

# **EXPLORATION OF ERROR MITIGATION TECHNIQUES FOR 20-QUBIT QUANTUM SYSTEMS**

**Chen Jiaxin**

**XIAMEN UNIVERSITY MALAYSIA**

**2025**



**XIAMEN UNIVERSITY MALAYSIA**  
**廈門大學馬來西亞分校**

**FINAL YEAR PROJECT REPORT/ THESIS OF DEGREE**

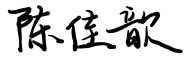
**EXPLORATION OF ERROR MITIGATION  
TECHNIQUES FOR 20-QUBIT QUANTUM SYSTEMS**

NAME OF STUDENT : CHEN JIAXIN  
STUDENT ID : PHY2109510  
SCHOOL/ FACULTY : XIAMEN UNIVERSITY MALAYSIA  
PROGRAMME : BACHELOR OF SCIENCE IN PHYSICS  
INTAKE : 2021/09  
  
SUPERVISOR : TOMASZ PATEREK  
PROFESSOR

**JUNE 2025**

## DECLARATION

I hereby declare that this project report is based on my original work except for citations and quotations which have been duly acknowledged. I also declare that it has not been previously and concurrently submitted for any other degree or award at Xiamen University Malaysia or other institutions.

Signature :   
Name : Chen Jiaxin  
ID No. : PHY2109510  
Date : June 30, 2025

## APPROVAL FOR SUBMISSION

I certify that this project report entitled "EXPLORATION OF ERROR MITIGATION TECHNIQUES FOR 20-QUBIT QUANTUM SYSTEMS" was prepared by CHEN JIAXIN has met the required standard for submission in partial fulfillment of the requirements for the award of Bachelor of PHYSICS at Xiamen University Malaysia.

Approved by,

Signature :   
Supervisor : Prof. Dr. Tomazs Paterek  
Date : 11 July 2025

Signature :  
Co-Supervisor : Prof. Dr. Rainer Dumke  
Date : \_\_\_\_\_

The copyright of this report belongs to the author under the terms of Xiamen University Malaysia copyright policy. Due acknowledgement shall always be made of the use of any material contained in, or derived from, this project report/ thesis.

©2025, Chen Jiaxin. All right reserved.

## **ACKNOWLEDGEMENTS**

I would like to thank all who have contributed to the successful completion of this project. I would like to express my gratitude to my research supervisor, Prof. Dr. Tomasz Paterek and external supervisor Prof. Dr. Rainer Dumke for their invaluable advice, guidance and enormous patience throughout the development of the research.

In addition, I would also like to express my gratitude to my loving parents and friends who have helped and given me encouragement.

## ABSTRACT

Quantum error mitigation (QEM) is essential for improving the reliability of noisy intermediate-scale quantum (NISQ) devices, particularly as current platforms scale toward 20 qubits where full quantum error correction remains infeasible. This project evaluates and compares three major QEM strategies—Zero-Noise Extrapolation (ZNE), Probabilistic Error Cancellation (PEC), and a machine learning-based framework (ML-QEM)—to assess their effectiveness and scalability. Through theoretical analysis and simulation study of ZNE under various noise models and depth, we show that while ZNE significantly enhances fidelity for shallow circuits, its effectiveness deteriorates rapidly in deeper or entangled circuits due to amplified decoherence and extrapolation instability. PEC provides an unbiased solution by inverting the noise channel but incurs prohibitive sampling and calibration costs in circuits with more than a few qubits, limiting its scalability. To address these limitations, we propose a modular ML-QEM approach that trains supervised models on 2-qubit circuits and applies them to larger circuits using a patch-based strategy. By incorporating hardware-aware and circuit-level features, the model aims to generalize noise behavior without relying on full noise tomography or ideal outputs during inference. While our ML implementation remains limited to simulated 4-qubit circuits due to time and resource constraints, results imply the viability of this framework as a scalable alternative. Rather than serving as a complete solution, this patch-based ML approach is positioned as a forward-looking data-driven direction. Overall, this study provides a comparative analysis of key QEM techniques and contributes a proof-of-concept ML-QEM workflow that bridges small-scale training with medium-scale application, offering practical potential into building adaptive error mitigation strategies for future superconducting quantum hardware.

**Keywords:** Quantum Error Mitigation (QEM), Zero-Noise Extrapolation (ZNE), Machine Learning (ML), Superconducting Qubits, Noisy Intermediate-Scale Quantum (NISQ), Patch-Based Modeling

## TABLE OF CONTENTS

|                                |             |
|--------------------------------|-------------|
| <b>DECLARATION</b>             | <b>ii</b>   |
| <b>APPROVAL FOR SUBMISSION</b> | <b>iii</b>  |
| <b>ACKNOWLEDGEMENTS</b>        | <b>vi</b>   |
| <b>ABSTRACT</b>                | <b>vii</b>  |
| <b>TABLE OF CONTENTS</b>       | <b>viii</b> |
| <b>LIST OF TABLES</b>          | <b>x</b>    |
| <b>LIST OF FIGURES</b>         | <b>xi</b>   |

## CHAPTER

|                                                             |           |
|-------------------------------------------------------------|-----------|
| <b>1 INTRODUCTION</b>                                       | <b>1</b>  |
| <b>2 BACKGROUND AND LITERATURE REVIEW</b>                   | <b>3</b>  |
| 2.1 Superconducting Qubits and Target 20-Qubit Architecture | <b>3</b>  |
| 2.2 Quantum Error Mitigation Strategy                       | <b>6</b>  |
| 2.2.1 Quantum Error Correction Theory                       | <b>6</b>  |
| 2.2.2 Two Foundational Quantum Error Mitigation Strategies  | <b>8</b>  |
| 2.2.3 Machine Learning-Based Quantum Error Mitigation       | <b>8</b>  |
| <b>3 QUANTUM NOISE MODEL</b>                                | <b>13</b> |
| 3.1 Open Quantum System and Dynamical Map                   | <b>13</b> |
| 3.1.1 Von Neumann equation from Kraus representation        | <b>15</b> |
| 3.2 Master Equation in Lindblad Formalism                   | <b>16</b> |
| 3.2.1 Markov approximation                                  | <b>16</b> |
| 3.2.2 Lindblad form                                         | <b>17</b> |
| 3.3 Examples for Quantum Noise Model                        | <b>19</b> |
| 3.3.1 Amplitude Damping                                     | <b>19</b> |

|                                                            |           |
|------------------------------------------------------------|-----------|
| 3.3.2 Depolarizing Noise                                   | 22        |
| 3.3.3 Phase Damping                                        | 22        |
| <b>4 RESEARCH METHODOLOGY</b>                              | <b>24</b> |
| 4.1 Zero Noise Extrapolation (ZNE) Method                  | 24        |
| 4.1.1 Theoretical Framwork                                 | 24        |
| 4.1.2 Simulation Workflow                                  | 27        |
| 4.2 Machine Learning-Based Quantum Error Mitigation Method | 29        |
| 4.2.1 Patch-Based ML-QEM Method                            | 30        |
| 4.2.2 Simulation Setup and Backend Details                 | 32        |
| <b>5 RESULTS AND DISCUSSION</b>                            | <b>35</b> |
| 5.1 MSE vs. Noise Strength across Noise Types              | 35        |
| 5.2 Fidelity vs. Gate Type under Mixed Noise               | 37        |
| 5.3 Fidelity vs. Circuit Depth under Mixed Noise           | 37        |
| 5.4 ML-based results                                       | 38        |
| 5.5 Comparison and Discussion of QEM Strategies            | 40        |
| <b>6 CONCLUSION</b>                                        | <b>43</b> |
| 6.1 Recommendation for Future Work                         | 44        |
| <b>REFERENCES</b>                                          | <b>45</b> |

## LIST OF FIGURES

|             |                                                                                                                                                                                     |    |
|-------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figure 2.1: | Josephson junction and energy levels . . . . .                                                                                                                                      | 6  |
| Figure 2.2: | Schematic of the IQM 20Q Garnet architecture. . . . .                                                                                                                               | 8  |
| Figure 2.3: | Structure of the surface code. . . . .                                                                                                                                              | 10 |
| Figure 2.4: | ML-QEM Framework. . . . .                                                                                                                                                           | 13 |
| Figure 3.1: | Open System Map . . . . .                                                                                                                                                           | 16 |
| Figure 3.2: | System Dynamic Map . . . . .                                                                                                                                                        | 16 |
| Figure 4.1: | ZNE recovery of $\langle X \rangle$ from a noisy Hadamard circuit under<br>amplitude damping noise. The extrapolated value closely ap-<br>proximates the ideal expectation. . . . . | 30 |
| Figure 5.1: | ZNE MSE vs Noise Parameter $p$ under different noise models.<br>ZNE reduces MSE especially under phase damping. . . . .                                                             | 36 |
| Figure 5.2: | ZNE fidelity across different single-qubit gates on a mixed<br>noise model. ZNE improves fidelity consistently across all<br>tested gates. . . . .                                  | 37 |
| Figure 5.3: | Fidelity vs circuit depth on real-device noise model with and<br>without ZNE. ZNE improves shallow circuits but degrades in<br>deeper ones. . . . .                                 | 37 |
| Figure 5.4: | ML-QEM performance analysis . . . . .                                                                                                                                               | 39 |
| Figure 5.4: | Comparison of ML error mitigation MSE across different gate<br>depth. . . . .                                                                                                       | 40 |

## LIST OF TABLES

|            |                                             |    |
|------------|---------------------------------------------|----|
| Table 5.1: | Feasures across different patches . . . . . | 39 |
|------------|---------------------------------------------|----|

## LIST OF SYMBOLS / ABBREVIATIONS

|                                |                                                           |
|--------------------------------|-----------------------------------------------------------|
| $G$                            | specific mass flow rate, kg/s                             |
| $h$                            | height, m                                                 |
| $\epsilon$                     | error rate (gate or measurement)                          |
| $\mathcal{N}$                  | noise channel (Kraus or Lindblad formalism)               |
| $\rho$                         | quantum state (density matrix)                            |
| $\phi$                         | phase difference across Josephson junction                |
| $E_J$                          | Josephson energy, $E_J = \frac{\hbar I_c}{2e}$            |
| $I_c$                          | critical current of Josephson junction                    |
| $E_C$                          | charging energy, $E_C = \frac{e^2}{2(C_S + C_J)}$         |
| $n$                            | number of Cooper pairs                                    |
| $\varphi$                      | superconducting phase difference                          |
| $\alpha$                       | qubit anharmonicity, $\alpha = E_{21} - E_{10}$           |
| $T_1$                          | energy relaxation time                                    |
| $T_2$                          | dephasing time                                            |
| $T_\phi$                       | pure dephasing time                                       |
| $\epsilon_0$                   | base error rate                                           |
| $\gamma$                       | error scaling exponent                                    |
| $ 0\rangle,  1\rangle$         | computational basis states                                |
| $ +\rangle,  -\rangle$         | Hadamard basis states                                     |
| $\langle \hat{O} \rangle$      | expectation value of observable $\hat{O}$                 |
| $X_j$                          | Pauli- $X$ (bit-flip) operator on qubit $j$               |
| $Z_j$                          | Pauli- $Z$ (phase-flip) operator on qubit $j$             |
| $A_v$                          | Vertex stabilizer (surface code)                          |
| $B_p$                          | Plaquette stabilizer (surface code)                       |
| $\rho$                         | density matrix                                            |
| $\mathcal{H}_S, \mathcal{H}_B$ | system/environment Hilbert spaces                         |
| $\gamma$                       | amplitude damping probability                             |
| $\Gamma$                       | amplitude damping rate, $\gamma = \Gamma \Delta t$        |
| $\xi$                          | thermal population of excited state (GAD)                 |
| $\beta$                        | inverse temperature, $\beta = 1/k_B T$                    |
| $\varepsilon$                  | energy gap, $\varepsilon = \varepsilon_2 - \varepsilon_1$ |
| $p$                            | depolarizing probability                                  |
| $v$                            | phase damping probability                                 |
| $\Upsilon$                     | phase damping rate, $v = \Upsilon \Delta t$               |

