaling factor $\lambda = 1$ (green diamonds), $\lambda = 3$ (blue triangles), and $\lambda = 5$ (grey pentagons), we perform 16 different randomly twirled quantum circuits, each with 625 shots, which gives 10,000 shots per scaling factor in total. The variation observed in the data points for the different scaling factors can be explained by the use of different twirling gates, as well as by statistical fluctuations and variations introduced by the backend used. The error bars (standard deviation) of the extrapolated values at zero noise ($\lambda = 0$ or $\epsilon = 0$, respectively) are calculated from the covariance matrix of the fit parameters provided by the curve fitting procedure. (a) In sZNE, the measured expectation values $\langle A_{Grover} \rangle$ are plotted versus the noise scaling factor λ. An exponential fit leads to the extrapolated value $\langle A_{Grover} \rangle = 0.970 \pm 0.0213$, which is clearly smaller than the ideal value $\langle A_{Grover} \rangle_{ideal} = 1$ (horizontal dotted line). (b) In IC-ZNE, the error strength ϵ of each circuit is determined by the method of inverted circuits (using 625 additional shots per circuit). The same expectation values $\langle A_{Grover} \rangle$ as in (a) are now plotted as a function of ϵ instead of λ, resulting in a more precise extrapolated value of $\langle A_{Grover} \rangle = 0.994 \pm 0.009$ obtained by a first-order linear fit.*

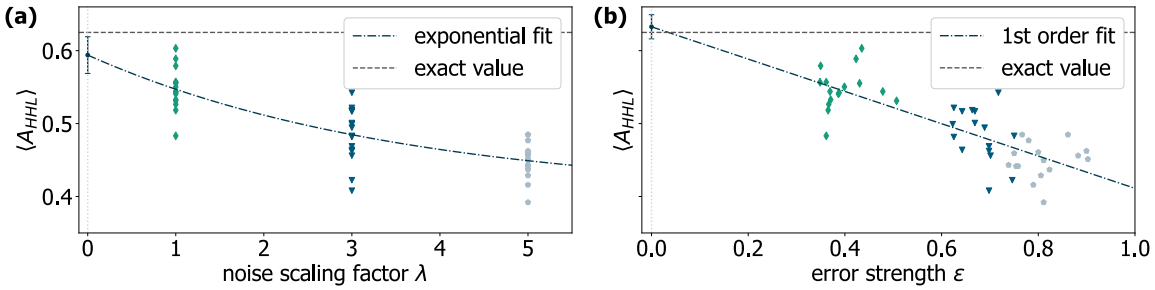


Figure 4.5: *Same as Figure 4.4, but for the HHL circuit instead of the Grover circuit. Again, the extrapolation using IC-ZNE (b) yields better agreement of the extrapolated value $\langle A_{HHL} \rangle = 0.594 \pm 0.025$ for sZNE and 0.633 ± 0.017 for IC-ZNE with the exact value $5/8 = 0.625$ than sZNE (a). Note that, due to a larger number of CNOT gates, the measured values of the error strengths ϵ are higher than in case of the Grover circuit in Figure 4.4 (b).*

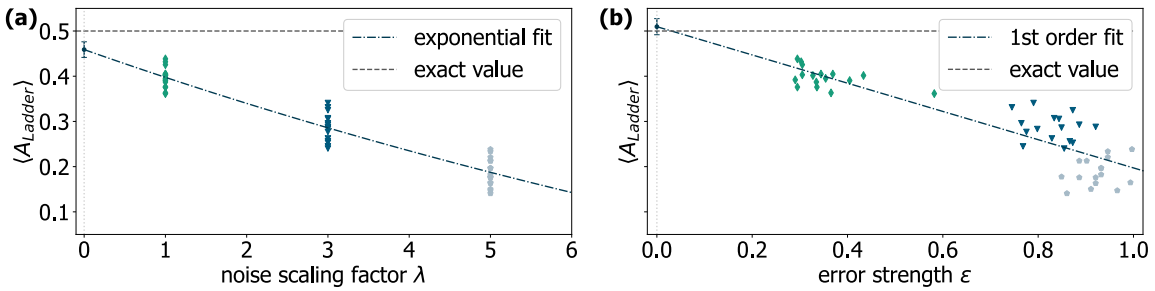


Figure 4.6: *Same as Figure 4.4, but for the H-Ladder circuit instead of the Grover circuit. Again, the extrapolation using IC-ZNE (b) yields better agreement of the extrapolated value $\langle A_{Ladder} \rangle = 0.459 \pm 0.017$ for sZNE and 0.510 ± 0.018 for IC-ZNE, with the exact value 0.5 than sZNE (a). The H-Ladder circuit shows the strongest error increase at higher amplification factors ($\lambda = 3, 5$). See main text for discussion of idle-time decoherence, leakage, and correlated noise.*

The H-Ladder circuit shows a higher error strength for the larger amplification factors $\lambda = 3$ and $\lambda = 5$, although randomized compiling converts coherent gate errors into effectively stochastic noise. This behavior can be explained by several effects that are not removed by twirling. The H-Ladder circuit operates on eight qubits, and its structure increases the number of idle qubits during gate execution and therefore enhances decoherence due to finite T_1 and T_2 times. These idle periods can be mitigated by inserting identity operations in the form of dynamical decoupling sequences [36, 113–115], which help to suppress decoherence effects during idle times. In addition, leakage out of the computational subspace and correlated noise between qubits can accumulate over the course of the computation. Small residual correlations in the readout, which remain even after readout-error mitigation, may further contribute to the observed error. These mechanisms scale with the number of qubits and the total circuit duration and are thus more pronounced for the H-Ladder circuit compared to the smaller Grover and HHL circuits. Since IC-ZNE reduces the error by a similar fraction for all circuits, here by 77%, the remaining noise appears to be largely stochastic, while the larger circuit width of the H-Ladder leads to an overall higher error strength, when folding is applied. Taken together, these results indicate that IC-ZNE provides a reliable reduction of stochastic errors across different circuit types and remains effective even in the presence of additional noise sources that are not fully captured by sZNE.

In summary, we observe that IC-ZNE not only provides extrapolated values closer to the exact results but also reduces the statistical uncertainty in most cases. These improvements stem from the fact that IC-ZNE measures the actual error strength of each circuit instance, rather than assuming an ideal scaling behavior with λ, as sZNE does. To gain a deeper understanding of this distinction, we now turn to a systematic investigation of how both methods perform as a function of the underlying error strength.

4.6 Influence of error strength on the performance of sZNE and IC-ZNE

Building on the previous results, we now analyze how the performance of sZNE and IC-ZNE depends on the underlying gate error strength. To this end, we systematically vary the noise level within a depolarizing noise model and examine the resulting extrapolation behavior. All simulations are performed using a custom implementation in *Qiskit* [116]. Simulations were performed using a depolarizing noise model parameterized by the average gate infidelity r, which is related to the depolarizing probability p by $r = \frac{2^q-1}{2^q}p$ [117–119], see Appendix A.2 for details. The single-qubit errors are set to one order of magnitude smaller than the two-qubit errors (10%). The differences in single- and two-qubit errors resemble those in *IBM*'s quantum hardware. The average gate infidelity r is varied within the range $0.005 \leq p \leq 0.05$. It is important to note that in this error model, other gates such as I, RZ and *Reset* gates, as well as initialization and measurement operations are assumed to be perfect, and decoherence processes acting on idle qubits are neglected. The simulation is conducted for the Grover, HHL and H-Ladder circuits (see Appendix A.4 for a description), utilizing 10 000 shots, amplification factors $\lambda \in \{1, 3, 5\}$ and repeated $M = 50$ times to ensure statistical reliability. For comparison, we also present the raw data obtained from the unscaled circuit without applying sZNE. It is

important to note that randomized compiling is not utilized in these simulations, as it does not affect a depolarizing channel.

Figures 4.7 to 4.9 show box plots (explanation see Appendix A.3) for the extrapolated expectation values of the depolarizing error model explained above. The accompanying plots at the bottom display the root-mean-square error (RMSE) relative to the ideal value, defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\Big(\langle A\rangle_i - \langle A\rangle_{\mathrm{ideal}}\Big)^2}, \tag{4.31}$$

where M is the number of runs ($i = 1, 2, ..., M$), $\langle A\rangle_i$ the i$^{\mathrm{th}}$ expectation value and $\langle A\rangle_{\mathrm{ideal}}$ the ideal expectation value.

From Figures 4.7 and 4.8, it is evident that, both for the Grover and the HHL circuit, and for all four error ranges, IC-ZNE demonstrates superior performance compared to sZNE. The smallest RMSE is reached in all but one instance, which corresponds to HHL with largest error rate.

For the Grover circuit (see Figure 4.7), IC-ZNE yields extrapolated values that are systematically larger than the corresponding $\langle A_{\mathrm{Grover}}\rangle$ for the two highest error rates, panels (c) and (d). In spite of this bias, IC-ZNE achieves RMSE values at least 1.6 times lower than those of sZNE. As shown in Figure 4.7 (b), the improvement reaches a factor of 3.2 when directly comparing RMSE values for 1% average gate infidelity. The bias toward larger values is not fatal because the RMSE advantage shows robustness. This means IC-ZNE is preferable even if the raw estimates appear strongly shifted.

For the HHL circuit, Figures 4.8 (a)–(c) indicate that sZNE produces results with larger variability than IC-ZNE. At the lowest error rate shown in (a), IC-ZNE performs 2.5 times better, attributable to its reduced variations and outliers. In the case of average gate infidelity reaching 5%, as depicted in Figure 4.8 (d), the extrapolated values for both sZNE and IC-ZNE significantly diverge from the exact value represented by the horizontal line. This observation highlights a fundamental limitation of ZNE techniques, which cannot be relied upon to produce accurate results when the noise level in the original circuit is excessively high. This is particularly evident in the increased number of CNOT gates relative to the Grover circuit. The higher gate count makes the technique less effective at large error rates, which can be accounted to a broader limitation of sZNE as it scales poorly with depth and noise. Taken together, these findings suggest that IC-ZNE is most valuable in regimes where noise is modest but non-negligible. These are exactly the conditions that are encountered in present noisy intermediate-scale devices.

The result for the H-Ladder circuit with 2% average gate infidelity in Figure 4.9 (c) indicate that IC-ZNE and sZNE exhibit comparable RMSE values. Notably, for average gate infidelities of $\leq 2\%$, the distribution of sZNE is broader than that of IC-ZNE. In Figure 4.9 (a) and (b), IC-ZNE demonstrates a minimum improvement of 1.7 times in terms of RMSE compared to sZNE. However, at higher error rates, the extrapolated values diverge from the exact value represented by the horizontal dotted line, highlighting the limitations of ZNE techniques under increased noise conditions. The advantage of IC-ZNE is most evident at low error rates. However, once noise dominates, neither sZNE nor IC-ZNE performs better. This defines a practical boundary of the approach.

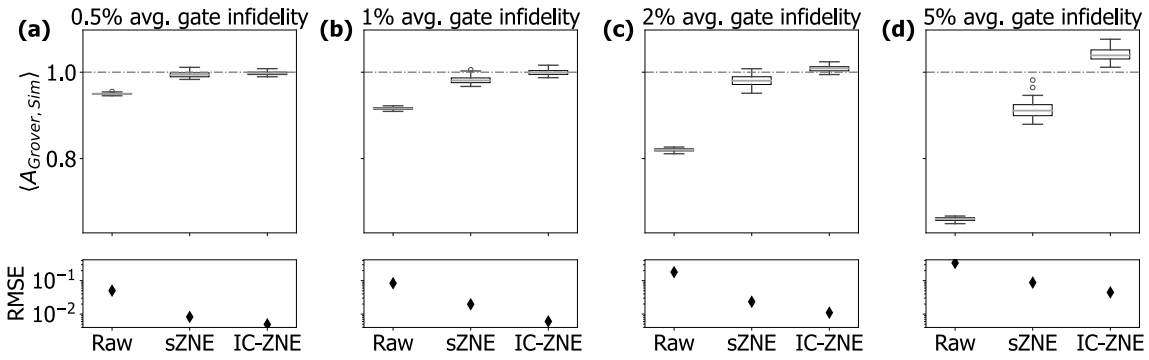


Figure 4.7: *Comparison of IC-ZNE, sZNE and unmitigated raw data (raw) for 50 simulation runs of the Grover circuit using a single- and two-qubit depolarizing error model with different average gate infidelities: (a) 0.5%, (b) 1%, (c) 2%, and (d) 5%. The error rate for single-qubit gates is set to one-tenth of the CNOT error rate. A box plot has been employed to present the extrapolated values $\langle A_{Grover} \rangle$ for each method (raw, sZNE and IC-ZNE). The exact value $\langle A_{Grover} \rangle_{ideal}$ is represented by a dashed horizontal line. The root-mean-square error (RMSE), which reflects the deviation from this ideal value, is depicted in the accompanying plot below on a logarithmic scale. Notably, in all cases, the IC-ZNE demonstrates the smallest RMSE.*

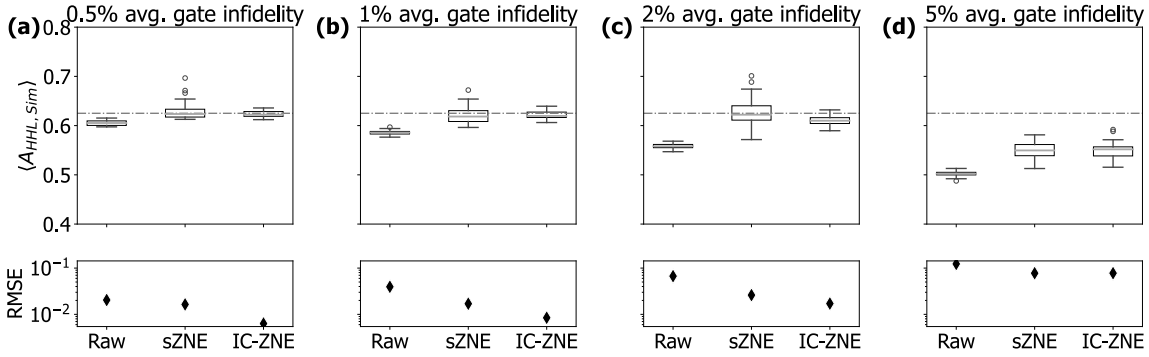


Figure 4.8: *Analogous to Figure 4.7, but for the HHL circuit instead of the Grover circuit. For average gate infidelities of $\leq 2\%$ ((a), (b), and (c)), the distribution for sZNE is broader compared to that of IC-ZNE and IC-ZNE outperforms sZNE by at least a factor of 1.6 in terms of RMSE. However, at higher error rates, the extrapolated values deviate significantly from the exact value indicated by the horizontal line. This suggests that, due to the greater number of CNOT gates relative to the Grover circuit, the errors in (d) are too substantial for all sZNE methods to function effectively.*

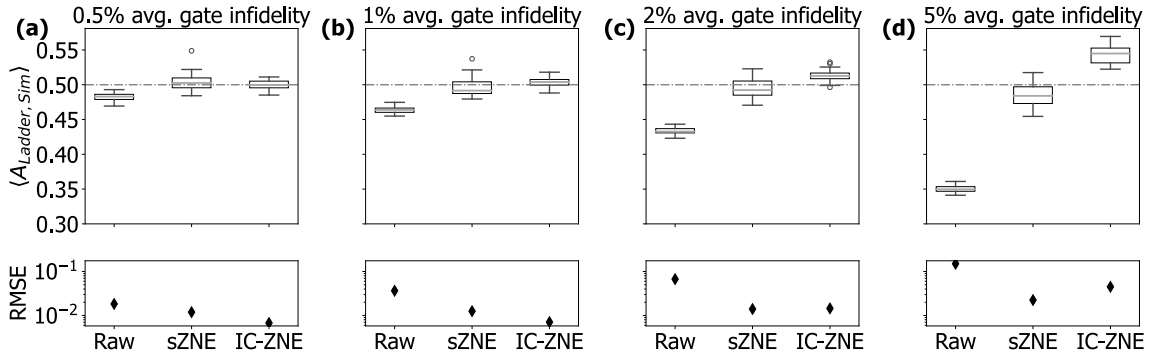


Figure 4.9: *Same as Figure 4.7, but for the H-Ladder instead of the Grover circuit. For average gate infidelities $\leq 2\%$, the distribution for sZNE is broader than for IC-ZNE. IC-ZNE performs at a minimum of 1.7 times (in terms of RMSE) better than sZNE in (a) and (b). For higher error rates ((c) and (d)), the obtained extrapolated values differ from the exact value (horizontal dotted line).*

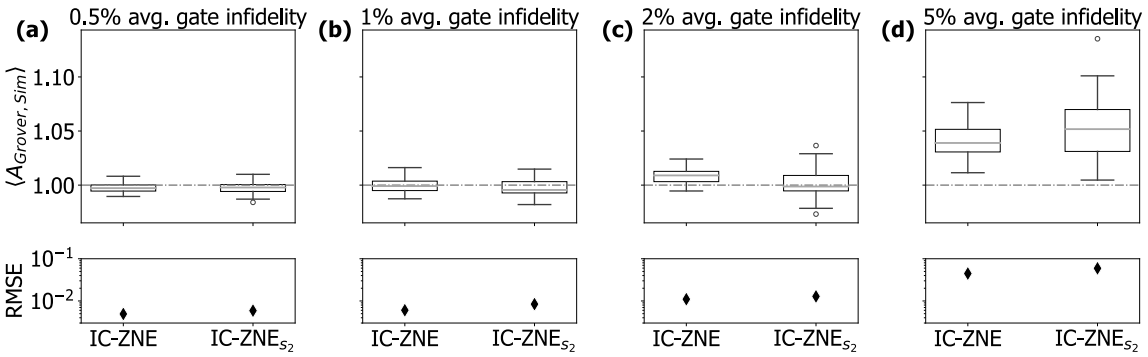


Figure 4.10: *Simulation results for IC-ZNE with reduced circuit repetition numbers and different average gate infidelities. Instead of 60 000 measurement repetitions, only 30 000 repetitions (IC-ZNE$_{s_2}$) were used. The smaller number of repetitions leads to a broader distribution, but the differences in RMSE values are minor.*

In summary, the analysis across various circuits, including the Grover, HHL and the H-Ladder circuit, consistently demonstrates the superior performance of IC-ZNE compared to sZNE, particularly at lower error rates. While both methods exhibit similar RMSE values at higher error rates, IC-ZNE consistently outperforms sZNE, particularly in scenarios with low noise. However, as error rates increase, the effectiveness of both techniques diminishes, emphasizing a critical limitation: ZNE methods may not yield reliable results in circuits with high noise levels due to the increased number of CNOT gates. In such cases, the measured result is completely dominated by noise, leaving no suitable information for ZNE. It must also be stated that these findings are obtained from simulations with a depolarizing noise model. Real devices may show different performance characteristics because hardware noise is often not fully captured by depolarizing models. Nevertheless, the overall trends indicate that IC-ZNE has practical value in present-day quantum computing experiments, in particular where noise is present but not completely dominant.

Simulations with reduced number of circuit repetitions

Finally, we analyze the additional resources required by IC-ZNE in comparison to sZNE. Since IC-ZNE needs extra circuits for determining the error strength, the overall resource usage is about twice as high. In our experiments, 10 000 measurement repetitions were used for the unmitigated raw data, 30 000 for sZNE, and 60 000 for IC-ZNE. One could also design experiments with equal overall resources by either doubling the number of repetitions per circuit for sZNE or halving them for IC-ZNE. This would slightly affect the statistical variance, but we expect that the advantage of IC-ZNE on real quantum hardware would remain. The reason is that IC-ZNE, especially when combined with randomized compiling, produces estimates with essentially negligible bias because systematic deviations are suppressed by a more accurate determination of the noise scaling.

This assumption is confirmed by the simulation results in Figure 4.10. Here IC-ZNE with 30 000 repetitions (IC-ZNE$_{s_2}$), the same number as for sZNE, is compared to the usual 60 000 repetitions. While fewer repetitions increase the variance of statistical fluctuations, the RMSE values remain almost unchanged.

In conclusion, although IC-ZNE requires more complex circuits and a higher repetition count, its ability to suppress systematic errors and deliver unbiased estimates justifies the additional resources. The robustness observed even with reduced repetition numbers underlines its practical relevance for current devices. These results, obtained on simulators with depolarizing noise, motivate further research on how the combination of IC-ZNE with randomized compiling can improve the mitigation of more complex noise in real hardware experiments.

4.7 Influence of error randomization on (Inverted-Circuit) Zero-Noise Extrapolation

In the preceding sections, we introduced sZNE and the IC-ZNE, with their methodological foundations and presenting numerical simulations to assess their performance under depolarizing noise with various strengths. However, on real quantum hardware the errors are more complex and coherent errors can arise alongside incoherent ones. A central assumption of real quantum hardware error mitigation is that randomization techniques, such as Pauli twirling via single-qubit gate insertion (see Subsection 2.4.2), convert coherent noise processes, including those affecting CNOT gates, into incoherent Pauli channels. This transformation substantially improves the accuracy of extrapolation schemes that assume incoherent, gate-independent noise.
However, the structure of IC-ZNE deviates from this framework by measuring directly the error strength through an inverted circuit (see Section 4.3). As discussed in Section 4.2, this approach implicitly assumes that the errors are effectively stochastic, such that $a = \mathrm{Tr}[\sigma\tilde{\sigma}] \ll 1$ (see Equation (4.22)). If this condition holds, randomization is not strictly necessary, since the noise estimation circuits already capture the relevant stochastic error strength. In contrast, for purely coherent noise the errors in U and $U^\dagger$ could partially or even completely compensate each other, leading to $a \approx 1$ and an underestimation of the actual noise level. In such cases, additional randomization, for instance by applying randomized compiling, can be beneficial to ensure that the measured error strength represents the true average behavior rather than coherent cancellation effects. This chapter investigates the implications of applying randomized compiling to IC-ZNE and seeks to clarify if twirling contributes to the efficacy of the IC-ZNE protocol.

We compare results from 50 independent extrapolation runs obtained on *IBM* quantum hardware (*ibmq_ehningen*), with and without randomized compiling. For each circuit, noise scaling factors $\lambda \in 1, 3, 5$ were applied, with 10 000 shots (measurement repetitions) per scaling factor. For IC-ZNE, each circuit was additionally measured to calculate ϵ with an additional number of 10 000 repetitions per scaling factor. The fitting procedure follows the approach described in Section 4.3. For IC-ZNE, a linear fit is applied, whereas for sZNE an exponential fit is used. In the absence of twirling, the same circuit is repeated 16 times, while with twirling 16 circuits are executed with randomly chosen Pauli twirls, as described in Subsection 2.4.2. The circuits used and device properties can be found in Appendices A.4 and A.1.

The analysis across the Grover, HHL, and H-Ladder circuits demonstrates consistent trends in the performance of IC-ZNE compared to sZNE and unmitigated data (Figures 4.11 to 4.13). In all cases, IC-ZNE achieves lower RMSE and fewer outliers,

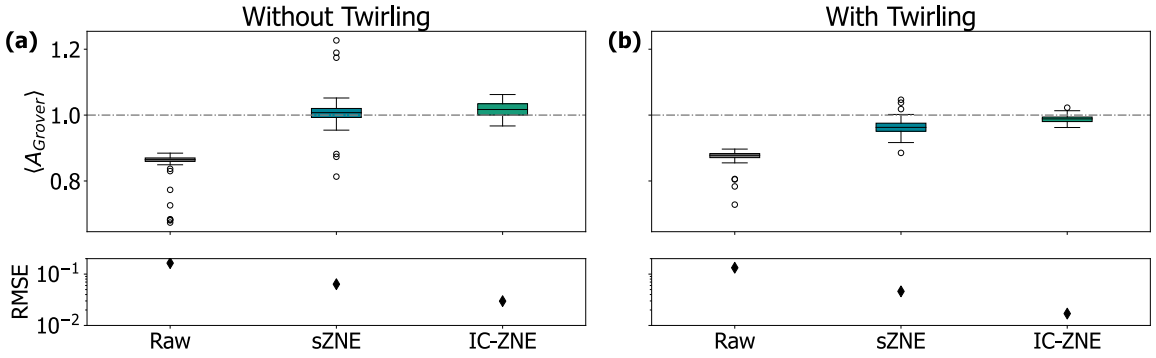


Figure 4.11: *Comparison of unmitigated raw data (Raw), sZNE and IC-ZNE for 50 runs of the Grover circuit the IBM quantum system ibmq_ehningen without (a) and with (b) Pauli twirling. The expectation values $\langle A_{Grover} \rangle$ are shown in a box plot. The root-mean-square errors (RMSE) representing the deviations from the ideal value $\langle A_{Grover} \rangle_{ideal} = 1$ (indicated by the horizontal dashed line) are shown in the plots below. The most precise result, characterized by the lowest RMSE and fewer statistical outliers (open circles), is achieved with IC-ZNE and twirling.*

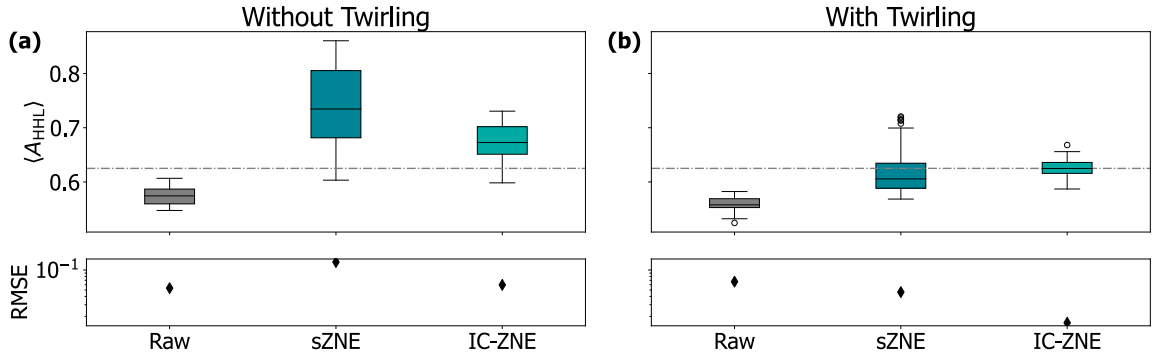


Figure 4.12: *Similar to Figure 4.11, but for the HHL circuit instead of the Grover circuit. (a) Data collected without twirling. The expectation values from both mitigation techniques do not show any improvement compared to the unmitigated raw data. (b) With twirling, IC-ZNE enhances the results, resulting in a root-mean-square error that is three times smaller than that of sZNE.*

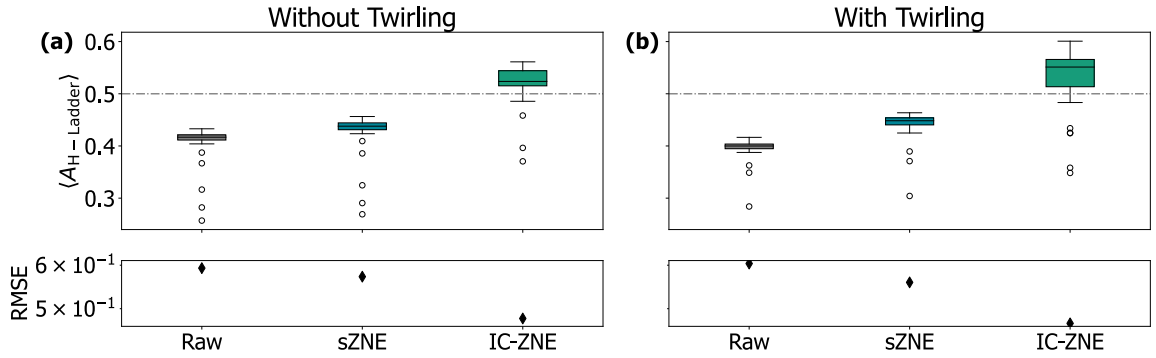


Figure 4.13: *Results for the H-Ladder circuit, comparing the performance with and without twirling. (a) Data collected without twirling gates shows that the expectation values from sZNE do not significantly improve compared to the unmitigated raw data. IC-ZNE exhibits a more substantial improvement but leads to an overestimation of the expectation value, while producing fewer outliers than sZNE. (b) With twirling, the interquartile range for IC-ZNE increases, yet it still delivers better results compared to the scenario without twirling. sZNE demonstrates a reduction in outliers along with a slight performance enhancement.*

indicating more reliable and unbiased estimates. Twirling gates further enhance IC-ZNE performance by reducing both the spread and bias of the extrapolated expectation values. By contrast, sZNE benefits less from twirling. While the number of outliers decreases, the median values may shift slightly, resulting in a minimal improvement in RMSE overall. This limited improvement may come from a more fundamental issue with sZNE, namely its assumption that error scales linearly with the number of identity insertions (e.g., three insertions should triple the error). Since this assumption does not hold in practice (see Section 4.4), sZNE remains inaccurate even when coherent errors are converted into Pauli errors by twirling. IC-ZNE, on the other hand, directly measures the effective error strength and therefore continues to benefit from twirling, as previously unmeasured coherent errors are transformed into Pauli errors and properly accounted for.

For example, in the Grover circuit (Figure 4.11), IC-ZNE provides the lowest RMSE both with and without twirling, and the number of outliers is markedly smaller compared to sZNE and raw data. Twirling further improves IC-ZNE by narrowing the distribution and stabilizing the estimates. Similarly, in the HHL circuit (Figure 4.12), the inclusion of twirling aligns the median of IC-ZNE with the exact expectation value, yielding an RMSE three times lower than that of sZNE. The H-Ladder circuit (Figure 4.13) exhibits comparable trends, with IC-ZNE consistently outperforming sZNE and mitigating outliers effectively, even when circuit-specific sensitivities amplify certain error types, such as crosstalk.

These observations highlight the robustness of IC-ZNE across different quantum circuits. The method reliably suppresses systematic errors and produces unbiased estimates, and its effectiveness is further enhanced when combined with randomized twirling. Overall, IC-ZNE provides a practical and broadly applicable noise mitigation strategy, independent of specific circuit architectures, while sZNE shows more limited and circuit-dependent improvements.

Despite the robustness of IC-ZNE's architecture-agnostic design, limited experimental resources may prevent its use. In the following chapter, we introduce adaptive scaling techniques that dynamically select feasible noise scales, offering an improvement over sZNE schedules.

4.8 Zero-Noise Extrapolation with Adaptive Scaling Factors

Even though IC-ZNE's design provides good extrapolation results compared to sZNE, there might be not enough resources to use it in experiments. The statistical variance of the extrapolated zero-noise value grows rapidly with increasing noise amplification [46] when using sZNE, limiting practical error mitigation. This effect is particularly pronounced when using predefined scaling factors, which do not account for circuit-specific noise characteristics.

To enhance the accuracy of ZNE, an optimized selection of noise scaling factors is introduced. This approach refines existing methods [38, 86, 120] by identifying a more effective spacing strategy that reduces the extrapolation error. Additionally, circuit-specific noise characteristics are incorporated into the scaling factor selection process,

ensuring that the chosen noise levels better reflect the actual error behavior of the system. The methods used are heuristic and can be applied to any quantum algorithm that aims to evaluate expectation values in a manner comparable to ZNE. We validate the effectiveness of this refined approach through experimental results on superconducting quantum hardware with different quantum circuits.

ASF provides a favorable trade-off: it can achieve error reductions at significantly lower quantum cost. When sufficient quantum resources are available and high accuracy is required, IC-ZNE remains the recommended choice. Under resource constraints, ASF offers a practical alternative that can substantially improve sZNE with few resources. Subsection 4.8.1 details how amplification factors should be selected based on the magnitude of the baseline error strength ϵ_0. As it potentially fluctuates, it should therefore be measured more accurately. Subsection 4.8.2 demonstrates how ZNE can be improved using this strategy. For reference, we also evaluate IC-ZNE combined with ASF to benchmark potential gains.

4.8.1 Equally spaced scaling factors versus exponentially spaced scaling factors

ZNE relies on executing a quantum circuit at artificially increased noise levels and then extrapolating the results back to the zero-noise limit. This requires a systematic way of scaling the noise in the circuit. As demonstrated in Subsection 2.4.1, such scaling can be achieved by repeating gates, which naturally leads to error scalings restricted to odd integer factors. In order to extend this approach to arbitrary scaling factors that deviate from odd integers, we adopt an implementation inspired by Refs. [46, 75, 121]. A more general expression than Equation (2.34) is

$$\lambda = \frac{1}{N_c}\sum_{i=1}^{N_c}(2n_i + 1) = 1 + \frac{2}{N_c}\sum_{i=1}^{N_c} n_i, \tag{4.32}$$

where N_c denotes the number of CNOT gates in the original, unscaled circuit and n_i is the number of identity insertions associated with the i-th CNOT gate.

Equation (4.32) makes clear that λ can only take values on a discrete grid with step size $2/N_c$, since the n_i's are integers. Therefore, not every scaling factor λ' can be realized exactly. The finite number of CNOT gates imposes a restriction on the set of achievable scaling factors.

To approximate a desired target value λ', one must project it onto the closest realizable value on this discrete grid. For this purpose, we employ a rounding procedure based on the floor function. This yields

$$\lambda = \frac{2\left\lfloor \frac{\lambda'-1}{2}N_c + \frac{1}{2} \right\rfloor}{N_c} + 1. \tag{4.33}$$

The floor function corresponds to the total number of identity pairs ($\sum_i n_i$) that can be inserted consistently with a target scaling factor λ'. Division by N_c converts the total

back to an average per CNOT gate, and the final addition of +1 restores the baseline contribution of the original circuit. In this way, Equation (4.33) provides the realizable scaling factor closest to λ'.

For each circuit, there exist several possible permutations of identity insertions across the N_c gates. To describe this, it is convenient to define the number of as

$$n = \frac{\lambda - 1}{2} = \frac{1}{N_c} \sum_{i=1}^{N_c} n_i. \tag{4.34}$$

Any choice of nonnegative integers $\{n_i\}$ satisfying Equation (4.34) is valid. To distribute insertions as evenly as possible, we choose $\lfloor n \rfloor$ or $\lfloor n \rfloor + 1$. Different distributions of these values among the N_c CNOT gates correspond to different valid circuit realizations. In practice, the assignment of n_i is chosen randomly for each run, and the final result is obtained by averaging over the outcomes of multiple such realizations.

As an example, consider $\lambda' = 4.2$ with $N_c = 3$. According to Equation (4.33), the closest realizable scaling factor is $13/3 \approx 4.33$. In this case, the mean number of identity insertions per CNOT is $n = (\lambda - 1)/2 = 5/3 \approx 1.67$. Therefore, the possible choices for (n_1, n_2, n_3) are

$$(n_1, n_2, n_3) \in \{(1, 2, 2), (2, 1, 2), (2, 2, 1)\}. \tag{4.35}$$

Each configuration leads to the same effective scaling factor $\lambda = 13/3$, which is close to the desired $\lambda' = 4.2$. By averaging over such randomized distributions, one obtains an unbiased estimate consistent with the intended noise scaling.

After having established the ability to select scaling factors beyond the constraints of odd integer scaling, the subsequent focus shifts to a proper choice of these scaling factors. In practice, ZNE requires evaluating the noisy circuit at several different scaling factors, and the distribution of these values plays a crucial role, as shown in [122], and determines the quality of the extrapolation. This holds not only in the presence of Markovian noise, which is typically assumed, but also under non-Markovian noise. The latter arises when a quantum system interacts with an environment whose correlation functions decay on time scales comparable to, or longer than, those of the system itself [87, 123]. Understanding and managing this spacing is therefore crucial, as it directly impacts both the stability and accuracy of the extrapolation. The subsequent discussion will address the implications of equally and exponentially spaced scaling factors.

We compare two noise-amplification schedules for ZNE. On the one hand, the linear or standard spacing, employs uniformly spaced factors, e.g. $\lambda = \{1, 3, 5\}$, see Subsection 2.4.1. On the other hand, the exponential schedule which arranges the amplification factors according to a geometric progression. This gives smaller noise amplification factors more statistical weight and larger ones less. This has two key effects. First, it reduces the variance in the extrapolated expectation value compared to placing too much weight on large amplification factors, because statistical uncertainty increases sharply for larger amplification factors [121, 122]. Second, exponential spacing helps to control bias induced by circuit depth or high physical error rates. It does so by giving more influence to points with small amplifications, where noise expansion is more reliable (lower-order terms dominate). This reduces the likelihood that the assumptions of polynomial extrap-

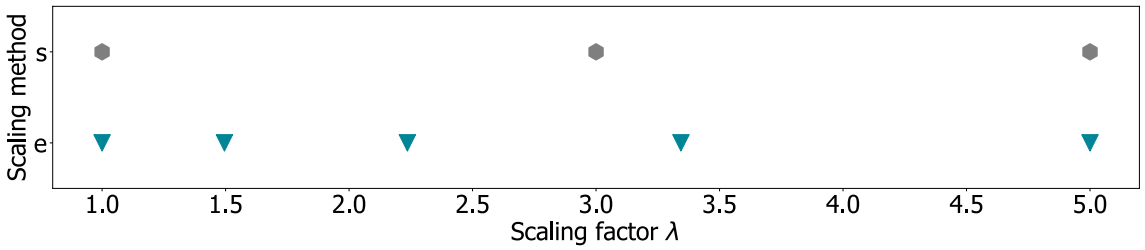


Figure 4.14: *The figure illustrates the two different scaling approaches considered. The standard equidistant scaling (s) and the exponential scaling (e). The latter is depicted with five scaling factors to emphasize the variation in methods.*

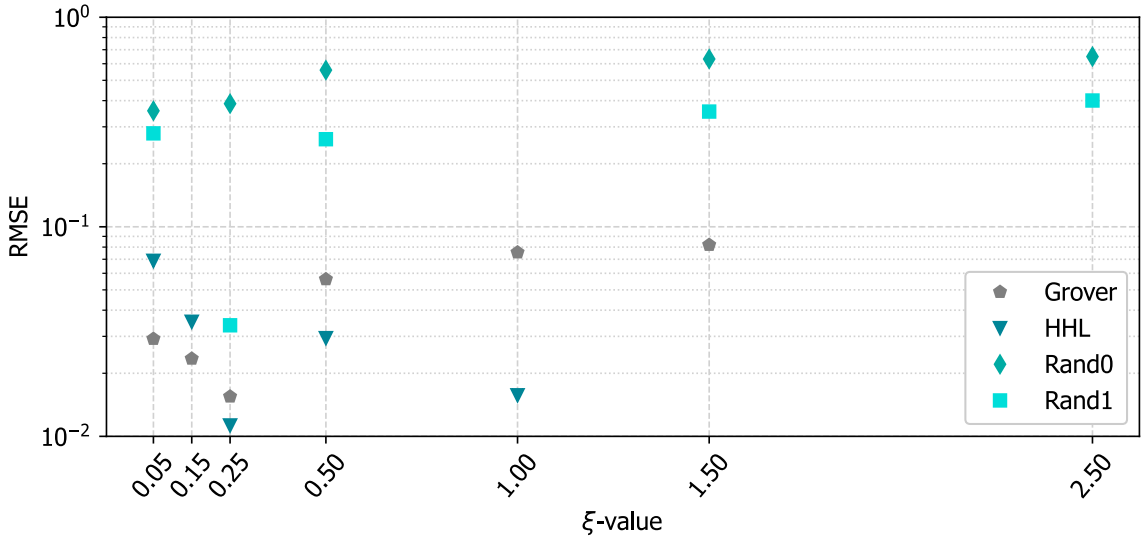


Figure 4.15: *Empirical investigation to ascertain the optimal ξ-values for the Grover, HHL, and two random circuits. The experimental findings indicated that values ranging from 0.05 to 0.25 yielded the lowest RMSE values for the aforementioned circuits.*

olation will break down. [31, 46]. However, there is a trade-off. While including larger amplification factors and more amplification factors can, in principle, better approximate the zero-noise limit, doing so tends to increase variance [122]. Furthermore, equidistant nodes experience significant shot-noise amplification as the number of nodes increases. In contrast, strategies such as exponential spacing yield lower bias for a given variance and superior overall mitigation performance [122]. Thus, exponential spacing is generally preferred when error rates are high. However, one must carefully balance and tune the number of amplification factors to avoid excessive statistical uncertainty. An example of the two different spacing is shown in Figure 4.14. The two spacing methods are illustrated with three (s - standard spacing) and five (e - exponential spacing) scaling factors to emphasize the variation between the methods.

The number and placement of amplification factors must be chosen with care, since they directly determine the accuracy and reliability of the extrapolation. A suboptimal choice, such as values clustered too closely together or not extending over a sufficiently wide range, can introduce additional estimation error. To avoid this, we adopt exponential

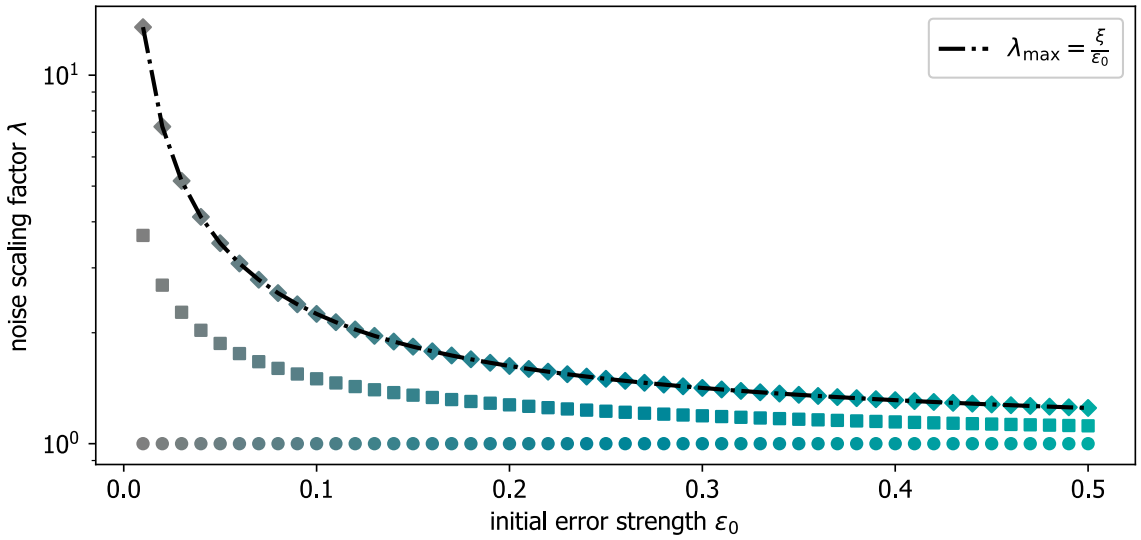


Figure 4.16: *The scaling factors λ, are plotted as a function of the initial error strengths ϵ_0 utilizing exponential spacing (see Equation (4.38)). The black dotted line represents the maximum scaling factor λ_{max}, defining the upper limit of amplification achievable under the given conditions. The data points are categorized and represented as follows: λ_1-circles, λ_2-squares, λ_3-diamonds. With 100 CNOT gates the difference between the achievable and original amplification factors (λ and λ') is negligible.*

spacing and define the maximal amplification factor λ_{max} as

$$\lambda_{max} = \frac{\xi}{\epsilon_0}, \tag{4.36}$$

where ϵ_0 is the initial error strength of the unscaled original circuit and ξ is a dimensionless tuning parameter that fixes how big the noise amplification is relative to the unscaled circuit. In this way, $\xi = \lambda_{\max}\epsilon_0$ gives an estimate of the maximum error strength of the scaled circuits. This estimate is obtained by multiplying the error strength of the original circuit by the maximum scaling factor. For ZNE to operate reliably, the maximum error strength must remain well below 1. At the same time, ξ should not be set too small, since the extrapolation requires a sufficiently broad range of noise strengths. To calibrate a practical value of ξ, we ran benchmarking circuits of Grover, HHL, and two random circuits with $\xi \in \{0.05, 0.15, 0.25, 0.50, 1.00, 1.50, 2.50\}$. For each ξ we performed 20 experiments and 10 000 shots per experiment. The computed RMSE values of ξ are presented in Figure 4.15. They exhibit a common minimum in the interval $0.05 \leq \xi \leq 0.25$. We fix $\xi = 1/8 = 0.125$ for all subsequent experiments. This value lies well inside the empirically optimal window and yields stable low RMSE values across all benchmarked circuits.

The determination of ϵ_0 is possible by either estimation of the error strength, which can be calculated by summing over the CNOT gate errors

$$\epsilon_{0,M} = \sum_{i=1}^{N_C} \epsilon_{0,i}, \tag{4.37}$$

provided by calibration values from the quantum device, or by direct measurement with inverted circuits as described in Section 4.2. The bounds for the adaptive scaling factors are $\lambda_{ASF} \in [1, \lambda_{max}]$. And the exponential target spacings can then be calculated by

$$\lambda'_j = (\lambda_{max})^{\frac{j-1}{k-1}}, \tag{4.38}$$

where k is the total number of amplification factors. The correlation between ASF and initial error strength ϵ_0 is shown in Figure 4.16. It is evident that, as the initial error rate diminishes, the range of scaling factors increases, signifying enhanced flexibility in amplification strategies. The data presented corresponds to a configuration with three amplification factors and a total of 100 CNOT gates and $0.01 \leq \epsilon_0 \leq 0.5$, with the methodology applicable to varying numbers of gates and nodes. The difference between λ and λ' with the high number of CNOT gates is negligible due to rounding for this large number of CNOT gates.

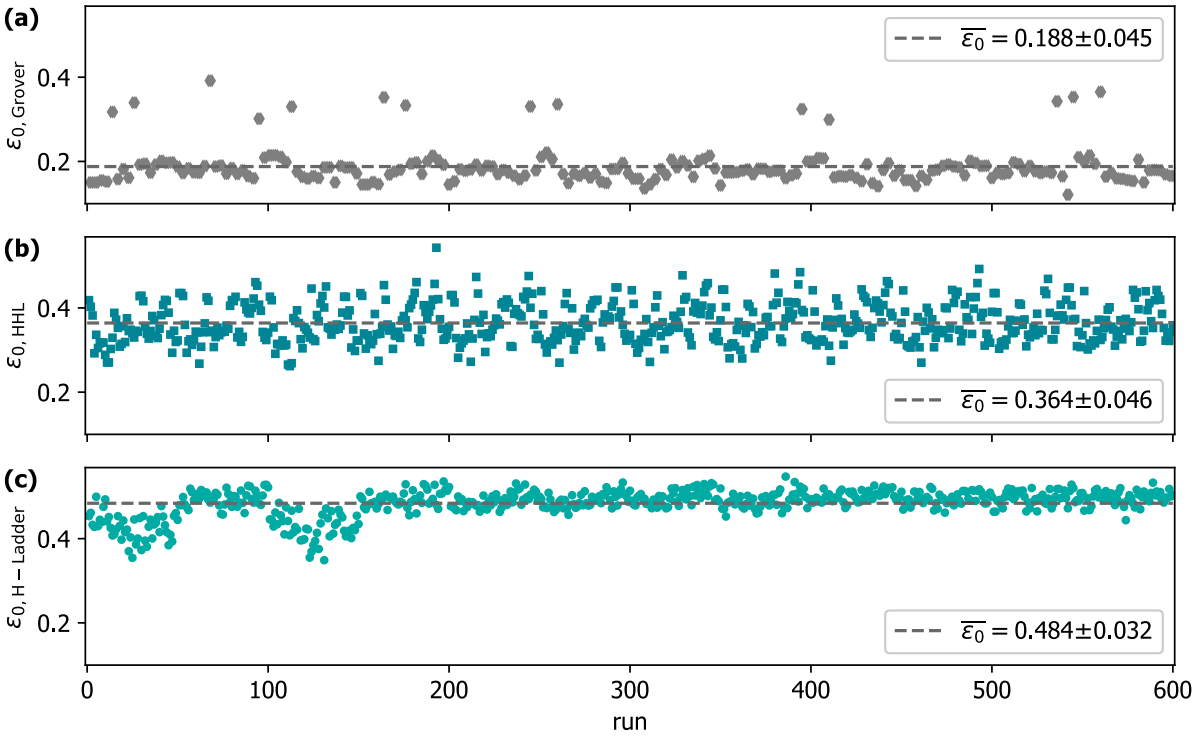


Figure 4.17: *Variation in error strength across multiple backend executions for the (a) Grover, (b) HHL, and (c) H-Ladder circuits. Each point corresponds to the measured initial error strength ϵ_0 of a single run. Grover exhibits the lowest average error $\overline{\epsilon_0}$ with occasional outliers, while HHL shows a higher but comparably variable error rate. The H-Ladder circuit has the highest average error and displays a two minima in the early runs before stabilizing after the 150th run. These results highlight that error characteristics differ markedly between circuits and fluctuate over time, underscoring the need to monitor backend stability when applying error mitigation.*

Whether ZNE should use ASF also depends on the temporal stability of the physical error rate ϵ_0. The analysis in Figure 4.17 provides valuable insight into this stability across different algorithmic structures. Grover achieves the lowest average error strength ($\overline{\epsilon_0} = 0.19$) but with notable shot-to-shot fluctuations and sporadic outliers. HHL operates at a significantly higher average error level (0.36) with similar variability. The H-Ladder circuit shows the highest mean error strength (0.48) and, unlike the other

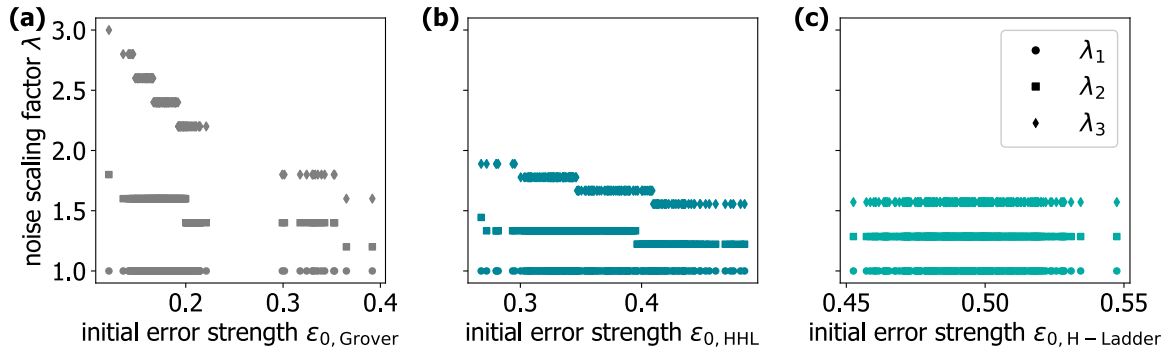


Figure 4.18: *Dependence of noise scaling factors on initial error strengths for the (a) Grover (10 CNOT gates), (b) HHL (18 CNOT gates), and (c) H-Ladder (7 CNOT gates) circuits. Each marker corresponds to one of the realized scaling factors $\lambda_1, \lambda_2, \lambda_3$ from Equation (4.33). For Grover and HHL, the plots show the expected inverse relationship between the baseline error strength ϵ_0 and the attainable noise scaling factor, consistent with the definition of $\lambda_{\max}$ in Equation (4.36). The available scaling factors also reflect the number of CNOT gates: circuits with fewer entangling gates allow larger amplification ranges. In contrast, the H-Ladder circuit exhibits a systematically higher ϵ_0, which restricts $\lambda_{\max}$ to values too small for generating finer-grained scaling. Consequently, only three distinct λ values appear, independent of variations in ϵ_0.*

cases, two minima over the first ~ 150 runs, followed by stabilization. This pattern may reflect temporal variations in backend calibration, the circuit's sensitivity to specific error channels, or chip heating.

Taken together, these results emphasize that error strengths are not only circuit dependent but also temporally dynamic. Robust ZNE strategies therefore require continuous tracking of ϵ_0 to select amplification schedules that remain within the noise-tolerant regime while still spanning a useful extrapolation range.

The dependence of the noise scaling factors on the initial error strength and the circuit structure is shown in Figure 4.18. As predicted by Equation (4.36), larger baseline error strengths correspond to smaller realizable scaling factors, thereby limiting the amplification range available for extrapolation. This effect is visible in both Grover and HHL, where ϵ_0 fluctuates but still permits multiple λ values. For the H-Ladder, however, the elevated ϵ_0 reduces $\lambda_{\max}$ so strongly that only three discrete amplification levels are feasible. From an error mitigation perspective, this limitation directly affects the resolution of the extrapolation: with fewer scaling points, bias reduction is weaker and variance grows more pronounced. These results highlight that the choice of the circuit not only influences ϵ_0 but also constrains the amplification schedule available for ZNE, underlining the need to adapt error mitigation strategies to algorithmic structure.

In the following chapter, we turn to two complementary methods for determining the initial error strength, a key parameter that directly impacts the effectiveness of ZNE with adaptive scaling factors.

Table 4.3: *Percentage reduction of RMSE for Grover, HHL, and H-Ladder circuits with adaptive scaling factors. ASF M refers to calibration data based scaling, ASF B to backend based scaling. Data is taken from Figure 4.19. The error reduction is always calculated relative to the corresponding baseline method, that is, for ZNE+ASF with respect to sZNE and for IC-ZNE+ASF with respect to IC-ZNE. Entries containing a hyphen indicate that the method proved ineffective and resulted in suboptimal outcomes.*

	ZNE+ASF M	sZNE+ASF B	IC-ZNE+ASF M	IC-ZNE+ASF M
Grover	4.87 %	–	–	–
HHL	3.37 %	12.60 %	–	–
H-Ladder	–	23.64 %	–	–

4.8.2 Impact of circuit error estimation and the choice of scaling factors

In the previous chapter, the advantages of ASF in quantum devices were examined, with particular emphasis on the fluctuations of error strength. The scaling factors are determined based on the error strength. This determination can be achieved through two distinct methodologies. This section provides a comparison of the two approaches. As described in Section 4.2, the initial error strength ϵ_0 can be measured directly on the backend at the beginning of each ZNE cycle using the unscaled circuit ($\lambda = 1$). The adaptive scaling factors obtained in this manner are referred to as ASF B, indicating that they are backend based. Compared to sZNE, this method has minimal additional resource requirement because it only needs one extra circuit to determine the initial error strength of the unscaled circuit. This makes the method highly efficient in terms of resource usage.
Alternatively, ϵ_0 can be estimated using the backend calibration data provided.In this case, the errors of all CNOT gates present in the circuit are summed. The scaling factors obtained in this way are referred to as ASF M, indicating that they are model based. This approach requires even fewer resources because no additional circuits need to be executed. Instead, the calibration data supplied by the provider can be employed directly.

The results of 50 runs on *IBM*'s quantum device *ibmq_ehningen* for each circuit are summarized in Figure 4.19. The figure compares ZNE and IC-ZNE with ASF, where the initial error strength is either derived from calibration data (ASF M) or measured directly on the device (ASF B).

For the Grover algorithm (Figure 4.19 (a)), ASF M performs better than ASF B. This can be attributed to the shallow circuit depth, the relatively small number of CNOT gates, and the simple qubit connectivity. Under these conditions, calibration data provides a reliable estimate of the actual device behavior, while direct measurements lead to a broader distribution.

For the HHL algorithm (Figure 4.19 (c)), both methods yield comparable results, but ASF B performs slightly better. In this case, backend calibration data become less representative, whereas direct measurements capture the effective noise more accurately.

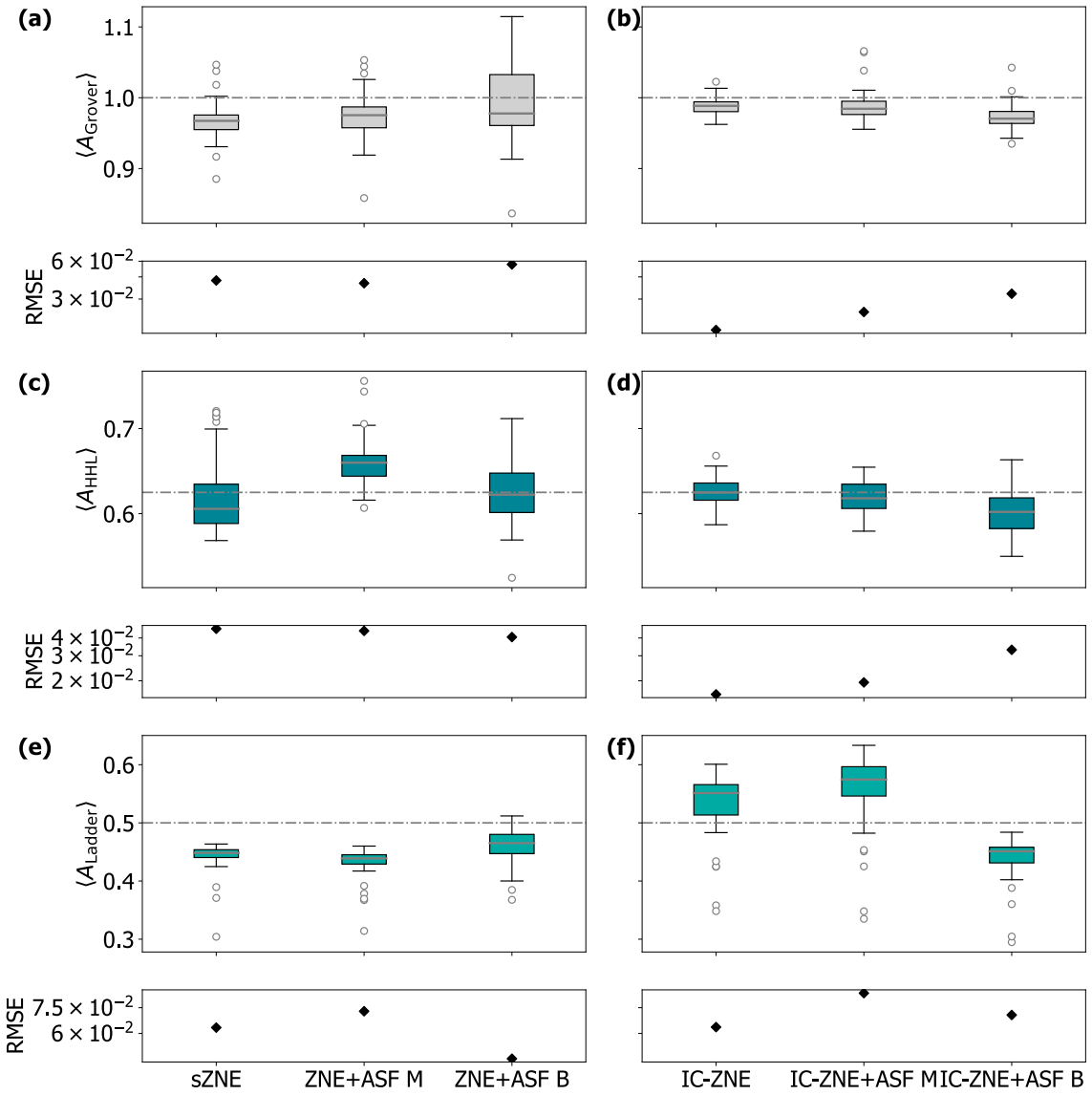


Figure 4.19: *Extrapolation results for Grover ((a),(b)), HHL ((c),(d)), and H-Ladder ((e),(f)) circuits using ZNE (left column) and IC-ZNE (right column) with ASF. The initial error strength is either estimated from calibration data (ASF M) or measured directly on the device (ASF B). Each subfigure contains a graph of the extrapolated zero-noise expectation values and, below it, a graph of the corresponding RMSE values. For the Grover algorithm (a), the estimated error strength (ASF M) appears sufficient to enhance the extrapolation results. In contrast, the HHL (c) and H-Ladder (e) circuits exhibit lower RMSE values when using measured error strengths (ASF B), which may be attributed to the larger fluctuations and higher error-strength values observed in this setting. For IC-ZNE, no clear improvement in RMSE values is visible when using either ASF M or ASF B.*

For the H-Ladder circuit (Figure 4.19 (e)), ASF B clearly outperforms ASF M. The circuit is substantially deeper and contains many entangling operations, which amplify error propagation. Consequently, calibration data fails to reflect the effective error strength, while direct measurements provide a more reliable estimate.

IC-ZNE combined with ASF does not show an improvement in these settings (Figure 4.19 (b), (d) and (f)). This can be explained by the fact that IC-ZNE explicitly measures the error strength for every extrapolation step and can therefore compensate

for varying error levels. However, it should be noted that the IC-ZNE RMSE values, as shown in Figure 4.19, are lower than the sZNE RMSE values in most cases. The percentage error reductions for all circuits and ASF methods are reported in Table 4.3, where reductions are always calculated relative to the corresponding baseline method. The highest reduction of 23.64 % is achieved for the H-Ladder circuit with sZNE and ASF B.

Although IC-ZNE achieves the the highest accuracy across most algorithms (with the exception of H-Ladder combined with ASF B), its associated cost must be considered. ASF requires a maximum of one additional circuit execution per extrapolation. In contrast, IC-ZNE introduces an overhead of a factor two since the error strength of each circuit, the original and the scaled ones, is measured separately. ASF therefore provides a favorable trade-off: it achieves noticeable error reductions at a significantly lower quantum cost.
As illustrated in Figure 4.17, the initial error strength exhibits fluctuations across different experimental runs, with some circuits showing stronger variability than others. This variability suggests that selectively discarding particularly noisy data could enhance the reliability of extrapolated zero-noise estimates, motivating the post-processing strategies examined in the next chapter.

4.9 Filtering methods

Although error mitigation, and ZNE in particular, have been widely investigated, the systematic use of classical filtering techniques has received comparatively little attention. The following section demonstrates how filtering can improve ZNE and IC-ZNE by identifying and discarding measurements that contribute disproportionately to extrapolation error. This idea is inspired by Ref. [3], where results deviating more than 2σ from the expected distribution are removed. While that work highlights the benefit of discarding statistically extreme outcomes, its filtering strategy is based on a single-variable threshold applied to expectation values only. In contrast, we extend this concept and investigate filtering strategies that exploit additional structure available in ZNE and IC-ZNE experiments. Specifically, we consider criteria based not only on observable values but also on error strengths and correlations between quantities, enabling a more fine-grained identification of unreliable data. The following subsections introduce these approaches and analyze their impact on extrapolation performance.

The filtering methods examined in this work are

- ϵ_0-filtering,
- global (ellipse) filtering,
- local (ellipse) filtering,
- a combination of ϵ_0 and local (ellipse) filtering.

In the following subchapters, we introduce the different methods, their theory, and their methodological approach. In Subsection 4.9.6, we discuss the experimental outcomes of the different post-processing methods.

4.9.1 ϵ_0-filtering

The filter method primarily uses the measurement of error strength, which has not yet been used as a filter criterion. As described in Section 4.8, the initial error strength ϵ_0 is measured for each circuit prior to noise amplification. This provides a baseline value for the error strength, allowing systematically rejecting unreliable experiments. Under normal ideal conditions, it is assumed that ϵ_0-values follow a Gaussian normal distribution

$$p(\epsilon_0) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(\epsilon_0 - \mu)^2}{2\sigma^2}\right), \tag{4.39}$$

where μ is the mean error and σ the standard deviation.

An empirical analysis of the experimental data revealed the emergence of a secondary Gaussian component in some cases. This behavior indicates that the overall distribution may be composed of multiple Gaussian contributions, plausibly associated with distinct sources of noise. Although, in principle, additional Gaussian components could arise from further noise mechanisms, the data most often exhibit no more than two resolvable contributions. From a theoretical standpoint, one would expect each independent noise process to generate an additional Gaussian peak. However, overlaps between components can obscure their individual contributions, thereby complicating unambiguous identification. Importantly, the presence of a secondary Gaussian does not compromise the robustness of the filtering procedure, as the analysis consistently excludes all values outside the 2σ interval of the primary Gaussian.

An example of the occurrence of a secondary Gaussian is shown in Figure 4.20 (a), where the initial error strength ϵ_0 from 50 independent ZNE runs of the Grover circuit is analyzed. In the figure, a larger primary distribution (orange) and a smaller secondary distribution (blue) are visible.

The second Gaussian distribution is identified using a Gaussian Mixture Model (GMM) [124], a statistical model that represents a data distribution as a mixture of multiple Gaussian components. The probability density function of a GMM with T components is defined as

$$p(x) = \sum_{t=1}^{T} \pi_t \cdot \mathcal{N}(x|\mu_t, \sigma_t^2), \tag{4.40}$$

where π_t is the mixing coefficient of the t-th component (with $\sum_{t=1}^{T} \pi_t = 1$), and $\mathcal{N}(x|\mu_t, \sigma_t^2)$ is a Gaussian distribution (as in Equation (4.39)) with mean μ_t and variance σ_t^2. The parameters μ_t, π_t, σ_t are estimated using the Expectation-Maximization (EM) algorithm, which iteratively refines the assignment of data points to components by maximizing the likelihood of the observed data. We employ the implementation provided by the *Scikit-learn* Python library [125].

Outliers are removed if they satisfy the criterion

$$|\epsilon_0 - \mu_1| > 2\sigma_1, \tag{4.41}$$

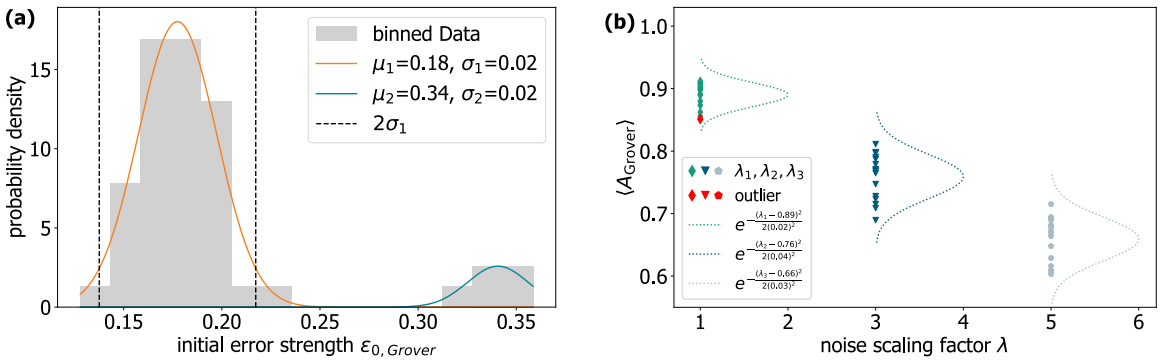


Figure 4.20: *Comparison of ϵ_0-filtering and local filtering. (a) Filtering applied to the initial error strength ϵ_0 of 50 experiments using the Grover circuit and a one-dimensional Gaussian Mixture Model (GMM). The primary and secondary distributions are indicated by orange and blue, respectively, and outliers are excluded based on a $2\sigma_1$ threshold, which refers to the first Gaussian distribution. (b) Local filtering applied to expectation values of the Grover circuit obtained in a single experimental run. Each noise scaling factor, λ_i, is indicated with its corresponding marker and fitted Gaussian envelope (parameters shown in the legend). Data points deviating by more than 2σ from the mean are highlighted in red and classified as outliers.*

where μ_1 and σ_1 denote the mean and standard deviation of the primary Gaussian component. The 2σ threshold is illustrated in Figure 4.20 (a) with vertical black dotted lines. Details of the binning procedure are provided in Appendix A.8. Once identified via Equation (4.41), entire runs corresponding to outliers are excluded from further analysis. This two-step process ensures that only reliable data is used for extrapolation.