|                                |                                                     |
|--------------------------------|-----------------------------------------------------|
| $\lambda$                      | decay constant (GAD)                                |
| $\sigma_x, \sigma_y, \sigma_z$ | Pauli matrices                                      |
| $\sigma_+$                     | raising operator, $\sigma_+ =  1\rangle\langle 0 $  |
| $\sigma_-$                     | lowering operator, $\sigma_- =  0\rangle\langle 1 $ |
| $\mathcal{L}$                  | Liouvillian superoperator                           |
| $\mathbb{I}$                   | identity operator                                   |
| $\{A, B\}$                     | anti-commutator, $AB + BA$                          |
| $K_n$                          | Kraus operator (for channel $\mathcal{E}$ )         |
| $L_k$                          | Lindblad operator (dissipative processes)           |
| $H_{\text{tot}}$               | Total Hamiltonian ( $H_S + H_B + H_{\text{int}}$ )  |
| $H_{\text{eff}}$               | Effective system Hamiltonian                        |
| $U_{SB}$                       | Unitary operator for system-environment evolution   |
| $\ell_2$                       | Euclidean norm                                      |
| Pauli- $X/Y/Z$                 | Pauli spin operators                                |
| $T_1$                          | energy relaxation time                              |
| $T_2$                          | dephasing time                                      |
| CNOT                           | Controlled-NOT gate                                 |
| $\mathcal{L}_\lambda$          | noise-dependent Liouvillian superoperator           |
| $E(\lambda)$                   | noisy expectation value at noise strength $\lambda$ |
| $E_0$                          | ideal (noise-free) expectation value                |
| $\alpha$                       | noise decay constant                                |
| $\lambda$                      | noise strength parameter                            |
| $c$                            | noise scaling factor (gate folding count)           |
| $\epsilon$                     | statistical error or model bias                     |
| $\hat{E}_0$                    | ZNE-extrapolated zero-noise estimate                |
| $A, B, C$                      | exponential fitting parameters                      |
| $\mathbf{x}_p$                 | feature vector for patch $p$                        |
| $\mathbf{y}_p$                 | ideal output vector for patch $p$                   |
| $\theta_p$                     | model parameters for patch $p$                      |
| $\mathcal{F}_p$                | regression model for patch $p$                      |
| $\langle \hat{Z}_i \rangle$    | expectation value of Pauli-Z on qubit $i$           |
| $\epsilon_{\text{noisy}}$      | $\ell_2$ error of noisy outputs                     |
| $\epsilon_{\text{ML}}$         | $\ell_2$ error of ML-mitigated outputs              |
| $N$                            | number of qubits or patches                         |
| $U$                            | unitary gate operation                              |
| $\mathcal{N}_\lambda$          | noise channel with strength $\lambda$               |

|                            |                                              |
|----------------------------|----------------------------------------------|
| $H$                        | Hadamard gate                                |
| $R_x, R_y, R_z$            | single-qubit rotation gates                  |
| $S, T$                     | phase gates                                  |
| Tr                         | trace operation                              |
| Var                        | variance                                     |
| $F(\rho, \sigma)$          | quantum state fidelity                       |
| $R^2$                      | Coefficient of Determination                 |
| Pearson $r$                | Pearson Correlation Coefficient              |
| Fidelity                   | Quantum State Fidelity ( $F(\rho, \sigma)$ ) |
| $\ell_2$                   | Euclidean Norm Error                         |
| CX                         | Controlled-X (CNOT) Gate                     |
| NISQ                       | Noisy Intermediate-Scale Quantum             |
| QEC                        | Quantum Error Correction                     |
| QEM                        | Quantum Error Mitigation                     |
| ZNE                        | Zero-Noise Extrapolation                     |
| PEC                        | Probabilistic Error Cancellation             |
| ML                         | Machine Learning                             |
| IQM 20Q Gar-net            | A specific quantum processor (20-qubit chip) |
| Qiskit                     | Quantum Information Software Kit             |
| crosstalk                  | unwanted qubit interactions                  |
| gate fidelity              | measure of quantum gate accuracy             |
| GST                        | Gate Set Tomography                          |
| MWPM                       | Minimum-Weight Perfect Matching (decoder)    |
| RF                         | Random Forest (ML model)                     |
| GNN                        | Graph Neural Network                         |
| SEM                        | Scanning Electron Microscope                 |
| LC oscillator              | Inductor-Capacitor harmonic oscillator       |
| CPTP                       | Completely Positive Trace-Preserving (map)   |
| GAD                        | Generalized Amplitude Damping                |
| $k_B$                      | Boltzmann constant                           |
| MSE                        | Mean Squared Error                           |
| $p$                        | Noise Strength Parameter                     |
| $\gamma$                   | PEC Sampling Overhead Factor                 |
| $\mathcal{E}_{\text{dep}}$ | Depolarizing Noise Channel                   |
| $\mathcal{E}^{-1}$         | Inverse Noise Channel                        |
| $U$                        | Ideal Unitary Operation                      |

|                 |                                  |
|-----------------|----------------------------------|
| $I/2$           | Maximally Mixed State            |
| Patch           | Local Qubit Subset (e.g., (0,1)) |
| $\mathcal{F}_p$ | Patch-Specific Regression Model  |
| $\theta_p$      | Model Parameters for Patch $p$   |

# CHAPTER 1

## INTRODUCTION

Quantum computation promises significant advantages over classical computing in solving certain classes of problems, such as integer factorization, quantum simulation, and combinatorial optimization. However, the practical realization of this advantage faces the challenge of unavoidable errors in physical quantum devices. The current period is what John Preskill has termed the “Noisy Intermediate-Scale Quantum” (NISQ) era (Preskill, 2018). These hardware limitations pose critical challenges to performing accurate and scalable quantum computations.

In the long term, achieving fully fault-tolerant quantum computers is considered the ultimate goal. This requires the implementation of Quantum Error Correction (QEC), which corrects errors in real time, ensuring that computations remain stable and accurate. However, these schemes typically demand gate fidelities exceeding 99.9%, low crosstalk, and hundreds to thousands of physical qubits per logical qubit. These requirements exceed the capabilities of today’s quantum hardware, where gate fidelities, connectivity, and qubit counts are still limited (Cai et al., 2023).

In light of these constraints, an alternative line of research has emerged: quantum error mitigation (QEM). Unlike error correction, which encodes and protects quantum information throughout the computation, error mitigation seeks to post-process or restructure the computation to suppress the impact of noise, without requiring additional qubit resources. Early foundational work by (Temme et al., 2017) introduced two principal QEM approaches: zero-noise extrapolation (ZNE) and probabilistic error cancellation (PEC). These methods have since been extensively developed and tested across a variety of hardware platforms.

Nevertheless, while QEM techniques have demonstrated clear improvements in shallow, low-qubit-number circuits, their scalability remains in question. PEC, for instance, requires prior noise characterization and an exponential number of samples with increasing qubit count, making it intractable for medium-sized devices (Song et al., 2019). ZNE, though more lightweight in implementation, suffers from numerical instability and fitting uncertainties as circuit depth or system size grows (Khan et al., 2024). These limitations are now well documented in the literature: (Takagi et al., 2022) and (Tsubouchi et al., 2023) both derive rigorous lower bounds showing exponential sampling cost scaling in

general Markovian noise settings, while (Qin et al., 2023) further illustrate sublinear error suppression limits even with optimal protocols.

As modern quantum processors begin to reach the 20–100 qubit scale, there is growing interest in hybrid or learning-based QEM approaches (Liao et al., 2024). These methods leverage machine learning (ML) to statistically model the relationship between noisy circuit outcomes and ideal observables, often without requiring complete noise reconstruction. This trend reflects a broader shift in NISQ-era mitigation research: from analytically structured methods toward data-driven strategies that may offer better scalability and practical deployment on noisy quantum hardware.

With the ultimate target at the IQM 20Q Garnet chip’s hardware, our goal is twofold: to evaluate the effectiveness of classical QEM methods such as ZNE, and to explore the potential of supervised ML models to learn noise-aware corrections across realistic multi-qubit settings. To this end, we first establish a theoretical framework for describing noise using Kraus and Lindblad formalisms, and then implement mitigation protocols across simulated circuits with various depths, gate configurations, and noise models. The project contributes to the ongoing effort in making scalable, near-term quantum computation more reliable by assessing the capabilities and limitations of different QEM strategies under realistic constraints.

Although all simulations are performed on Qiskit’s noise-aware simulator using realistic models, the study contribute to a deeper understanding of error mitigation in realistic superconducting architectures and inform strategies for future deployment on mid-scale quantum devices.

## CHAPTER 2

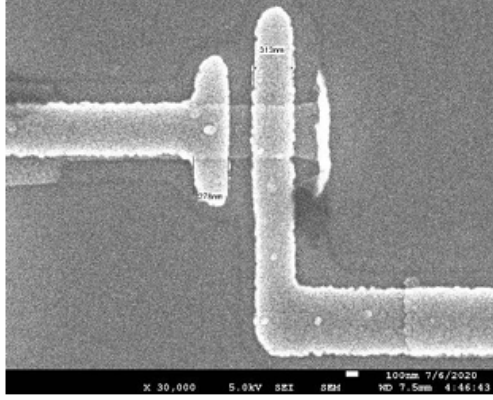
### BACKGROUND AND LITERATURE REVIEW

In this section, prior work was surveyed on quantum error mitigation (QEM) with a particular focus on its applicability to medium-scale superconducting quantum processors. This section first review the characteristics of basic superconducting qubit and our target architecture, then analyze mainstream QEM strategies—Zero Noise Extrapolation (ZNE), Probabilistic Error Cancellation (PEC), and Machine Learning-based QEM (ML-QEM)—in terms of principles, scalability, and limitations.

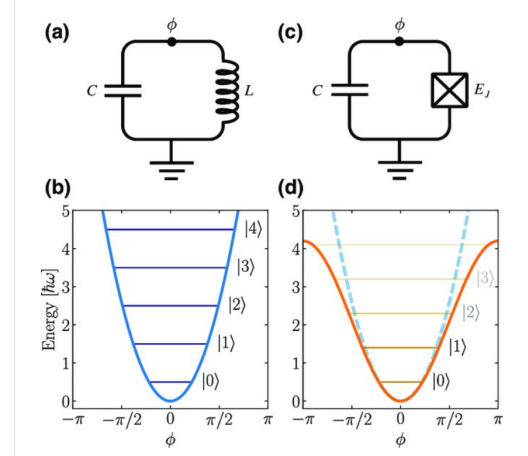
#### 2.1 Superconducting Qubits and Target 20-Qubit Superconducting Architecture

Superconducting qubits are among the most mature platforms for quantum computation, widely adopted by industrial systems such as IBM Q, Google Sycamore, and IQM’s Gar-net architecture. These qubits are realized using electrical circuits based on Josephson junctions, which introduce nonlinear inductance, called Josephson junction, essential for defining discrete and controllable energy levels (Clarke & Wilhelm, 2008; Koch et al., 2007).

With their core shown in Fig.2.1, unlike a simple harmonic oscillator, which has equally spaced energy levels, the anharmonicity in superconducting qubits provides a nonlinear inductance that breaks the harmonicity of the oscillator, ensuring uneven energy levels. The SEM image and circuit diagram of a typical superconducting qubit is shown in Fig.2.1. This property allows the qubit to isolate two specific levels, typically denoted as ground state and excited state (used as the computational  $|0\rangle$  and  $|1\rangle$  states), to act as the computational basis states.



(1)



(2)

Figure 2.1: **Josephson junction and energy levels**

(1) The SEM image showcases the physical realization of a Josephson junction.

(2) Equivalent circuit models and energy spectra:

(a) LC oscillator with harmonic potential

(b) Equally spaced energy levels

(c) Josephson junction creates anharmonicity

(d) Unequal level spacing enables qubit control

(Blue dashed curve: harmonic approximation)

The effective Hamiltonian of the circuit can be derived by quantizing the phase variable  $\phi$  across the junction. The potential energy of the circuit is described by the Josephson potential:

$$U(\phi) = -E_J \cos(\phi), \quad (1)$$

where  $\phi$  is the phase difference across the junction, and  $E_J = \frac{\hbar I_c}{2e}$  is the Josephson energy, with  $I_c$  being the critical current of the junction.

The Hamiltonian of a superconducting qubit can be approximated as:

$$H = 4E_C n^2 - E_J \cos \varphi, \quad (2)$$

where  $E_C = \frac{e^2}{2(C_S + C_J)}$  is the charging energy,  $E_J$  is the Josephson energy,  $n$  is the number of Cooper pairs, and  $\varphi$  is the superconducting phase difference across the Josephson junction.

By expanding the cosine term to second order around its minimum, the Hamiltonian approximates that of a quantum harmonic oscillator:

$$H \approx 4E_C n^2 + \frac{1}{2} E_J \varphi^2. \quad (3)$$

The energy levels of this harmonic oscillator are equidistant. However, the higher-order terms from the cosine expansion introduce anharmonicity, leading to non-equidistant energy levels:

$$E_m = m\hbar\omega - \frac{\alpha}{2}m(m-1), \quad (4)$$

where  $m$  is the energy level index,  $\omega$  is the qubit frequency, and  $\alpha$  is the anharmonicity. The anharmonicity is defined as:

$$\alpha = E_{21} - E_{10}, \quad (5)$$

where  $E_{21}$  and  $E_{10}$  are the energy differences between the first two transitions.

This anharmonicity is essential for preventing leakage into higher energy levels during qubit operations. It enables precise control of the  $|0\rangle \leftrightarrow |1\rangle$  transition, which is the basis for quantum gates. However, the anharmonicity also makes these qubits sensitive to environmental noise, such as electromagnetic interference and thermal fluctuations, which must be carefully mitigated to achieve high-fidelity operations.

The operational reliability of superconducting qubits is characterized by two coherence times:  $T_1$  (energy relaxation time) and  $T_2$  (dephasing time). These parameters are influenced by noise from material defects, stray electromagnetic fields, and control line imperfections (Kjaergaard et al., 2020). The relationship between them is:

$$\frac{1}{T_2} = \frac{1}{2T_1} + \frac{1}{T_\phi}, \quad (6)$$

where  $T_\phi$  is the pure dephasing time caused by low-frequency noise. These coherence limitations directly impact the fidelity of quantum gates and the accumulation of noise in multi-qubit circuits.

### Target Device: 20-Qubit Superconducting Architecture

This work focuses on a representative 20-qubit superconducting architecture as our target platform, the IQM 20Q Garnet chip. The figure.2.2 can see the key components: qubits in red area are coupled through bus resonators that enable interactions during multi-qubit gates. Qubit drive lines control the state of individual qubits, while readout resonators measure the states of the qubits via shared readout lines. Readout lines further allow information from multiple readout resonators to be extracted through a common path.