4.9.2 Local filtering

In scenarios where only a limited number of experiments are available, relying on a single run can make robust statistical analysis difficult. In such cases, filtering techniques that discard individual data points rather than entire experiments can be advantageous. Local filtering uses a similar principle as ϵ_0-filtering but is applied within a single experiment: here, each experiment consists of three amplification factors with 16 twirled circuits each. By analyzing these twirled circuit outcomes, local filtering identifies and removes anomalous points without discarding full experiments. In our setting, the use of 16 twirling circuits per amplification factor provides enough samples to apply this method while maintaining meaningful statistics. Increasing this number further would reduce the measurement shots available per circuit and therefore increase statistical noise. Consequently, local filtering can isolate unreliable results even within a single experimental execution and enhance the robustness of the data analysis. Specifically, any expectation values $\langle A \rangle$ fulfilling the condition

$$|\langle A \rangle - \mu_A| > 2\sigma_A, \tag{4.42}$$

where μ_A and σ_A refer to the mean and standard deviation, are neglected.

Local filtering does not use a GMM, as there are only 16 data points per λ_i, which is a too small sample size to provide reliable results for a GMM. In Figure 4.20 (b), the expectation values are shown for three different noise scaling factors λ_i, each indicated

by a distinct marker. For every λ_i, the distribution of data points is compared to a fitted Gaussian (dashed curve) with mean and variance estimated from the corresponding subset. For illustrative purposes, the prefactor from Equation (4.39) is omitted. Points lying beyond 2σ from the mean are highlighted in red, marking them as outliers. This visualization illustrates how local filtering operates on each λ_i independently, ensuring that fluctuations are removed without affecting the remaining data.

4.9.3 Global filtering

Another filtering strategy is global filtering, in which the distribution of expectation values $\langle A \rangle$ is evaluated for each noise scaling factor λ_i across several independent runs. In our setup, this analysis is based on 50 runs, meaning that all expectation values corresponding to a given λ_i are grouped together. As in local filtering, data points outside the criterion in Equation (4.42) are excluded.

Under ideal conditions, one primary Gaussian distribution is expected for each scaling factor λ_i. This distribution captures the statistics of all twirled circuits associated with each λ_i (here 50x16 circuits per scaling factor). As with ϵ_0-filtering, a secondary Gaussian component may appear, therefore, a GMM is used for separation.

The key difference to local filtering is that outlier detection is based on statistics aggregated across multiple runs rather than within a single run. Compared to ϵ_0-filtering, global filtering rejects individual data points instead of entire runs. For ZNE, no additional runs are needed to measure the error strength. Therefore, fewer hardware resources are needed.

4.9.4 Local and global ellipse filtering

When filtering is applied to IC-ZNE, it must be considered that the expectation values $\langle A \rangle$ and error strengths ϵ_0 are two different measurements. Therefore, a two-dimensional Gaussian distribution can be assumed, as these two values are correlated. Hence, an ellipse is fitted to the values and used to filter outliers beyond the 2σ bound. The two-dimensional Gaussian is defined by

$$p(x,y) = \frac{1}{2\pi\sigma_x\sigma_x\sqrt{1-\rho_{x,y}^2}} \exp\left(-\frac{1}{2(1-\rho_{x,y}^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]\right), \tag{4.43}$$

where x and y are $\langle A \rangle$ and ϵ_0. The correlation coefficient $\rho_{x,y}$ can be calculated through the covariance or the standard deviation with

$$\mathrm{Cor}(x,y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(y)}} = \frac{\sigma_{x,y}}{\sigma_x\sigma_y} =: \rho_{x,y}. \tag{4.44}$$

This method can also be applied to compute confidence ellipses for identifying and excluding outlier data points. As shown in Figure 4.21 (a), a single Grover experiment

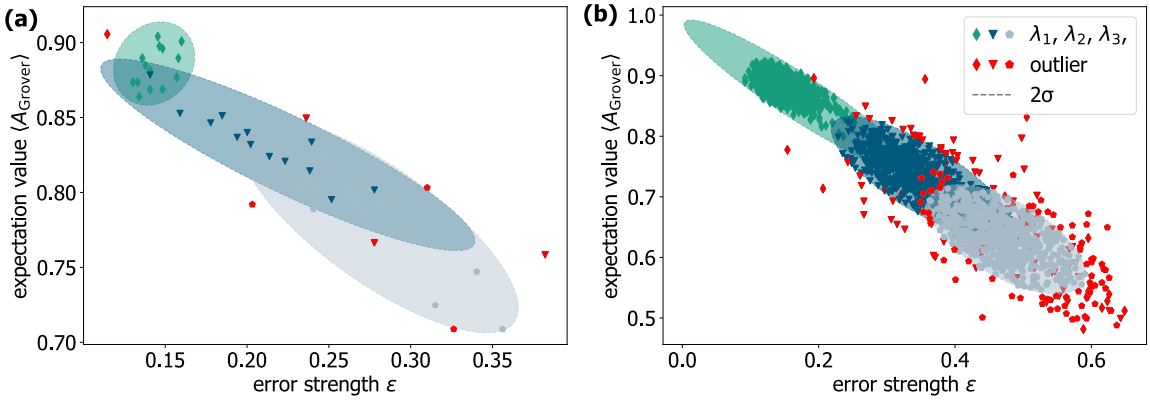


Figure 4.21: *(a) Local filtering based on confidence ellipses obtained from a two-dimensional Gaussian distribution. Each point represents a twirled circuit instance of a single Grover experiment for one of three amplification factors (markers: diamonds, $\lambda_1 = 1$; triangles, $\lambda_2 = 3$; pentagons, $\lambda_3 = 5$). Two-sigma confidence ellipses are shown as dashed contours, with outliers outside the ellipses highlighted in red. (b) Global filtering applied across 50 Grover experiments, using the same marker conventions as in (a).*

with three amplification factors and 16 twirled circuits is depicted. Two-sigma confidence ellipses for each amplification factor are indicated by dashed lines, with outliers beyond the ellipses highlighted in red. Using this technique, individual data points can be discarded, potentially improving the extrapolation result. However, as discussed in Subsection 4.9.2, if too few data points remain, the reliability of the statistical estimates may degrade.

For global filtering, multiple runs are analyzed jointly, as illustrated in Figure 4.21 (b). Data points corresponding to the same amplification factors across different runs are compared and excluded according to the same statistical criterion. The procedure for calculating the ellipse parameters is described in detail in Appendix A.8.

A key advantage of the global ellipse filtering approach is that outliers can be identified more robustly by leveraging statistics across many experiments, reducing the impact of spurious fluctuations in individual runs. On the other hand, global filtering requires a larger amount of data and may inadvertently remove valid points if systematic shifts occur between runs, potentially biasing the extrapolation. For this purpose, local filtering would be more suitable than global filtering.

4.9.5 Combined ϵ_0- and local (ellipse) filtering

Outliers may originate from both systematic sources, such as hardware instabilities or calibration drifts, and statistical fluctuations inherent to the measurement process. To mitigate the influence of both types of noise, we implement a two-level filtering strategy. At the run level, entire runs with anomalously high initial error strengths ϵ_0 are excluded (see Subsection 4.9.1). At the intra-run level, local filtering (or local ellipse filtering in case of IC-ZNE, see Subsections 4.9.2 and 4.9.4) identifies and removes outliers within each amplification factor.

This hierarchical approach separates fundamentally inconsistent runs from isolated statistical deviations. ϵ_0-filtering thereby suppresses systematic errors, while local filtering reduces the impact of random fluctuations on the extrapolation fits. By successively eliminating both classes of artifacts, the method is supposed to increase the robustness of the analysis and to ensure that the extrapolated trends reflect the underlying physical behavior rather than noise.

4.9.6 Hardware evaluation of filtering methods

Having established the theoretical foundations in the previous chapters, we now turn to the experimental validation. Our goal is to analyze hardware runs using the different filtering techniques and identify the most effective ones. In particular, we compare sZNE, ASF, IC-ZNE, and their combinations with different filtering techniques.

For each method or combination, we perform 50 experiments with three scaling factors and 16 twirling circuits per scaling factor. Each experiment consists of 10 000 measurement shots. The circuits and ZNE methods are applied as described in the preceding sections. For the ASF method, we employ the variant that yielded the best results in Section 4.8, using ϵ_0 measured either on hardware (ASF B) or from the model (ASF M).

To illustrate the impact of filtering, Figure 4.22 reports the percentage error reduction for the Grover, HHL, and H-Ladder circuits when using the four filtering techniques in combination with three ZNE methods. The percentage is calculated relative to results without filtering.

For the Grover circuit, the most significant error reduction is obtained using ASF M and ϵ_0-filtering (see Figure 4.22 (b)). This improvement can be attributed to the enhanced error estimation and adaptive nature of the scaling factors, which may support a more effective filtering process. In contrast, IC-ZNE yields the least favorable results, as shown in Figure 4.22 (c). Since IC-ZNE already produces a robust error estimate during execution, additional filtering offers little to no benefit for the extrapolated outcomes. Global filtering can even worsen the results by up to 12.1%, as it fails to account for the unique error distributions inherent to individual runs. A slight improvement for IC-ZNE is observed with local ellipse filtering, which removes outliers originating from single runs. The effectiveness of global, ϵ_0-, and ϵ_0+local-filtering methods suggests that sources of error in the Grover circuit are occasional outliers rather than drifts, which these methods suppress effectively. For IC-ZNE, when the underlying error measurement is already reliable, additional filtering tends to provide limited to no benefit and may distort the data distribution.

For the HHL circuit, the best error reduction of 39.1% is achieved with sZNE combined with local filtering. The second most effective error reduction is observed for IC-ZNE under global filtering. Interestingly, the combined filtering method performs worse for sZNE despite the individual efficacy of local and ϵ_0-filtering, which leads a minimal increase in error reduction of only 7.9%. Notably, unlike the Grover circuit, all filtering methods except local filtering reduce the error for IC-ZNE. This contrast suggests that HHL is not dominated by isolated outliers, but rather by a broader noise distribution (see Figure 4.17). Local filtering is therefore most beneficial for sZNE because it eliminates runs with poor performance. IC-ZNE's error strength measurement can in this case

be improved by filtering, which helps to stabilize the extrapolation against variability, although IC-ZNE already takes this variability into account through the error strength measurement.

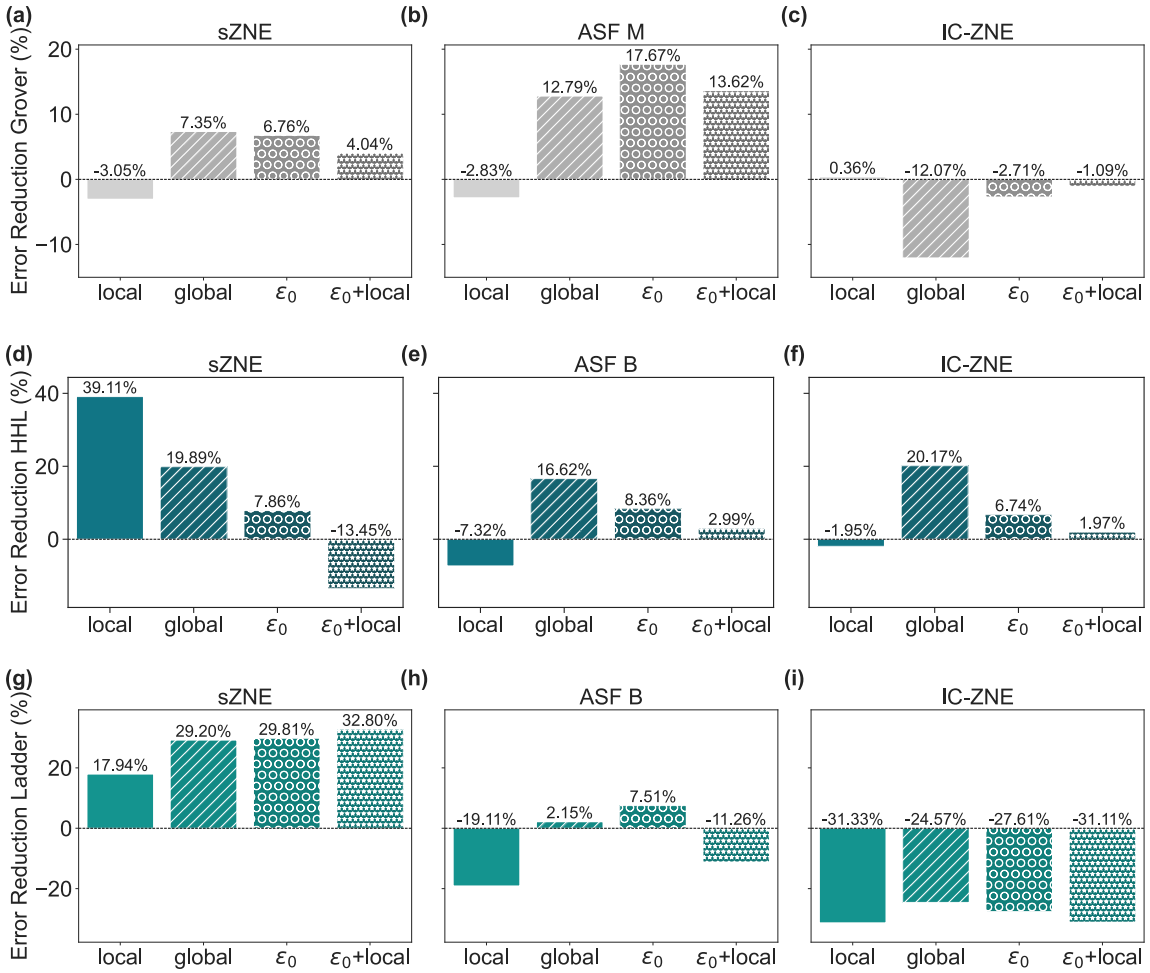


Figure 4.22: *Error reduction for the Grover [(a)–(c)], HHL [(d)–(f)], and H-Ladder [(g)–(i)] circuits using three error mitigation methods (sZNE, ASF, and IC-ZNE) under four different filtering strategies. (a)–(c) Grover: (a) sZNE shows the most significant improvement with global filtering. (b) ASF achieves the best error reduction using ϵ_0-filtering. (c) IC-ZNE yields only minor improvement, with local filtering offering the best (though limited) effect. (d)–(f) HHL: (d) sZNE benefits most from local filtering. (e) ASF shows maximum error reduction with global filtering. (f) IC-ZNE improves significantly under global filtering. (g)–(i) H-Ladder: (g) sZNE achieves the greatest error reduction through the combination of ϵ_0+local-filtering. (h) ASF improves only with global and ϵ_0-filtering. (i) IC-ZNE offers no enhancement with any filtering strategy. Overall, sZNE and ASF gain the most from filtering across all circuits. IC-ZNE remains comparatively weak and is only effective for the HHL circuit with global filtering. These results indicate that IC-ZNE might already capture a substantial part of the relevant error behaviour during execution, which may limit the extent to which filtering can further enhance the extrapolated outcomes. The effectiveness of a filtering strategy therefore depends on the mitigation method used and on the structure of the underlying circuit.*

The H-Ladder circuit exhibits the best results with sZNE combined with ϵ_0+local filtering, achieving an error reduction of 32.8%, shown in Figure 4.22 (g). Similar to the Grover circuit, the filtering methods offer no improvement for IC-ZNE and in fact tend to worsen the results. For the ASF runs, only ϵ_0-filtering results in a modest improvement of 7.5%.

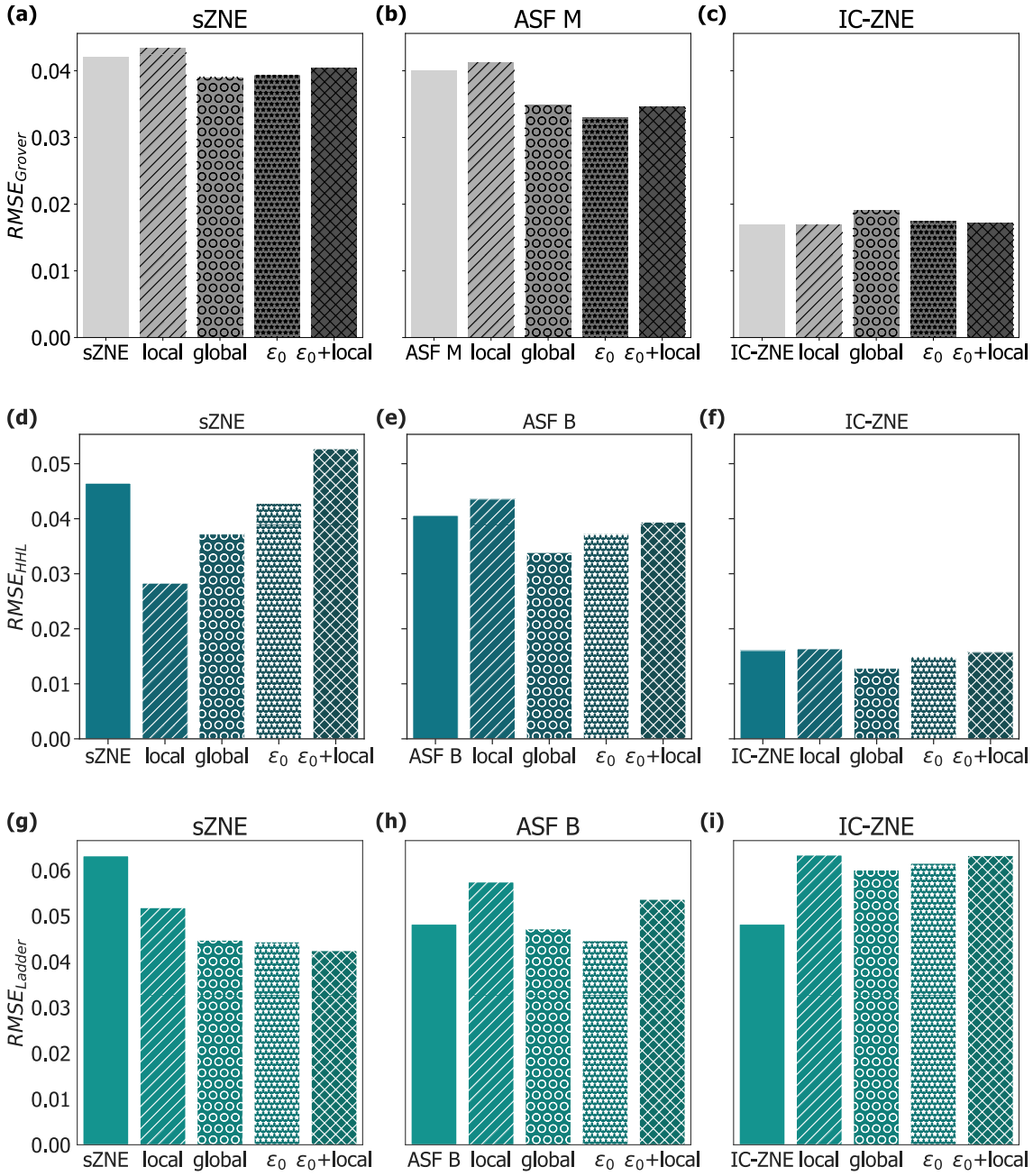


Figure 4.23: *RMSE values for Grover (a)–(c), HHL (d)–(f), and H-Ladder (g)–(i) circuits under various filtering techniques. Each subplot compares the performance of sZNE, ASF, and IC-ZNE when combined with no filtering, local filtering, global filtering, ϵ_0-filtering, and combined ϵ_0+local filtering. The RMSE value is calculated with respect to the ideal zero-noise result.*

Although the H-Ladder circuit contains only 7 CNOT gates, the initial error strength is high (see Figure 4.17). This may limit the effectiveness of ASF and filtering does not provide a substantial improvement. For sZNE, filtering reduces the impact of occasional outliers through ϵ_0-filtering and and mitigates drift effects through local filtering, so their combination yields the strongest error reduction. Once again, IC-ZNE does not benefit, which reinforces the idea that its error-strength-measurement-based extrapolation

already captures most of the relevant error behavior. This leaves little scope for additional filtering to improve the results.

In conclusion, no single filtering method consistently outperforms the others across all circuits. Global filtering often provides an improvement, although its effectiveness depends strongly on the circuit and the mitigation method. For IC-ZNE, filtering is generally not beneficial and may even degrade performance, since the method already incorporates a robust error-strength estimate.

A comparison of RMSE values for the different filtering methods, shown in Figure 4.23, provides an additional perspective on their performance and allows a direct comparison with IC-ZNE. In several cases, such as in subfigures (f), (g) and (h), filtering combined with sZNE or ASF achieves lower RMSE values than IC-ZNE. In subfigure (f), global filtering even improves IC-ZNE itself, which is the only instance where filtering benefits this method. Despite these exceptions, IC-ZNE still yields the lowest RMSE values in nearly all scenarios, indicating that its error-strength measurement generally captures the dominant noise behaviour.

Across the circuits, global and ϵ_0 filtering reduce RMSE most consistently for sZNE and ASF, while local and ϵ_0+local filtering show more variable performance. The data do not support the assumption that combining local and ϵ_0 filtering systematically improves the mitigation process, as this approach can also remove excessive data or distort the distribution.

Notably, for the H-Ladder circuit, ASF B combined with ϵ_0 filtering outperforms IC-ZNE, showing that filtering strategies can exceed IC-ZNE performance when the circuit structure and noise characteristics are favorable. Nevertheless, in most cases IC-ZNE remains the most robust method in terms of RMSE. Filtering with sZNE or ASF therefore represents a practical lower-cost alternative that can approach, and occasionally surpass, the performance of IC-ZNE when full IC-ZNE runs are not feasible.

Table 4.4 shows the fraction of outliers discarded for each circuit, mitigation method, and filtering strategy. Of all the approaches, the combination of ϵ_0-filtering and local filtering is the most aggressive. This method eliminates the largest share of data, up to 60.9% in the case of HHL with IC-ZNE and similarly high percentages across other circuits. In contrast, global and local filtering alone are far more conservative, typically discarding only a few percent of the data (usually below 5% for sZNE and ASF and slightly higher for IC-ZNE). Notably, the extrapolated results often improve even when large fractions of runs are removed, indicating that the filtering criteria identify runs with substantially higher noise levels without degrading the quality of the remaining data.

These filtering statistics provide additional insight into the behavior of current quantum hardware. The fact that removing more than half of the data can still improve the result suggests that a significant portion of runs is affected by transient or calibration-related fluctuations rather than stable noise characteristics. Filtering therefore not only removes inconsistent data but also highlights the more reliable measurement regime within the device's operation. At the same time, the results underline the importance of choosing an appropriate filtering strategy: while global filtering benefits from increased statistical robustness across many experiments, it relies on large data sets and may inadvertently

discard valid data in the presence of systematic run-to-run shifts. In such cases, local filtering approaches may offer a more suitable alternative.

Table 4.4: *The percentage of outliers removed for each circuit (Grover, HHL, and H-Ladder) is shown for different error mitigation methods (sZNE, ASF, and IC-ZNE). The results are reported for four filtering strategies: global, local, and two strategies with ϵ_0-filtering: without and with additional local filtering. Lower percentages indicate fewer discarded experiments, whereas higher percentages indicate a higher number of outliers.*

Circuit	Method	Filtering	Outlier (%)
Grover	sZNE	global	3.7
		local	4.3
		ϵ_0	42
		ϵ_0+local	44.5
	ASF	global	4.7
		local	4.3
		ϵ_0	38
		ϵ_0+local	40.8
	IC-ZNE	global	9.4
		local	11.5
		ϵ_0	44
		ϵ_0+local	50.7
HHL	sZNE	global	4.2
		local	4.4
		ϵ_0	58
		ϵ_0+local	59.7
	ASF	global	6.4
		local	4.3
		ϵ_0	62
		ϵ_0+local	63.7
	IC-ZNE	global	13.4
		local	10.9
		ϵ_0	56
		ϵ_0+local	60.9
H-Ladder	sZNE	global	4.8
		local	3.6
		ϵ_0	58
		ϵ_0+local	59.2
	ASF	global	4.5
		local	4.0
		ϵ_0	56
		ϵ_0+local	57.7
	IC-ZNE	global	13.5
		local	10.2
		ϵ_0	48
		ϵ_0+local	53.4

Having established the effectiveness of different filtering strategies, we will now examine the influence of folding schemes. Since folding determines how noise is amplified, the choice of folding scheme directly affects the performance of ZNE variants. In the following, we compare global and local folding within adaptive and error-measurement-based settings to assess their practical impact.

4.10 Analysis of folding schemes

Noise amplification in ZNE can be achieved using either global or local folding. As discussed in Section 2.4, the chosen folding strategy can influence the mitigation performance. To better comprehend the practical implications of these strategies, this section presents a detailed analysis of their impact on the performance of ZNE and IC-ZNE when combined with ASF.

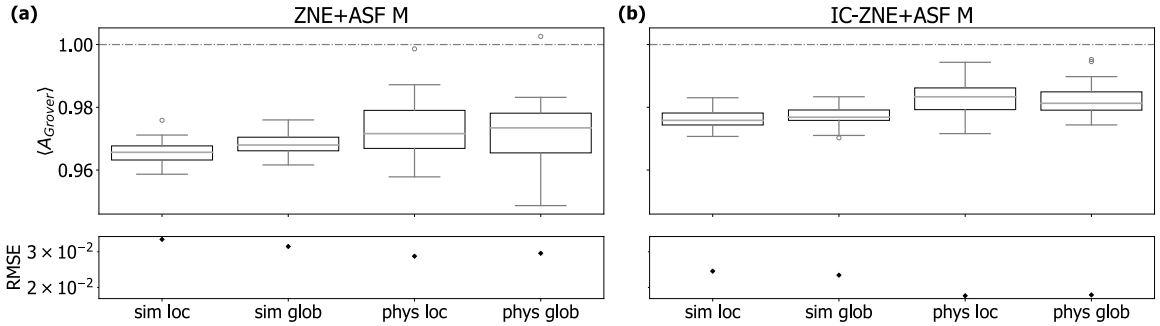


Figure 4.24: *Comparison of local and global folding in ZNE and IC-ZNE with ASF M applied to the Grover circuit. (a) ZNE+ASF M and (b) IC-ZNE+ASF M. The upper panels show box plots of extrapolated $\langle A_{Grover} \rangle$ from 30 simulation runs (sim-loc/glob) and corresponding hardware experiments (phys-loc/glob). The horizontal dotted line resembles the ideal value $\langle A_{Grover} \rangle_{ideal} = 1$. The lower panels report the root-mean-square error (RMSE) of each method. Local folding achieves lower RMSE on hardware, while global folding performs better in simulations.*

A key motivation for restricting the comparison of global and local folding to the ASF setting is that ASF already modifies the effective noise-scaling procedure. Since ASF aims to adaptively adjust the amplification to match the noise characteristics of the device, the interplay with different folding strategies becomes more subtle and potentially more impactful than in other settings. Previous works [84, 126] have shown, for example, that global folding tends to perform better under time-correlated noise, suggesting that the combination with ASF could further improve robustness. By testing both global and local folding within ASF, we can directly assess whether adaptive scaling leverages the strengths of one scheme over the other, and whether these observations extend to IC-ZNE as well.

In ASF with global folding, we decompose

$$n = n_1 + n_2, \tag{4.45}$$

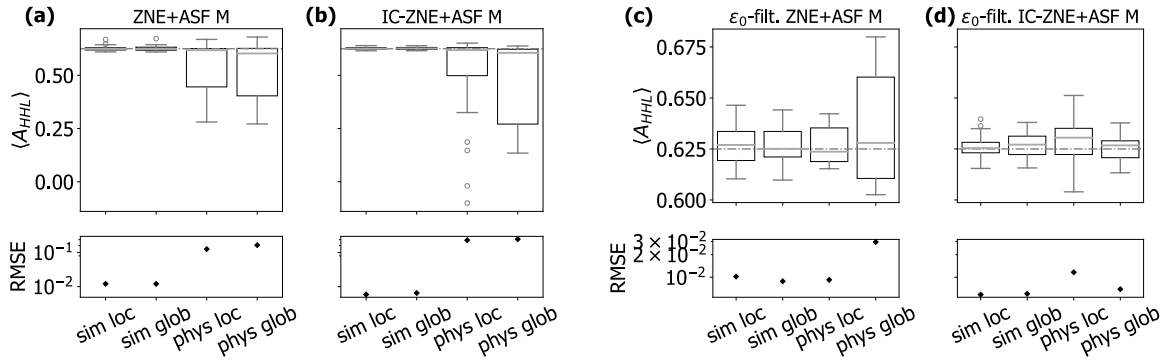


Figure 4.25: *Comparison of local and global folding in ZNE and IC-ZNE with ASF M applied to the HHL circuit. (a)–(b) show results for HHL without filtering, and (c)–(d) for HHL with ϵ_0-filtering. The subfigure layout mirrors that of Figure 4.24. The horizontal dotted line resembles the ideal value $\langle A_{HHL} \rangle_{ideal} = 5/8 = 0.625$. Without filtering, hardware runs exhibit high variability, with IC-ZNE producing many outliers (points in the box plots). Applying ϵ_0-filtering improves hardware performance, and the combination of IC-ZNE with global folding yields the lowest RMSE.*

where $n_1 \in \mathbb{Z}_{\geq 0}$ and $n_2 \in [0, 1)$. The integer part n_1 is realized through unitary folding of the entire circuit, $(UU^\dagger)^{n_1}$, which corresponds to a base scaling factor

$$\lambda = 2n_1 + 1. \tag{4.46}$$

The fractional part n_2 is implemented by local folding as described in Subsection 4.8.1. The achievable fractional resolution is limited by the available CNOT gates. Therefore, the realized fractional scaling may differ slightly from the requested value. This leads to a small discrepancy between the desired scaling factor λ and the implemented one λ'.

Each experiment is based on 30 runs, performed with Pauli twirling and ASF using ϵ_0 from the noise model. The runs are carried out both on hardware and in simulation. The simulations use a device-specific error model that is designed and provided by *IBM*, designed to approximate the behaviour of the corresponding device. The parameters and description of the error model can be found in Appendix A.1. For HHL runs, outliers were excluded using ϵ_0-filtering (see Section 4.9) so that the analysis reflects stable trends rather than being dominated by extreme data points. Five amplification factors are used in the folding experiments. Although the number could be reduced for consistency with prior sections, doing so would significantly alter the extrapolation behavior, particularly for the HHL algorithm, and is therefore avoided. Each experiment uses 10 000 shots for ZNE with ASF and 20 000 shots for IC-ZNE with ASF, the latter requiring twice as many measurements because of the additional circuit used to estimate the error strength. As the confidence interval scales with $1/\sqrt{N}$, halving the number of shots increases statistical uncertainty by approximately $\sqrt{2}$. Simulations, however, indicate that this increased uncertainty has a negligible effect on the outcome (see Figure 3.2) and sampling noise is not a dominant error source. The results for Grover and HHL are shown in Figures 4.24 and 4.25, respectively.

For Grover's algorithm, the simulation and hardware results align qualitatively. In Figure 4.24 (a), expectation values cluster slightly below the ideal value, with global folding performing better in the simulation than on the physical device. Local folding on

hardware gives the lowest RMSE value and more stable outcomes, while global folding introduces a broader distribution. With IC-ZNE in Figure 4.24 (b), the extrapolated values lie closer to the ideal, and hardware runs achieve lower RMSE values than simulations, although with wider distributions. The overall lowest RMSE is obtained for IC-ZNE with ASF M on hardware using local folding.

For HHL, the discrepancies between simulation and hardware are more pronounced. In Figure 4.25 (a), simulation underestimates the errors observed on the backend where ZNE+ASF M achieves better RMSE values than IC-ZNE+ASF M (see Figure 4.25 (b)). On hardware, local folding slightly better than global folding for both methods, although both exhibit wide distributions and the difference is small. In Figure 4.25 (b), local folding reduces the RMSE value but introduces more outliers. These observations suggest that hardware behaviour is influenced by effects not captured by the simplified error model, such as Two-Level Systems (TLS) [127, 128], qubit crosstalk, non-Markovian noise, or other correlated error sources. In this setting, filtering becomes advantageous.

After applying ϵ_0-filtering to the HHL results shown in Figures 4.25 (c) and (d), the hardware distributions improved significantly. In Figure 4.25 (c) for ZNE+ASF M, local folding leads to the lowest RMSE value, whereas global folding on hardware shows the largest variability. In simulation, no folding scheme appears more favorable than the other. For IC-ZNE with filtering, shown in Figure 4.25 (d), the variability decreases notably, and filtering significantly improves extrapolation quality. On hardware, global folding achieves the lowest RMSE value, while in simulation local folding performs best. This indicates that filtering interacts differently with each folding strategy and can, under certain conditions, make global folding advantageous despite its larger gate overhead.

The discrepancies observed in Figures 4.24 and 4.25 (a) and (b) are likely due to physical effects not fully represented in the hardware error model. As a result, simulations tend to underestimate hardware noise, and hardware outcomes can deviate more strongly. For Grover, noise correlations in the device may incidentally improve extrapolation stability, resulting in lower hardware RMSE values than in simulation. For HHL, the deeper circuit depth and higher number of CNOT gates makes the method more sensitive to error amplification. In this example the excessive variability in data, can be reduced by filtering but not fully removed.

Across Grover and HHL we do not identify a single preferred folding scheme. IC-ZNE (with ϵ_0-filtering) is the most robust configuration overall, as it improves extrapolation for both circuits and often yields the lowest RMSE value. Comparing global and local folding across algorithms and methods shows no consistent advantage of either scheme. In the device-specific noise model, uniform amplification via global folding can enhance extrapolation, consistent with more homogeneous or weakly correlated noise. On hardware, however, the noise is more complex and often correlated, so global folding can amplify these effects beyond what extrapolation can compensate. Local folding is not inherently better in this regard, as it can similarly magnify circuit-dependent noise mechanisms.

4.11 Summary of Inverted-Circuit ZNE and mitigation refinements

This chapter presented our findings in the further development and evaluation of error mitigation techniques, with a focus on Zero-Noise Extrapolation (ZNE). The critical question in ZNE is how the error strength can be controlled and scaled as accurately as possible. With the introduction of the new Inverted-Circuit ZNE (IC-ZNE) technique, this challenge is addressed based on a precise definition of the error strength ϵ. This technique allows the error strength to be measured directly on the device without significant additional effort. Then, the extrapolation can be performed using the measured error instead of an assumed error scaling factor, which is only approximately valid.

Across the investigated settings, IC-ZNE demonstrated the highest degree of robustness. By avoiding explicit assumptions about noise scaling and instead using the measured error strength, IC-ZNE achieved the lowest Root-Mean-Square Error (RMSE) values in almost all experiments. In one case, the extrapolation error was reduced by more than 80% compared to standard ZNE. While IC-ZNE requires an additional circuit execution to measure the error strength, effectively doubling the number of shots, this overhead was compensated by the improved extrapolation accuracy.

Pauli twirling reduced coherent contributions and led to more stable behaviour, particularly for IC-ZNE on hardware. Pauli twirling randomizes coherent errors by inserting random Pauli gates that transform them into an effective stochastic (Pauli) noise channel upon averaging. Standard ZNE benefited to a lesser extent, as twirling alone did not eliminate its sensitivity to model mismatch.

As a further aspect, we investigated strategies for selecting and optimizing error scaling factors. While circuit folding with fixed scaling factors of 1, 3, and 5 is used by default, Adaptive Scaling Factors (ASF) allow the scaling to be tailored to the backend noise strength, with the largest benefit observed for standard ZNE. Adaptive scaling selects noise amplification factors based on an estimate of the initial circuit error strength ϵ_0 and distributes the scales exponentially. Two variants were considered. ASF M uses error estimates obtained from the noise model, while ASF B measures the error strength on hardware. No clear preference between the two variants was observed across the studied circuits. Both reduced variance and improved extrapolation stability when the underlying error estimate was sufficiently accurate, achieving error reductions of up to 24% at minimal additional cost. Their impact on IC-ZNE was limited, as IC-ZNE already measures the quantities required for its extrapolation.

Filtering served as an additional tool to remove highly noisy runs or individual outliers. Global filtering and ϵ_0-filtering most consistently reduced RMSE in standard ZNE and ZNE combined with ASF. Local filtering and combined ϵ_0+local filtering were more aggressive and occasionally removed a substantial fraction of the samples. However, these methods could still improve the fit when noise fluctuations dominated. The highest error reduction of 39% was observed for standard ZNE combined with local filtering for the Harrow-Hassidim-Lloyd circuit. Filtering rarely improved IC-ZNE, consistent with the method already capturing much of the relevant error behavior.

The comparison between global and local folding showed no universal preference. Simulations with device-level noise models indicated an advantage for global folding with

the Grover circuit, consistent with more homogeneous or weakly correlated noise. On hardware, where correlations and temporal variations are stronger, neither folding scheme dominated consistently. Under certain conditions, such as the Harrow-Hassidim-Lloyd circuit with filtering, global folding produced lower RMSE, but these cases remained specific to individual circuits.

While the presented results demonstrate that refined ZNE-based methods can substantially reduce bias and variance for individual circuits, they also highlight that mitigation performance depends sensitively on stability and reproducibility across repeated executions. This aspect becomes particularly relevant for variational algorithms, which rely on many circuit evaluations within an iterative optimization loop and can amplify residual bias or fluctuations into degraded convergence behavior. In the following chapter, we therefore study the application of error mitigation techniques in variational settings and analyze how their performance influences optimization dynamics and solution quality.

5 Quantum error mitigation for variational algorithms

The results presented in this chapter were obtained in collaboration with Rinor Kelmendi (Karlsruhe Institute of Technology, KIT) and Tobias Nauck (Fraunhofer Institute for Applied Solid State Physics, IAF). The algorithms were implemented by Rinor Kelmendi (GM-QAOA) and Tobias Nauck (VQE), while all error mitigation procedures, data analysis, and interpretation were performed by the author.

In the previous chapters, we examined noise and how its impact can be mitigated by Quantum Error Mitigation (QEM) techniques, using circuits with a single known observable. Building on this foundation, we will now turn to near-term variational algorithms. These algorithms require the solution to be determined through optimization and can be implemented on Noisy Intermediate-Scale Quantum (NISQ) devices.

This chapter focuses on two important hybrid quantum-classical algorithms: the Quantum Approximate Optimization Algorithm (QAOA) [96], already introduced in Section 2.5, which is used for combinatorial optimization problems, and the Variational Quantum Eigensolver (VQE) [99], which estimates molecular ground-state energies. These algorithms provide valuable benchmarks for assessing performance, noise resilience, and scalability of NISQ-era hardware.

To systematically assess the performance of these variational algorithms, we perform simulations of both QAOA and VQE under ideal conditions and with controlled quantum hardware noise models. We also investigate how ZNE and IC-ZNE improve the quality of results and discuss how these insights can support the preparation of practical experiments on real quantum devices.

This chapter is structured as follows. In Section 5.1, we investigate the Grover-mixer QAOA (GM-QAOA) [98] and present simulation results across different circuit depths, both in the noise-free setting and with noise mitigation using ZNE and IC-ZNE. In Section 5.2, we examine VQE simulations for the dihydrogen (H_2) and potassium hydride (KH) molecules and compare the effect of applying QEM either during the entire optimization or only in the final evaluation step (inspired by Ref. [129]). For KH, we additionally employ a scaled noise model to emulate realistic near-future hardware performance. Finally, we analyze the double-minimum behavior observed for KH and discuss whether it originates from molecular properties or from limitations of the ansatz and optimizer.

5.1 Error Mitigation for the Quantum Approximate Optimization Algorithm

As a first application, the Grover-mixer QAOA (GM-QAOA) is applied to the Traveling Salesman Problem (TSP) [130] as a benchmark for combinatorial optimization. Building on the theoretical background presented in Subsection 2.5.1, this section examines the impact of hardware noise and the effectiveness of error-mitigation methods on the algorithm's output quality.

The use case considered here is the TSP. Given a set of cities and pairwise distances, the task is to find a closed tour that visits each city exactly once and has minimal total length. As a classical baseline, we implement a dynamic programming algorithm following the Held–Karp approach [131]. This method computes the optimal tour length by recursively building up minimal costs for partial paths using the distance matrix, and is used here to obtain reference solutions for small TSP instances. Since the TSP instance was generated randomly, the cost of the optimal route C_{min} was not evaluated. Therefore, comparisons are made relative to the noise-free GM-QAOA reference.

For the quantum approach, we encode a three-city TSP instance into four-qubits using a problem-specific binary encoding in which each qubit represents a particular segment of the tour. The state-preparation circuit generates a superposition over all valid candidate assignments encoded in this basis. The cost Hamiltonian encodes the path lengths and includes penalty terms that assign higher costs to invalid configurations, such as visiting two cities at the same time step or revisiting the same city, so that these configurations receive unfavorable cost values. As mixer we use a Grover-type reflection around the prepared state [98], replacing the standard mixer of QAOA. This mixer amplifies low-cost configurations within the encoded subspace. The expectation value $\langle C \rangle$ of the cost Hamiltonian corresponds to the average tour cost of the output distribution, where invalid configurations contribute with their assigned penalty values. By optimizing the QAOA parameters to minimize $\langle C \rangle$, probability is concentrated on bitstrings representing the shortest valid tour. A detailed description can be found in Appendix A.5.

First, we analyze the optimal case to establish a noise-free reference, then we introduce a hardware noise model (here *FakeCairo*, see Ref. [132] for device parameters) to examine the resulting performance degradation. To counteract these effects, we apply ZNE with Adaptive Scaling Factors measured on the backend noise model (ASF B) and IC-ZNE+ASF B (following the procedure described in Section 4.8), and evaluate their influence on the expectation values at different GM-QAOA depths p.
The optimization of the GM-QAOA parameters is performed using the *COBYLA* [133] algorithm from the *SciPy* optimization library [134], which is a gradient-free method suitable for noisy objective functions. The optimizer iteratively updates the parameters β and γ until the change in the expectation value of the cost Hamiltonian falls below its internal convergence threshold (approximately 10^{-6}), or until the internal iteration limit (1 000) is reached. In each optimization step, the quantum circuit is executed and evaluated with 10 000 measurement shots to estimate the expectation value $\langle C \rangle$. The GM-QAOA depth is varied using $p = 1, 2, 3, 4$, and 5. The variational parameters β and γ are initialized as a uniform vector of ones and then optimized.
To ensure statistical robustness, 25 independent optimization runs were performed to

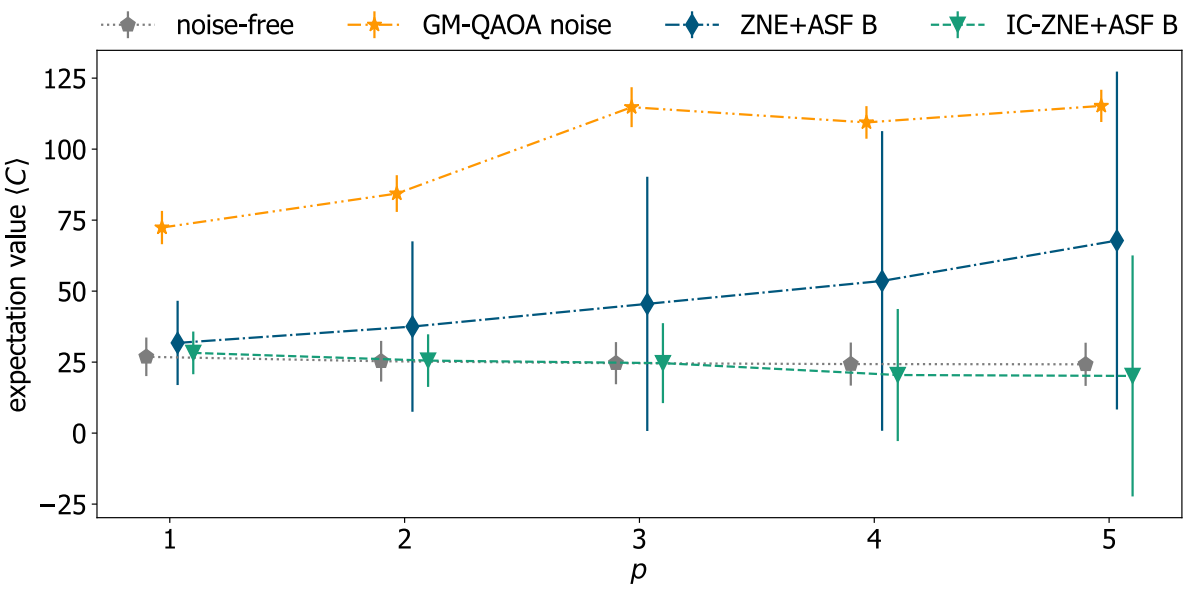


Figure 5.1: *Comparison of the mean expectation values and standard deviations over 25 optimization runs for GM-QAOA with increasing depth p. Results are shown for the noise-free evaluation case (gray hexagons), noisy GM-QAOA execution without error mitigation (orange stars), and with error mitigation using ZNE+ASF B (blue diamonds) and IC-ZNE+ASF B (green triangles). Data points are connected by lines and slightly offset for better visibility.*

account for variations in noisy realizations. For the mitigation runs (ZNE and IC-ZNE), we reuse the β and γ values optimized on the noisy backend and do not re-optimize, which reduces the computational overhead. The underlying assumption is that ZNE removes the systematic bias introduced by noise: if the noise-optimized parameters β and γ are sufficiently close to the ideal optimum and the circuit's error rate is low enough for ZNE to remain effective, the extrapolated expectation value approaches the noise-free value. The noisy value is further from the noise-free one due to decoherence and readout bias, and ZNE corrects this drift. This reuse of the noise-optimized parameters in combination with ZNE corresponds to what we refer to as Final Point (FP) mitigation. FP mitigation means applying the error-mitigation procedure only once at the final evaluation step, using the parameters obtained from the noisy optimization without re-optimizing under noise scaling.
ZNE+ASF B is implemented using a quadratic extrapolation fit, a method that is selected because it is more stable than an exponential fit in this case and does not require boundary constraints on the fit parameters. ASF B factors are automatically adapted by the ϵ_0 measurement for ZNE and IC-ZNE as described in Section 4.8. While the circuit-level (see Section 4.8) benchmarks showed no clear benefit from using ASF together with IC-ZNE, the QAOA results discussed below demonstrate that IC-ZNE+ASF B provides the highest accuracy. The number of CNOT gates increases with p and also depends on the circuit-to-hardware mapping. It ranges from 59 for $p = 1$ to 271 for $p = 5$.

Figure 5.1 displays the mean and standard deviation of the expectation value $\langle C \rangle$ over 25 optimization runs. The corresponding numerical values are reported in Table A.3 in Appendix A.5. Local and global folding do not differ significantly, as discussed in Section 4.10. Therefore only the local folding results are shown for clarity. Moreover, standard ZNE and standard IC-ZNE (i.e., without Adaptive Scaling Factors) did not

yield improvements in our tests (results not shown here), and ASF M even increased deviations from the ideal results in both methods, which indicates that the calibration data do not fully reflect the actual backend noise model. The results include the TSP results evaluated under noise-free conditions (gray hexagons), with GM-QAOA and simulated hardware noise but without error mitigation (orange stars), and with final point error mitigation using ZNE+ASF B (blue diamonds) and IC-ZNE+ASF B (green triangles). To improve visual clarity, the data points are connected by lines and slightly shifted along the horizontal axis to avoid overlap.

The noise-free simulations show a low standard deviation, with an expectation value of 26.87 ± 6.79 at $p = 1$, which decreases slightly with increasing circuit depth, reaching 24.23 ± 7.6 at $p = 5$. The noisy implementation yields consistently higher expectation values compared to the noise-free case. The deviation increases with GM-QAOA depth, ranging from 72.38 ± 5.86 at $p = 1$ to 115.23 ± 5.68 at $p = 5$, while maintaining a comparable variance. This reflects the accumulation of noise under repeated applications of the mixer and cost Hamiltonian. The ZNE+ASF B results initially lie close to the noise free values, at 31.76 ± 14.83 at $p = 1$, and transition towards the noisy regime as p increases (67.81 ± 59.49 at $p = 5$). Both, the expectation value and the standard deviation, increase with circuit depth, which is consistent with the enhanced effective noise. An exponential fitting model would be more appropriate (see Subsection 2.4.1), but it was not usable here because several runs failed to converge to a stable fit.

In contrast, the mean values of IC-ZNE+ASF B remain closest to the noise-free values across all depths. For $p = 4$ at 20.45 ± 23.26 and for $p = 5$ at 20.14 ± 42.44, the mean value even falls slightly below the noise-free reference. This suggests an overestimation of the noise level, as also noted in Section 4.7, although the results remain comparable with statistical fluctuations. For IC-ZNE+ASF B the standard deviation increases with circuit depth and noise intensity. The improved performance of IC-ZNE+ASF B may result from the explicit ϵ_0 measurement of each scaled circuit, which can provide a more accurate estimate of the underlying noise. While error rates for $p > 3$ are typically too large for meaningful extrapolation and the amplified expectation values differ only marginally, IC-ZNE+ASF B still yields unexpectedly good results, even for noise strengths as high as $\epsilon > 0.8$. This indicates that the system has not yet collapsed to a completely mixed state.

For small problem sizes such as this example, both ZNE+ASF B and IC-ZNE+ASF B provide a clear benefit over the noisy baseline. However, it remains uncertain whether these methods will continue to be effective for larger problem instances with more qubits and deeper circuits, where the noise levels are substantially higher. This requires further investigation. Experiments on real quantum hardware would provide additional insight into the performance of both mitigation schemes. For ZNE the exponential fit should be stabilized to ensure consistent behavior, since several runs were aborted due to fitting failures. The influence of filtering procedures could also be examined in future work, although this would require an automated integration into the optimization loop. While QAOA benefits from ZNE+ASF B and IC-ZNE+ASF B when expectation values are evaluated, the methods cannot be applied when only bitstrings are used, which is the case in many practical QAOA applications. These are important limitations.

In summary, our results show that both ZNE+ASF B and IC-ZNE+ASF B can partially recover the noise-free expectation values for small TSP instances, with IC-ZNE providing

consistently better accuracy at higher circuit depths. Despite not being part of the optimization loop, FP mitigation improves the results, supporting the hypothesis that the noise-optimized parameters lie close to the ideal ones. Thus, a single mitigation step at the final evaluation often suffices in this setting. However, the growing variance at large noise levels suggest that these techniques do not scale straightforwardly to larger problems. Improved hardware fidelity will be required to extend these findings beyond small benchmarks.

5.2 Error Mitigation for the Variational Quantum Eigensolver

In this section, we investigate the performance of the Variational Quantum Eigensolvers (VQE) when combined with quantum error mitigation techniques. Focusing on the H_2 and KH molecules, we analyze how noise affects the accuracy of the computed ground-state energies and assess to what extent sZNE and IC-ZNE can compensate for these errors. Aside from that, we investigate whether mitigation must be applied throughout the full optimization procedure or if it is sufficient to incorporate it only in the final iteration. For the more noise-sensitive KH simulations, we use a noise model with a tunable noise strength, where the original error rates are uniformly scaled down. This allows us to place the system in a noise regime where ZNE still produces meaningful extrapolations and reflects noise levels expected for future devices.

H_2 is often used for benchmarking in quantum chemistry calculations because it allows direct comparison to exact solutions. As a result, it provides a clear way to assess the strengths and weaknesses of quantum algorithms. KH was chosen because it is a chemically relevant diatomic molecule with a more complex electronic structure than H_2, presenting a greater challenge for VQE-type algorithms. This helps to illustrate how the accuracy and resource requirements of VQE scale with problem complexity and electron count.

We implement the VQE modules as described in Subsection 2.5.2 and Appendix A.6. For our calculations we use the UCCSD method provided by *Qiskit Nature* [135] together with PySCF [136] and SPSA [137] for classical parameter optimization. All simulations are carried out with the STO-3G basis set [138–140]. All experiments use a total of 30 000 measurement shots, which are distributed across the three (sZNE) or five (IC-ZNE) noise scaling factors, and measurement error mitigation is performed using Mthree [51]. The maximum number of iterations in the SPSA optimizer is set to 750. A more detailed description of the VQE implementation is provided in Appendix A.6.

For sZNE, we apply the noise scaling factors $\lambda = [1, 3, 5]$. A single untwirled circuit yields one data point per scaling factor, limiting the sample size and, in turn, the confidence in the extrapolated trend. As described in Section 5.1, we use a quadratic fit for VQE as it is more stable than an exponential fit in this context and does not require boundary constraints on the fit parameters. For IC-ZNE we use the ASF M method described in Section 4.8. In the benchmarks presented in Section 4.8, which examined circuits and not variational algorithms, ASF did not improve the performance of IC-ZNE. The VQE results below, however, demonstrate that IC-ZNE+ASF M provides the highest accuracy.

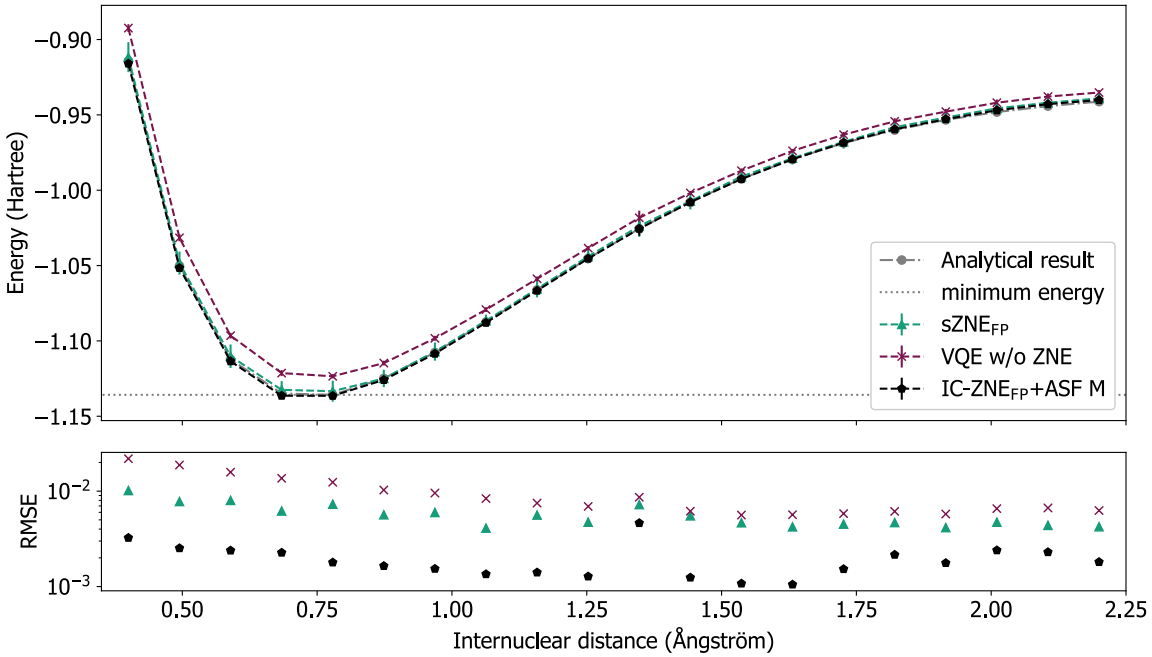


Figure 5.2: *Distribution of 50 optimization runs per internuclear distance for the H_2 molecule, used to estimate the ground-state energy. The horizontal dashed line indicates the minimum energy obtained from the analytical solution. Four methods are shown: the analytical reference (gray circles), sZNE$_{FP}$ (green triangles), VQE without error mitigation (red crosses), and IC-ZNE$_{FP}$ (black hexagons). For clarity, data points associated with the same method are connected by dashed lines, serving as a visual guide. The lower graph shows the corresponding RMSE values with respect to the analytical reference. IC-ZNE$_{FP}$ consistently yields the lowest RMSE across all bond lengths. A pronounced RMSE outlier is observed at 1.34 Å, most likely due to variability in the classical optimization procedure.*

These two approaches yield the best results among the different methods we tested. After adapting the circuits to the hardware we obtain between two and four CNOT gates per circuit for H_2. For KH the number of CNOT gates lies between 183 and 217. The qubits used are q_0 and q_1 for H_2, and q_0, q_1, q_2, and q_3 for KH. The simulations are performed with the noise model of *ibmq_ehningen* as described in Appendix A.2. To obtain physically meaningful VQE results for the KH molecule, we found it necessary to rescale the hardware error model. In our simulations, reducing all errors to 10% of their calibrated values was sufficient to achieve stable convergence and accurate observables, while avoiding regimes in which ZNE becomes unstable.

As introduced conceptually in Section 5.1, we consider the Final Point (FP) mitigation. A related final-step strategy was introduced by Belaloui et al. [141] for readout mitigation using Twirled Readout Error Extinction [56]. After completing the optimization without applying ZNE, the final parameters are reused for a single ZNE evaluation. Noise scaling is applied only at the final measurement, and re-optimization is not performed at each scaling factor. The main benefits are reduced quantum resource usage and shorter wall-clock time. We denote FP variants with the subscript 'FP' (e.g. sZNE$_{\text{FP}}$, IC-ZNE$_{\text{FP}}$).

Figure 5.2 shows the distribution of energies obtained from 50 independent VQE optimization runs for the H_2 molecule at various internuclear distances, obtained using simulations with the *ibmq_ehningen* noise model. The accompanying plot below depicts

the corresponding Root-Mean-Square Error (RMSE) values relative to the analytical reference. Across all bond lengths, the application of error-mitigation techniques improves accuracy. In particular, IC-ZNE$_{\mathrm{FP}}$ consistently yields the lowest RMSE values, indicating a robust reduction of both bias and variance compared to unmitigated VQE and sZNE$_{\mathrm{FP}}$. Although fluctuations are visible, the relative performance of the different methods remains qualitatively similar over the full range of internuclear distances.

We restrict the subsequent analysis to both molecules at their equilibrium bond distances, the chemically most relevant cases, to provide a clean comparison of mitigation methods. For H_2 the distance is set to 0.735 Å and for KH we use a distance of 2.24 Å. The corresponding ground-state energies are -1.1373 Hartree for H_2 and -593.579 Hartree for KH. Each experiment is executed 100 times. In the following, we benchmark FP against the corresponding ZNE protocols that apply mitigation at every optimization step, providing a direct comparison.

To assess whether FP preserves accuracy while reducing cost, we benchmark it first on the H_2 molecule (KH is discussed below). Figure 5.3 shows 100 independent optimization runs obtained using simulations with the *ibmq_ehningen* noise model. For each mitigation method, we compare the version that applies ZNE at every optimization step with its FP variant, which applies mitigation only at the final point.

For sZNE, the two approaches lead to almost indistinguishable outcomes (see Figure 5.3 (a)). The distributions of the estimated energies strongly overlap and the extracted mean values and standard deviations are nearly identical (-1.1359 ± 0.0066 Hartree compared to -1.1358 ± 0.0059 Hartree, see Table 5.1). The RMSE values (definition see Equation (4.31)) are also close to each other with 6.7 mHartree and 6.1 mHartree for sZNE and sZNE$_{\mathrm{FP}}$ respectively. This suggests that the enhanced accuracy of sZNE is already captured by applying the mitigation only at the end of the optimization.

A similar behaviour is observed for IC-ZNE combined with ASF M (see Figure 5.3 (b)). Both the standard method and its final point variant yield energy estimates that lie closer to the reference value than those obtained with sZNE with RMSE values of 1.8 mHartree for IC-ZNE+ASF M and 1.9 mHartree for IC-ZNE$_{\mathrm{FP}}$+ASF M. In addition, the standard deviations are smaller with 1.7 mHartree and 1.6 mHartree.

Since FP mitigation performs only one final ZNE evaluation, it does not yield a significant classical speedup compared to an optimization with many function evaluations, i.e., it does not reduce the number of classical optimization steps or function evaluations (here, 750 with SPSA). However, the relevant cost on real quantum hardware is dominated by the number of circuit executions, i.e., the total number of measurement shots. FP mitigation avoids repeatedly evaluating the circuit at multiple noise scaling factors throughout the optimization, and instead applies mitigation only once at the final parameter point. As a result, FP mitigation greatly reduces the shot overhead compared to ZNE or IC-ZNE at every step, even though the classical optimization runtime (number of optimizer iterations) remains similar. In practice, this means that ZNE$_{\mathrm{FP}}$ offers a substantial reduction in quantum-hardware runtime compared to mitigation at every optimization step, while maintaining comparable accuracy.

Figure 5.4 (a) and (b) show analogous trends for KH with a scaled noise model, where all error probabilities are uniformly scaled to 10% of their original values. The presence of a secondary distribution can be seen with both mitigation methods and is most

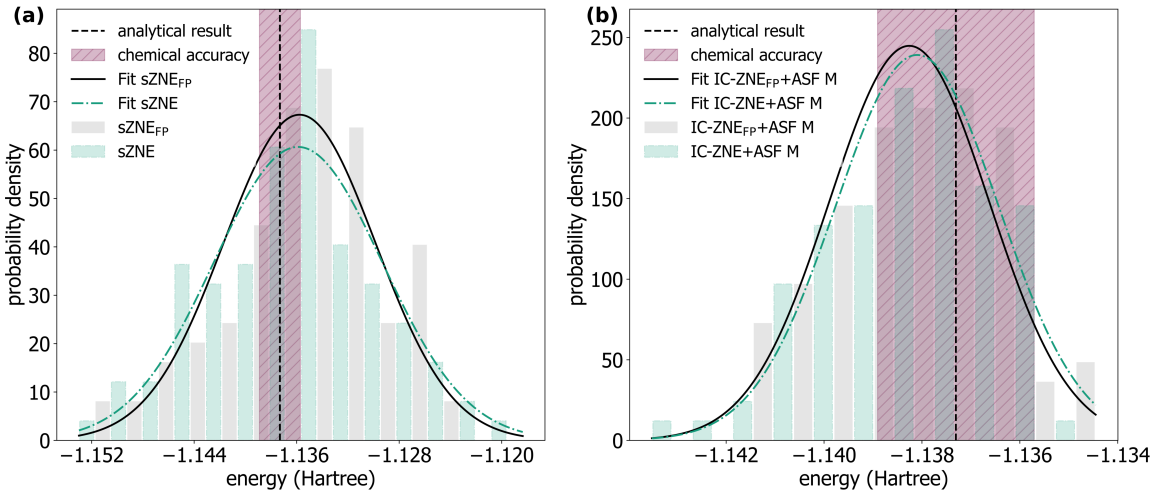


Figure 5.3: *Distribution of 100 optimization runs for the H_2 molecule to determine the ground-state energy. The vertical dashed line indicates the analytical reference result. The red hatched region marks the range of chemical accuracy, which corresponds to an energy error below 4.2 kJ/mol or 1.6×10^{-3} Hartree [142]. The final point method is shown in gray and the standard methods in green. Gaussian fits are drawn for each dataset and the underlying histograms of the energy distributions are shown in the corresponding colors. All simulations are performed with the noise model of ibmq_ehningen. (a) Results for sZNE and sZNE$_{FP}$. No significant differences between the two methods are visible. Their means and standard deviations are also very similar, see Table 5.1. (b) Results for IC-ZNE$_{\mathrm{FP}}$+ASF M and IC-ZNE+ASF M. Again no relevant differences are observed in the plots or in the statistical parameters. Both IC-ZNE methods perform better than sZNE (note the different energy-axis scale), with results closer to the exact energy of -1.1373 Hartree and with lower standard deviations.*
These observations show that, for this variational algorithm, it is sufficient to apply the mitigation only in the final iteration. This reduces runtime while preserving accuracy.

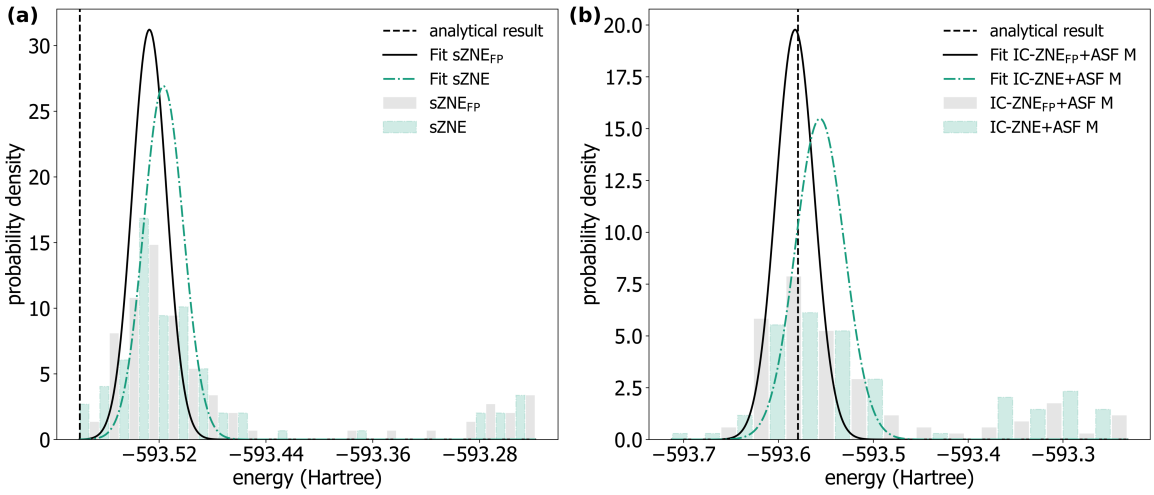


Figure 5.4: *Same as Figure 5.3, but for KH with 10% error using the noise model of ibmq_ehningen. Chemical accuracy is not shown here because the corresponding interval would be too narrow to be visible. A secondary distribution is visible in the data, and it is excluded by applying a filtering step. The overall trends are similar to those observed for H_2. The final point method yields results comparable to the standard approach, and IC-ZNE+ASF M again performs better than sZNE.*

Table 5.1: *Summary of mean μ, standard deviations σ, and RMSE values (all reported in Hartree) for all molecules and mitigation methods used. The error cases correspond to simulations in which the underlying hardware error model is scaled to 10% or 5% of its nominal value. The final column reports the mixture weights w_1 and w_2 associated with the first and second Gaussian components of the GMM. For H_2 only a single Gaussian distribution is observed, while for KH two Gaussian components are present.*

Molecule (error)	Method	$\mu_1 \pm \sigma_1$	RMSE	$\mu_2 \pm \sigma_2$	Weights w_1, w_2
H_2	sZNE	-1.1359 ± 0.0066	0.0067	–	–
	sZNE_{FP}	-1.1358 ± 0.0059	0.0061	–	–
	IC-ZNE+ASF M	-1.1381 ± 0.0017	0.0018	–	–
	$\text{IC-ZNE}_{\text{FP}}$+ASF M	-1.1383 ± 0.0016	0.0019	–	–
KH (10%)	sZNE	-593.516 ± 0.029	0.064	-593.266 ± 0.031	0.88, 0.12
	sZNE_{FP}	-593.525 ± 0.025	0.054	-593.278 ± 0.037	0.83, 0.17
	IC-ZNE+ASF M	-593.555 ± 0.046	0.035	-593.304 ± 0.041	0.75, 0.26
	$\text{IC-ZNE}_{\text{FP}}$+ASF M	-593.580 ± 0.039	0.020	-593.339 ± 0.066	0.81, 0.19
KH (5%)	VQE without ZNE	-593.504 ± 0.012	0.075	-593.258 ± 0.025	0.78, 0.22
	sZNE_{FP}	-593.563 ± 0.016	0.017	-593.298 ± 0.023	0.78, 0.22
	$\text{IC-ZNE}_{\text{FP}}$+ASF M	-593.572 ± 0.024	0.014	-593.320 ± 0.044	0.77, 0.22
KH	noise-free	-593.579 ± 0.002	0.001	-593.308 ± 0.017	0.62, 0.38

likely caused by convergence to a local minimum of the VQE landscape. Given the chosen ansatz and optimizer, it can be assumed that the VQE landscape contains several stationary points.