The performance of the IQM 20Q Garnet chip is limited by various noise sources inherent to the qubit design and environmental interactions. These errors arise from interactions with the environment and limitations in the hardware. Amplitude damping and dephasing,

for example, are sensitive to environmental factors such as electromagnetic fluctuations and low-frequency flux noise, which can induce energy relaxation and coherence loss. The qubit drive lines and readout resonators, essential for qubit state manipulation and measurement, can also introduce imperfections in signal delivery, contributing to depolarizing noise and exacerbating thermal fluctuations, leading to generalized amplitude damping. By mitigating various noise sources in the design, this project aims to achieve higher fidelity operations and establish a foundation for implementing advanced noise mitigation strategies.

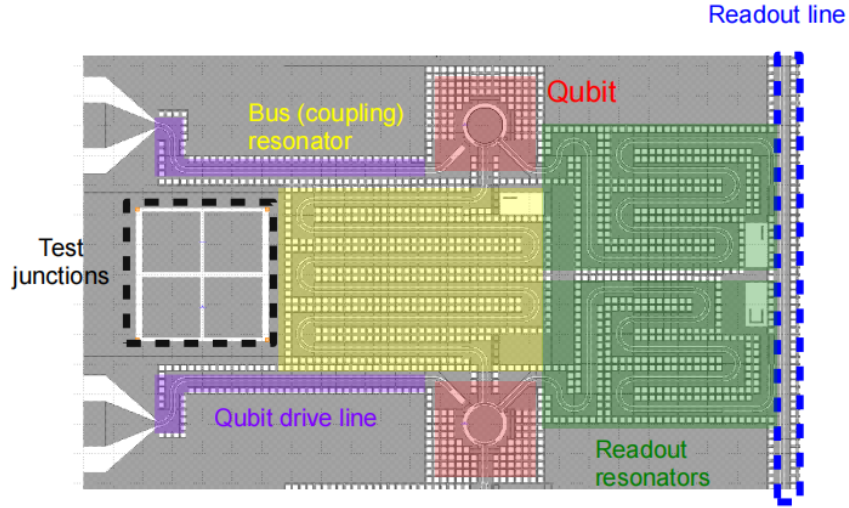


Figure 2.2: **Schematic of the IQM 20Q Garnet architecture:** Red blocks represent qubits; horizontal bus resonators enable entangling gates; readout and control lines are shared across multiple elements.

Given the severe limitations imposed by such noise processes on circuit depth and fidelity, especially in mid-scale superconducting devices, quantum error mitigation strategies—an emerging class of techniques designed to enhance computational accuracy without requiring full fault tolerance—are now reviewed in detail.

## 2.2 Quantum Error Mitigation Strategy

### 2.2.1 Quantum Error Correction

Quantum error correction (QEC)(Nielsen & Chuang, 2010) is a foundational framework that enables fault-tolerant quantum computation by actively detecting and correcting errors during computation. Physical qubits are inherently noisy, suffering from bit-flip ( $|0\rangle \leftrightarrow |1\rangle$ ) and phase-flip ( $|+\rangle \leftrightarrow |-\rangle$ ) errors due to interactions with their environment, such

as amplitude damping, dephasing, or crosstalk. Without mitigation, these errors rapidly accumulate and corrupt quantum information.

To counteract this, QEC encodes a logical qubit into multiple physical qubits. A simple example is the three-qubit repetition code, which protects against bit-flips via redundancy. More advanced codes, like the Shor code and Steane code, correct both bit-flip and phase-flip errors using a combination of stabilizer measurements and classical decoding. These codes follow the stabilizer formalism, wherein the quantum state is projected onto a subspace stabilized by a commuting set of Pauli operators. Any deviation from this subspace indicates the occurrence and location of an error.

**The Surface Code: A Scalable QEC Architecture** Among all error-correcting codes, the *surface code* is widely regarded as the most scalable and hardware-friendly scheme for near-term fault-tolerant quantum computation (Devitt et al., 2013; Fowler et al., 2012). As shown in figure 2.3, it arranges physical qubits on a two-dimensional lattice, where local stabilizer measurements detect both bit-flip and phase-flip errors. Its geometric locality makes it compatible with superconducting qubit architectures, which often feature planar layouts.

The surface code defines two types of stabilizers:

- **Vertex operators** (or  $X$ -type) detect bit-flip errors:

$$A_v = \prod_{j \in \text{star}(v)} X_j$$

- **Plaquette operators** (or  $Z$ -type) detect phase-flip errors:

$$B_p = \prod_{j \in \text{boundary}(p)} Z_j$$

A surface code of lattice size  $d \times d$  defines a single logical qubit using  $d^2$  physical qubits, achieving fault-tolerance thresholds near 1% error rates under realistic noise. Errors are identified by measuring stabilizer eigenvalues, known as the *syndrome*, and are corrected using classical decoding algorithms such as Minimum-Weight Perfect Matching (MWPM) or machine learning-based decoders.

Despite its theoretical elegance, full-fledged QEC remains out of reach for current noisy intermediate-scale quantum (NISQ) devices. Implementing surface code error correction requires:

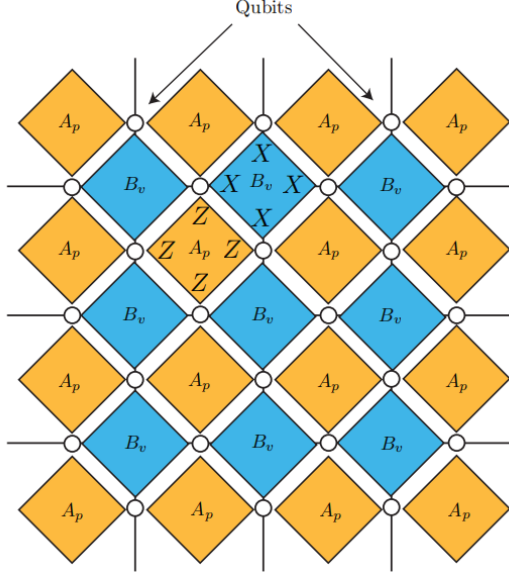


Figure 2.3: **Structure of the surface code:** Qubits are positioned on edges, with vertex ( $X$ -type) and plaquette ( $Z$ -type) stabilizers defined on nodes and faces respectively. Local measurements detect and localize bit-flip and phase-flip errors (Devitt et al., 2013).

- Hundreds to thousands of high-fidelity physical qubits per logical qubit;
- Fast and repeated mid-circuit stabilizer measurements;
- Real-time feedback and decoding.

These demands exceed the capabilities of today’s quantum hardware, where gate fidelities, connectivity, and qubit counts are still limited. Moreover, even basic QEC implementations often require hardware-level customizations not yet supported on general-purpose quantum platforms like IBM Q or Rigetti.(Murali et al., 2019)

Consequently, quantum error mitigation (QEM) has emerged as a practical alternative for near-term quantum computation. Unlike QEC, QEM does not require redundancy or mid-circuit operations, but instead relies on statistical and algorithmic techniques to reduce noise effects during post-processing. This study focus on QEM as a more viable strategy for improving the accuracy of computations on 20-qubit superconducting platforms. The following sections review existing QEM techniques, including zero noise extrapolation (ZNE), probabilistic error cancelation (PEC), and machine learning-based approaches, along with their strengths and limitations in realistic settings.

### 2.2.2 Foundational

Two foundational QEM techniques—Zero-Noise Extrapolation (ZNE) and Probabilistic Error Cancellation (PEC)—were introduced by Temme (Temme et al., 2017). ZNE estimates the zero-noise limit of an observable by executing the same circuit at scaled noise levels and performing Richardson extrapolation or some other fitting models to infer the ideal result. PEC, in contrast, performs quasiprobabilistic resampling over a reconstructed inverse noise model, effectively canceling noise through statistical inversion of quantum channels.

These methods are notable for requiring no additional qubits or error-correcting codes. Because it operates at the level of circuit scheduling and measurement post-processing, it can be deployed on virtually any quantum platform without requiring modifications to the device or extensive noise characterization. Experimental studies on real quantum devices, such as those by Khan (Khan et al., 2024), have demonstrated that ZNE can significantly improve measurement accuracy, especially in shallow circuits and for simple observables.

However, both ZNE and PEC suffer from fundamental scalability issues. As originally noted by Temme (Temme et al., 2017), the sampling cost for accurate extrapolation or cancellation grows exponentially with the number of qubits and circuit depth. Theoretical studies by Takagi et al. (Takagi et al., 2022) and Tsubouchi et al. (Tsubouchi et al., 2023) have rigorously demonstrated that the variance of mitigated observables grows exponentially under general Markovian noise. Even in favorable settings like global depolarizing noise, they showed that extrapolation becomes prohibitively expensive without strong noise structure assumptions. From a statistical standpoint, (Qin et al., 2023) demonstrated that while mitigation reduces bias, it does not eliminate variance: the residual error after extrapolation scales sublinearly as  $O(\epsilon_0 N^\gamma)$ , with  $\gamma \approx 0.5$ , where  $N$  is the gate count. This behavior, rooted in the law of large numbers, implies diminishing returns for deeper circuits even under ideal extrapolation.

In the context of this project, these scalability issues are particularly salient. Our simulations confirm that while ZNE can recover expectation values with reasonable accuracy for circuits involving fewer than 10 qubits or with depth less than 15 layers, its performance degrades rapidly beyond this regime. For instance, when applied to 20-qubit random circuits with moderate depth, the extrapolated values exhibit strong fluctuations, and fitting becomes unreliable regardless of the extrapolation order. In some cases, the extrapolated results even fall outside the physically plausible range (e.g.,  $[-1, 1]$ ) for Pauli expectation values), further highlighting the numerical instability.

In practice, additional complications arise due to the assumption that artificial noise scaling—such as repeated gate insertions—faithfully reflects the underlying logical noise. This assumption breaks down in real devices where noise is non-static, context-dependent, and may include correlated crosstalk or drift over time. For example, hardware studies such as those by Khan et al. (Khan et al., 2024) have reported non-monotonic or erratic extrapolation behavior in mid-depth circuits, attributable to such non-idealities.

Probabilistic error cancellation offers theoretical advantages: it can, in principle, exactly invert Markovian noise processes. However, PEC requires complete knowledge of the device’s noise profile, typically obtained via gate set tomography (GST) or randomized benchmarking. Song et al. (Song et al., 2019) demonstrated PEC on 1- and 2-qubit superconducting circuits, successfully reconstructing noise channels and performing quasiprobability resampling. While error suppression was achieved, even these minimal implementations required considerable sampling resources and calibration overhead. The sampling cost of PEC also grows exponentially with system size, making its application to circuits with more than 10 qubits practically infeasible.

Given that our target system is a 20-qubit processor with only simulated access and no detailed noise characterization, PEC is not a feasible choice for this study. PEC thus was excluded from our simulations and focused instead on scalable alternatives.

These theoretical and practical insights underscore the need for alternative or hybrid QEM strategies like combining with readout error mitigation, dynamical decoupling, or machine learning. (Cai et al., 2023)) This motivates the core of our study: exploring the use of supervised machine learning for quantum error mitigation in 20-qubit settings. Unlike analytical methods, ML-based QEM learns the statistical relationship between noisy and ideal outputs directly from data, offering potential scalability and generalization to complex, hardware-specific noise that traditional methods struggle to address.

A comprehensive review by Cai (Cai et al., 2023) categorizes QEM techniques into three primary classes: extrapolation-based (e.g., ZNE), probabilistic (e.g., PEC), and machine learning-based approaches. The review emphasizes that while PEC is theoretically optimal for certain noise models, its practical scalability is hindered by the exponential growth in required samples and noise characterization complexity. In contrast, ZNE requires less overhead but becomes unreliable in deeply layered or structurally irregular circuits. Importantly, the authors highlight machine learning (ML) as a potentially scalable QEM strategy, owing to its ability to learn from statistical correlations in data without requiring explicit knowledge of the full noise model.

### 2.2.3 Machine Learning Approaches to Quantum Error Mitigation

Motivated by the limitations of conventional QEM methods, machine learning-based quantum error mitigation (ML-QEM) has emerged as a promising alternative, particularly for large-scale or structurally irregular quantum circuits. A prominent framework in this direction is proposed by Liao et al. (Liao et al., 2024), who introduce a general ML-QEM pipeline consisting of three stages: encoding, execution, and prediction (Fig.2.4). In the encoding stage, circuit-level features (e.g., gate types, depth, topology) and hardware-specific metadata (e.g.,  $T_1$ ,  $T_2$ , gate fidelities) are extracted as input to a learning model. During execution, the circuit is run on a noisy quantum processor to collect observables such as expectation values  $\langle \hat{O} \rangle_{\text{noisy}}$ . A supervised ML model—typically a random forest or neural network—is then trained to predict the ideal outcome  $\langle \hat{O} \rangle_{\text{target}}$ , using noiseless simulations of small-scale circuits as reference labels. Once trained, this model can mitigate errors without further circuit repetitions, significantly reducing runtime and calibration overhead. Their results show that random forest (RF) models offer a favorable trade-off between accuracy and efficiency, outperforming ZNE in several benchmark settings while requiring fewer resources.