To investigate whether this secondary distribution is influenced by mitigation or by other effects, we performed one hundred additional simulations with a further reduced error model of five percent (results in Table 5.1). A second distribution also appears in noise-free runs and in VQE runs without mitigation. It therefore appears plausible that the secondary distribution originates from limitations of the chosen ansatz, rather than from the mitigation itself.

The analysis of the Gaussian Mixture Model (GMM, see Subsection 4.9.1) weights shows that the noise-free data assign $w_1 = 0.62$ to the first component, while the noisy data yield substantially higher values (e.g. $w_1(\text{sZNE}_{\text{FP}}) = 0.78$). This shift indicates that noise redistributes the sampling statistics toward the first component, which may effectively reduce the probability of the optimization becoming trapped in competing local minima. Future work should examine whether alternative ansätze or optimizer settings can prevent the optimization from becoming confined to such ansatz-limited subspaces.

We apply the filtering ansatz of Section 4.9, to identify the secondary distribution using a Gaussian mixture model and retain only the main distribution that contains the majority of results. The RMSE values for sZNE with 6.4 mHartree and sZNE_{FP} with 5.4 mHartree are further away from the analytical reference than the RMSE values obtained with IC-ZNE+ASF M which yields 3.5 mHartree and with $\text{IC-ZNE}_{\text{FP}}$+ASF M which yields 2.0 mHartree as shown in Table 5.1. Although the standard deviations are larger for the IC-ZNE based variants, we see that for molecules that are more complex than H_2, applying mitigation only in the final iteration remains feasible.

As for H_2, the FP variants for KH consist of one noisy VQE optimization and a single final mitigated ZNE evaluation at fixed parameters. While this does not yield a major classical speedup, FP avoids mitigation at every optimization step and therefore offers a significant reduction in shot overhead on real quantum hardware, especially for IC-ZNE.

A systematic analysis over all bond distances is left for future research. The optimization process almost always reaches the iteration limit, indicating further improvements are possible. Alternative optimizer configurations or ansätze may also help to avoid preventing convergence to restricted, ansatz-defined subspaces. A more robust procedure for obtaining exponential fits for ZNE without strong constraints would also be desirable.

In summary, our results demonstrate that combining VQE with ZNE-based error mitigation remains effective even in challenging noisy settings and for molecular systems beyond the simplest test cases. In particular, we find that applying the mitigation only once at the final iteration is sufficient to achieve the same level of accuracy as full per-iteration mitigation. This behaviour holds for both sZNE and IC-ZNE across the investigated systems. The Final Point strategy therefore provides a practical and resource-efficient alternative to standard ZNE protocols, enabling substantial reductions in runtime without any measurable loss of precision. These findings underline that cost-effective mitigation schemes can be incorporated into variational quantum algorithms and will remain relevant for near-term devices as well as future noise-reduced architectures.

5.3 Summary of error mitigation for variational algorithms

This chapter evaluated quantum error mitigation for variational algorithms using simulations of the Quantum Approximate Optimization Algorithm (QAOA) for the solution of a travelling salesman problem and the Variational Quantum Eigensolver (VQE) applied to two different molecules. Standard ZNE and IC-ZNE were compared, the impact of Adaptive Scaling Factors (ASF) was examined, and a resource-efficient Final Point (FP) strategy was tested, in which mitigation is applied only once after noisy optimization.

For QAOA, ZNE reduced bias at shallow circuit depth but became increasingly ineffective as circuit depth p grew. At $p = 1$, ZNE+ASF B reduces the deviation from the noisy baseline by approximately 56%. However, this improvement continuously decreases with depth, finally dropping to approximately 41% at $p = 5$. Here, percentage reductions refer to the fraction of the noise-induced bias removed relative to the unmitigated noisy QAOA result at the same circuit depth. In contrast, IC-ZNE with backend-calibrated scaling factors remained closest to the noise-free reference up to high circuit depths, although with increased variance. Across the investigated depths, IC-ZNE+ASF B reduces the deviation from the noisy baseline by approximately 61% at $p = 1$ and by 82.5% at $p = 5$, demonstrating a substantially stronger suppression of noise-induced bias than ZNE. These results indicate that IC-ZNE is well suited for near-term QAOA applications while also highlighting limitations for deeper circuits.

For VQE, FP mitigation achieved the same accuracy as per-iteration mitigation while substantially reducing shot overhead. ZNE improved the VQE results to a limited extent, whereas IC-ZNE with scaling factors derived from the noise model yielded the

lowest Root-Mean-Square Error (RMSE) and the smallest variance. For the H_2 molecule, IC-ZNE reduces the RMSE by 73.13% compared to sZNE, and the FP variant yields a statistically comparable accuracy, with an RMSE that differs by only 5.56% from the corresponding per-iteration IC-ZNE result (1.9 vs 1.8 mHartree). For the more complex KH molecule at 10% scaled noise, IC-ZNE+ASF M with FP lowers the RMSE by about 63% relative to sZNE with FP, and at 5% noise the total RMSE reduction exceeds 80% compared to unmitigated VQE. Although ASF provided limited benefits for the circuits considered in Chapter 4, the combination of IC-ZNE with ASF M delivered the lowest RMSE values in this setting. Additional filtering further improved the VQE outcomes.

These results collectively demonstrate that IC-ZNE is a reliable and precise error mitigation method for variational algorithms. By directly measuring the error strength ϵ on the device, IC-ZNE removes the need for approximate error-scaling assumptions and consistently achieves the lowest RMSE across a wide range of settings. Despite a moderate increase in shot cost, this overhead is compensated by a substantial improvement in accuracy and stability, making IC-ZNE a particularly effective and reliable approach to Zero-Noise Extrapolation in noisy quantum devices.

With these variational algorithm-level analysis completed, the following chapter places the results of this work into a broader perspective, summarizing the main contributions and outlining open challenges and directions for future research.

6 Conclusion and Outlook

This thesis addressed a central challenge of the Noisy Intermediate-Scale Quantum (NISQ) era, namely how to reduce and manage errors on transmon-based quantum processors in order to enable meaningful computations without full fault tolerance. The work combined hardware-level calibration, gate-level performance analysis and circuit-level error mitigation. New techniques for these purposes have been developed and validated on *IBM*'s *ibmq_ehningen* device as well as on simulators with different noise models.

The first part of the thesis concerned calibration of single qubit gates, as well as different implementations of two qubit gates. Cross-point amplitude calibration was shown to outperform the established ping-pong method across all relevant criteria. It provided higher precision, higher signal-to-noise ratio and faster convergence while exhibiting fewer fit instabilities. Repeated runs approached the shot-noise limit, and the direct zero-crossing fit made the method more robust against drift and fit noise than ping-pong calibration, which remained sensitive to initial guesses and fluctuations. For the two-qubit gate, a Hamiltonian tomography of the cross-resonance interaction confirmed the validity of a reduced six-term model that can be used for calibrating CNOT gates based on echoed cross-ressonance gates. A comparison of this CNOT implementation with an alternative, direct implementation showed that direct CNOTs are shorter and achieve lower median error rates, although at the cost of higher calibration sensitivity and occasional outliers. Echoed CNOTs remained more robust. Algorithmic benchmarks based on Quantum Approximate Optimization Algorithm (QAOA) MaxCut and Portfolio Optimization demonstrated that the relative advantage of each gate is problem-dependent, and that hybrid strategies can be favorable.

The second part of the thesis introduced Inverted-Circuit Zero-Noise Extrapolation (IC-ZNE). Its central idea is to measure the effective circuit error strength ϵ by appending the inverse circuit and then measuring the probability of recovering the initial state. Subsequently, extrapolation is performed with respect to the actually measured error strength ϵ rather than on relying on simple, approximate noise scaling assumptions as in standard ZNE. On hardware, IC-ZNE delivers extrapolated observables that were consistently closer to ideal references than standard ZNE, with equal or smaller uncertainty, across Grover, HHL and H-Ladder circuits. On the tested circuits the protocol achieved up to 77% error reduction. By measuring the effective noise of each circuit instance directly, IC-ZNE becomes more robust to fluctuations in error strength. Pauli twirling further improved both IC-ZNE and standard ZNE. Although IC-ZNE requires extra circuits to measure the error strength, its advantage remains even when the overall number of circuit executions (i.e., measurement shots) is kept equal to that used for standard ZNE.

Further refinements made the mitigation more resource-aware. Adaptive Scaling Factors (ASF) set the largest scaling factor proportional to the initial error strength and use

exponential spacing. This improves the extrapolation result and reduces variance at negligible cost, with the best choice depending on the circuit. Simple post-processing filters based on the initial error strength and global or local outlier detection improved the results on hardware, with gains up to 39% in selected cases. Global and ϵ_0-based filtering worked reliably, while combining filters did not always help. The experimental results on folding schemes demonstrated that there is no statistically significant difference in performance between the two folding methods on hardware, suggesting that both methods are equally effective.

Applications to representative variational quantum algorithms, which iteratively adjust parameters to minimize a cost function, confirmed the practical value and limitations of these techniques. In simulations, IC-ZNE combined with ASF remained closest to the noise-free references for a variant of the QAOA, which seeks good solutions to combinatorial optimization problems, applied to the traveling salesman problem up to a high circuit depth, whereas standard ZNE with ASF improved shallow circuits but its performance degraded as depth increased. Together, these results illustrate both the potential and the limits of mitigation as circuits grow deeper. For the Variational Quantum Eigensolver (VQE), a variational method used in quantum chemistry to estimate molecular ground state energies, applied to the H_2 and KH molecules, IC-ZNE with ASF achieved the lowest deviation from the analytical result. A final-point variant, where mitigation is applied only to the last energy evaluation, matched the accuracy of per-iteration mitigation while reducing runtime. For KH, stable results required a scaled noise model with a reduced amount of noise. Furthermore, a secondary distribution of optimization outcomes was observed, corresponding to an additional minimum that did not represent the true ground state. This secondary minimum was traced to the chosen ansatz and optimizer rather than the mitigation method.

The analysis also revealed important limitations. IC-ZNE can underestimate purely coherent errors, when applying the inverse of the circuit cancels those errors, and Pauli twirling, although helpful, does not remove all coherent errors or crosstalk. Backend noise can be non-Markovian and time-varying, and mitigation benefits depend strongly on factors such as circuit structure, idle-time decoherence, and readout errors. If the baseline error is too large, neither ZNE nor IC-ZNE is reliable. In VQE, the choice of ansatz and optimizer can dominate performance and create secondary minima in the optimization landscape that are unrelated to the mitigation scheme.

Overall, this thesis demonstrates that cross-point calibration, careful two-qubit gate selection and IC-ZNE or ZNE extended by ASF, twirling and filtering form a practical and effective toolbox for improving accuracy on present-day transmon hardware at reasonable cost. These methods extend the useful regime of NISQ devices and provide actionable pathways to lower-bias estimation in the absence of fault tolerance.

Outlook

The reduction of errors in quantum computing through a comparison of two calibration techniques and the development of an improved version of ZNE was considerably advanced in this thesis. The next steps involve deepening the theoretical understanding of IC-ZNE, and expanding experimental integration to demonstrate the practical advantages of cross-point calibration, particularly regarding its translation into device performance.

In calibration and control, cross-point calibration needs to be incorporated into the fine-calibration loop and into automated implementations on hardware. A crucial task is the empirical verification of whether cross-point calibration enhances single-qubit benchmarking results. Investigating Stark-shift compensation and testing DRAG-based pulse optimization represent additional avenues toward improved gate performance. A further important aspect is the explicit assessment of the method's robustness to SPAM offsets. Moreover, the calibration process can be extended to two-qubit gates to determine whether the observed improvements generalize to entangling operations.

In error mitigation, the combination of IC-ZNE with other techniques, such as probabilistic error cancellation or probabilistic error amplification, promises potential performance gains. Local folding remains the preferred strategy, while global folding is applicable primarily when the error distribution is uniform and low. Global and ϵ_0 filters benefit from standardized, conservative thresholds and well-defined abort criteria to avoid unstable extrapolations. It is also valuable to quantify how many measurement shots and twirling circuits are required for statistically meaningful results for IC-ZNE, and to evaluate the trade-off between sampling cost and achievable accuracy.

On the theoretical side, IC-ZNE merits further analysis under coherent and non-Markovian noise models. Two empirically chosen factors (a, Equation (4.21) in Eq. 4.2 and ξ, Eq. 4.36 in Subsection 4.8.1) should be determined more rigorously and investigated systematically. Developing a unified pipeline that integrates randomized compiling, dynamical decoupling and IC-ZNE with reliable outlier control offers a promising direction toward more robust circuit-level mitigation. In addition, a clearer rationale for selecting ASF-M versus ASF-B modes, as well as for the various filtering techniques, remains to be established.

In algorithmic applications, QAOA and VQE can be executed on hardware to assess the impact of realistic noise on the final optimized parameters. For VQE, broader exploration of ansätze and extensions to molecular systems beyond H_2 and KH becomes feasible as gate fidelities improve.

A broader perspective involves replicating the presented methods on newer transmon generations and on alternative hardware platforms, as well as conducting systematic comparisons with other emerging mitigation strategies. The techniques developed in this work also contribute to the path toward fault tolerance, for instance as pre-QEC diagnostic tools or as post-selection and verification mechanisms for shallow logical circuits.

The findings presented in this thesis make the calibration and mitigation techniques a compelling candidate for improving the reliability of near-term quantum processors across multiple platforms. In summary, precise single- and two-qubit control, error-aware extrapolation and adaptive mitigation strategies such as ASF, filtering, local folding and twirling offer a practical route to higher-quality results on current transmon devices. While not a substitute for full fault-tolerant quantum computation, these methods remain valuable and complementary tools on the way toward scalable quantum technologies.

Bibliography

[1] R. P. Feynman, "Simulating physics with computers," *International Journal of Theoretical Physics*, vol. 21, pp. 467–488, June 1982.

[2] P. Shor, "Algorithms for quantum computation: discrete logarithms and factoring," in *Proceedings 35th Annual Symposium on Foundations of Computer Science*, pp. 124–134, 1994.

[3] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. van den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme, and A. Kandala, "Evidence for the utility of quantum computing before fault tolerance," *Nature*, vol. 618, pp. 500–505, June 2023.

[4] S. Ebadi, T. T. Wang, H. Levine, A. Keesling, G. Semeghini, A. Omran, D. Bluvstein, R. Samajdar, H. Pichler, W. W. Ho, S. Choi, S. Sachdev, M. Greiner, V. Vuletić, and M. D. Lukin, "Quantum phases of matter on a 256-atom programmable quantum simulator," *Nature*, vol. 595, pp. 227–232, July 2021.

[5] P. W. Shor, "Scheme for reducing decoherence in quantum computer memory," *Physical Review A*, vol. 52, pp. R2493–R2496, Oct. 1995.

[6] A. R. Calderbank and P. W. Shor, "Good quantum error-correcting codes exist," *Physical Review A*, vol. 54, pp. 1098–1105, Aug. 1996.

[7] A. M. Steane, "Error correcting codes in quantum theory," *Physical Review Letters*, vol. 77, pp. 793–797, July 1996.

[8] F.-M. L. Régent, "Awesome quantum computing experiments: Benchmarking experimental progress towards fault-tolerant quantum computation," 2025. arXiv:2507.03678.

[9] J. Clarke and F. K. Wilhelm, "Superconducting quantum bits," *Nature*, vol. 453, pp. 1031–1042, June 2008.

[10] G.-Q. Liu and X.-Y. Pan, "Quantum information processing with nitrogen–vacancy centers in diamond*," *Chinese Physics B*, vol. 27, p. 020304, Feb. 2018.

[11] J. Wrachtrup and F. Jelezko, "Processing quantum information in diamond," *Journal of Physics: Condensed Matter*, vol. 18, p. S807, May 2006.

[12] J. I. Cirac and P. Zoller, "Quantum computations with cold trapped ions," *Physical Review Letters*, vol. 74, pp. 4091–4094, May 1995.

[13] H.-J. Briegel, T. Calarco, D. Jaksch, J. I. Cirac, and P. Zoller, "Quantum computing with neutral atoms," *Journal of Modern Optics*, vol. 47, no. 2-3, pp. 415–451, 2000.

[14] T. F. Watson, S. G. J. Philips, E. Kawakami, D. R. Ward, P. Scarlino, M. Veldhorst, D. E. Savage, M. G. Lagally, M. Friesen, S. N. Coppersmith, M. A. Eriksson, and L. M. K. Vandersypen, "A programmable two-qubit quantum processor in silicon," *Nature*, vol. 555, p. 633–637, Feb. 2018.

[15] P. Cerfontaine, T. Botzem, D. P. DiVincenzo, and H. Bluhm, "High-fidelity single-qubit gates for two-electron spin qubits in gaas," *Physical Review Letters*, vol. 113, p. 150501, Oct. 2014.

[16] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.

[17] J. M. Martinis and K. Osborne, "Superconducting qubits and the physics of josephson junctions," 2004. arXiv:cond-mat/0402415.

[18] 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," *Physical Review A*, vol. 76, p. 042319, Oct. 2007.

[19] D. Aharonov and M. Ben-Or, "Fault-tolerant quantum computation with constant error," in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '97, (New York, NY, USA), p. 176–188, Association for Computing Machinery, 1997.

[20] A. Y. Kitaev, "Quantum computations: algorithms and error correction," *Russian Mathematical Surveys*, vol. 52, p. 1191, Dec. 1997.

[21] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation," *Physical Review A*, vol. 86, p. 032324, Sept. 2012.

[22] I. D. Kivlichan, C. Gidney, D. W. Berry, N. Wiebe, J. McClean, W. Sun, Z. Jiang, N. Rubin, A. Fowler, A. Aspuru-Guzik, H. Neven, and R. Babbush, "Improved fault-tolerant quantum simulation of condensed-phase correlated electrons via trotterization," *Quantum*, vol. 4, p. 296, July 2020.

[23] IBM Quantum, "Expanding the IBM Quantum roadmap to anticipate the future of quantum-centric supercomputing." https://www.ibm.com/quantum/blog/ibm-quantum-roadmap-2025, 2025. Accessed: 2025-12-04.

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

[25] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta, "Characterizing errors on qubit operations via iterative randomized benchmarking," *Physical Review A*, vol. 93, p. 012301, Jan. 2016.

[26] N. Wittler, F. Roy, K. Pack, M. Werninghaus, A. S. Roy, D. J. Egger, S. Filipp, F. K. Wilhelm, and S. Machnes, "Integrated tool set for control, calibration, and characterization of quantum devices applied to superconducting qubits," *Physical Review Applied*, vol. 15, p. 034080, Mar. 2021.

[27] K. N. Smith, G. S. Ravi, T. Alexander, N. T. Bronn, A. R. R. Carvalho, A. Cervera-Lierta, F. T. Chong, J. M. Chow, M. Cubeddu, A. Hashim, L. Jiang, O. Lanes, M. J. Otten, D. I. Schuster, P. Gokhale, N. Earnest, and A. Galda, "Programming physical quantum systems with pulse-level control," *Frontiers in Physics*, vol. 10, 2022.

[28] Y. Xu, G. Huang, J. Balewski, A. Morvan, K. Nowrouzi, D. I. Santiago, R. K. Naik, B. Mitchell, and I. Siddiqi, "Automatic qubit characterization and gate optimization with QubiC," *ACM Transactions on Quantum Computing*, vol. 4, Oct. 2022.

[29] D. Fan, G. Liu, S. Li, M. Gong, D. Wu, Y. Zhang, C. Zha, F. Chen, S. Cao, Y. Ye, Q. Zhu, C. Ying, S. Guo, H. Qian, Y. Wu, H. Deng, G. Wu, C.-Z. Peng, X. Ma, X. Zhu, and J.-W. Pan, "Calibrating quantum gates up to 52 qubits in a superconducting processor," *npj Quantum Information*, vol. 11, Feb. 2025.

[30] V. Tripathi, D. Kowsari, K. Saurav, H. Zhang, E. M. Levenson-Falk, and D. A. Lidar, "Benchmarking quantum gates and circuits," *Chemical Reviews*, vol. 125, pp. 5745–5775, June 2025.

[31] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, "Quantum error mitigation," *Reviews of Modern Physics*, vol. 95, p. 045005, Dec. 2023.

[32] A. J. Goldschmidt, J. Dubois, S. L. Brunton, and J. N. Kutz, "Model predictive control for robust quantum state preparation," *Quantum*, vol. 6, p. 837, 2022.

[33] P. Cerfontaine, R. Otten, and H. Bluhm, "Self-consistent calibration of quantum-gate sets," *Physical Review Applied*, vol. 13, Apr. 2020.

[34] M. Werninghaus, D. Egger, and S. Filipp, "High-speed calibration and characterization of superconducting quantum processors without qubit reset," *PRX Quantum*, vol. 2, p. 020324, May 2021.

[35] N. Kanazawa and K. Heya, "Hardware-efficient calibration framework for quantum computing devices," June 2023. United States Patent, US11687818B2.

[36] D. A. Lidar, "Review of decoherence-free subspaces, noiseless subsystems, and dynamical decoupling," in *Quantum Information and Computation for Chemistry*. Advances in Chemical Physics. Wiley, 2014.

[37] D. Suter and G. A. Álvarez, "Colloquium: Protecting quantum information against environmental noise," *Reviews of Modern Physics*, vol. 88, p. 041001, Oct. 2016.

[38] K. Temme, S. Bravyi, and J. M. Gambetta, "Error mitigation for short-depth quantum circuits," *Physical Review Letters*, vol. 119, p. 180509, Nov. 2017.

[39] S. McArdle, X. Yuan, and S. Benjamin, "Error-mitigated digital quantum simulation," *Physical Review Letters*, vol. 122, p. 180501, May 2019.

[40] X. Bonet-Monroig, R. Sagastizabal, M. Singh, and T. E. O'Brien, "Low-cost error mitigation by symmetry verification," *Physical Review A*, vol. 98, p. 062339, Dec. 2018.

[41] J. R. McClean, M. E. Kimchi-Schwartz, J. Carter, and W. A. de Jong, "Hybrid quantum-classical hierarchy for mitigation of decoherence and determination of excited states," *Physical Review A*, vol. 95, p. 042308, Apr. 2017.

[42] J. R. McClean, Z. Jiang, N. C. Rubin, R. Babbush, and H. Neven, "Decoding quantum errors with subspace expansions," *Nature Communications*, vol. 11, p. 636, Jan. 2020.

[43] T. Takeshita, N. C. Rubin, Z. Jiang, E. Lee, R. Babbush, and J. R. McClean, "Increasing the representation accuracy of quantum simulations of chemistry without extra quantum resources," *Physical Review X*, vol. 10, Jan. 2020.

[44] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, "Virtual distillation for quantum error mitigation," *Physical Review X*, vol. 11, p. 041036, Nov. 2021.

[45] Y. Li and S. C. Benjamin, "Efficient variational quantum simulator incorporating active error minimization," *Physical Review X*, vol. 7, p. 021050, June 2017.

[46] T. Giurgica-Tiron, Y. Hindy, R. LaRose, A. Mari, and W. J. Zeng, "Digital zero noise extrapolation for quantum error mitigation," in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pp. 306–316, Oct. 2020.

[47] J. J. Wallman and J. Emerson, "Noise tailoring for scalable quantum computation via randomized compiling," *Physical Review A*, vol. 94, p. 052325, Nov. 2016.

[48] Y. Kim, C. J. Wood, T. J. Yoder, S. T. Merkel, J. M. Gambetta, K. Temme, and A. Kandala, "Scalable error mitigation for noisy quantum circuits produces competitive expectation values," *Nature Physics*, vol. 19, pp. 752–759, May 2023.

[49] M. M. Wilde, *The Noisy Quantum Theory*, p. 97–140. Cambridge University Press, 2013.

[50] 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. arXiv:311.05933.

[51] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, "Scalable mitigation of measurement errors on quantum computers," *PRX Quantum*, vol. 2, p. 040326, Nov. 2021.

[52] F. B. Maciejewski, Z. Zimborás, and M. Oszmaniec, "Mitigation of readout noise in near-term quantum devices by classical post-processing based on detector tomography," *Quantum*, vol. 4, p. 257, Apr. 2020.

[53] S. Bravyi, S. Sheldon, A. Kandala, D. C. Mckay, and J. M. Gambetta, "Mitigating measurement errors in multiqubit experiments," *Physical Review A*, vol. 103, p. 042605, Apr. 2021.

[54] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, p. 242–246, Sept. 2017.

[55] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, "Robust shadow estimation," *PRX Quantum*, vol. 2, p. 030348, Sept. 2021.

[56] E. van den Berg, Z. K. Minev, and K. Temme, "Model-free readout-error mitigation for quantum expectation values," *Physical Review A*, vol. 105, p. 032620, Mar. 2022.

[57] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, "Demonstration of a fundamental quantum logic gate," *Physical Review Letters*, vol. 75, pp. 4714–4717, Dec. 1995.

[58] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien, "Quantum computers," *Nature*, vol. 464, p. 45–53, Mar. 2010.

[59] 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, June 2019.

[60] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, "Superconducting qubits: Current state of play," *Annual Review of Condensed Matter Physics*, vol. 11, p. 369–395, Mar. 2020.

[61] M. AbuGhanem, "IBM quantum computers: evolution, performance, and future directions," *The Journal of Supercomputing*, vol. 81, p. 687, Apr. 2025.

[62] 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.

[63] J. B. Hertzberg, E. J. Zhang, S. Rosenblatt, E. Magesan, J. A. Smolin, J.-B. Yau, V. P. Adiga, M. Sandberg, M. Brink, J. M. Chow, and J. S. Orcutt, "Laser-annealing josephson junctions for yielding scaled-up superconducting quantum processors," *npj Quantum Information*, vol. 7, Aug. 2021.

[64] J. M. Chow, A. D. Córcoles, J. M. Gambetta, C. Rigetti, B. R. Johnson, J. A. Smolin, J. R. Rozen, G. A. Keefe, M. B. Rothwell, M. B. Ketchen, and M. Steffen, "Simple all-microwave entangling gate for fixed-frequency superconducting qubits," *Physical Review Letters*, vol. 107, p. 080502, Aug. 2011.

[65] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta, "Procedure for systematically tuning up cross-talk in the cross-resonance gate," *Physical Review A*, vol. 93, p. 060302, June 2016.

[66] E. Magesan and J. M. Gambetta, "Effective hamiltonian models of the cross-resonance gate," *Physical Review A*, vol. 101, p. 052308, May 2020.

[67] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta, "Efficient z gates for quantum computing," *Physical Review A*, vol. 96, p. 022330, Aug 2017.

[68] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, "Simple pulses for elimination of leakage in weakly nonlinear qubits," *Physical Review Letters*, vol. 103, p. 110501, Sept. 2009.

[69] J. M. Gambetta, F. Motzoi, S. T. Merkel, and F. K. Wilhelm, "Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator," *Physical Review A*, vol. 83, p. 012308, Jan. 2011.

[70] N. Kanazawa, D. J. Egger, Y. Ben-Haim, H. Zhang, W. E. Shanks, G. Aleksandrowicz, and C. J. Wood, "Qiskit experiments: A python package to characterize and calibrate quantum computers," *Journal of Open Source Software*, vol. 8, no. 84, p. 5329, 2023.

[71] R. Manenti and M. Motta, *Quantum Information Science*. Oxford University Press, 2023.

[72] C. Rigetti and M. Devoret, "Fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies," *Physical Review B*, vol. 81, p. 134507, Apr. 2010.

[73] N. Sundaresan, I. Lauer, E. Pritchett, E. Magesan, P. Jurcevic, and J. M. Gambetta, "Reducing unitary and spectator errors in cross resonance with optimized rotary echoes," *PRX Quantum*, vol. 1, p. 020318, Dec. 2020.

[74] T. Alexander, N. Kanazawa, D. J. Egger, L. Capelluto, C. J. Wood, A. Javadi-Abhari, and D. C McKay, "Qiskit pulse: programming quantum computers through the cloud with pulses," *Quantum Science and Technology*, vol. 5, p. 044006, Aug. 2020.

[75] A. He, B. Nachman, W. A. de Jong, and C. W. Bauer, "Zero-noise extrapolation for quantum-gate error mitigation with identity insertions," *Physical Review A*, vol. 102, p. 012426, July 2020.

[76] S. J. Devitt, W. J. Munro, and K. Nemoto, "Quantum error correction for beginners," *Reports on Progress in Physics*, vol. 76, p. 076001, June 2013.

[77] C. Gidney and M. Ekerå, "How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits," *Quantum*, vol. 5, p. 433, Apr. 2021.

[78] B. Koczor, "Exponential error suppression for near-term quantum devices," *Physical Review X*, vol. 11, p. 031057, Sept. 2021.

[79] 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, pp. 491–495, Mar. 2019.

[80] A. Kandala, K. X. Wei, S. Srinivasan, E. Magesan, S. Carnevale, G. A. Keefe, D. Klaus, O. Dial, and D. C. McKay, "Demonstration of a high-fidelity CNOT gate for fixed-frequency transmons with engineered ZZ suppression," *Physical Review Letters*, vol. 127, p. 130501, Sept. 2021.

[81] E. Magesan, J. M. Gambetta, and J. Emerson, "Scalable and robust randomized benchmarking of quantum processes," *Physical Review Letters*, vol. 106, p. 180504, May 2011.

[82] E. Magesan, J. M. Gambetta, and J. Emerson, "Characterizing quantum gates via randomized benchmarking," *Physical Review A*, vol. 85, p. 042311, Apr. 2012.

[83] A. Hashim, R. K. Naik, A. Morvan, J.-L. Ville, B. Mitchell, J. M. Kreikebaum, M. Davis, E. Smith, C. Iancu, K. P. O'Brien, I. Hincks, J. J. Wallman, J. Emerson, and I. Siddiqi, "Randomized compiling for scalable quantum computing on a noisy superconducting quantum processor," *Physical Review X*, vol. 11, p. 041039, Nov. 2021.

[84] R. Majumdar, P. Rivero, F. Metz, A. Hasan, and D. S. Wang, " Best Practices for Quantum Error Mitigation with Digital Zero-Noise Extrapolation ," in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, (Los Alamitos, CA, USA), pp. 881–887, IEEE Computer Society, Sept. 2023.

[85] T. Kurita, H. Qassim, M. Ishii, H. Oshima, S. Sato, and J. Emerson, "Synergetic quantum error mitigation by randomized compiling and zero-noise extrapolation for the variational quantum eigensolver," *Quantum*, vol. 7, p. 1184, Nov. 2023.

[86] Z. Cai, "Multi-exponential error extrapolation and combining error mitigation techniques for NISQ applications," *npj Quantum Information*, vol. 7, p. 80, May 2021.

[87] Z. Cai, "A practical framework for quantum error mitigation," 2023. arXiv:2110.05389.

[88] D. Gottesman, "Theory of fault-tolerant quantum computation," *Physical Review A*, vol. 57, pp. 127–137, Jan. 1998.

[89] P. Zhao, K. Linghu, Z. Li, P. Xu, R. Wang, G. Xue, Y. Jin, and H. Yu, "Quantum crosstalk analysis for simultaneous gate operations on superconducting qubits," *PRX Quantum*, vol. 3, p. 020301, Apr. 2022.

[90] K. Rudinger, C. W. Hogle, R. K. Naik, A. Hashim, D. Lobser, D. I. Santiago, M. D. Grace, E. Nielsen, T. Proctor, S. Seritan, S. M. Clark, R. Blume-Kohout, I. Siddiqi, and K. C. Young, "Experimental characterization of crosstalk errors with simultaneous gate set tomography," *PRX Quantum*, vol. 2, p. 040338, Nov. 2021.