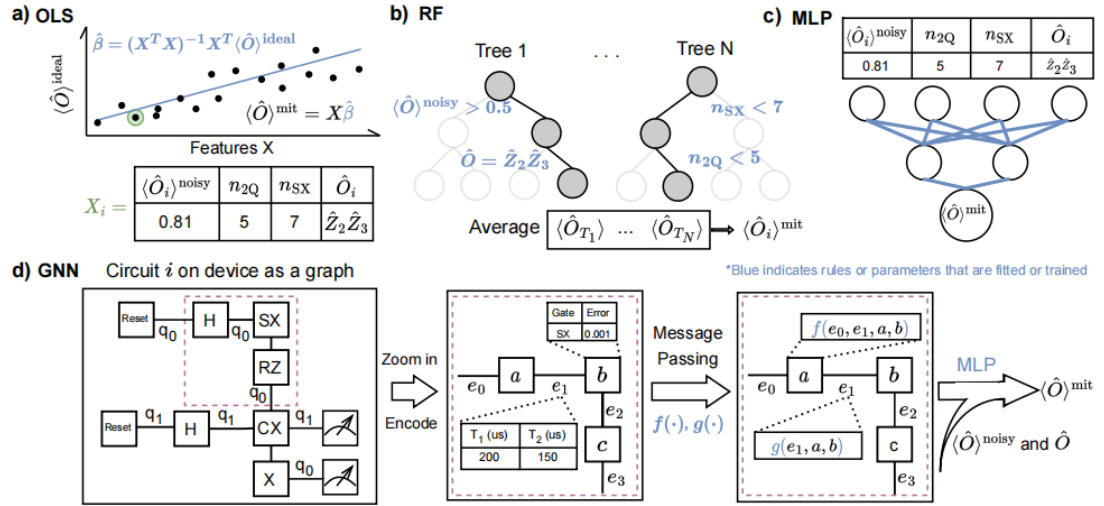


Figure 2.4: **ML-QEM Framework:**

A quantum circuit is encoded into circuit and hardware features. These, along with the noisy expectation value  $\langle \hat{O} \rangle_{\text{noisy}}$ , are input into a trained machine learning model, which predicts the ideal outcome  $\langle \hat{O} \rangle_{\text{mit}}$ . The model is trained on small circuits with known  $\langle \hat{O} \rangle_{\text{target}}$ .

Our work follows this direction but with additional constraints: since ideal data from large-scale circuits are unavailable, a local-to-global strategy inspired by Czarnik et al. (Czarnik et al., 2021) was adopted. Specifically, this study trained a regression model on small-scale circuits whose ideal results can be efficiently simulated under a realistic noise model.

These are then used to train regression models (e.g., random forest) and deploy them on structurally similar large circuits. This modular strategy bypasses the need for ideal outputs on large test circuits while still capturing key noise behaviors. Such a decoupling of training and inference also enhances portability across devices with similar characteristics.

Other notable contributions reinforce the growing diversity of ML-QEM strategies Kim (Kim et al., 2022) applied neural networks to mitigate readout errors by learning nonlinear mappings from raw measurement outcomes to ideal bitstrings. Lee and Park (Lee & Park, 2023) introduced transfer learning techniques to extend trained models from 7- to 13-qubit systems, relying on conditional independence assumptions. Meanwhile, Srushti et al. (Patil et al., 2025) proposed a graph neural network (GNN) approach that encodes hardware topology and gate noise into graph structures, enabling generalization across different device layouts using shallow subcircuits for training.

Together, these works illustrate a unifying theme: ML-QEM methods exploit local structure, statistical regularities, and minimal supervision to overcome the cost barriers of traditional techniques. They do so by learning mappings either from raw data (e.g., observables) or from engineered features, avoiding the need for explicit inversion of quantum channels.

This study prioritize interpretability and practicality. Rather than mimicking prior mitigation results (as done in Liao’s mimicry variant), our models are trained to directly predict ideal observables from noisy outputs and encoded circuit/device features. This reduces reliance on auxiliary methods like ZNE or PEC and enables simpler deployment across simulated architectures. Furthermore, by selecting small, classically simulable subcircuits for training, this study ensure compatibility with current NISQ-era simulators while laying the groundwork for extension to real hardware.

## CHAPTER 3

### QUANTUM NOISE MODEL

Quantum error mitigation strategies, such as those discussed above, rely fundamentally on the statistical relationship between noisy and ideal quantum measurements. To model and mitigate these errors effectively—especially in the absence of full device calibration—one must understand the underlying structure of quantum noise. The primary noise sources affecting the IQM 20Q chip include amplitude damping, generalized amplitude damping, dephasing, and depolarizing noise. These errors are often described by noise channels, which can be characterized using either Kraus operators or Lindblad master equations, depending on the nature of the system.

#### 3.1 Open Quantum Systems and Dynamical Map

For open quantum systems, the environment (or reservoir) interacts with the system, leading to the exchange of quantum information—such as populations, coherence, and correlations—between the two, as shown in Figure 3.1. This interaction causes:

- Dissipative effects: Quantum information, including energy and coherence, is irreversibly lost from the system to the reservoir.
- Non-Markovian effects: Part of this lost information temporarily flows back from the reservoir to the system, restoring coherence or reducing entropy. In this work, we adopt the information backflow definition of non-Markovianity.

After such an interaction, the pure initial state of a system transitions to a statistical mixture of pure states with respective probabilities. This transition, where a pure state evolves into a mixed state, is known as *decoherence*.

To analyze the dynamics of an open quantum system, one considers it as a subsystem of a larger closed system  $S + B$ , where  $B$  represents the environment (or bath) adjoined to the principal system  $S$ . The total Hilbert space is given by  $\mathcal{H} = \mathcal{H}_S \otimes \mathcal{H}_B$ . The initial state of the total system is assumed to be separable:

$$\rho_{SB} = \rho_S \otimes \rho_B, \quad (7)$$

which is an important assumption for the subsequent CPTP map. The purity is without

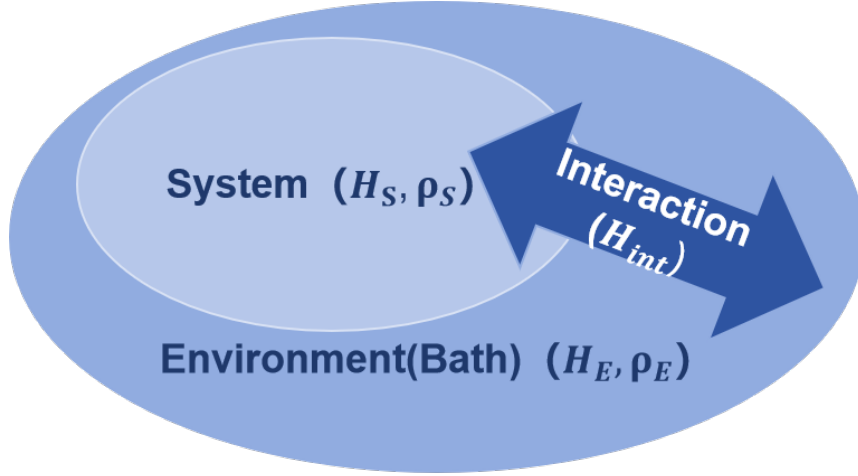


Figure 3.1: Open System Map

loss of generality because one can always purify the environmental state by introducing an even bigger environment and this does not affect the dynamics of the principal system itself.

Since the total system is closed, the total system evolves unitarily:

$$\rho(t) = U_{SB}(\rho_S \otimes \rho_B)U_{SB}^\dagger.$$

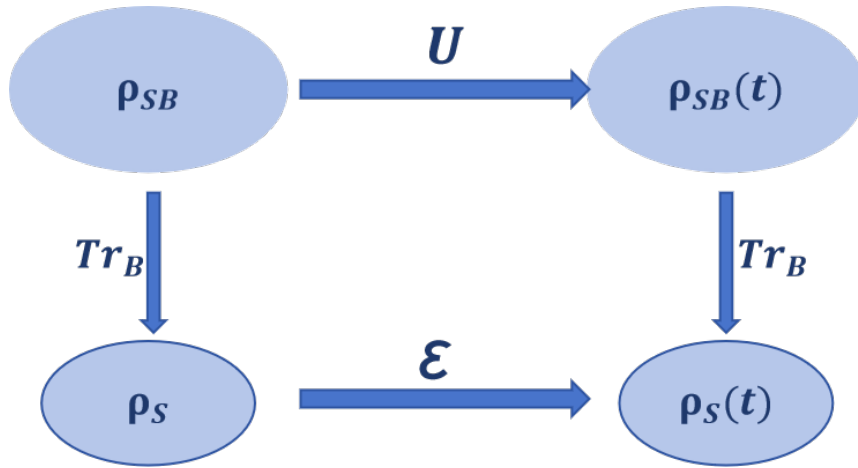


Figure 3.2: System Dynamic Map

To obtain the evolution of the principle system  $S$  only, the environmental degrees of freedom will be traced out and then the reduced density matrix of the system can be represented by the *Kraus operator-sum representation*:

$$\rho_S(t) = \text{Tr}_B(U(\rho \otimes |e_B\rangle \langle e_B| U^\dagger)) = \sum_n K_n \rho K_n^\dagger.$$

The Kraus operators  $K_n$  acting on the system's Hilbert space  $\mathcal{H}_S$  are defined as  $K_n = \langle k_B | U_{SB} | e_B \rangle$  where  $\{|k_B\rangle\}$  is an orthogonal basis for the environment's Hilbert space  $\mathcal{H}_B$ , and  $|e_B\rangle$  is the environment's initial state.

As the fig.3.2, a quantum channel  $\mathcal{E}$  describes the principle system's evolving under the influence of the environment. It is represented as a completely positive trace-preserving (CPTP) map:

$$\mathcal{E}(\rho) = \sum_n K_n \rho K_n^\dagger. \quad (8)$$

where the Kraus operators encode the non-unitary effects of environmental interactions on the system, including decoherence and dissipation. And to ensure that the evolution of the open system is physically valid, the Kraus operators must satisfy two key conditions:

1. *Trace preservation*: The map preserves the trace of the density matrix, ensuring that probabilities remain normalized:

$$\sum_n K_n^\dagger K_n = \mathbb{I}, \quad (9)$$

where  $\mathbb{I}$  is the identity operator.

2. *Complete positivity*: The map preserves positivity not only for the system's state but also for any larger composite system. Physically, if  $\mathcal{E}$  is applied on a subsystem, it produces a semi-positive definite physical state. Mathematically, for any extended Hilbert space  $\mathcal{H}_S \otimes \mathcal{H}_E$ , the map  $\mathcal{E} \otimes \mathbb{I}_E$  is positive:

$$\mathcal{E} \otimes \mathbb{I}_E(\rho_{SE}) \geq 0, \quad \forall \rho_{SE} \geq 0. \quad (10)$$

### 3.1.1 Von Neumann equation from Kraus representation

(Chia et al., 2016) The evolution of closed quantum systems is governed by the Von Neumann equation, which is a generalization of the Schrödinger equation for mixed states. While the Schrödinger equation describes the evolution of pure states, the Von Neumann equation captures the time evolution of the density matrix of the system,  $\rho(t)$ , under a Hamiltonian  $H$ .

Starting from Kraus evolution over an infinitesimal time interval  $\Delta t$ :

$$\rho(t + \Delta t) = \sum_n K_n(\Delta t) \rho(t) K_n^\dagger(\Delta t),$$

where  $\sum_n K_n^\dagger(\Delta t) K_n(\Delta t) = \mathbb{I}$  and assume that there is only one Kraus operator given by a unitary matrix.

**Unitary Evolution (First Order):** For a closed system, the presence of only one Kraus operator is assumed, represented by a unitary matrix:

$$K_0(\Delta t) = U(\Delta t) = e^{-iH\Delta t} \approx \mathbb{I} - iH\Delta t + \mathcal{O}(\Delta t^2),$$

Substituting this into the Kraus representation:

$$\rho(t + \Delta t) = K_0 \rho(t) K_0^\dagger = \rho(t) - i\Delta t [H, \rho(t)],$$

Dividing by  $\Delta t$  and taking the limit to zero, we arrive at the Von Neumann equation:

$$\dot{\rho}(t) = -i[H, \rho(t)].$$

## 3.2 Master Equation in Lindblad Formalism

While the Von Neumann equation provides a foundation for understanding the evolution of closed systems, real-world quantum systems are rarely isolated. Interactions with the environment introduce noise and dissipation, requiring a more general framework to describe the non-unitary evolution of open systems.

The dynamics of open quantum systems can be effectively described by the master equation, which provides a framework for understanding both coherent and dissipative effects. This section explores the master equation in the context of the Lindblad formalism, beginning with the Markov approximation, followed by a brief discussion of the Kraus representation as a foundation for the Lindblad form.

### 3.2.1 Markov approximation

(OpenCourseWare, 2012) In many physical scenarios, the interaction between the principle system and its environment is weak, and the environment's correlation time is very short. This allows the use of the *Markov approximation*, which assumes that the system's evolution is memoryless and local in time. Under this approximation, the environment rapidly "forgets" the information it acquires from the system, preventing any significant feedback. This condition holds if the correlation time of the environment,  $\delta t_E$ , is much shorter than the coarse-grained time scale  $\delta t_{\text{coarse}}$  over which the system's evolution is observed, which in turn must be much shorter than the system's damping time  $\delta t_{\text{damp}}$ , i.e.,  $\delta t_E \ll \delta t_{\text{coarse}} \ll \delta t_{\text{damp}}$ . When this separation of time scales is satisfied, the system's evolving can be described by a master equation, which is a first-order differential equation

governing the time evolution of the density matrix  $\rho(t)$ :  $\frac{d\rho}{dt} \approx \frac{\rho(t+\Delta t) - \rho(t)}{\Delta t}$ . This description captures both the coherent Hamiltonian dynamics and the dissipative effects caused by the environment.

### 3.2.2 Lindblad form

The reduced density matrix  $\rho(t)$  of the system evolves according to the master equation:

$$\frac{d\rho}{dt} = \mathcal{L}[\rho],$$

where  $\mathcal{L}$  is the Liouvillian superoperator. For Markovian dynamics, the master equation takes the *Lindblad form*, a general form consistent with complete positivity and trace preservation:

$$\mathcal{L}[\rho] = -\frac{i}{\hbar}[H, \rho] + \sum_j \left( L_j \rho L_j^\dagger - \frac{1}{2} \{L_j^\dagger L_j, \rho\} \right). \quad (11)$$

Here:

- $H$  is the Hermitian Hamiltonian describing a coherent evolution of the system.
- $L_j$  are Lindblad operators representing dissipative processes due to the environment.
- $\{x, y\} = xy + yx$  denotes the anti-commutator.

### Short derivation of the Lindblad form

The total Hamiltonian of the system, including the principal system ( $S$ ), the environment ( $B$ ), and their interaction (int), is written as:

$$H_{\text{tot}} = H_S + H_B + H_{\text{int}},$$

where:

$$H_{\text{int}} = \sum_k S_k \otimes B_k,$$

and  $S_k$  are operators acting on the system's Hilbert space  $\mathcal{H}_S$ , while  $B_k$  act on the environment's Hilbert space  $\mathcal{H}_B$ .

The unitary evolution operator for the total system is expanded to the second order in  $\Delta t$ :

$$\mathcal{E}_{\text{tot}}(\Delta t) = e^{-iH_{\text{tot}}\Delta t} \approx \mathbb{I} - iH_{\text{tot}}\Delta t - \frac{1}{2}(H_{\text{tot}}\Delta t)^2 + \mathcal{O}(\Delta t^3).$$

Using the second-order expansion of  $\mathcal{E}_{\text{tot}}(\Delta t)$ , the total density matrix evolves as:

$$\begin{aligned}
\rho_{\text{tot}}(t + \Delta t) &= \mathcal{E}_{\text{tot}}(\Delta t) \rho_{\text{tot}}(t) \mathcal{E}_{\text{tot}}^\dagger(\Delta t) \\
&= \rho_{\text{tot}}(t) \\
&\quad - i\Delta t [H_{\text{tot}}, \rho_{\text{tot}}(t)] \\
&\quad - \frac{\Delta t^2}{2} (H_{\text{tot}} [H_{\text{tot}}, \rho_{\text{tot}}(t)] + [H_{\text{tot}}, \rho_{\text{tot}}(t)] H_{\text{tot}}) \\
&\quad + \mathcal{O}(\Delta t^3).
\end{aligned}$$

where each line represent the contributions to the state at time  $t$  in terms of increasing order. To study the dynamics of the principal system alone, the trace is taken over the bath:

$$\rho_S(t + \Delta t) = \text{Tr}_B [\rho_{\text{tot}}(t + \Delta t)].$$

**For the first-order term:**

$$\text{Tr}_B (-i\Delta t [H_{\text{tot}}, \rho_{\text{tot}}(t)]) = -i\Delta t \text{Tr}_B [H_S + H_B + H_{\text{int}}, \rho_{\text{tot}}(t)] = -i\Delta t [H_{\text{eff}}, \rho_S].$$

For a large environment, one expects its state to be close to a thermal state  $\rho_B = \frac{1}{Z} e^{-H_B/kT}$  and to be decorrelated from the principal system, i.e.  $\rho_{\text{tot}}(t)$  approx  $\rho_{S(t)}$  tensor  $\rho_B$ . In such a case  $[H_B, \rho_B] = 0$  and  $H_{\text{eff}} = \dots$  (Details omitted here, as this term simply encodes the static average effect of the environment.)

Here,  $\text{Tr}_B [H_B, \rho_{\text{tot}}] = 0$  assuming no initial correlations, and the remaining effective system Hamiltonian becomes  $H_{\text{eff}} = H_S + \sum_k \beta_k S_k$ , where  $\beta_k = \text{Tr}_E [B_k \rho_E]$ .

**For the second-order term:**

$$\text{Tr}_B \left( -\frac{\Delta t^2}{2} [H_{\text{tot}} [H_{\text{tot}}, \rho_{\text{tot}}(t)]] \right) = -\frac{\Delta t^2}{2} \text{Tr}_B [H_{\text{int}}^2 \rho_{\text{int}}(t) - 2H_{\text{int}} \rho_{\text{int}}(t) H_{\text{int}}^\dagger + \rho_{\text{int}}(t) H_{\text{int}}^2] \quad (12)$$

$$= \Delta t^2 \sum_k \left( L_k \rho_S L_k^\dagger - \frac{1}{2} \{L_k^\dagger L_k, \rho_S\} \right) \quad (13)$$

Weak coupling and the Markov approximation permit the omission of the second-order term  $\text{Tr}_B \left( -\frac{\Delta t^2}{2} [H_S, [H_S, \rho_S]] \right) = 0$ . Coarse-graining over  $\delta t_{\text{coarse}} \gg \delta t_E$  further enforces rapid decay of reservoir correlations, eliminating memory effects ( $\text{Tr}_B [B_k'^\dagger(t') B_k(t) \rho_B] \propto \delta_{kk'} \delta(t - t')$ ).

As a result, the interaction term will be:

$$H_{\text{int}}^2 = H_{\text{int}}^\dagger H_{\text{int}} = \left( \sum_k S_{k'}^\dagger \otimes B_{k'}^\dagger \right) \left( \sum_k S_k \otimes B_k \right) = \sum_{k,k'} (S_{k'}^\dagger S_k \otimes B_{k'}^\dagger B_k).$$

Combining this result with the first-order term,

$$\rho_S(t + \Delta t) = \rho_S(t) - i\Delta t[H_{\text{eff}}, \rho_S] + \Delta t^2 \sum_k \left( L_k \rho_S L_k^\dagger - \frac{1}{2} \{L_k^\dagger L_k, \rho_S\} \right),$$

For the entire evolution:

$$\rho(t + \Delta t) = K_0 \rho(t) K_0^\dagger + \sum_k K_n \rho K_n^\dagger,$$

Assume the no-jump evolution  $K_0$  corresponding to the coherent part of the evolution

$$K_0 = \mathbb{I} + \Delta t(-iH + K) + \mathcal{O}(\Delta t^2),$$

where  $H = H_S$  and  $K = -\frac{1}{2} \sum_k L_k^\dagger L_k$ , with  $L_k$  being the Lindblad operators describing transitions due to reservoir interactions.

The jump evolution  $K_n$  for higher order ( $n > 0$ ) describe dissipative processes:  $K_n = \sqrt{\Delta t} L_k + \mathcal{O}(t^{3/2})$

Taking the limit  $\Delta t \rightarrow 0$ , the Lindblad equation is obtained, consisting of the unitary term and dissipative term:

$$\dot{\rho}(t) = \mathcal{L}[\rho] = -i[H, \rho(t)] + \sum_k \left( L_k \rho(t) L_k^\dagger - \frac{1}{2} \{L_k^\dagger L_k, \rho(t)\} \right).$$

### 3.3 Examples for Quantum Noise Models

The following sections explore four fundamental noise models(Havel, 2003)(Nielsen & Chuang, 2010): amplitude damping, generalized amplitude damping, phase damping, and depolarizing noise. These models describe common physical noise processes. The noise process channel is represented as a completely positive trace-preserving (CPTP) map:

$$\mathcal{E}(\rho) = \sum_n K_n \rho K_n^\dagger,$$

where  $K_n$  are the Kraus operators satisfying completeness condition:  $K_0^\dagger K_0 + K_1^\dagger K_1 = \mathbb{I}$

### 3.3.1 Amplitude Damping

Amplitude damping is a model that describes energy loss from a quantum system to its environment. For example, it models processes like spontaneous emission, where a quantum system transitions from the excited state  $|1\rangle$  to the ground state  $|0\rangle$ .

$$|\psi_1\rangle \rightarrow |\psi_0\rangle \quad (\text{e.g., decay } |1\rangle \rightarrow |0\rangle)$$

**System-Environment Interaction:** Let the system be coupled to the environment in the initial state  $|0_E\rangle$ . The joint evolution under a operation  $\mathcal{E}$  is:

$$\begin{aligned}\mathcal{E}|0\rangle_S|0\rangle_E &= |0\rangle_S|0\rangle_E \\ \mathcal{E}|1\rangle_S|0\rangle_E &= \sqrt{1-\gamma}|1\rangle_S|0\rangle_E + \sqrt{\gamma}|0\rangle_S|1\rangle_E,\end{aligned}$$

where  $\gamma$  is the probability of decay during the interaction.

**Kraus Operators:** To describe the system's reduced dynamics, project the environment's state and trace it out. The Kraus operators satisfying the completeness relation are:

$$K_0 = \langle 0_E | \mathcal{E} | 0_E \rangle = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, \quad K_1 = \langle 1_E | \mathcal{E} | 0_E \rangle = \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}.$$

**Master Equation:** Assuming  $\gamma = \Gamma \Delta t$ , where  $\Gamma$  is the decay rate, the following approximation is made for small  $\Delta t$ :

$$\begin{aligned}K_0 &= \sigma_- \sigma_+ + \sqrt{1-\gamma} \sigma_+ \sigma_- \approx \mathbb{I} - \frac{\Gamma \Delta t}{2} \sigma_+ \sigma_-, \\ K_1 &\approx \sqrt{\Gamma \Delta t} \sigma_-\end{aligned}$$

where  $\sigma_+ = |1\rangle\langle 0|$  and  $\sigma_- = |0\rangle\langle 1|$ . The dynamic of the density matrix  $\rho$  is:

$$\rho(t + \Delta t) = \sum_k K_k \rho(t) K_k^\dagger.$$

Expanding to first-order in  $\Delta t$ , the Lindblad master equation is derived:

$$\frac{d\rho}{dt} = \Gamma \left( \sigma_- \rho \sigma_+ - \frac{1}{2} \{ \sigma_+ \sigma_-, \rho \} \right)$$

### Generalized Amplitude Damping (GAD)

The Generalized Amplitude Damping (GAD) channel models the interaction of a two-level quantum system (qubit) with a thermal bath at finite temperature. It captures both energy relaxation and thermal excitation, making it particularly useful for systems where the environment has a non-zero temperature. (de Oliveira et al., 2020)

**Kraus Operator for GAD channels** The generalized amplitude damping (GAD) channel can be described by considering two distinct probabilities:

- Probability  $1 - \xi$  (ground-state dominance):

$$\begin{aligned}\mathcal{E}|0\rangle_S|0\rangle_E &= |0\rangle_S|0\rangle_E, \\ \mathcal{E}|1\rangle_S|0\rangle_E &= \sqrt{1-\gamma}|1\rangle_S|0\rangle_E + \sqrt{\gamma}|0\rangle_S|1\rangle_E.\end{aligned}$$

- Probability  $\xi$  (excited-state dominance):

$$\begin{aligned}\mathcal{E}|1\rangle_S|1\rangle_E &= |1\rangle_S|1\rangle_E, \\ \mathcal{E}|0\rangle_S|1\rangle_E &= \sqrt{1-\gamma}|0\rangle_S|1\rangle_E + \sqrt{\gamma}|1\rangle_S|0\rangle_E.\end{aligned}$$

The interaction of the qubit with a thermal bath is characterized by the thermal population  $\xi$  of the excited state  $|1\rangle$  and the ground state  $|0\rangle$ . The Kraus operators, satisfying the completeness relation  $\sum_{i=0}^3 K_i^\dagger K_i = \mathbb{I}$ , for the GAD channel are defined as follows:

$$\begin{aligned}\hat{K}_0 &= \sqrt{1-\xi} \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\gamma} \end{pmatrix}, & \hat{K}_1 &= \sqrt{1-\xi} \begin{pmatrix} 0 & \sqrt{\gamma} \\ 0 & 0 \end{pmatrix}, \\ \hat{K}_2 &= \sqrt{\xi} \begin{pmatrix} 0 & 0 \\ \sqrt{1-\gamma} & 0 \end{pmatrix}, & \hat{K}_3 &= \sqrt{\xi} \begin{pmatrix} \sqrt{\gamma} & 0 \\ 0 & 0 \end{pmatrix}.\end{aligned}$$

Here:

- $\gamma = 1 - e^{-\lambda t}$  is the decay probability over time  $t$ , where  $\lambda$  is the decay constant.
- $\xi \in [0, 0.5]$  is the thermal population of the excited state, defined as:

$$\xi = \frac{1}{1 + e^{\beta\varepsilon}},$$

where  $\beta = \frac{1}{k_B T}$ ,  $k_B$  is the Boltzmann constant, and energy gap  $\varepsilon = \varepsilon_2 - \varepsilon_1$ .