[91] A. Ketterer and T. Wellens, "Characterizing crosstalk of superconducting transmon processors," *Physical Review Applied*, vol. 20, p. 034065, Sept. 2023.

[92] H. Perrin, T. Scoquart, A. Shnirman, J. Schmalian, and K. Snizhko, "Mitigating crosstalk errors by randomized compiling: Simulation of the bcs model on a superconducting quantum computer," *Physical Review Research*, vol. 6, Feb. 2024.

[93] S. Lloyd, "Universal quantum simulators," *Science*, vol. 273, no. 5278, pp. 1073–1078, 1996.

[94] A. W. Harrow, A. Hassidim, and S. Lloyd, "Quantum algorithm for linear systems of equations," *Physical Review Letters*, vol. 103, p. 150502, Oct. 2009.

[95] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, "Variational quantum algorithms," *Nature Reviews Physics*, vol. 3, p. 625–644, Aug. 2021.

[96] E. Farhi, J. Goldstone, and S. Gutmann, "A quantum approximate optimization algorithm," 2014. arXiv:1411.4028.

[97] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, "Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices," *Physical Review X*, vol. 10, p. 021067, June 2020.

[98] A. Bartschi and S. Eidenbenz, "Grover mixers for QAOA: Shifting complexity from mixer design to state preparation," in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, p. 72–82, IEEE, Oct. 2020.

[99] 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, p. 4213, July 2014.

[100] P. Jordan and E. Wigner, "Über das paulische äquivalenzverbot," *Zeitschrift für Physik*, vol. 47, pp. 631–651, Sep 1928.

[101] S. B. Bravyi and A. Y. Kitaev, "Fermionic quantum computation," *Annals of Physics*, vol. 298, no. 1, pp. 210–226, 2002.

[102] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, "A comparison of the bravyi–kitaev and jordan–wigner transformations for the quantum simulation of quantum chemistry," *Journal of Chemical Theory and Computation*, vol. 14, pp. 5617–5630, Nov. 2018.

[103] 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," *Physical Review X*, vol. 6, p. 031007, July 2016.

[104] Y. Shen, X. Zhang, S. Zhang, J.-N. Zhang, M.-H. Yung, and K. Kim, "Quantum implementation of the unitary coupled cluster for simulating molecular electronic structure," *Physical Review A*, vol. 95, p. 020501, Feb. 2017.

[105] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, "Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz," *Quantum Science and Technology*, vol. 4, p. 014008, Oct. 2018.

[106] W. J. Huggins, J. R. McClean, N. C. Rubin, Z. Jiang, N. Wiebe, K. B. Whaley, and R. Babbush, "Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers," *npj Quantum Information*, vol. 7, Feb. 2021.

[107] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid quantum-classical algorithms and quantum error mitigation," *Journal of the Physical Society of Japan*, vol. 90, p. 032001, Mar. 2021.

[108] J. K. L. MacDonald, "Successive approximations by the rayleigh-ritz variation method," *Physical Review*, vol. 43, pp. 830–833, May 1933.

[109] M. Reed and B. Simon, *Methods of Modern Mathematical Physics, IV: Analysis of Operators*, vol. 4 of *Methods of Modern Mathematical Physics*. New York: Academic Press, 1978.

[110] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, pp. 1627–1639, 1964.

[111] P. D. Nation and M. Treinish, "Suppressing quantum circuit errors due to system variability," *PRX Quantum*, vol. 4, p. 010327, Mar. 2023.

[112] R. Jozsa, "Fidelity for mixed quantum states," *Journal of Modern Optics*, vol. 41, no. 12, pp. 2315–2323, 1994.

[113] L. Viola, E. Knill, and S. Lloyd, "Dynamical decoupling of open quantum systems," *Physical Review Letters*, vol. 82, pp. 2417–2421, Mar. 1999.

[114] J. Bylander, S. Gustavsson, F. Yan, F. Yoshihara, K. Harrabi, G. Fitch, D. G. Cory, Y. Nakamura, J.-S. Tsai, and W. D. Oliver, "Noise spectroscopy through dynamical decoupling with a superconducting flux qubit," *Nature Physics*, vol. 7, p. 565–570, May 2011.

[115] N. Ezzell, B. Pokharel, L. Tewala, G. Quiroz, and D. A. Lidar, "Dynamical decoupling for superconducting qubits: A performance survey," *Physical Review Applied*, vol. 20, p. 064027, Dec. 2023.

[116] A. Javadi-Abhari, M. Treinish, K. Krsulich, C. J. Wood, J. Lishman, J. Gacon, S. Martiel, P. D. Nation, L. S. Bishop, A. W. Cross, B. R. Johnson, and J. M. Gambetta, "Quantum computing with Qiskit," 2024. arXiv:2405.08810.

[117] M. A. Nielsen, "A simple formula for the average gate fidelity of a quantum dynamical operation," *Physics Letters A*, vol. 303, p. 249–252, Oct. 2002.

[118] J. Emerson, R. Alicki, and K. Życzkowski, "Scalable noise estimation with random unitary operators," *Journal of Optics B: Quantum and Semiclassical Optics*, vol. 7, p. S347–S352, Sept. 2005.

[119] R. Harper and S. T. Flammia, "Estimating the fidelity oftgates using standard interleaved randomized benchmarking," *Quantum Science and Technology*, vol. 2, p. 015008, Mar. 2017.

[120] L. Hour, S. Heng, M. Go, and Y. Han, "Improving zero-noise extrapolation for quantum-gate error mitigation using a noise-aware folding method," 2024. arXiv:2401.12495.

[121] V. R. Pascuzzi, A. He, C. W. Bauer, W. A. de Jong, and B. Nachman, "Computationally efficient zero-noise extrapolation for quantum-gate-error mitigation," *Physical Review A*, vol. 105, p. 042406, Apr. 2022.

[122] M. Krebsbach, B. Trauzettel, and A. Calzona, "Optimization of richardson extrapolation for quantum error mitigation," *Physical Review A*, vol. 106, p. 062436, Dec. 2022.

[123] H.-P. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*. Great Clarendon Street: Oxford University Press, Jan. 2007.

[124] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[125] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, p. 2825–2830, Nov. 2011.

[126] K. Schultz, R. LaRose, A. Mari, G. Quiroz, N. Shammah, B. D. Clader, and W. J. Zeng, "Impact of time-correlated noise on zero-noise extrapolation," *Physical Review A*, vol. 106, p. 052406, Nov. 2022.

[127] P. Klimov, J. Kelly, Z. Chen, M. Neeley, A. Megrant, B. Burkett, R. Barends, K. Arya, B. Chiaro, Y. Chen, A. Dunsworth, A. Fowler, B. Foxen, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, E. Lucero, J. Mutus, O. Naaman, C. Neill, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. White, S. Boixo, R. Babbush, V. Smelyanskiy, H. Neven, and J. Martinis, "Fluctuations of energy-relaxation times in superconducting qubits," *Physical Review Letters*, vol. 121, p. 090502, Aug. 2018.

[128] M. Carroll, S. Rosenblatt, P. Jurcevic, I. Lauer, and A. Kandala, "Dynamics of superconducting qubit relaxation times," 2022.

[129] J. Ehrlich, D. F. Urban, and C. Elsässer, "Variational quantum-algorithm based self-consistent calculations for the two-site dmft model on noisy quantum computing hardware," *Journal of Physics: Condensed Matter*, vol. 37, p. 225901, may 2025.

[130] E. L. Lawler, J. K. Lenstra, A. H. G. R. Kan, and D. B. Shmoys, eds., *The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization*. Wiley-Interscience Series in Discrete Mathematics and Optimization, Chichester / New York: John Wiley & Sons, 1985.

[131] M. Held and R. M. Karp, "A dynamic programming approach to sequencing problems," in *Proceedings of the 1961 16th ACM National Meeting*, ACM '61, (New York, NY, USA), p. 71.201–71.204, Association for Computing Machinery, 1961.

[132] Qiskit IBM Runtime Team, "Fakecairo backend." https://github.com/Qiskit/qiskit-ibm-runtime/tree/stable/0.43/qiskit_ibm_runtime/fake_provider/backends/cairo, 2024. Accessed 2025-11-15.

[133] M. J. D. Powell, "A direct search optimization method that models the objective and constraint functions by linear interpolation," in *Advances in Optimization and Numerical Analysis* (S. Gomez and J.-P. Hennart, eds.), pp. 51–67, Dordrecht: Springer, 1994.

[134] P. Virtanen, R. Gommers, T. E. Oliphant, *et al.*, "Scipy 1.0: Fundamental algorithms for scientific computing in python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[135] The Qiskit Nature developers and contributors, "Qiskit nature 0.6.0." Zenodo, Apr. 2023.

[136] Q. Sun, T. C. Berkelbach, N. S. Blunt, G. H. Booth, S. Guo, Z. Li, J. Liu, J. McClain, E. R. Sayfutyarova, S. Sharma, S. Wouters, and G. K.-L. Chan, "The python-based simulations of chemistry framework (PySCF)." arXiv:1701.08223, 2017.

[137] J. Spall, "Multivariate stochastic approximation using a simultaneous perturbation gradient approximation," *IEEE Transactions on Automatic Control*, vol. 37, no. 3, pp. 332–341, 1992.

[138] W. J. Hehre, R. F. Stewart, and J. A. Pople, "Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals," *The Journal of Chemical Physics*, vol. 51, pp. 2657–2664, Sept. 1969.

[139] W. J. Hehre, R. Ditchfield, R. F. Stewart, and J. A. Pople, "Self-Consistent Molecular Orbital Methods. IV. Use of Gaussian Expansions of Slater-Type Orbitals. Extension to Second-Row Molecules," *The Journal of Chemical Physics*, vol. 52, pp. 2769–2773, Mar. 1970.

[140] W. J. Pietro, B. A. Levi, W. J. Hehre, and R. F. Stewart, "Molecular orbital theory of the properties of inorganic and organometallic compounds. 1. STO-NG basis sets for third-row main-group elements," *Inorganic Chemistry*, vol. 19, pp. 2225–2229, Aug. 1980.

[141] N. E. Belaloui, A. Tounsi, A. R. Khamadja, H. Benkadour, M. M. Louamri, A. Benslama, and M. T. Rouabah, "Improving vqe parameter quality on noisy quantum processors with cost-effective readout error mitigation," 2025.

[142] K. A. Peterson, D. Feller, and D. A. Dixon, "Chemical accuracy in ab initio thermochemistry and spectroscopy: current strategies and future challenges," *Theoretical Chemistry Accounts*, vol. 131, p. 1079, Jan. 2012.

[143] 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.

[144] D. E. Knuth, *The art of computer programming, volume 3: (2nd ed., 20. print) sorting and searching*. USA: Addison Wesley Longman Publishing Co., Inc., 2005.

[145] S. Brandhofer, I. Polian, and K. Krsulich, " Optimal Qubit Reuse for Near-Term Quantum Computers ," in *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, (Los Alamitos, CA, USA), pp. 859–869, IEEE Computer Society, Sept. 2023.

[146] Y. Ji, K. F. Koenig, and I. Polian, "Optimizing quantum algorithms on bipotent architectures," *Physical Review A*, vol. 108, Aug. 2023.

[147] R. Kelmendi, "Evaluation and Comparison of State of the Art QAOA Implementations." Master's thesis computer science, University of Tübingen, Tübingen, Germany, 2023.

[148] J. S. Baker and S. K. Radha, "Wasserstein solution quality and the quantum approximate optimization algorithm: A portfolio optimization case study." arXiv:2202.06782, 2022.

[149] S. Brandhofer, D. Braun, V. Dehn, G. Hellstern, M. Hüls, Y. Ji, I. Polian, A. S. Bhatia, and T. Wellens, "Benchmarking the performance of portfolio optimization with QAOA," *Quantum Information Processing*, vol. 22, p. 25, Dec. 2022.

[150] D. J. Egger, C. Gambella, J. Marecek, S. McFaddin, M. Mevissen, R. Raymond, A. Simonetto, S. Woerner, and E. Yndurain, "Quantum computing for finance: State-of-the-art and future prospects," *IEEE Transactions on Quantum Engineering*, vol. 1, pp. 1–24, 2020.

[151] N. N. Hegade, P. Chandarana, K. Paul, X. Chen, F. Albarrán-Arriagada, and E. Solano, "Portfolio optimization with digitized counterdiabatic quantum algorithms," *Physical Review Research*, vol. 4, p. 043204, Dec. 2022.

[152] F. Glover, G. Kochenberger, R. Hennig, and Y. Du, "Quantum bridge analytics i: a tutorial on formulating and using QUBO models," *Annals of Operations Research*, vol. 314, pp. 141–183, July 2022.

[153] J. Wurtz and P. Love, "Maxcut quantum approximate optimization algorithm performance guarantees for $p > 1$," *Physical Review A*, vol. 103, p. 042612, Apr. 2021.

[154] P. Chandarana, N. N. Hegade, K. Paul, F. Albarrán-Arriagada, E. Solano, A. del Campo, and X. Chen, "Digitized-counterdiabatic quantum approximate optimization algorithm," *Physical Review Research*, vol. 4, p. 013141, Feb. 2022.

[155] A. Szabo and N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Books on Chemistry, Mineola, New York: Dover Publications, Inc., 1st edition, revised ed., 1996.

[156] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-

C. Kwek, and A. Aspuru-Guzik, "Noisy intermediate-scale quantum algorithms," *Reviews of Modern Physics*, vol. 94, p. 015004, Feb. 2022.

[157] B. O. Roos, R. Lindh, P. Å. Malmqvist, V. Veryazov, and P.-O. Widmark, "The Multiconfigurational (MC) Self-Consistent Field (SCF) Theory," *Multiconfigurational Quantum Chemistry*, Chapter 9, pp.103-130, John Wiley & Sons, Ltd, 2016.

[158] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, "Tapering off qubits to simulate fermionic hamiltonians," 2017.

[159] P. K. Barkoutsos, J. F. Gonthier, I. Sokolov, N. Moll, G. Salis, A. Fuhrer, M. Ganzhorn, D. J. Egger, M. Troyer, A. Mezzacapo, S. Filipp, and I. Tavernelli, "Quantum algorithms for electronic structure calculations: Particle-hole hamiltonian and optimized wave-function expansions," *Physical Review A*, vol. 98, Aug. 2018.

[160] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, "Quantum computational chemistry," *Reviews of Modern Physics*, vol. 92, p. 015003, Mar. 2020.

[161] D. Freedman and P. Diaconis, "On the histogram as a density estimator: L2 theory," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, vol. 57, no. 4, pp. 453–476, 1981.

Appendix A
Experimental details, algorithmic implementations, and theoretical derivations

A.1 Error rates of the *ibmq_ehningen* quantum device

The used hardware in this work is one of *IBM*'s cloud devices *ibmq_ehningen*. All backend properties are taken as they are reported by *IBM*. Table A.1 presents the single-qubit device properties and readout errors. Since the R_z-gate is implemented in software [67], it does not result in errors. The error rates for each CNOT gate possible for each qubit pair are presented in Table A.2. Note that these error rates only partially describe the noise occurring on the actual device. For example crosstalk effects are not included, as discussed in Ref. [91].

A.2 Noise models for quantum hardware and depolarizing noise

In this appendix, we describe the noise models used in our simulations. First, we outline the backend noise model generated by *Qiskit* for the *ibmq_ehningen* device, which includes gate, readout, and decoherence errors based on hardware calibration data. Second, we introduce a manually constructed depolarizing noise model used to isolate the effect of gate-level depolarizing channels. Together, these models allow us to compare hardware noise with a simplified and analytically tractable noise setting.

A.2.1 Quantum hardware Noise Model

Qiskit provides noise models that simulate the behaviour of the quantum hardware. These models reproduce the characteristic errors of a specific device. A noise model consists of error channels that are applied to qubits during gate execution and measurements. For *IBM* quantum processors these channels are derived from calibration data of the backend and are defined for each qubit and each gate.

Single qubit and two qubit gates are affected by quantum channel errors. *Qiskit* represents these errors by Kraus operators, see Equation (2.5). The most relevant error types

Table A.1: *Table of the single-qubit device properties of ibmq_ehningen (as of August 18th, 2023). The table displays the frequencies, T1 and T2, gate errors of rz-, sx-, and x-gates, as well as readout errors for the all 27 qubits, as reported by IBM.*

Qubit	Frequency GHz	T1 µs	T2 µs	RZ error	SX error	X error	Readout error
Q0	4.961	142.59275	95.47855	0	0.00018	0.00018	0.011
Q1	5.18191	181.00386	70.72969	0	0.00021	0.00021	0.0096
Q2	5.12694	95.06903	9.06901	0	0.0003	0.0003	0.009
Q3	5.26815	99.89562	28.579	0	0.00017	0.00017	0.0229
Q4	5.05357	159.58091	70.14323	0	0.00068	0.00068	0.0123
Q5	5.07116	104.48761	8.15402	0	0.00032	0.00032	0.024
Q6	4.89006	143.66941	122.05597	0	0.00035	0.00035	0.014
Q7	4.97776	123.99653	113.28929	0	0.00021	0.00021	0.0076
Q8	5.17419	84.87928	69.68872	0	0.00104	0.00104	0.0245
Q9	4.9925	112.5169	62.87694	0	0.00573	0.00573	0.0175
Q10	4.83511	191.01486	162.14782	0	0.00039	0.00039	0.0092
Q11	5.11944	92.86756	127.55782	0	0.00027	0.00027	0.017
Q12	4.72549	169.03301	238.81786	0	0.00015	0.00015	0.0117
Q13	4.92598	155.88103	225.73924	0	0.00022	0.00022	0.008
Q14	5.17671	94.24501	258.67336	0	0.00033	0.00033	0.0077
Q15	4.89299	93.54811	176.56123	0	0.00019	0.00019	0.0105
Q16	5.02214	194.67907	191.30807	0	0.00028	0.00028	0.0078
Q17	5.13566	146.76432	22.47776	0	0.00052	0.00052	0.0073
Q18	4.99642	130.38322	228.58117	0	0.00024	0.00024	0.0128
Q19	4.7841	172.28179	77.40924	0	0.00032	0.00032	0.0137
Q20	5.04235	204.49692	220.82418	0	0.00053	0.00053	0.025
Q21	4.93974	126.28787	202.81237	0	0.00026	0.00026	0.0076
Q22	4.72513	58.42632	34.32104	0	0.00025	0.00025	0.0131
Q23	4.80479	190.68715	238.40096	0	0.00043	0.00043	0.0084
Q24	5.07449	233.74912	337.44479	0	0.00015	0.00015	0.0075
Q25	4.95019	198.39525	439.17036	0	0.00018	0.00018	0.0076
Q26	5.15132	190.73366	24.68715	0	0.00015	0.00015	0.0079

Table A.2: *Table displaying the gate errors associated with the available two-qubit gate pairs for ibmq_ehningen (as of August 18th, 2023).*

CX Gate Pair	Gate error	CX Gate Pair	Gate error
25_22	0.00565	10_12	0.00849
22_25	0.00565	12_10	0.00849
20_19	0.00602	7_10	0.02103
19_20	0.00602	10_7	0.02103
14_16	0.00847	19_22	0.00993
16_14	0.00847	22_19	0.00993
18_17	0.00896	21_23	0.00961
17_18	0.00896	23_21	0.00961
14_11	0.00959	13_12	0.00629
11_14	0.00959	12_13	0.00629
12_15	0.00726	26_25	0.00684
15_12	0.00726	25_26	0.00684
8_5	0.0101	3_2	0.00358
5_8	0.0101	2_3	0.00358
1_0	0.01138	7_6	0.00595
0_1	0.01138	6_7	0.00595
1_2	0.00681	16_19	0.00676
2_1	0.00681	19_16	0.00676
8_11	0.01272	18_21	0.00607
11_8	0.01272	21_18	0.00607
24_23	0.00885	4_7	0.00542
23_24	0.00885	7_4	0.00542
4_1	0.00655	14_13	0.00667
1_4	0.00655	13_14	0.00667
8_9	0.03875	24_25	0.00839
9_8	0.03875	25_24	0.00839
18_15	0.00544	3_5	0.00588
15_18	0.00544	5_3	0.00588

are: depolarizing errors, coherent gate errors, readout errors and thermal relaxation errors. The depolarizing noise channel is described with Equation (2.12). This models random noise that acts independently of the quantum state. For CNOT gates a separate depolarizing probability is extracted for each physical qubit pair. This probability is usually 10-times larger than for single qubit gates. Gate errors are modeled by associating noisy quantum channels, such as depolarizing or thermal-relaxation noise, with every gate operation. In the simulations, each gate is represented as an ideal unitary composed with its corresponding noise channel, effectively applying the error immediately after the gate.

Each backend provides a readout assignment matrix A for every qubit. This matrix describes the probability that a measured value is misidentified. For a single qubit it has the form

$$A = \begin{pmatrix} P(0|0) & P(0|1) \\ P(1|0) & P(1|1) \end{pmatrix}, \tag{A.1}$$

where $P(a|b)$ denotes the probability of measuring a given that the ideal value was b. These errors are applied directly at measurement. The readout error probabilities are taken from real calibration data of the device.

Thermal noise is described by T_1 (energy relaxation time) and T_2 (dephasing time). For a gate with duration t the simulator applies amplitude damping and phase damping channels derived from these time constants $\gamma_1 = 1 - e^{-t/T_1}$, $\gamma_\phi = 1 - e^{-t/T_2}$. Relaxation errors are applied during the execution of each gate according to its duration. Longer gates and especially two qubit gates contribute more strongly to these errors.

Backend noise models contain a stored snapshot of calibration data from a real device. This includes gate error probabilities, single- and two-qubit depolarizing rates, readout assignment matrices, relaxation times T_1 and T_2, gate durations, and the device connectivity and native gate set. Depolarizing channels are parameterized such that, when applied to each gate, they reproduce the average error rates reported for the corresponding hardware operations.

The simulator uses stored parameters to construct the backend noise model. During the simulation each gate in the quantum circuit is followed by its corresponding noise channel. Measurement operations apply the readout assignment errors and idle times between gates accumulate thermal relaxation noise.

In this way backend based noise models approximate the behaviour of current superconducting quantum devices and allow an evaluation of algorithms under near experimental conditions.

A.2.2 Depolarizing Noise Model

To account for gate imperfections, we simulated the circuits using depolarizing noise. Independent depolarizing channels were applied to all single- and two-qubit gates, such that each gate operation is followed by a corresponding depolarizing channel. The depolarizing channel is defined in Equation (2.12). The average gate infidelity (or average

error rate) r with respect to the ideal gate is defined as [117–119]

$$r = \frac{2^q - 1}{2^q} p, \tag{A.2}$$

where q is the number of qubits and p the probability that a depolarizing event occurs.For a two-qubit gate, the depolarizing probability is therefore $p_{2q} = 4/3\, r$. To reflect typical hardware ratios, we take 10% of this error probability for single-qubit gates, yielding $p_{1q} = 2r \cdot 0.1 = r/5$. This defines the 1- and 2-qubit depolarizing noise model used in this work. All other noise sources, such as measurement or thermal relaxation errors, were set to zero. Setting the average gate infidelity to 0.02 (2%) results in single-qubit and two-qubit depolarizing probabilities of $p_{1q} = 0.4\%$ and $p_{2q} = 2.7\%$, respectively. For the 2-qubit depolarizing noise model, the single-qubit errors were set to zero.

A.3 Box plot definition

From the values obtained in simulations and hardware runs, we construct box plots to summarize the distribution of extrapolated expectation values. A box plot visualizes the five-number summary consisting of the first quartile (Q1), the median, the third quartile (Q3), and the whisker endpoints. The box extends from Q1 to Q3, with the central line indicating the median; its height (here length)corresponds to the interquartile range (IQR = Q3 - Q1). The whiskers reach the most extreme data points within 1.5 IQR from the quartiles (i.e., down to Q1 - 1.5 IQR and up to Q3 + 1.5 IQR). Data points outside these bounds are classified as outliers and plotted separately as circles. An example is shown in Figure A.1.

A.4 Quantum algorithms and circuits implemented in this work

In this chapter, the quantum algorithms and circuits investigated in this work are presented. Specifically, we describe the Grover algorithm, the Harrow-Hassidim-Lloyd (HHL) algorithm, the Hadamard-Ladder (H-Ladder) circuit, as well as two random circuits. These circuits each serve a different purpose in order to gain insights into the performance of the analyzed error mitigation techniques.

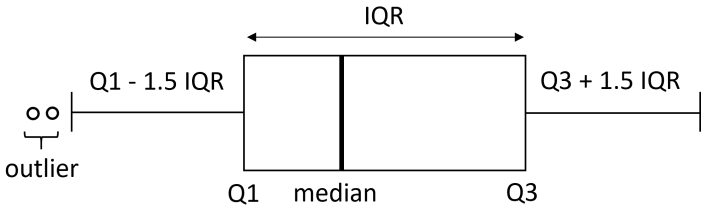


Figure A.1: *Box plot showing the median (central line), interquartile range (box), whiskers at Q1 - 1.5 IQR and Q3 + 1.5 IQR, and outliers (circles).*

A.4.1 Grover's search algorithm

Grover's search algorithm [143], in short Grover's algorithm, accelerates the search for a specific element within an unstructured set of elements with a quadratic speedup over classical approaches.

Search problems can be arranged into two different categories, structured and unstructured. A structured search involves searching for a word in an alphabetically ordered list, for example. This can easily be solved by a classical computer [144]. In an unstructured search, however, there is no order within the data set. An element can be found using a brute force approach: checking one element after another by using a linear search [144]. The worst-case time complexity is $\mathcal{O}(N)$, where N is the number of elements in the data set. However, Grover's algorithm has a time complexity of only $\mathcal{O}(\sqrt{N})$. It can significantly speed up a search.

Grover's algorithm amplifies the probabilities of the amplitudes of the target states, while suppressing those of the other states through repeated application of the oracle. The latter marks the target states by flipping the sign of their amplitudes. Afterwards, the diffusion operator is applied, inverting the amplitudes and further amplifying the marked state. For a data set with M target states, number of oracle queries is $R \simeq \frac{\pi}{4}\sqrt{\frac{M}{N}}$.

In this work, we consider the case where two states or binary strings are to be found in a 3-qubit system, $|101\rangle$ and $|011\rangle$. This means that we have $N = 8$ possible states and $M = 2$ target states. The corresponding observable we are looking for is $A = |101\rangle\langle 101| + |011\rangle\langle 011|$.

In the general case, we want to find the right binary string x_0 such that

$$f(x) = \begin{cases} 1 & \text{if } x = x_0 \\ 0 & \text{if } x \neq x_0. \end{cases} \tag{A.3}$$

The individual steps required to obtain x_0 with Grover's algorithm are described below.

1. **Initialization:**
   Prepare the uniform superposition of all three qubits, by applying Hadamard gates to $|\Psi\rangle = |000\rangle : H^{\otimes 3}|000\rangle$,

$$|\Psi_H\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle = \frac{1}{\sqrt{8}} \sum_{x=0}^{7} |x\rangle \tag{A.4}$$

2. **Grover Iterations:**
   Application $R \simeq 1.57$-times (one iteration suffices):

   (a) Apply oracle O
   The oracle O flips the sign of the amplitudes for $|101\rangle$ and $|011\rangle$ to mark the target states. The phase inversion can be written as

$$|\Psi_O\rangle = O|\Psi_H\rangle = \begin{cases} -|x\rangle & \text{if } x = 101 \text{ or } x = 011, \\ |x\rangle & \text{otherwise.} \end{cases} \tag{A.5}$$

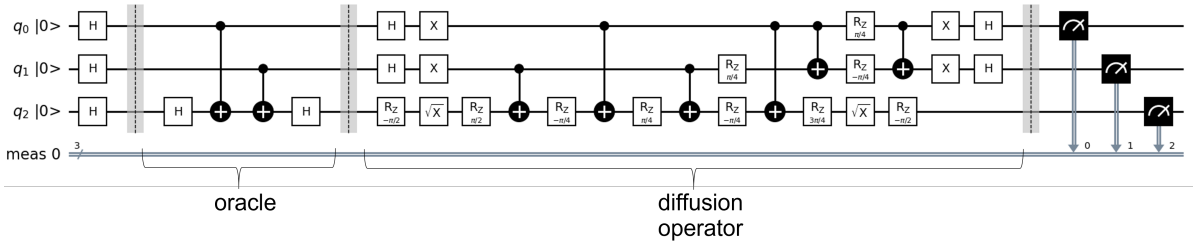


Figure A.2: *Representation of the Grover circuit for evaluating the performance of Zero-Noise Extrapolation (ZNE). It highlights the distinct phases of initialization, oracle execution, and diffusion application.*

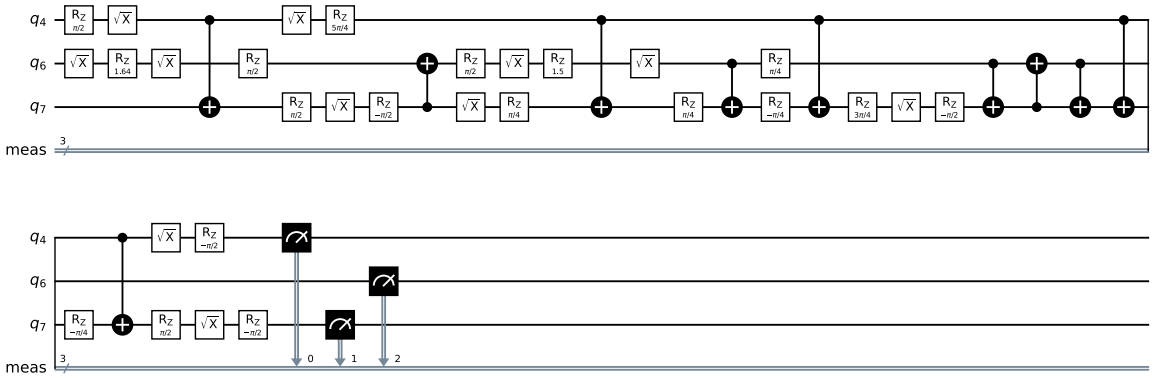


Figure A.3: *Representation of the Grover circuit. The circuit is designed to align with IBM's quantum device architecture through adaptation to the hardware, which necessitates one swap gate, which is implemented using three CNOT gates.*

(b) Apply diffusion operator D
The diffusion operator

$$D = 2|\Psi_H\rangle\langle\Psi_H| - \mathbb{I} \tag{A.6}$$

reflects the amplitudes about the mean μ (inversion about the mean), increasing the probabilities of $|101\rangle$ and $|011\rangle$. The final state after diffusion is then $D|\Psi_O\rangle = \frac{1}{\sqrt{2}}(|101\rangle + |011\rangle)$.

3. Measurement:
Measuring the state yields either $|101\rangle$ or $|011\rangle$ with probability $p(011) = p(101) = \left|\frac{1}{\sqrt{2}}\right|^2 = 0.5$, so that the total probability to measure one of the states is $p_{total} = p_{101} + p_{011} = 1$.

Following the explanation of the Grover algorithm, the circuit is shown in Figure A.2, which illustrates the original gate set, including marked oracle and diffusion operators. This representation also includes non-native gates that require conversion to comply with *IBM*'s quantum device architecture during hardware adaption. Figure A.3 illustrates the Grover circuit utilized in this work. It is specifically tailored to conform to *IBM*'s quantum device architecture through adaptation to the hardware. The circuit diagram incorporates one swap gate, necessitated by the constraints of the device architecture, thereby augmenting the total count of CNOT gates from 8 to 10.

A.4.2 Harrow-Hassidim-Lloyd algorithm

The Harrow-Hassidim-Lloyd (HHL) algorithm [94] is a quantum algorithm for numerically solving a linear system of equations with an exponential speedup under certain conditions. Classical methods scale as $\mathcal{O}(Ns\kappa \log(\frac{1}{\epsilon}))$ [94],where $\kappa = \frac{\delta_{max}}{\delta_{min}}$ is the condition number, where δ_i are the eigenvalues of the matrix B, s is the sparsity of matrix B, and ϵ is the error tolerance. In contrast, the HHL algorithm achives $\mathcal{O}(\frac{\kappa^2 s^2}{\epsilon} log(N))$ [94].

The algorithm solves $B\vec{x} = \vec{b}$ for $B \in \mathbb{C}^{N\times N}$ and a vector $\vec{b} \in \mathbb{C}^N$ to find $\vec{x} \in \mathbb{C}^N$ encoding the solution $\vec{x} = B^{-1}\vec{b}$ in a quantum state $|x\rangle$.

Consider the linear system defined by

$$B = \begin{pmatrix} 1 & -1/3 \\ -1/3 & 1 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{A.7}$$

The exact solution is $\vec{x} = (9/8, 3/8)$, with the norm $||\vec{x}|| = \frac{3\sqrt{10}}{8}$. The algorithmic steps to compute the ideal expectation value $\langle A_{\text{HHL}}\rangle_{\text{ideal}}$ is described below and an overview of the different algorithmic steps is shown in Figure A.4. The registers are abbreviated as follows: ancilla register (r_a), clock register (r_c) and input register (r_b).

1. **Load data** b
   Perform transformation

   $$|0\rangle \rightarrow |b\rangle_{r_b} = \sum_{i=0}^{N-1} b_i|i\rangle \text{ for } \vec{b} = (1,0)^T. \tag{A.8}$$

   State preparation in this example is trivial since $|b\rangle = |0\rangle$.