**Master Equation** The dynamics of the GAD channel can also be expressed in the Lindblad form of the master equation:

$$\frac{d\rho}{dt} = -i[H, \rho] + \sum_k (L_k \rho L_k^\dagger - \frac{1}{2}\{L_k^\dagger L_k, \rho\}),$$

where the second term is the Lindblad dissipator, and  $L_k$  are the Lindblad operators:

$$\begin{aligned}L_1 &= \sqrt{\Gamma(1-\xi)}|0\rangle\langle 1|, \\ L_2 &= \sqrt{\Gamma\xi}|1\rangle\langle 0|,\end{aligned}$$

where  $\gamma = \Gamma \Delta t$  is the decay rate.

Substituting these operators into the master equation yields:

$$\frac{d\rho}{dt} = -i[H, \rho] + \Gamma(1 - \xi) \left( \sigma_- \rho \sigma_+ - \frac{1}{2} \{ \sigma_- \sigma_+, \rho \} \right) + \Gamma \xi \left( \sigma_+ \hat{\rho} \sigma_- - \frac{1}{2} \{ \sigma_+ \sigma_-, \rho \} \right).$$

The first term represents coherent evolution under the system Hamiltonian  $\hat{H}$ , while the remaining terms describe relaxation and thermal excitation induced by the thermal bath.

**Long-Time Limit** In the long-time limit ( $t \rightarrow \infty$ ,  $p \rightarrow 1$ ), the system relaxes to the thermal state:

$$\hat{\rho}_{\text{thermal}} = \begin{pmatrix} 1 - \xi & 0 \\ 0 & \xi \end{pmatrix}.$$

This state corresponds to thermal equilibrium with the environment.

### 3.3.2 Depolarizing Noise

Depolarizing noise describes a uniform distribution of errors, such as bit-flip, phase-flip, or both. This model leads to a complete mixing of the quantum state with the maximally mixed state.

**Kraus Operators:**

$$K_0 = \sqrt{1-p} \mathbb{I}, \quad K_1 = \sqrt{\frac{p}{3}} \sigma_x, \quad K_2 = \sqrt{\frac{p}{3}} \sigma_y, \quad K_3 = \sqrt{\frac{p}{3}} \sigma_z,$$

where  $p$  is the depolarization probability, and  $\sigma_x, \sigma_y, \sigma_z$  are Pauli matrices.

**Master Equation:**

$$\frac{d\rho}{dt} = \frac{\lambda}{3} \sum_{i=x,y,z} \left( \sigma_i \rho \sigma_i^\dagger - \frac{1}{2} \{ \sigma_i^\dagger \sigma_i, \rho \} \right).$$

This equation highlights the uniform distribution of errors across the state space.

### 3.3.3 Phase Damping

Phase damping describes the change of quantum phase between states, leading the loss of quantum coherence without energy exchange, such as random phase noise in the environment. This process does not affect the populations (diagonal elements of the density matrix) but reduces the coherence (off-diagonal elements).

**Kraus Operators:** The phase damping process can be described by Kraus operators that implement an action proportional to  $v \square [0, 1]$ , which is described as the probability of dephasing, on  $|1\rangle$  state:

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-v} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & 0 \\ 0 & \sqrt{v} \end{pmatrix}.$$

Input a general single-qubit density matrix:  $\rho_{\text{in}} = \begin{pmatrix} a & c \\ c^* & 1-a \end{pmatrix}$  the evolution under the phase damping channel becomes:

$$\rho = \begin{pmatrix} a & \sqrt{1-v}c \\ \sqrt{1-v}c^* & 1-a \end{pmatrix}$$

**Master Equation:** For small  $\Delta t$ , assuming  $v = \Upsilon \Delta t$ , where  $\Upsilon$  is the dephasing rate, we approximate  $\sqrt{1-v} \approx 1 - \frac{\Upsilon \Delta t}{2}$

The density matrix evolves as:

$$\frac{\rho(t + \Delta t) - \rho(t)}{\Delta t} = \begin{pmatrix} 0 & -\frac{\Upsilon}{2}c \\ -\frac{\Upsilon}{2}c^* & 0 \end{pmatrix}.$$

For small  $\Delta t$ , we approximate the Lindblad equation:

$$\frac{d\rho}{dt} \approx \frac{\rho(t + \Delta t) - \rho(t)}{\Delta t} = \frac{\Upsilon}{2} (\sigma_z \rho \sigma_z - \rho),$$

$\sigma_z$  is the Pauli-Z operator.

# CHAPTER 4

## RESEARCH METHODOLOGY

### 4.1 Zero Noise Extrapolation (ZNE)

#### 4.1.1 Theoretical Framework

In noisy intermediate-scale quantum (NISQ) devices, fully error-corrected computation remains infeasible due to limited qubit count and low gate fidelity. An alternative approach, Zero Noise Extrapolation (ZNE), attempts to estimate noise-free expectation values by deliberately amplifying noise and extrapolating the results to the zero-noise limit.

To understand the feasibility of this method, we begin with the Lindblad master equation, which describes Markovian open quantum system dynamics [11]:

$$\frac{d\rho(t)}{dt} = \mathcal{L}_\lambda(\rho(t)) = -\frac{i}{\hbar}[H, \rho(t)] + \sum_j \left( L_j \rho(t) L_j^\dagger - \frac{1}{2} \{L_j^\dagger L_j, \rho(t)\} \right), \quad (14)$$

where  $\mathcal{L}_\lambda$  is a superoperator encoding both the coherent evolution and decoherence effects. For common noise processes, this yields exponential decay in observable expectation values:

$$E(\lambda) = \text{Tr}[A\rho(T)] = E_0 e^{-\alpha\lambda}, \quad (15)$$

with  $E_0$  representing the ideal (noise-free) result and  $\alpha$  a model-dependent decay constant.

We now provide explicit derivations for three canonical noise channels to support this exponential ansatz.

**Depolarizing Noise** In Qiskit, the single-qubit depolarizing noise channel is defined as:

$$\mathcal{E}_\lambda(\rho) = (1 - \lambda)\rho + \lambda \frac{I}{2}, \quad (16)$$

where  $\lambda \in [0, 1]$  is the depolarizing error probability (param\_phase in Qiskit), and  $I/2$  denotes the maximally mixed state. Its Lindblad form which was derived in Eq. (3.3.2) is:

$$\frac{d\rho}{dt} = \lambda \sum_{j=1}^3 (\sigma_j \rho \sigma_j - \rho) = \lambda \mathcal{L}(\rho), \quad (17)$$

where we define the Liouvillian superoperator  $\mathcal{L}(\rho) = \sum_{j=1}^3 \sigma_j \rho \sigma_j - 3\rho$ . To explicitly analyze the decay of observable expectation values, consider a generic initial state  $\rho(0) =$

$\begin{bmatrix} \rho_{00} & \rho_{01} \\ \rho_{10} & \rho_{11} \end{bmatrix}$ . Under the Lindblad equation, the off-diagonal elements decay as  $\frac{d\rho_{01}}{dt} = \frac{d\rho_{10}^*}{dt} = -4\lambda\rho_{01}$ , and the diagonal elements evolve according to  $\frac{d\rho_{00}}{dt} = \lambda(1 - 2\rho_{00})$ ,  $\frac{d\rho_{11}}{dt} = -\frac{d\rho_{00}}{dt}$ . Solving these first-order differential equations yields:

$$\rho(t) = \begin{bmatrix} \frac{1}{2} + (\rho_{00}(0) - \frac{1}{2})e^{-2\lambda t} & \rho_{01}(0)e^{-4\lambda t} \\ \rho_{10}(0)e^{-4\lambda t} & \frac{1}{2} - (\rho_{00}(0) - \frac{1}{2})e^{-2\lambda t} \end{bmatrix}. \quad (18)$$

which implies that for coherence-sensitive observables like  $X$  and obtains:

$$\langle X(t) \rangle = \langle X(0) \rangle e^{-4\lambda t}. \quad (19)$$

**Phase Damping** The phase damping channel in Qiskit captures dephasing processes, where coherence is lost without energy relaxation. It is described by the Kraus operators:

$$A_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-b} \end{bmatrix}, \quad A_1 = \begin{bmatrix} 0 & 0 \\ 0 & \sqrt{b} \end{bmatrix},$$

where  $b \in [0, 1]$  is the dephasing probability (param\_phase in Qiskit).

Given dephasing rate  $\Upsilon = \frac{\lambda}{\Delta t}$  with the effective discrete noise strength  $\lambda$  applied per unit time step  $\Delta t$ , we compute the time evolution using the Lindblad equation of phase damping model 3.3.3. The off-diagonal elements decay exponentially:

$$\frac{d\rho(t)}{dt} = \begin{bmatrix} 0 & -\Upsilon\rho_{01}(t) \\ -\Upsilon\rho_{10}(t) & 0 \end{bmatrix} = \frac{1}{\Delta t} \begin{bmatrix} 0 & -\lambda\rho_{01}(t) \\ -\lambda\rho_{10}(t) & 0 \end{bmatrix}, \quad (20)$$

which leads to the solution:

$$\rho(t) = \begin{bmatrix} \rho_{00} & \rho_{01}e^{-\Upsilon t} \\ \rho_{10}e^{-\Upsilon t} & \rho_{11} \end{bmatrix}. \quad (21)$$

where  $\Upsilon$  is the dephasing rate. This similarly results in exponential decay of  $X$ -expectation values:

$$\langle X(t) \rangle = \langle X(0) \rangle e^{-\Upsilon t} = E_0 e^{-\alpha\lambda}. \quad (22)$$

where  $\alpha = t/\Delta t$  and  $E_0 = \langle X(0) \rangle$ . This confirms that under phase damping, the  $X$ -observable decays exponentially with respect to the noise strength  $\lambda$ .

**Amplitude Damping** In Qiskit, the amplitude damping channel is described by a set of Kraus operators modeling energy relaxation into a thermal equilibrium. For a qubit initially in state  $\rho$ , the channel acts as:

$$\mathcal{E}_{a,p_1}(\rho) = \sum_{j=0}^1 A_j \rho A_j^\dagger + \sum_{j=0}^1 B_j \rho B_j^\dagger, \quad (23)$$

where:

$$\begin{aligned} A_0 &= \sqrt{1-p_1} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-a} \end{bmatrix}, & A_1 &= \sqrt{1-p_1} \begin{bmatrix} 0 & \sqrt{a} \\ 0 & 0 \end{bmatrix}, \\ B_0 &= \sqrt{p_1} \begin{bmatrix} \sqrt{1-a} & 0 \\ 0 & 1 \end{bmatrix}, & B_1 &= \sqrt{p_1} \begin{bmatrix} 0 & 0 \\ \sqrt{a} & 0 \end{bmatrix}, \end{aligned}$$

where  $a$  is the damping strength parameter (denoted `param_amp` in Qiskit), and  $p_1$  is the excited-state population, representing the asymptotic thermal population of the  $|1\rangle$  state.

We write the density matrix in standard form by using the Lindblad master equation 3.3.1:

$$\rho(t) = \begin{bmatrix} \rho_{00}(t) & \rho_{01}(t) \\ \rho_{10}(t) & \rho_{11}(t) \end{bmatrix}.$$

Under Eq. (3.3.1), the time evolution of the matrix elements is:

$$\frac{d\rho_{00}}{dt} = \Gamma\rho_{11}, \quad \frac{d\rho_{11}}{dt} = -\Gamma\rho_{11}, \quad \frac{d\rho_{01}}{dt} = -\frac{\Gamma}{2}\rho_{01}, \quad \frac{d\rho_{10}}{dt} = -\frac{\Gamma}{2}\rho_{10}. \quad (24)$$

This implies that:

$$\rho(t) = \begin{bmatrix} 1 - \rho_{11}(0)e^{-\Gamma t} & \rho_{01}(0)e^{-\Gamma t/2} \\ \rho_{10}(0)e^{-\Gamma t/2} & \rho_{11}(0)e^{-\Gamma t} \end{bmatrix}. \quad (25)$$

The expectation value of the Pauli- $X$  operator is:

$$\langle X(t) \rangle = \langle X(0) \rangle e^{-\Gamma t/2}, \quad (26)$$

These derivations justify the use of exponential fitting in ZNE. Since all three noise models preserve populations but degrade coherence, we select  $\langle X \rangle$  as our primary observable. This quantity is sensitive to the real part of the off-diagonal density matrix elements ( $\rho_{01} + \rho_{10}$ ), making it ideal for capturing noise-induced decoherence. In contrast,  $\langle Z \rangle$  depends only on populations and is thus less informative under dephasing and amplitude damping.

**Gate Folding and Extrapolation** To implement noise amplification, we adopt the *gate folding* strategy. Any unitary operation  $U$  is replaced by a folded version:

$$U \rightarrow U \otimes \mathbb{I} = UU^\dagger U, \quad (27)$$

which is equivalent to the identity operation in the noiseless case, but introduces additional noise when applied on hardware or in simulation. For Markovian noise, each application of a gate followed by noise channel  $\mathcal{N}_\lambda$  contributes a fixed noise strength  $\lambda$ . Repeating the folded gate  $c = 2r + 1$  times therefore results in a total noise strength  $(2r + 1)\lambda$ . Repeating the folding multiple times effectively increases the noise by a scaling factor  $c$ , resulting in:

$$\rho \xrightarrow{U} U\rho U^\dagger \xrightarrow{\mathcal{N}} \mathcal{N}(U\rho U^\dagger) \xrightarrow{\mathcal{N}} \dots \xrightarrow{\text{folding with } c \text{ times}} (\mathcal{N}_\lambda)^c (U\rho U^\dagger), \quad (28)$$

where  $\mathcal{N}_\lambda$  is the noise channel associated with a single application of  $U$ .

Assuming the noise per gate remains constant, the total effective noise strength scales linearly:

$$E(c\lambda) = E_0 e^{-\alpha c\lambda} + \epsilon, \quad (29)$$

with  $\epsilon$  representing statistical error or model bias. Extrapolation to the zero-noise limit  $\hat{E}_0$  is then performed by fitting measurements at multiple fold factors  $c \in \{1, 3, 5\}$  using:

$$f(c) = Ae^{-Bc} + C, \quad \hat{E}_0 = \lim_{c \rightarrow 0} f(c) = A + C. \quad (30)$$

#### 4.1.2 Simulation Workflow

To evaluate the practical effectiveness of Zero Noise Extrapolation (ZNE), we perform a series of numerical experiments using Qiskit. Our simulation framework is designed to examine the impact of ZNE across multiple noise models, gate types, and circuit depths. All circuits are executed with 10,000 shots per configuration, and each setup is repeated 300 times to ensure statistical stability.

**Visualizing ZNE Behavior via Noise Scaling** We begin by visualizing the core mechanism of ZNE using a minimal single-qubit circuit. A Hadamard gate  $H$  is applied to the initial state  $|0\rangle$ , ideally resulting in an output of  $\langle X \rangle = 1$ . Under amplitude damping noise, the measured expectation decays with increasing noise strength. To amplify noise artificially, we apply gate folding at scale factors  $c \in \{1, 2, 3, 5, 7, 9\}$ , resulting in longer circuit duration and increased decoherence, while preserving logical functionality.

This setup demonstrates how expectation values degrade under noise and how extrapolation can recover the ideal value. The variance due to finite sampling is approximately  $\text{Var}[\tilde{E}] \approx \frac{1}{10000}$ , ensuring reliable statistical convergence.

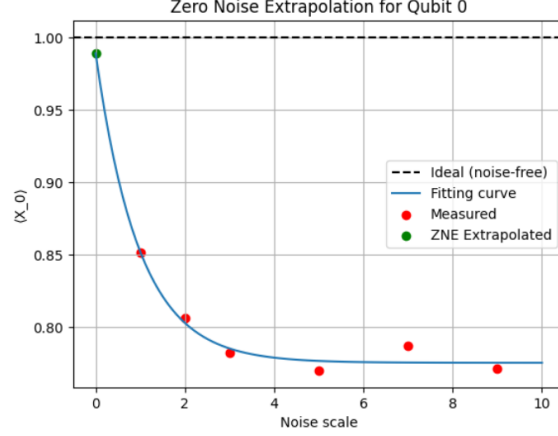


Figure 4.1: ZNE recovery of  $\langle X \rangle$  from a noisy Hadamard circuit under amplitude damping noise. The extrapolated value closely approximates the ideal expectation.

**(1) ZNE Accuracy vs. Noise and Circuit Depth** To quantitatively assess ZNE accuracy, we measure the mean squared error (MSE) between the extrapolated and ideal values across a variety of noise models and circuit depths. Random single-qubit circuits are constructed with depths from 1 to 4 using both Clifford gates ( $H$ ,  $S$ ,  $X$ ,  $Y$ ,  $Z$ ) and non-Clifford rotations ( $R_x(\pi/4)$ ,  $R_y(\pi/4)$ ,  $R_z(\pi/4)$ ). For each depth and noise model, we simulate the circuit with and without ZNE, and compute the MSE:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{E}_i - E_{\text{ideal}})^2,$$

where  $\hat{E}_i$  is the ZNE-estimated observable and  $E_{\text{ideal}}$  is the exact noiseless result. Here,  $N$  is the number of independently sampled random circuits (e.g., 300 repetitions) per noise setting, providing statistically averaged error estimates.

We consider the following canonical noise models, which have been discussed before:

- **Depolarizing noise:** Simulates uniform Pauli error with probability  $p$ .
- **Amplitude damping:** Models spontaneous relaxation from  $|1\rangle \rightarrow |0\rangle$ .
- **Phase damping:** Models dephasing with no energy loss.

This evaluation tests how ZNE performance scales with both noise strength and circuit complexity.

**(2) Gate-Specific Fidelity under Mixed Noise** Beyond circuit-wide expectation values, we examine the fidelity of ZNE mitigation for individual quantum gates. Each gate ( $X$ ,

$Y, Z, H, S, T, R_x, R_y, R_z$ ) is applied to  $|0\rangle$ , and the resulting noisy state  $\rho$  is compared against the ideal output  $\sigma$  using the state fidelity:

$$F(\rho, \sigma) = \left( \text{Tr} \sqrt{\sqrt{\sigma} \rho \sqrt{\sigma}} \right)^2.$$

A composite noise model (combining depolarizing, amplitude, and phase damping) is applied. Gate folding is used to generate a sequence of noise-amplified circuits with  $c \in \{1, 2, 3, 4, 5\}$ , followed by exponential fitting to estimate the zero-noise output. Simulations use Qiskit’s ‘statevector’ backend to compute fidelity exactly, avoiding classical sampling noise. This analysis evaluates whether ZNE generalizes across gate types with varying noise sensitivities.

**(3) Fidelity vs. Circuit Depth under Backend Noise** To study scalability, we analyze how ZNE fidelity varies with increasing circuit depth. Random single-qubit gate sequences of increasing length are generated, with each gate drawn from the combined Clifford and non-Clifford gate set. A composite Lindblad noise model is applied uniformly across all layers.

For each depth, we compare the fidelity of the noisy and ZNE-mitigated output states against the ideal target. This depth-dependent analysis reveals how circuit complexity interacts with noise accumulation and ZNE recovery capacity.

Together, these three evaluation schemes—MSE scaling, gate-specific fidelity, and depth-resolved performance—provide a comprehensive characterization of ZNE effectiveness under diverse and realistic quantum noise settings.

## 4.2 Machine Learning-Based Quantum Error Mitigation

To address the challenge of quantum errors in near-term devices, we adopt a supervised learning-based quantum error mitigation (QEM) strategy. This method is inspired by the ML-QEM framework proposed by (Liao et al., 2024), which fits noisy circuit outputs to their corresponding ideal results using data from small-scale circuits. While their study demonstrated the effectiveness of learning-based mitigation for shallow random circuits, it did not explicitly address generalization to larger circuits—though such scalability was highlighted as a promising direction. Our approach extends this idea by incorporating realistic device characteristics from superconducting quantum processors and structural

features of quantum circuits, enabling effective error mitigation without relying on large-scale ideal simulations.

Ultimately, our goal is to apply this method to large-scale circuits executed on real superconducting quantum processors. However, for such circuits, the corresponding ideal outputs cannot be efficiently obtained due to the exponential growth of the Hilbert space. To validate the generalization ability of our model in a controlled setting, we simulate both training and testing circuits using Qiskit’s noise-aware backends. Specifically, we train our model using small 2-qubit random circuits and evaluate its performance on larger 4-qubit circuits, both simulated using the same realistic backend (FakeAthensV2) that mimics the behavior of actual QPUs. This simulation framework allows us to emulate the noise characteristics of a real quantum processor while still accessing ground-truth ideal results via noise-free simulation, thereby enabling rigorous evaluation of the mitigation performance.

#### 4.2.1 Patch-based ML-QEM Method

Our quantum error mitigation method is based on supervised learning, aiming to predict ideal measurement expectation values from noisy outputs of quantum circuits executed on superconducting quantum processors. Inspired by prior works (Liao et al., 2024), we model the mitigation task as a regression problem that maps noisy observations and relevant circuit/device features to their ideal counterparts. To extend the applicability of the model to larger circuits, we adopt a local patch model approach, which decomposes large circuits into overlapping small qubit subsets (patches), applies trained local models on these patches, and then aggregates their outputs to form a global prediction, which can be summarized in two stages: training and inference.

**Training Phase.** We construct a training dataset using randomly generated 2-qubit circuits (patches). Each training circuit is simulated twice: once under a realistic noise model to obtain the noisy output, and once ideally (without noise) to obtain the ideal output. The training process proceeds as follows:

1. Patch Circuit Generation: We generate a large number of 2-qubit circuits with varying depths, gate configurations, and noise-affected qubit pairs.
2. Feature Extraction: For each patch  $p$ , we extract:
  - Noisy measurement outputs:  $\langle \hat{Z}_i \rangle_{\text{noisy}}$  for each qubit in the patch;
  - Device noise features:  $T_1, T_2$ , gate error rates, readout error of each qubit in the patch;

- Circuit structure features: depth, total number of gates, average gates per qubit, two-qubit gate density, etc.

These features are concatenated into a feature vector  $\mathbf{x}_p$ , while the corresponding ideal expectation values  $\mathbf{y}_p = (\langle \hat{Z}_i \rangle_{\text{ideal}})$  serve as training labels.

3. Model Training: We train a supervised regression model  $\mathcal{F}_p : \mathbf{x}_p \mapsto \mathbf{y}_p$ , which minimizes the mean squared error loss:

$$\min_{\theta_p} \sum_{k=1}^N \|\mathcal{F}_p(\mathbf{x}_p^{(k)}; \theta_p) - \mathbf{y}_p^{(k)}\|_2^2$$

where  $\theta_p$  are the learnable parameters of the model. In this work, we employ *Random Forest Regression* as  $\mathcal{F}_p$ , motivated by its strong performance and robustness in prior studies (Liao et al., 2024).

**Inference and Aggregation.** To apply the trained patch model to a larger circuit with  $N > 2$  qubits, we divide the circuit into overlapping patches of two adjacent qubits. For instance, a 4-qubit circuit would be divided as:

$$\mathcal{Q}_1 = \{q_0, q_1\}, \quad \mathcal{Q}_2 = \{q_1, q_2\}, \quad \mathcal{Q}_3 = \{q_2, q_3\}$$

The inference proceeds as follows:

1. For each patch  $p$ , extract the same feature vector  $\mathbf{x}_p$  as used during training, based on the noisy execution of the full circuit.
2. Feed  $\mathbf{x}_p$  into the trained model to obtain the patch-wise predicted ideal outputs:

$$\hat{\mathbf{y}}_p = \mathcal{F}_p(\mathbf{x}_p)$$

3. Since each qubit may appear in multiple patches, we average the predictions from all patches covering that qubit. Let  $N$  denote the number of patches that include qubit  $i$ , the final predicted ideal expectation value is:

$$\langle \hat{Z}_i \rangle_{\text{ML}} = \frac{1}{N} \sum_p [\hat{\mathbf{y}}_p]$$

where the summation is taken over all patches  $p$  containing qubit  $i$ , and  $[\hat{\mathbf{y}}_p]$  denotes the component corresponding to qubit  $i$  within patch  $p$ .

This patch-based structure leverages the locality of noise in quantum processors and enables the use of small-scale, accurately simulatable circuits to mitigate errors in larger circuits.

**Error Metric.** To quantitatively assess the performance of the mitigation method, we compute the discrepancy between the predicted and ideal observables using the  $\ell_2$  norm error. Given the noisy outputs  $\langle \hat{\mathbf{Z}} \rangle_{\text{noisy}}$ , the ML-mitigated outputs  $\langle \hat{\mathbf{Z}}_{\text{ML}} \rangle$ , and the ideal outputs  $\langle \hat{\mathbf{Z}} \rangle_{\text{ideal}}$ , we define:

$$\epsilon_{\text{noisy}} = \left\| \langle \hat{\mathbf{Z}} \rangle_{\text{noisy}} - \langle \hat{\mathbf{Z}} \rangle_{\text{ideal}} \right\|_2, \quad \epsilon_{\text{ML}} = \left\| \langle \hat{\mathbf{Z}}_{\text{ML}} \rangle - \langle \hat{\mathbf{Z}} \rangle_{\text{ideal}} \right\|_2,$$

where  $\langle \hat{\mathbf{Z}} \rangle = (\langle \hat{Z}_0 \rangle, \langle \hat{Z}_1 \rangle, \dots, \langle \hat{Z}_{N-1} \rangle)^\top$ .

A reduction in  $\epsilon_{\text{ML}}$  relative to  $\epsilon_{\text{noisy}}$  indicates successful error mitigation by the model.

Although we report the prediction error using the  $\ell_2$  norm, this quantity can be easily converted to the commonly used mean squared error (MSE). Specifically, if a circuit has  $N$  qubits, then the MSE is simply the squared  $\ell_2$  norm divided by the number of qubits:

$$\text{MSE} = \frac{\epsilon^2}{N}$$

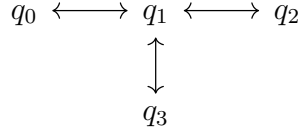
where  $\epsilon$  is the  $\ell_2$  error. This allows our results to be compared fairly with methods such as zero-noise extrapolation (ZNE), which are often evaluated using MSE.

#### 4.2.2 Simulation Setup and Backend Details

In a real quantum error mitigation scenario, the ultimate aim is to correct outputs from large quantum circuits executed on noisy quantum hardware, where ideal outputs are not available. To evaluate the patch-based ML-QEM approach, a realistic simulation pipeline using Qiskit 2.0 and noise-aware backend models was constructed. In particular, we use the FakeEssexV2 backend, which emulates a IBM's 5-qubit superconducting processor with qubit-specific relaxation times ( $T_1$ ), dephasing times ( $T_2$ ), gate fidelities, and readout errors. On top of these hardware characteristics, we define a custom noise model combining depolarizing, amplitude damping, and phase damping channels to simulate mixed noise effects. These are used consistently in training and evaluation.