2. **Quantum Phase Estimation**
   Apply Quantum Phase Estimation (QPE) [16] with $U = e^{iBt}$ to estimate eigenvalues of B. Here, $t = \frac{3\pi}{4}$,

   $$U = e^{iBt} = \sum_{j=0}^{N-1} e^{i\delta_j t}|u_j\rangle\langle u_j|, \tag{A.9}$$

   where $\delta_0 = \frac{2}{3}$, $\delta_1 = \frac{4}{3}$ are eigenvalues of B and $|u_j\rangle$ the eigenvectors. After QPE, the state becomes

   $$\sum_{j=0}^{N-1} \beta_j|\tilde{\delta_j}\rangle|u_j\rangle_{r_b}, \tag{A.10}$$

   where $\beta_j = \langle u_j|b\rangle$ and $|\tilde{\delta_j}\rangle$ encodes $|\delta_j\rangle$.

3. **Inversion of eigenvalues**
   Perform a controlled rotation on an ancilla qubit, conditioned on $|\tilde{\delta_j}\rangle$

   $$\sum_{j=0}^{N-1} \beta_j|\tilde{\delta_j}\rangle_{r_c}|u_j\rangle_{r_b}(\sqrt{1 - \frac{C^2}{\delta_j^2}}|0\rangle + \frac{C}{\delta_j}|1\rangle), \tag{A.11}$$

   where $C \leq \min|\delta_j| = \frac{2}{3}$. In this example $C = \frac{1}{4}$.

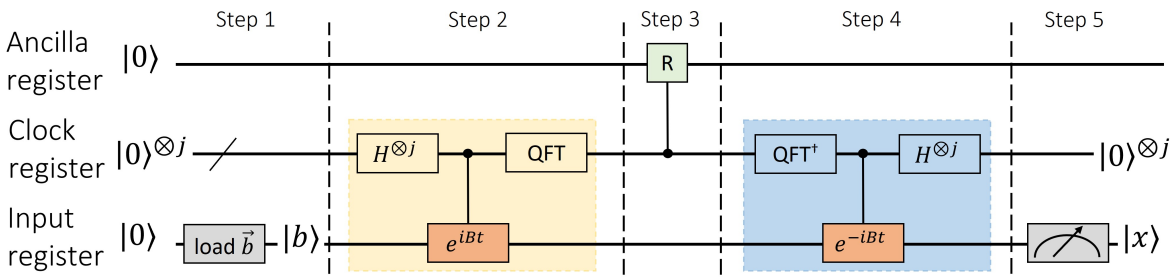


Figure A.4: *Schematic of the HHL algorithm. The different steps are explained in the main text. In step 2 the QPE is represented by a yellow box with the application of Hadamard gates to the clock register, a controlled rotation and the quantum Fourier transform (QFT). In step 4 the inverese QPE is represented by a blue box with the application of the inverse QFT, a controlled rotation and Hadamard gates.*

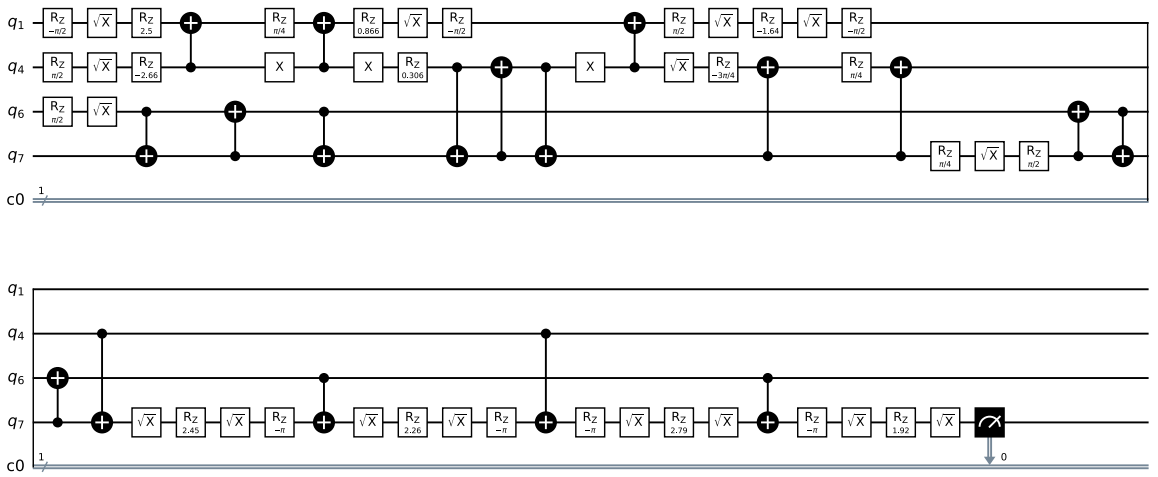


Figure A.5: *Representation of the HHL circuit for evaluating the performance of Zero-Noise Extrapolation (ZNE). Only the last qubit (q_7) is measured in case of HHL.The circuit is adapted to IBM's hardware layout, which requires three swap gates, each realised by a sequence of three CNOT gates.*

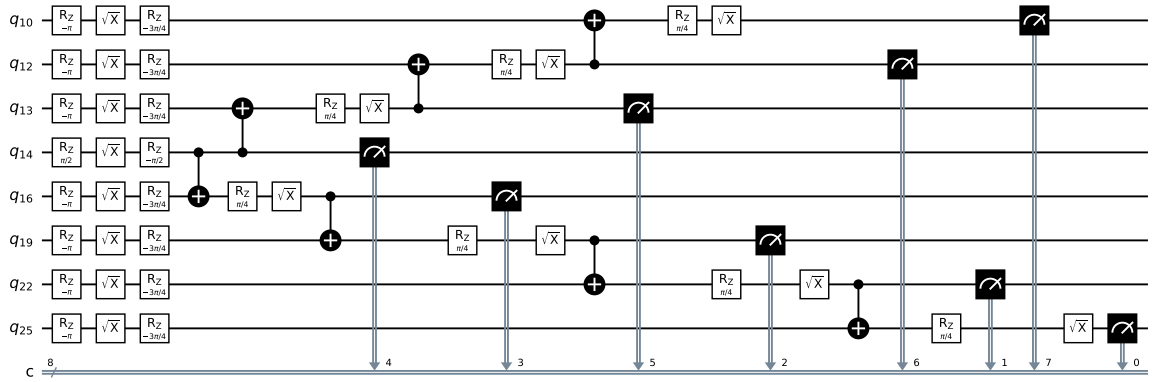


Figure A.6: *H-Ladder circuit adapted to align with IBM's device architecture.*

4. **Inverse quantum phase estimation**
   Uncompute the QPE registers to

$$\sum_{j=0}^{N-1} \beta_j |0\rangle_{r_c} |u_j\rangle_{r_b} \left(\sqrt{1 - \frac{C^2}{\delta_j^2}}|0\rangle + \frac{C}{\delta_j}|1\rangle\right). \tag{A.12}$$

5. **Measurement**
   Only the last qubit is measured, and the corresponding expectation value of the observable

$$A_{\mathrm{HHL}} = \mathbb{I} \otimes \mathbb{I} \otimes \mathbb{I} \otimes |1\rangle\langle 1| \tag{A.13}$$

   i.e., the probability to measure the last qubit in state $|1\rangle$, yields the norm of $\vec{x}$ as follows:

$$||\vec{x}|| = \frac{3}{2}\sqrt{\langle A_{\mathrm{HHL}}\rangle_{\mathrm{ideal}}}. \tag{A.14}$$

   In our case, $\langle A_{\mathrm{HHL}}\rangle_{\mathrm{ideal}} = (\frac{2}{3}||\vec{x}||)^2 = 5/8$, in accordance with the correct solution $\vec{x} = (9/8, 3/8)$ of the above linear system.

The four qubit circuit adapted to the hardware layout is presented in Figure A.5. It contains 18 CNOT gates in total. The qubits are equivalent to those in Figure A.4: q_1: ancilla register, q_4 and q_6:clock register, q_7: input register.

A.4.3 Hadamard-Ladder Circuit

The Hadamard-Ladder (H-Ladder) circuit, originally introduced in [51] and later employed in [91, 145], serves as a compact yet insightful benchmark for investigating noise effects in error mitigation techniques like Zero-Noise Extrapolation (ZNE).

This circuit prepares a uniform superposition state over all computational basis states, making it particularly suitable for investigating various types of noise. The circuit can exhibit strong sensitivity to coherent noise effects such as crosstalk and gate errors. These features make the H-Ladder circuit a valuable tool for probing non-local error mechanisms and their effect on ZNE.

Balancing between circuit depth and structural complexity, it is neither too shallow to be trivial nor too deep to be dominated by decoherence effects. This makes it especially well-suited for current noisy intermediate-scale quantum devices.

In this work, we consider the expectation value

$$\langle A_{\mathrm{Ladder}}\rangle = \langle 11111111|\rho_1|11111111\rangle, \tag{A.15}$$

which corresponds to the probability of measuring all qubits in $|1\rangle$. For the ideal, noiseless case, the expectation value is $\langle A_{\mathrm{Ladder}}\rangle_{\mathrm{ideal}} = 0.5$.

The H-Ladder circuit adapted to the hardware layout and used in this work is shown in Figure A.6.

A.4.4 Random circuits

In addition to algorithm-specific circuits, we also consider two random quantum circuits to further investigate the influence of noise and layout-dependent gate mapping on error mitigation processes. Both random circuits consist of four qubits and follow a similar structural template but differ in the number of entangling gates due to variations introduced during the adaptation to the hardware layout. Both random circuits are constructed such that each is followed by its exact inverse. Ideally, ensuring a return to the initial state $|0000\rangle$.

The first circuit $rand_0$ comprises 31 CNOT gates, while the second $rand_1$ contains 37 CNOT gates. Each circuit was designed to have 16 CNOT gates before being adapted to the hardware architecture. The discrepancy in the number of CNOT gates stems from the use of different qubit layouts during this adaptation, specifically $\{q_1, q_4, q_7, q_6\}$ for $rand_0$ and $\{q_6, q_7, q_4, q_1\}$ for $rand_1$. These layouts influence how the logical circuit is mapped onto the physical hardware, affecting the gate count.

For both random circuits, we consider the expectation value

$$\langle A_{\mathrm{rand}_{0,1}} \rangle = \langle 0000|\rho_0|0000\rangle, \tag{A.16}$$

which corresponds to the probability of measuring all qubits in state $|0\rangle$. For the ideal, noiseless case, the expectation value is $\langle A_{\mathrm{rand}_{0,1}} \rangle_{ideal} = 1.0$ The hardware adapted circuits are shown in Figure A.7 for $rand_0$ and in Figure A.8 for $rand_1$.

A.5 Quantum Approximate Optimization Algorithm implementations and problem definitions

This appendix provides the quantum circuit implementations and problem formulations used for the QAOA studies in this work. The Maximum Cut and Portfolio Optimization circuits follow the descriptions in Ref. [146]. The formulation of the Traveling Salesman Problem is based on the method introduced in Ref. [147]. These implementations are reproduced here for completeness and to maintain consistency with the original approaches.

A.5.1 Portfolio Optimization

Portfolio Optimization (PortOpt) aims to select the optimal portfolio to maximize expected returns while minimizing financial risk. The QAOA algorithm is a quantum algorithm that can be used for this purpose [148–151]. The PortOpt problem is first transformed into a quadratic unconstrained binary optimization (QUBO) [152] problem. Given n available assets and a selection budget of B assets, the cost function C can be formulated in terms of quantum operators [146, 149]

$$C = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{\lambda}{2}\Big(q\sigma_{ij} + A\Big) Z_i Z_j - \sum_{i=1}^{n} k_i Z_i, \tag{A.17}$$

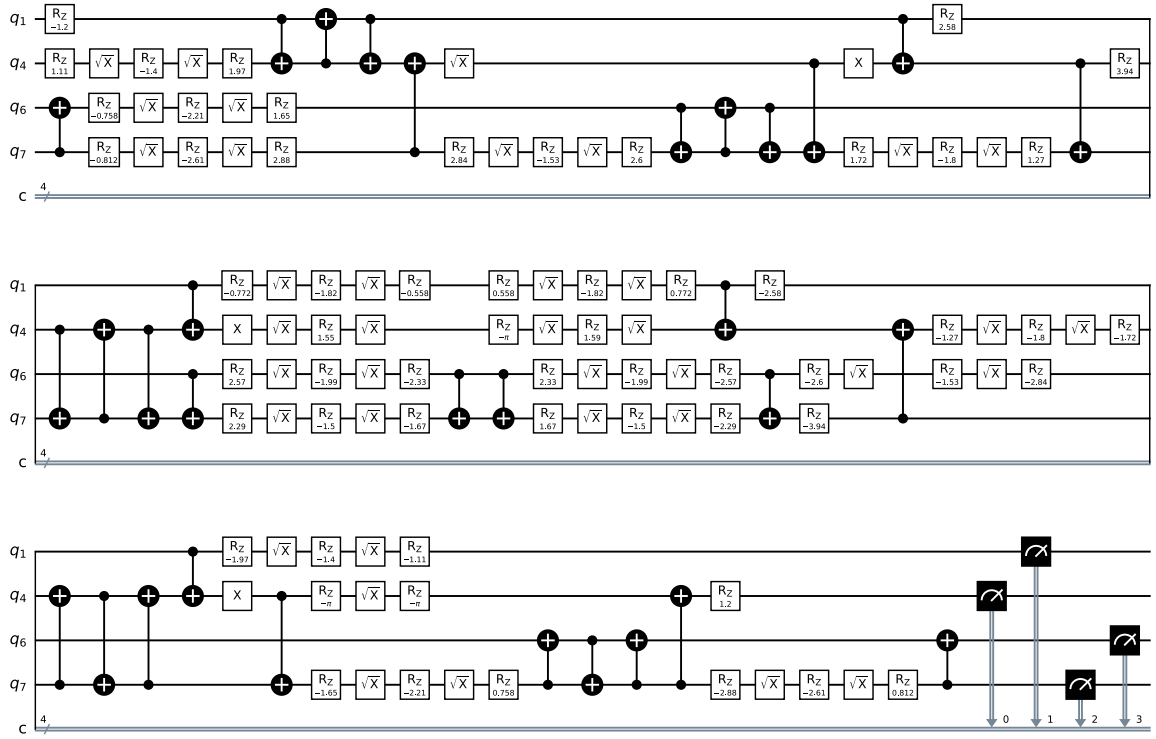


Figure A.7: *Random circuit adapted to hardware layout with 31 CNOT gates, based on layout* $[q_1, q_4, q_7, q_6]$*. The circuit is followed by its inverse, ideally returning the system to the initial state* $|0000\rangle$*.*

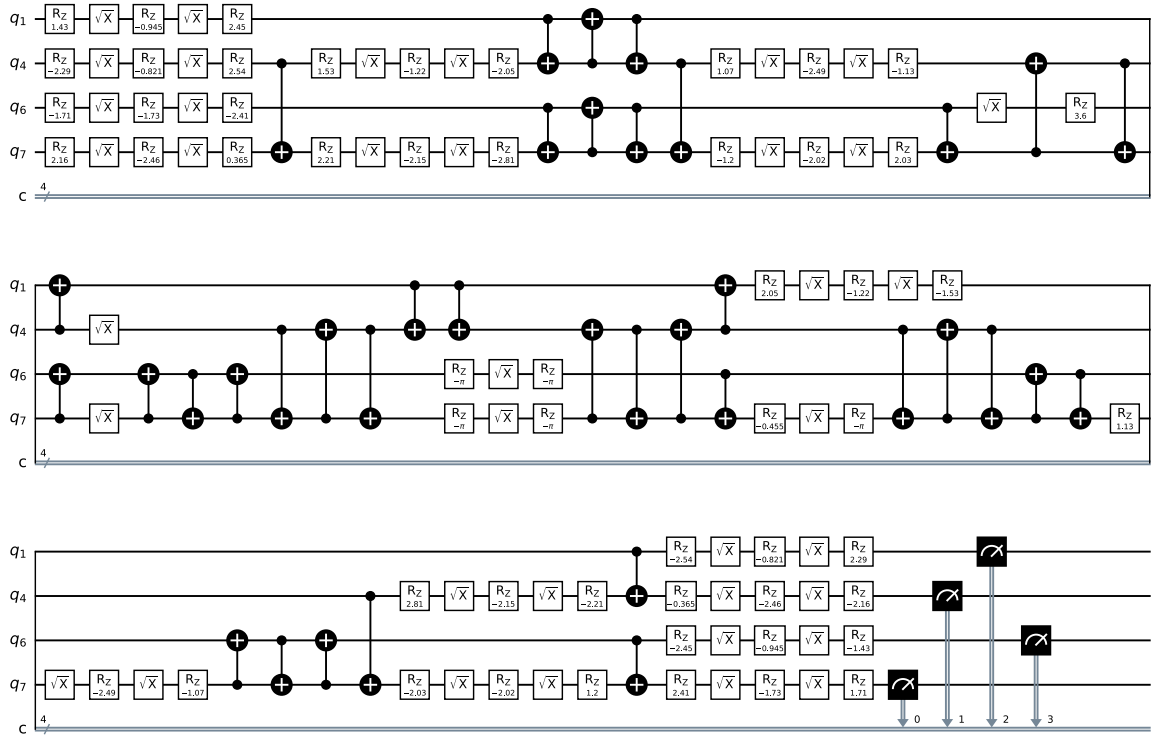


Figure A.8: *Random circuit adapted to hardware layout with 37 CNOT gates, based on layout* $[q_6, q_7, q_4, q_1]$*. The circuit is followed by its inverse, ideally returning the system to the initial state* $|0000\rangle$*.*

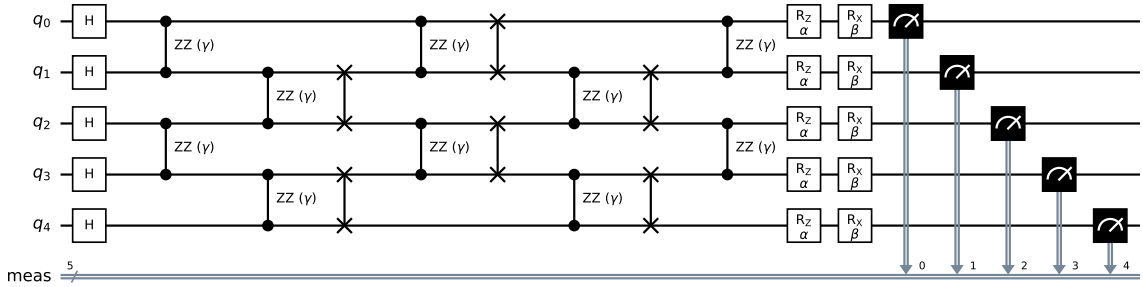


Figure A.9: *Representation of the portfolio optimization circuit for depth $p = 1$. The circuit is not adapted to the hardware basis gates. The layouts used on ibmq_ehningen are $[q_{16}, q_{14}, q_{11}, q_8, q_9]$ for echoed CNOT gates, $[q_6, q_7, q_4, q_1, q_2]$ for direct CNOT gates and $[q_{17}, q_{18}, q_{21}, q_{23}, q_{24}]$ for varied CNOT gates. The figure includes the controlled Z gate which is depicted by two filled dots connected by a line. It also includes the swap gate which is shown by two symbols with the letter x that are connected by a line. The swap gate is implemented through three CNOT gates. Other single qubit and standard gates follow the usual quantum circuit notation.*

with

$$k_i = \frac{\lambda}{2}\left[A(2B - n) + (1 - q)\mu_i - q\sum_{j=1}^{n}\sigma_{ij}\right], \tag{A.18}$$

where λ, q, σ_{ij}, A and μ_i are the global scaling factor, risk preference, covariance matrix, penalty factor, and expected return. Z_iZ_j and Z_i are the ZZ interaction on qubits i, j and Pauli Z operator acting on qubit i.

We use the QAOA algorithm with a cost Hamiltonian H_C, built from ZZ terms, as well as an X-mixer H_M, which are applied in p alternating layers. We use *COBYLA* [133] to optimize the QAOA parameters.

To assess the performance of QAOA for PortOpt we use the Approximation Ratio (AR) and the Success Probability (SP). A higher value of AR or SP indicates better performance. The post selected results that satisfy the budget constraint, which means all feasible solutions, are used to evaluate the AR for n assets $\{z_1, ..., z_n\}$. It is defined in the following way

$$r(z_1, \ldots, z_n) = C(z_1, \ldots, z_n)/C_{\text{opt}}, \tag{A.19}$$

where $C(z_1, ..., z_n)$ and C_{opt} denote the mean value obtained by QAOA and the optimal value respectively [146]. The SP corresponds to the probability of obtaining the optimal solution.

The circuit used in this work is shown in Figure A.9. It is presented in a form that is not yet adapted to the backend. On the quantum device, it would be decomposed into the basis gates of the backend. Three different layout implementations were applied in the experiments on *ibmq_ehningen*. These were used for the echoed CNOT gates $[q_{16}, q_{14}, q_{11}, q_8, q_9]$, the direct CNOT gates $[q_6, q_7, q_4, q_1, q_2]$ and the varied CNOT gates $[q_{17}, q_{18}, q_{21}, q_{23}, q_{24}]$. The figure shows the controlled Z gate and the swap gate, where the latter is composed of three CNOT gates. Each circuit evaluation is performed with 50 000 measurement shots.

A.5.2 Maximum Cut

Another application of QAOA is the determination of the maximum cut of a graph. The goal is to divide the vertices of the graph into two sets such that the number of edges between the two sets is maximized. This is known as the MaxCut problem [96, 153, 154].

The cost function C an be expressed by quantum operators in the following form

$$C = \frac{1}{2}\sum_{i,j}(1 - Z_i Z_j), \tag{A.20}$$

where vertices i, j share an edge. AR and SP are defined similar as for PortOpt above. For the MaxCut problem, we consider the complete graph which leads to the same connectivity requirements as the PortOpt circuit. Each circuit evaluation uses 50 000 measurement shots and the experiments are carried out on five qubits.

The circuit corresponds to the one shown in Figure A.9 with the only difference that no R_Z gates are applied. The layouts used for the five qubit MaxCut experiments on *ibmq_ehningen* were $[q_{16}, q_{14}, q_{11}, q_8, q_9]$ for echoed CNOT gates, $[q_6, q_7, q_4, q_1, q_0]$ for varied CNOT gates, and $[q_6, q_7, q_4, q_1, q_2]$ for direct CNOT gates.

A.5.3 Traveling Salesman Problem

The Traveling Salesman Problem (TSP) aims to find the shortest tour with a given set of cities, where each city is visited once. The QAOA algorithm can be used to solve this problem with a quantum computer.

The total cost of a tour can be encoded into a cost function that acts on the computational basis states

$$\begin{aligned} C(z) = &\sum_{i,j} w_{ij} \sum_t z_{i,t} z_{j,t+1} + A\sum_i \sum_t z_{i,t} + \frac{1}{2}A\sum_t \left(\sum_i \sum_{j>i} (-1)^{z_{i,t}+z_{j,t}} \right) \\ &+ \frac{1}{2}A\sum_i \left(\sum_t \sum_{\tau>t} (-1)^{z_{i,t}+z_{i,\tau}} \right), \end{aligned} \tag{A.21}$$

where $z = \{z_{i,t}\}$ specifies a candidate tour, $z_{i,t} \in \{0, 1\}$ indicates whether city i is visited at time step t, w_{ij} denotes the cost of traveling from city i to city j, $A > 0$ is a penalty coefficient, and τ labels a time step that occurs after t. The first term represents the travel cost between consecutive time steps. The second term provides an energy offset that ensures valid tours are energetically favored over invalid configurations, while the third term penalizes configurations in which more than one city is visited at the same time step. The final term penalizes tours in which a city is visited more than once at different time steps. By construction, valid tours minimize the cost function, while violations of the routing constraints incur a penalty proportional to A.

The Grover-mixer (GM-QAOA) [98] approach constructs a unitary U_S that prepares a uniform superposition over all valid tours

$$U_S|0\rangle_{\otimes q} = \frac{1}{\sqrt{n!}}\sum_\pi |\pi\rangle, \tag{A.22}$$

Table A.3: *Mean and standard deviation (mean ± std) over 25 optimization runs for GM-QAOA with increasing depth p. The corresponding results are visualized in Figure 5.1.*

p	Noise-free	GM-QAOA	ZNE+ASF B	IC-ZNE+ASF B
1	26.87 ± 6.79	72.38 ± 5.86	31.76 ± 14.83	28.25 ± 7.52
2	25.31 ± 7.16	84.35 ± 6.47	37.52 ± 30.00	25.53 ± 9.26
3	24.63 ± 7.42	114.77 ± 7.03	45.51 ± 44.76	24.64 ± 14.08
4	24.31 ± 7.58	109.40 ± 5.74	53.58 ± 52.76	20.45 ± 23.26
5	24.23 ± 7.60	115.23 ± 5.68	67.81 ± 59.49	20.14 ± 42.44

where the sum runs over all permutations π of a valid tour and q is the required number of qubits. This unitary ensures that every computational basis state in the superposition corresponds to one feasible tour. The implementation uses sequences of Hadamard, X and CNOT gates to activate exactly one qubit per time step and guarantees that the active qubits correspond to all permutations of the cities. The inverse unitary $U_S^\dagger$ is applied explicitly in the mixer construction. The mixer of GM-QAOA takes the form

$$U_M(\beta_k)e^{-i\beta_k|F\rangle\langle F|}, \tag{A.23}$$

where β_k is a variational parameter and $|F\rangle$ is the equal superposition of all feasible states. This mixer preserves the set of feasable tours and redistributes amplitudes within this subspace. The implementation uses a layer of X gates on all qubits, a multi controlled phase gate with angle proportional to β_k and a second layer of X gates.

The phase separator for one layer is the same as described in Subsection 2.5.1. A measurement in the computational basis produces a bitstring that represents one tour. The expectation value of the cost is

$$\langle C\rangle = \sum_z C(z)p(z), \tag{A.24}$$

where $p(z)$ is the observed probability of outcome z. The variational parameters are updated to minimize this expectation value. The final bitstring with the lowest total cost to the shortest tour found by the quantum algorithm. In this work, only the expectation value is evaluated.

The adapted circuit to the backend noise model of *FakeCairo* is shown in Figure A.10. The circuit includes 69 CNOT gates for $p = 1$.

The backend noise model used in this work is the *FakeCairo* backend provided in the *IBM Qiskit Runtime* package [116]. It is constructed from calibration data of the corresponding device as described in Appendix A.1. The backend contains 27 qubits and its native gate set is id, rz, sx, x, cx, reset. The qubit layout corresponds to the structure shown in Figure 2.2. The complete backend data are available in Ref. [132].

For completeness, Table A.3 summarizes the numerical values corresponding to the GM-QAOA results shown in Figure 5.1, reporting the mean and standard deviation over 25 optimization runs for each circuit depth p.

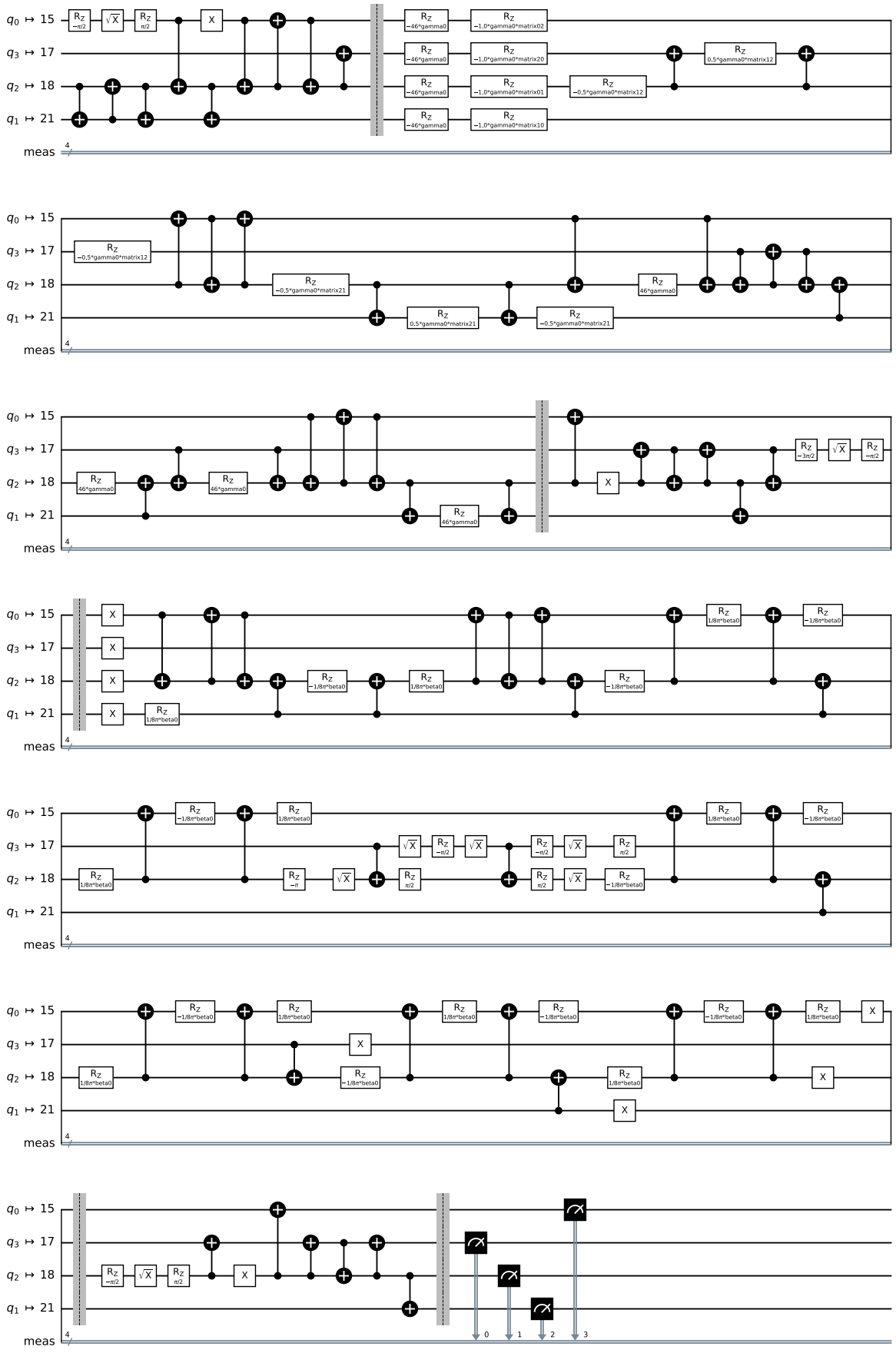


Figure A.10: *Representation of the TSP circuit. The circuit is adapted to the hardware layout of FakeCairo. The circuit includes 69 CNOT gates for $p = 1$.*

A.6 Variational Quantum Eigensolver implementations and problem definitions

In the following, we outline the steps required to construct a quantum simulation of a molecular system. Several approaches exist for electronic structure preparation, orbital selection, fermion to qubit mappings, ansätze for variational circuits, and classical optimization. We focus on one representative workflow and describe the specific methods used in this work.

We formulate the electronic structure problem in the second quantized representation [155, 156], which provides a compact and symmetry preserving description of fermionic operators. The molecular orbitals and the corresponding one and two electron integrals are obtained with the Python based Simulations of Chemistry Framework (PySCF) [136]. These integrals enter the construction of the fermionic Hamiltonian.

Next we define an active space [157]. Core orbitals are treated at the mean field level and kept frozen. High lying virtual orbitals can be removed if they do not contribute significantly to the correlation energy. The active space then contains the selected molecular orbitals and the electrons for which we include electron correlation. By focusing on frontier orbitals and near degenerate states, we capture the essential static correlation with far fewer degrees of freedom. This reduces the number of qubits and the required circuit depth for execution on near term quantum devices. An active space that is too small, however, may miss dynamic correlation contributions.

We then map the fermionic Hamiltonian to qubits. For this step we use the parity mapping [158]. Since particle number and spin parity are known for the state of interest, we can fix the corresponding symmetry sectors and remove the associated qubits. This tapering procedure preserves the energy spectrum inside the chosen symmetry sector and further reduces the circuit width.

For quantum state preparation, we choose a physically motivated ansatz. For molecular systems, we use the Unitary Coupled Cluster Singles and Doubles (UCCSD) ansatz [105, 159]. The reference state is the Hartree Fock state, constructed by setting the qubits corresponding to occupied spin orbitals to $|1\rangle$ and the others to $|0\rangle$. This reference improves the stability of the optimization procedure and reflects the mean field structure of the molecule. The UCCSD ansatz is chemically motivated and captures important correlation effects in a size consistent manner [103–105].

For classical optimization we employ Simultaneous Perturbation Stochastic Approximation (SPSA) [137]. Each iteration requires only two energy evaluations independent of the number of variational parameters, and the method is robust with respect to device noise and shot fluctuations [160]. The gradient is estimated through random perturbations applied in all directions at once, and the parameters are updated according to this estimate. A callback records expectation values and parameter vectors during the optimization.

The hardware adapted circuits for H_2 and KH with the implemented UCCSD ansatz are shown in Figures A.11 and A.12. They were generated using standard reductions such as core freezing, active-space selection, and parity mapping. The H_2 circuit uses 2 qubits and requires 4 CNOT gates, while the KH circuit uses 4 qubits and requires 158

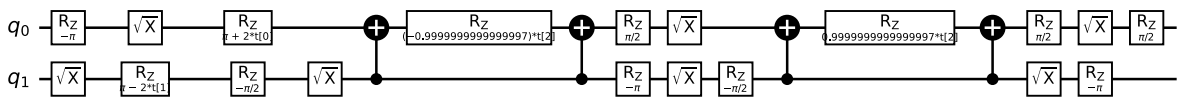


Figure A.11: *Hardware-adapted VQE circuit for H_2. The circuit uses two qubits and contains four CNOT gates after hardware adaption to the ibmq_ehningen coupling map. Only the logical qubits and active circuit elements are shown.*

CNOT gates. For KH the parametrized RZ rotations are displayed with their values set to zero to simplify the visual representation.

A.7 Derivation of the adjoint quantum channel property

We want to show that $\mathcal{E}_U = \mathcal{E}^\dagger{}_{U^\dagger}$, where

$$\mathcal{E}_U = \Lambda_p \mathcal{U}_U, \tag{A.25}$$

with $\Lambda_p(A) = (1-p)A + p\frac{\mathbb{I}}{d}\mathrm{A}$ [16] is the depolarizing noise channel, A an operator, p the probability that a qubit is depolarized and d the dimension of the Hilbert space. The channel corresponding to the unitary U acts on a density matrix ρ by [16]

$$\mathcal{U}_U(\rho) = U\rho U^\dagger, \tag{A.26}$$

and the channel corresponding to $U^\dagger$ accordingly by

$$\mathcal{U}_{U^\dagger}(\rho) = U^\dagger \rho U. \tag{A.27}$$

First, we show that $\Lambda_p \mathcal{U}_U = \mathcal{U}_U \Lambda_p$, meaning that the depolarizing channel commutes with the unitary evolution.

Proof:

$$\begin{aligned}
(\mathcal{U}_U \Lambda_p)(\rho) &= \mathcal{U}_U((1-p)\rho + p\frac{\mathbb{I}}{d}\mathrm{Tr}[\rho]) \\
&\underbrace{=}_{\mathcal{U}_U(\rho)=U\rho U^\dagger} (1-p)U\rho U^\dagger + p\frac{U\mathbb{I}U^\dagger}{d}\mathrm{Tr}[\rho] \\
&= (1-p)\underbrace{U\rho U^\dagger}_{\overline{\rho}=\mathcal{U}_U(\rho)} + p\frac{U\mathbb{I}U^\dagger}{d}\mathrm{Tr}[\rho\underbrace{U^\dagger U}_{\mathbb{I}}] \\
&= (1-p)\overline{\rho} + p\frac{\mathbb{I}}{d}\mathrm{Tr}[\underbrace{U\rho U^\dagger}_{\overline{\rho}}] \\
&= \Lambda_p(\overline{\rho}) \\
&= \Lambda_p(\mathcal{U}_U(\rho)) \\
&= (\Lambda_p \mathcal{U}_U)(\rho).
\end{aligned} \tag{A.28}$$

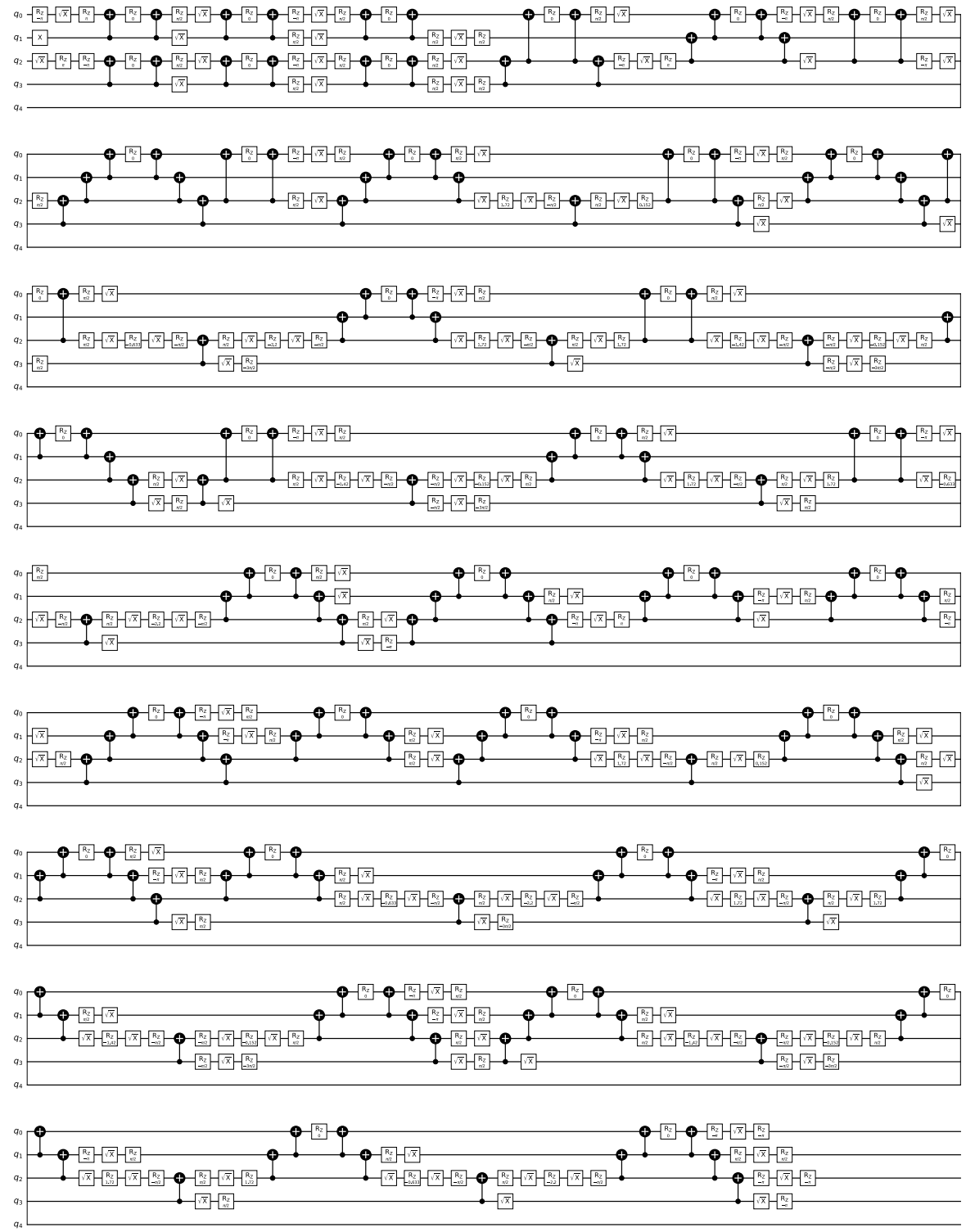


Figure A.12: *Hardware-adapted VQE circuit for KH. The circuit uses four qubits and contains 158 CNOT gates after adaption to the ibmq_ehningen device. The large number of entangling gates arises from the increased molecular complexity and depth of the UCCSD ansatz.*

For the channel $\mathcal{E}$, the adjoint channel $\mathcal{E}^\dagger$ is defined by

$$\mathrm{Tr}[A\mathcal{E}(B)] = \mathrm{Tr}\Big[\mathcal{E}^\dagger(A)B\Big], \tag{A.29}$$

for all linear operators A and B [16, 49].

Next, we show that the depolarizing channel is self-adjoint ($\Lambda_p^\dagger = \Lambda_p$). For arbitrary operators A, B

$$\begin{aligned}\mathrm{Tr}[A\Lambda_p(B)] &= \mathrm{Tr}[A(1-p)B + A\frac{p\mathbb{I}}{d}\mathrm{Tr}[B]] \\ &= (1-p)\mathrm{Tr}[AB] + \frac{p}{d}\mathrm{tr}[A]\mathrm{tr}[B], \qquad (A.30)\\ & \qquad (A.31)\end{aligned}$$

and similarly

$$\begin{aligned}\mathrm{Tr}[\Lambda_p(A)B] &= \mathrm{Tr}[(1-p)AB + \frac{p\mathbb{I}}{d}\mathrm{Tr}[A]B] \\ &= (1-p)\mathrm{Tr}[AB] + \frac{p}{d}\mathrm{tr}[A]\mathrm{tr}[B], \qquad (A.32)\\ & \qquad (A.33)\end{aligned}$$

so that, using the definition of the adjoint channel in Equation (A.29), we find $\Lambda_p^\dagger = \Lambda_p$.

Now, we show that $\mathcal{U}_U^\dagger = \mathcal{U}_{U^\dagger}$, for arbitrary operators A, B

$$\begin{aligned}\mathrm{Tr}[A\mathcal{U}_{U^\dagger}(B)] &= \mathrm{Tr}[AU^\dagger BU] \\ &= \mathrm{Tr}[UAU^\dagger B]\end{aligned}$$

with the cyclic property of trace

$$= \mathrm{Tr}[\mathcal{U}_U(A)B]. \tag{A.34}$$

By referring to the definition A.29 of the adjoint channel, we get

$$\mathrm{Tr}[A\mathcal{U}_{U^\dagger}(B)] = \mathrm{Tr}[\mathcal{U}^\dagger{}_{U^\dagger}(A)B]. \tag{A.35}$$

Comparing the two expressions, we conclude

$$\mathcal{U}_U = \mathcal{U}_{U^\dagger}^\dagger. \tag{A.36}$$

And finally we show that $\mathcal{U}_{U^\dagger}^\dagger\Lambda_p^\dagger = (\Lambda_p\mathcal{U}_{U^\dagger})^\dagger$, with

$$\begin{aligned}\mathrm{Tr}[A(\Lambda_p\mathcal{U}_{U^\dagger})(B)] &= \mathrm{Tr}[A\Lambda_p(\mathcal{U}_{U^\dagger}(B))] \\ &= \mathrm{Tr}[\Lambda_p^\dagger(A)\mathcal{U}_{U^\dagger}(B)] \\ &= \mathrm{Tr}[\mathcal{U}_{U^\dagger}^\dagger(\Lambda_p^\dagger(A))B] \\ &= \mathrm{Tr}[\underbrace{(\mathcal{U}_{U^\dagger}^\dagger\Lambda_p^\dagger)}_{(\Lambda_p\mathcal{U}_{U^\dagger})^\dagger}(A)B] \\ & \qquad (A.37)\end{aligned}$$

By the general property of adjoint channels

$$(\Lambda_p \mathcal{U}_{U^\dagger})^\dagger = \mathcal{U}_{U^\dagger}^\dagger \Lambda_p^\dagger. \tag{A.38}$$

Since $\Lambda_p^\dagger = \Lambda_p$ and $U_{U^\dagger}^\dagger = U_U$, we have

$$(\Lambda_p \mathcal{U}_{U^\dagger})^\dagger = \mathcal{U}_U \Lambda_p. \tag{A.39}$$

In summary we have

$$\mathcal{E}_U = \Lambda_p \mathcal{U}_U \overset{(A.28)}{=} \mathcal{U}_U \Lambda_p \overset{(A.39)}{=} (\Lambda_p \mathcal{U}_{U^\dagger})^\dagger = {\mathcal{E}^\dagger}_{U^\dagger}. \tag{A.40}$$

Thus, we have shown $\mathcal{E}_U = {\mathcal{E}^\dagger}_{U^\dagger}$ and therefore $\rho = \tilde{\rho}$ for depolarizing noise channels.

A.8 Details of the data-filtering procedure

This section provides the mathematical background and derivations underlying the data-filtering procedure, including the construction of histogram bins and confidence ellipses used for identifying and removing outliers.

The number of bins is determined using the Freedman–Diaconis rule [161], which balances sensitivity to data variation and robustness to outliers. The optimal bin width h is calculated as

$$h = 2\frac{\mathrm{IQR}}{n^{1/3}}, \tag{A.41}$$

where IQR is the interquartile range and n is the number of data points. The corresponding number of bins is then given by

$$b = \frac{x_{\max} - x_{\min}}{h}, \tag{A.42}$$

where $x_{\max}$and $x_{\min}$ are the maximum and minimum values in the data set, respectively.

To identify and exclude outliers based on bivariate (2D) distributions, we compute confidence ellipses that represent contours. The ellipses are derived from the empirical mean and covariance of the data, assuming approximately Gaussian statistics. Given a set of data points x_i, y_i, the procedure is as follows:

1. Compute the sample mean vector

$$\vec{\mu} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$$

2. Compute the covariance matrix

$$\Sigma = \begin{pmatrix} \mathrm{Var}(x) & \mathrm{Cov}(x,y) \\ \mathrm{Cov}(x,y) & \mathrm{Var}(y) \end{pmatrix}$$

3. Eigenvalue decomposition of covariance matrix yields
   - Eigenvalues δ_1, δ_2, which determine the squared axis lengths
   - Eigenvectors $\vec{v_1}, \vec{v_2}$, which determine the ellipse orientation
4. Construct ellipse

$$\vec{e}(\theta) = \vec{\mu} + \sqrt{5.99} \cdot A \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \end{pmatrix} \quad \text{for } \theta \in [0, 2\pi],$$

   where $A = \begin{pmatrix} \vec{v_1} \\ \vec{v_2} \end{pmatrix} \cdot \text{diag}(\sqrt{\delta_1}, \sqrt{\delta_2})$ is a transformation matrix composed of the eigenvectors $\vec{v_1}, \vec{v_2}$ and the square roots of the eigenvalues. The factor 5.99 resembles the confidence level 2σ.

Outliers are identified as points lying outside the desired confidence ellipse (e.g., beyond the 2σ contour). In our implementation, these ellipses are calculated per amplification factor for each run (local filtering), and optionally across multiple runs (global filtering), as described in Section 4.9.

Appendix B
Glossary

B.1 Acronyms

Acronym	Full name
AR	Approximation Ratio
ASF	Adaptive Scaling Factors
ASF B	Adaptive Scaling Factors, initial error strength measured on device
ASF M	Adaptive Scaling Factors, initial error strength derived from calibration data
COBYLA	Constrained Optimization BY Linear Approximations
CNOT	Controlled-NOT (gate)
CP	Cross-point (calibration)
CPTP	Completely Positive Trace-Preserving (map)
CR	Cross-Resonance (interaction/gate)
CX	Controlled-X (CNOT gate)
DRAG	Derivative Removal by Adiabatic Gate
ECR	Echoed Cross-Resonance (gate)
EM	Expectation-Maximization (algorithm)
FP	Final Point (mitigation)
GMM	Gaussian Mixture Model
GM-QAOA	Grover-Mixer Quantum Approximate Optimization Algorithm
H-Ladder	Hadamard-Ladder (circuit)
HHL	Harrow–Hassidim–Lloyd (algorithm)
IC-ZNE	Inverted-Circuit Zero-Noise Extrapolation
IQR	Interquartile Range
NISQ	Noisy Intermediate-Scale Quantum (era)
NV	Nitrogen Vacancy (centre)
PEC	Probabilistic Error Cancellation
PEA	Probabilistic Error Amplification
PortOpt	Portfolio Optimization
POVM	Positive Operator-Valued Measure
PP	Ping-Pong (calibration)
PySCF	Python-based Simulations of Chemistry Framework
QAMP	Qiskit Advocate Mentorship Program
QAOA	Quantum Approximate Optimization Algorithm

Acronym	Full name
QEC	Quantum Error Correction
QEM	Quantum Error Mitigation
QFT	Quantum Fourier Transform
QPE	Quantum Phase Estimation
QUBO	Quadratic Unconstrained Binary Optimization
RMSE	Root-Mean-Square Error
SNR	Signal-to-Noise Ratio
SP	Success Probability
SPAM	State Preparation And Measurement (error)
SPSA	Simultaneous Perturbation Stochastic Approximation
STO-3G	Slater-Type Orbital approximated by a linear combination of 3 primitive Gaussian functions
SX	Square-root-of-X gate
sZNE	standard Zero-Noise Extrapolation
TLS	Two-Level Systems
TSP	Traveling Salesman Problem
UCCSD	Unitary Coupled Cluster Singles and Doubles
VQA	Variational Quantum Algorithm
VQE	Variational Quantum Eigensolver
ZNE	Zero-Noise Extrapolation

B.2 List of symbols

Quantum States and Hilbert Spaces

q	Number of qubits in the quantum register
d	Dimension of the Hilbert space, $d = 2^q$
$\mathcal{H} = \mathcal{H}_2^{\otimes q}$	Hilbert space of a q-qubit system
$\otimes$	tensor (Kronecker) product
$\|0\rangle, \|1\rangle$	Logical basis states of a qubit
$\|\psi\rangle, \|\Psi\rangle$	Pure quantum states
ρ	Density matrix of a quantum state
ρ_*	Maximally mixed state, $\rho_* = I/2^q$

Operators, Channels, and Measurements

A	Observable (Hermitian operator)
$\langle A \rangle$	Expectation value of observable A
$\langle A \rangle_{\text{ideal}}$	Ideal (noise-free) expectation value
U	Ideal circuit unitary
$U^\dagger$	Adjoint (inverse) of U
$\mathcal{E}$	Quantum channel
E_k	Kraus operators of a quantum channel
$I, \mathbb{I}$	Identity operator (dimension 2×2, $2^q \times 2^q$)
Tr	Trace over the system Hilbert space

Quantum Gates and Pauli Operators

X, Y, Z	Pauli operators
$\sigma_x, \sigma_y, \sigma_z$	Pauli matrices in Hamiltonian notation
H	Hadamard gate
CNOT	Controlled-NOT (controlled-X) gate
$P_i = \|i\rangle\langle i\|$	Projector onto computational basis state i
$\mathcal{P}_i$	Multi-qubit Pauli operator

Noise Models and Errors

p	Depolarizing probability
Λ_p	Depolarizing noise channel with probability p
T_1	Energy relaxation time
T_2	Dephasing (coherence) time
γ_1	Relaxation parameter, $\gamma_1 = 1 - e^{-t/T_1}$
γ_ϕ	Dephasing parameter, $\gamma_\phi = 1 - e^{-t/T_2}$
t	Gate duration / evolution time in noise models

Calibration, Hardware and Pulse-Level Parameters

ω_{01}	Qubit transition frequency
ω_d	Drive frequency
$\Delta = \omega_{01} - \omega_d$	Drive detuning
Ω	Rabi frequency
A	Pulse amplitude
τ	Pulse duration
$I(t), Q(t)$	In-phase and quadrature pulse envelopes
α	Transmon anharmonicity
Γ	Spectroscopy linewidth (Full Width Half Maximum, FWHM)
f_{Rabi}	Rabi oscillation frequency
ϕ, φ	Phase offset
A_π	Pulse amplitude for a π rotation
Δ_A	Applied/scanned amplitude offset around a rough estimate A_0 (cross-point calibration)
Δ_A^*	Fitted cross-point offset on the Δ_A axis (intersection point); used in $A_{\text{new}} = A_0 + \Delta_A^*$
A_0	Rough initial amplitude estimate around which ΔA is scanned
A_{new}	Updated calibrated amplitude obtained from the cross-point estimate
$a_\pm$	Fit parameter cross-point calibration
θ_{ideal}	Ideal rotation angle
θ_{actual}	Implemented rotation angle
δ	Small coherent rotation error
β	Effective rotation angle per repetition (ping–pong fit)
n	Number of gate repetitions in calibration sequence
b	Offset parameter in calibration fit
a	oscillation amplitude
P	measured population
N	Number of measurement shots

Cross-Resonance Gate

H_{CR}	Effective cross-resonance Hamiltonian
a_{ab}	Coefficients of the reduced CR Hamiltonian
ϕ	Cross-resonance rotation angle
Φ	Relative CR drive phase
$\mathrm{CR}(\varphi)$	Cross-resonance unitary operation

Error Mitigation and (Inverted-Circuit) Zero-Noise Extrapolation

λ	Noise scaling factor
λ_i	Local noise scaling factor per gate
$\lambda_{\max}$	Maximum noise scaling factor
$\langle A\rangle_\lambda$	Expectation value at noise scale λ
$\langle A\rangle_*$	Expectation value of the maximally mixed state
F	State fidelity
$f(\lambda)$	Extrapolation fit function
r	Average gate infidelity
ϵ	Circuit error strength (1− fidelity)
ϵ_0	Initial (unscaled) circuit error strength
σ	Density matrix in $\rho = (1-\epsilon)\lvert\Psi\rangle\langle\Psi\rvert + \epsilon\sigma$
$\tilde{\sigma}$	Operator obtained by adjoint-channel evolution of the noise density matrix σ
a	Overlap parameter, $a = \mathrm{Tr}[\sigma\tilde{\sigma}]$
$\mathcal{F}_{\text{local, global}}$	local, global gate folding
M	Number of experimental runs
N_C	Number of CNOT gates
n_i	Number of identity insertions
ξ	Dimensionless tuning parameter for adaptive scaling

Randomized Compiling

π	Pauli permutation induced by Clifford conjugation
$\Lambda_{\mathcal{P}}$	Stochastic Pauli channel
p_i	Probability of Pauli error P_i

Variational Quantum Algorithms (QAOA and VQE)

H	Problem Hamiltonian
H_C	Cost (phase-separator) Hamiltonian
H_M	Mixer Hamiltonian
γ	QAOA phase-separator parameter
β	QAOA mixer parameter
p	QAOA circuit depth (context-dependent; distinct from depolarizing probability)
A	QAOA penalty factor
$\lvert\Psi(\boldsymbol{\beta},\boldsymbol{\gamma})\rangle$	QAOA variational state
$E(\boldsymbol{\theta})$	VQE energy expectation value
h_i	Coefficients of Pauli terms in the VQE Hamiltonian
b_i	VQE eigenvalue
$\boldsymbol{\theta}$	Vector of variational parameters
R	Internuclear distance
R_0	Equilibrium bond length

Filtering and Statistical Analysis

μ	Mean value of a data distribution
σ	standard deviation
σ^2	variance
$\rho_{x,y}$	correlation coefficient
Σ	Covariance matrix
δ_1, δ_2	Eigenvalues of covariance matrix
$\vec{v}_1, \vec{v}_2$	Eigenvectors of covariance matrix
$x_{\min}, x_{\max}$	Minimum and maximum data values in histograms
b	Number of histogram bins
h	Histogram bin width (Freedman–Diaconis rule)

Note: Some symbols (e.g. A, p, β) are used with different meanings in distinct contexts. Their interpretation is always clear from the surrounding discussion and chapter context.

B.3 List of figures

1.1 Coherence times of different qubit platforms . . . 1
1.2 Qubit numbers and entangling gate errors across platforms . . . 2

2.1 Gate symbols in quantum computing. . . . 7
2.2 Architecture of the *ibmq_ehningen* device. . . . 10
2.3 Additional gate symbols of *ibmq_ehningen*. . . . 10
2.4 Example data from the rough calibration procedure . . . 13
2.5 Schematic of the amplitude-calibration (ping-pong) error-amplification protocol . . . 15
2.6 Schematic of the amplitude-calibration (cross-point) error-amplification protocol . . . 17
2.7 Comparison of direct vs echoed CNOT gate . . . 19
2.8 Illustration of the Pauli twirling process. . . . 24
2.9 Structure of QAOA . . . 26
2.10 Structure of VQE . . . 28

3.1 Ping-pong and cross-point calibration experiment . . . 31
3.2 Box plot of X_π-pulse amplitude for different calibration methods and shot numbers . . . 32
3.3 $X_{\pi/2}$ amplitude for different calibration methods and shot numbers with identical rough values across experiments. . . . 35
3.4 Extracted X_π pulse amplitudes over four calibration rounds for each method, including the rough initial calibration. . . . 36
3.5 Expectation values of 15 two-qubit Pauli bases under CR gate evolution over varying pulse duration . . . 39
3.6 Comparison of the fitting result for the CNOT gate . . . 40
3.7 Architecture of *ibmq_ehningen* indicating two different implementations of the CNOT gate . . . 42
3.8 Error rates and gate times of direct and echoed CNOT gates . . . 43
3.9 Properties of *imbq_ehningen* . . . 45
3.10 QAOA runs on *imbq_ehningen* . . . 46

4.1 Analysis of noise strength ϵ_a as a function of qubit measurement probability P_0 . . . 55
4.2 Comparison of the Zero-Noise Extrapolation workflow with the Inverted-Circuit ZNE approach . . . 57

4.3 Scaled error strength as a function of the noise scaling factor 59
4.4 Comparison of a single run of sZNE and Inverted-Circuit ZNE for the Grover circuit 62
4.5 Comparison of a single run of sZNE and IC-ZNE for the HHL circuit . . 62
4.6 Comparison of a single run of sZNE and IC-ZNE for the H-Ladder circuit 62
4.7 Comparison of IC-ZNE, sZNE and unmitigated raw data for 50 simulation runs of the Grover circuit using a single- and two-qubit depolarizing error model 65
4.8 Comparison of IC-ZNE, sZNE and unmitigated raw data for 50 simulation runs of the HHL circuit using a single- and two-qubit depolarizing error model 65
4.9 Comparison of IC-ZNE, sZNE and unmitigated raw data for 50 simulation runs of the H-Ladder circuit using a single- and two-qubit depolarizing error model 65
4.10 Impact of number of circuit repetitions on IC-ZNE for different average gate infidelities. 66
4.11 Comparison of error mitigation techniques for Grover circuit on a *IBM* quantum device: unmitigated data, sZNE, and IC-ZNE 68
4.12 Comparison of error mitigation techniques for HHL circuit on a *IBM* quantum device: unmitigated data, sZNE, and IC-ZNE 68
4.13 Comparison of error mitigation techniques for H-Ladder circuit on a *IBM* quantum device: unmitigated data, sZNE, and IC-ZNE 68
4.14 Comparison of scaling approaches: standard equidistant scaling vs. exponential scaling with multiple factors 72
4.15 Empirical investigation of optimal ξ-values for Grover, HHL, and random circuits based on RMSE 72
4.16 Scaling factors λ as a function of initial error strengths ϵ_0 with exponential spacing 73
4.17 Variation in error strength across multiple backend executions for the Grover, HHL and H-Ladder circuits 74
4.18 Dependence of noise scaling factors on initial error strengths for the Grover, HHL and H-Ladder circuits 75
4.19 Extrapolation results based on adaptive scaling factors for Grover, HHL, and H-Ladder circuits using ZNE and IC-ZNE 77
4.20 Comparison of ϵ_0-filtering and local filtering 80
4.21 Local and global ellipse filtering 82
4.22 Error reduction for the Grover, HHL, and H-Ladder circuits using three error mitigation methods (sZNE, ASF, and IC-ZNE) under four different filtering strategies 84

4.23 Root Mean Square Error values for Grover, HHL, and H-Ladder circuits under various filtering techniques 85

4.24 Comparison of local and global folding in ZNE and IC-ZNE with ASF M applied to the Grover circuit. 88

4.25 Comparison of local and global folding in ZNE and IC-ZNE with ASF M applied to the HHL circuit. 89

5.1 Comparison of GM-QAOA results obtained under ideal and noisy conditions, with and without error-mitigation techniques. 95

5.2 Distribution of 50 optimization runs per internuclear distance for the H_2 molecule, used to estimate the ground-state energy. 98

5.3 Distribution of 100 optimization runs for the H_2 molecule to determine the ground-state energy. 100

5.4 Distribution of 100 optimization runs for the KH molecule to determine the ground-state energy. 100

A.1 Example box plot xvii

A.2 Representation of the Grover circuit xix

A.3 Representation of the Grover circuit aligned with *ibm*'s quantum device architecture xix

A.4 Representation of the HHL circuit xxi

A.5 Representation of the HHL circuit aligned with *ibm*'s quantum device architecture xxi

A.6 Representation of the H-Ladder circuit aligned with *ibm*'s quantum device architecture xxi

A.7 Representation of the random circuit 1 aligned with *ibm*'s quantum device architecture xxiv

A.8 Representation of the random circuit 2 aligned with *ibm*'s quantum device architecture xxiv

A.9 Representation of the portfolio optimization circuit xxv

A.10 Representation of the TSP optimization circuit xxviii

A.11 Hardware-adapted VQE circuit for H_2 xxx

A.12 Hardware-adapted VQE circuit for KH xxxi

Appendix C
List of publications

Journal articles

Kathrin Friederike Koenig, Finn Reinecke, Thomas Wellens
Adaptive folding and noise filtering for robust quantum error mitigation,
arXiv:2505.04463, 2025

Kathrin Friederike Koenig, Finn Reinecke, Walter Hahn, Thomas Wellens
Inverted-circuit zero-noise extrapolation for quantum-gate error mitigation,
Physical Review A, 110.042625, 2024, doi: 10.1103/PhysRevA.110.042625

Yanjun Ji, Kathrin Friederike Koenig, Ilia Polian
Improving the Performance of Digitized Counterdiabatic Quantum Optimization via Algorithm-Oriented Qubit Mapping,
Physical Review A, 110.032421, 2024, doi: 10.1103/PhysRevA.110.032421

Yanjun Ji, Kathrin Friederike Koenig, Ilia Polian
Optimizing quantum algorithms on bipotent architectures,
Physical Review A, 108.022610, 2023, doi: 10.1103/PhysRevA.108.022610

Andreas Ketterer, Kathrin König, Thomas Wellens
Error Characterization, Mitigation, and Correction,
Quantum Technologies, edited by R. Neugebauer, *Fraunhofer Verlag*, Stuttgart, p. 332-336, 2022, ISBN 978-3-8396-1850-9

Conference contributions

Kathrin Friederike Koenig, Finn Reinecke, Thomas Wellens
Inverted-circuit zero-noise extrapolation, *IQST early-career-researchers retreat*, Günzburg, Germany, May 22-24, 2024
Poster presentation

Kathrin Friederike Koenig, Finn Reinecke, Walter Hahn, Thomas Wellens
Inverted-circuit zero-noise extrapolation, *2024 APS March Meeting*, Minneapolis, USA, March 4-8, 2024
Oral presentation

Kathrin Friederike Koenig, Finn Reinecke, Thomas Wellens
Inverted-circuit zero-noise extrapolation for quantum error mitigation, *IBM Quandum Developer Forum*, Tarrytown, USA, April 23-25, 2023
Invited oral presentation

Kathrin Friederike Koenig, Finn Reinecke, Thomas Wellens
New techniques to improve zero-noise extrapolation on superconducting qubits, *26th Conference on Quantum Information Processing*, Ghent, Belgium, February 4-10, 2023
Poster presentation

Kathrin Friederike Koenig, Finn Reinecke, Thomas Wellens
New techniques to improve zero-noise extrapolation on superconducting qubits, *778. WE-Heraeus-Seminar: Coping with Errors in Scalable Quantum Computing Systems*, Bad Honnef, Germany, January 8-11, 2023
Poster presentation

Kathrin Friederike Koenig, Finn Reinecke, Thomas Wellens
Error Mitigation Techniques - Zero-Noise Extrapolation, *KQCBW - Developer Conference*, Ehningen, Germany, October 20-21, 2022
Oral presentation

Kathrin König, Thomas Wellens
Error Mitigation - Handling noisy quantum hardware, *DPG Meeting of the Condensed Matter Section*, Regensburg, Germany, September 4-9, 2022
Oral presentation

Kathrin König, Naoki Kanazawa, Shota Nakasuji
Analysis of calibration techniques for single-qubit gates on superconducting quantum computer, *DPG Meeting of the Condensed Matter Section*, Regensburg, Germany, September 4-9, 2022
Oral presentation

Acknowledgments

I would like to acknowledge the support of numerous individuals who have contributed to this work, both professionally and personally.

First and foremost, I would like to express my sincere gratitude to Prof. Dr. Dr. Oliver Ambacher, my supervisor, whose guidance and support made this PhD possible. I am also very grateful to PD Dr. Thomas Wellens, my day-to-day supervisor and second examiner, for his continuous support, insightful discussions, and valuable feedback throughout this thesis.

I would like to thank Prof. Dr. Rüdiger Quay for supporting my doctoral research at the institute. I am further thankful to all members of the Quantum Information group for the stimulating working environment and helpful discussions, as well as to my fellow PhD students for sharing both the challenges and successes of doctoral research.

I also gratefully acknowledge the support and collaboration of my colleagues at IAF and of collaborators involved in various projects.

On a personal level, I am deeply grateful to my family for their support and encouragement throughout this journey. I am also thankful to my friends, especially Ina, Lukas, Sebi, Mey, Patrick, and Conny, for their friendship and encouragement.
I would like to thank Tim, who has always been there for me and supported me in the best possible way. Especially over the past year, you have taken on so much, and I cannot express how grateful I am to have you by my side. Whether facing personal tragedies or challenges at work, I could always rely on you.
I would also like to thank Dirk. Although you are no longer here to see this, you were among those who always believed in me and supported me unconditionally. Even after your sudden death, you have continued to support me in spirit, and I know you would be extremely proud of what I have achieved here.

Quantum computing promises to address problems in simulation, optimization, and chemistry, but current hardware operates in the noisy intermediate-scale (NISQ) regime where coherence limits, gate imperfections, and readout errors restrict performance. Improving calibration and mitigating errors therefore remain essential.

This thesis analyzes error reduction on superconducting transmon qubits through calibration, gate analysis, and circuit-level mitigation. It compares single-qubit amplitude calibration methods and analyzes two-qubit gates based on the cross-resonance interaction, identifying differences in precision, convergence, and performance. In addition, it develops Inverted-Circuit Zero-Noise-Extrapolation (IC-ZNE) which measures effective error strength using inverted circuits. Compared to standard ZNE, IC-ZNE achieves results closer to ideal values and reduces observable errors. Overall, the presented methods improve the reliability of near-term quantum processors.

ISBN 978-3-8396-2199-8

Fraunhofer VERLAG