**Qubit Topology.** To simplify the experimental setup while retaining meaningful qubit interactions, we restrict the study to the first four physical qubits  $\{q_0, q_1, q_2, q_3\}$  of the

FakeEssexV2, which form a connected subgraph with the topology:



**Training Data Generation.** For each patch, we generate 1000 random 2-qubit circuits composed of single-qubit rotations (rx, rz) and CNOT gates. These logical circuits are transpiled() to the FakeEssexV2 backend, which ensures they are physically mapped to the correct qubit pair and respects the hardware’s gate directionality and native basis gates. Each 2-qubit circuit is executed under two conditions:

1. *Noisy simulation:* Using fake provider FakeEssexV2, yielding  $\langle Z_i \rangle_{\text{noisy}}$  measurements;
2. *Ideal simulation:* Using AerSimulator with noise\_model=None, yielding  $\langle Z_i \rangle_{\text{ideal}}$ .

We repeat this process across all selected patch pairs, collecting a training set of approximately 300 circuits per patch. For each training instance, we extract:

- **Device features:**  $T_1$ ,  $T_2$ , gate error, readout error for each qubit; CNOT gate error; connectivity flag
- **Circuit features:** circuit depth, gate counts, average gate per qubit;
- **Noisy outputs:** Pauli-Z expectation values from noisy simulation.

This yields supervised pairs  $(\mathbf{x}_p, \mathbf{y}_p)$ , where  $\mathbf{x}_p$  is the input feature vector, and  $\mathbf{y}_p$  is the ideal Pauli-Z expectation values for the patch.

**Testing on Simulated QPU.** We generate 200 random 4-qubit circuits that use only the physical qubits  $\{q_0, q_1, q_2, q_3\}$  on FakeEssexV2 to obtain noisy measurement outcomes  $\langle Z_i \rangle_{\text{noisy}}$ ;

To apply our patch-based ML-QEM model, we partition each 4-qubit circuit into overlapping 2-qubit patches based on the device topology:

$$\mathcal{Q}_1 = \{q_0, q_1\}, \quad \mathcal{Q}_2 = \{q_1, q_2\}, \quad \mathcal{Q}_3 = \{q_2, q_3\}$$

For each patch  $\mathcal{Q}_p$ , we extract local features  $\mathbf{x}_p$ , and apply the corresponding pre-trained 2-qubit model  $\mathcal{F}_p$  to obtain predictions  $\hat{\mathbf{y}}_p = \mathcal{F}_p(\mathbf{x}_p)$ . Finally, for each qubit  $i$ , we average the predictions from all patches containing  $i$  to obtain the global ML-mitigated estimate  $\langle \hat{Z}_i \rangle_{\text{ML}}$ .

In real experimental settings, ideal reference values  $\langle Z_i \rangle_{\text{ideal}}$  are not accessible due to the exponential cost of classical simulation. However, since we conduct our experiments using simulated backends, we are able to generate ideal outputs using a noiseless simulator. These ideal values allow us to quantitatively assess the accuracy and effectiveness of the ML-based mitigation.

**Evaluation and Visualization.** To assess the effectiveness of our patch-based ML-QEM method, we evaluate prediction accuracy using the  $\ell_2$ -distance between the mitigated and ideal Pauli-Z expectation values, as introduced earlier. We focus on comparing the raw noisy outputs  $\langle Z \rangle_{\text{noisy}}$  and the ML-mitigated outputs  $\langle Z \rangle_{\text{ML}}$  against the ideal baseline  $\langle Z \rangle_{\text{ideal}}$  obtained from noiseless simulation.

To visualize the performance across the 200 test circuits, we adopt multiple complementary methods:

- **Error distribution plots** display the histogram of errors before and after mitigation, illustrating how ML-QEM shifts the overall error distribution toward lower values.
- **Prediction scatter plots** compare the ML-predicted values  $\langle Z_i \rangle_{\text{ML}}$  directly against their ideal counterparts. A tight clustering along the diagonal  $y = x$  reflects accurate recovery, while deviations expose systematic biases or outliers.

Together, these metrics and visualizations provide both aggregate and per-instance insight into the performance of the ML-based mitigation strategy under realistic backend noise conditions.

## CHAPTER 5

### RESULTS AND DISCUSSION

In this section, we evaluate the performance of zero-noise extrapolation (ZNE) on a variety of quantum circuits subject to different types and strengths of noise. The results are analyzed using two standard performance metrics: mean squared error (MSE) between noisy and ideal observables, and state fidelity between noisy and ideal final states. We investigate how ZNE performs under different noise channels, gate types, and circuit depths, with a particular focus on its scalability limitations in deeper circuits.

#### 5.1 MSE vs. Noise Strength across Noise Types and Circuit Depths

As shown in Fig. 5.1, the application of ZNE consistently reduces the mean squared error (MSE) between noisy and ideal expectation values. This trend holds across all three tested noise models—depolarizing, amplitude damping, and phase damping. Among them, ZNE proves particularly effective against phase damping, which predominantly introduces dephasing without energy loss.

For all noise models, MSE grows monotonically with the noise strength parameter, as expected. As the circuit depth increases, noise accumulates, and both the raw MSE and the extrapolated error grow accordingly. Nevertheless, ZNE consistently suppresses MSE compared to the unmitigated results.

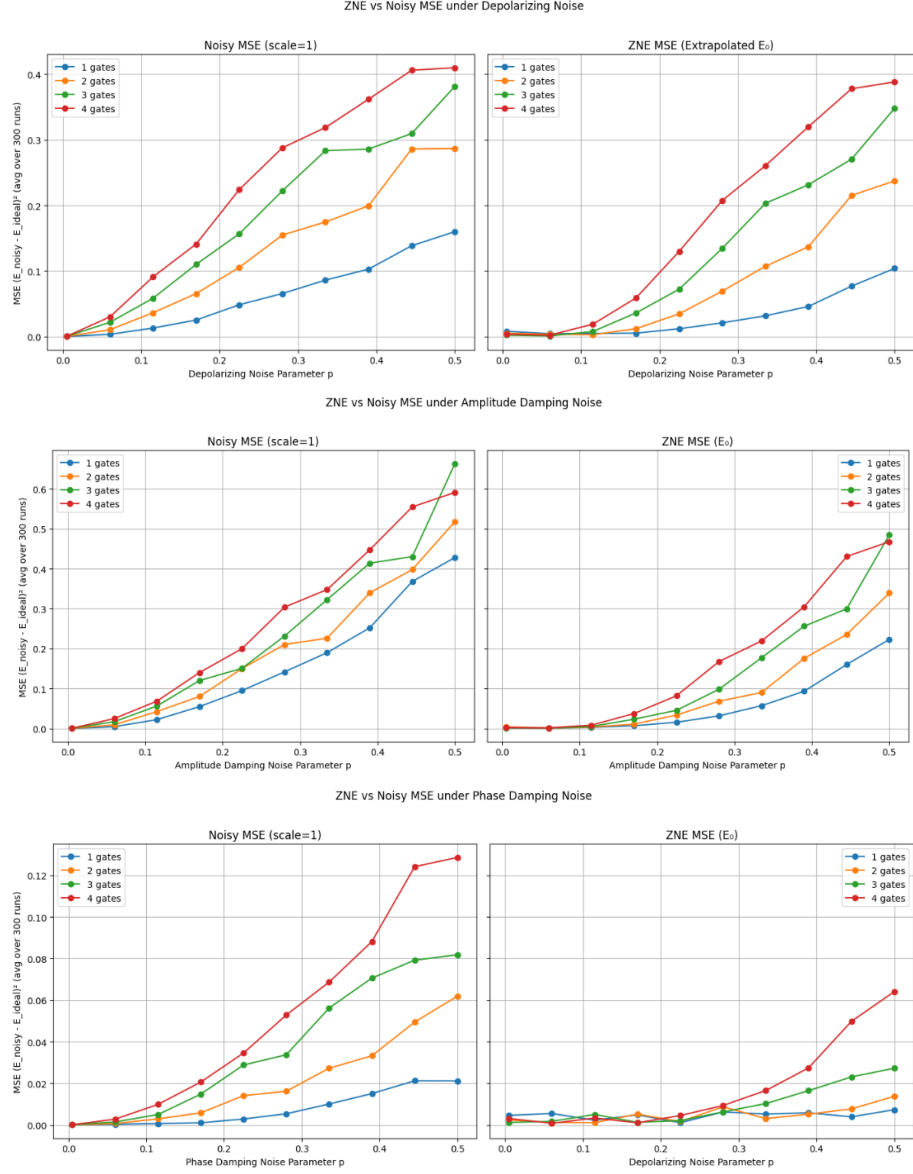


Figure 5.1: ZNE MSE vs Noise Parameter  $p$  under different noise models. ZNE reduces MSE especially under phase damping.

## 5.2 Fidelity vs. Gate Type under Mixed Noise

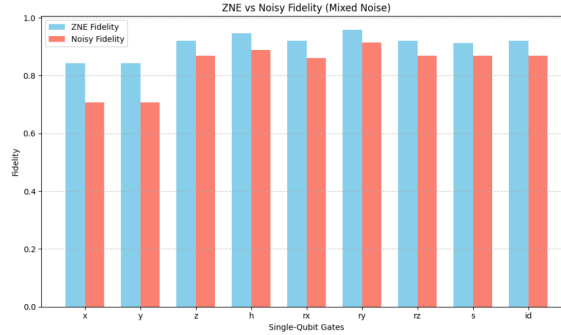


Figure 5.2: ZNE fidelity across different single-qubit gates on a mixed noise model. ZNE improves fidelity consistently across all tested gates.

To examine how ZNE interacts with different quantum gates, we simulate single-gate circuits under a mixed noise model combining depolarizing, amplitude damping, and phase damping noise. As illustrated in Fig. 5.2, ZNE provides fidelity improvements across all tested gate types, including  $X$ ,  $Y$ ,  $Z$ ,  $H$ ,  $S$ ,  $T$  and rotation gates. The most notable gains occur for Clifford gates like  $X$  and  $Y$ .

## 5.3 Fidelity vs. Circuit Depth under Mixed Noise

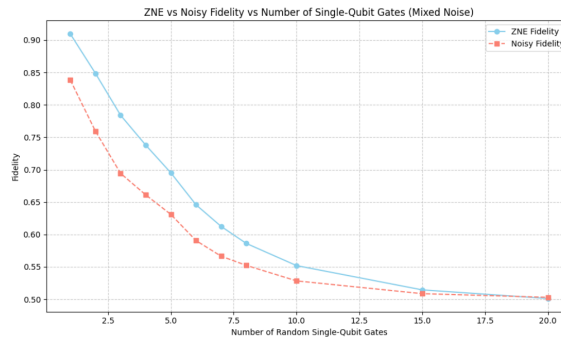


Figure 5.3: Fidelity vs circuit depth on real-device noise model with and without ZNE. ZNE improves shallow circuits but degrades in deeper ones.

To evaluate the scalability of ZNE, we simulate random single-qubit circuits with increasing numbers of gates, each drawn from a fixed gate set. These circuits are executed under the same mixed noise model used previously. The results, plotted in Fig. 5.3, reveal that

state fidelity declines monotonically with increasing circuit depth, reflecting the cumulative effects of quantum noise.

While ZNE is able to partially recover fidelity in shallow circuits—particularly those with fewer than 15 gates—its effectiveness deteriorates rapidly as circuit depth increases. In deeper circuits (e.g., 20+ gates), repeated applications of noisy operations drive the quantum state toward a maximally mixed state, effectively erasing any coherence present in the initial state. In this regime, the fidelity between the noisy state and the original pure state approaches its theoretical lower bound of 0.5. This results in numerical diminishing in extrapolating back to the zero-noise limit, especially when using high noise amplification factors or deep circuits. This behavior is consistent with theoretical studies (Takagi et al., 2022; Tsubouchi et al., 2023).

These scalability bottlenecks motivate the exploration of alternative mitigation strategies, such as machine learning-based QEM, which do not rely on analytic extrapolation and may be better suited to medium-scale noisy quantum devices. In the following sections, we investigate whether data-driven models can generalize from shallow circuit behavior to infer error-mitigated outcomes for larger and more complex circuits.

## 5.4 ML-based results

To evaluate the effectiveness of the machine learning-based quantum error mitigation (ML-QEM) approach, we analyzed both the error distribution before and after mitigation, as well as the correlation between mitigated and the ideal expectation values.

On average, fidelity improved from 0.8335 (noisy) to 0.8409 (after mitigation), indicating that the ML model slightly reduces the impact of noise. As shown in Fig.5.4, the left panel compares the error distributions before and after mitigation. The mitigated errors (green) exhibit a narrower distribution centered closer to zero.

The scatter plot in the right panel further supports this conclusion, showing the relationship between the mitigated predictions and the corresponding ideal values. While some dispersion remains—particularly for values near the extremes—the overall alignment along the diagonal demonstrates that the model captures the dominant trends in the noise-to-ideal mapping. This is further supported by a strong Pearson correlation coefficient of 0.89 between mitigated and ideal results.

Hardware-aware features, previously introduced in the methodology section play a central

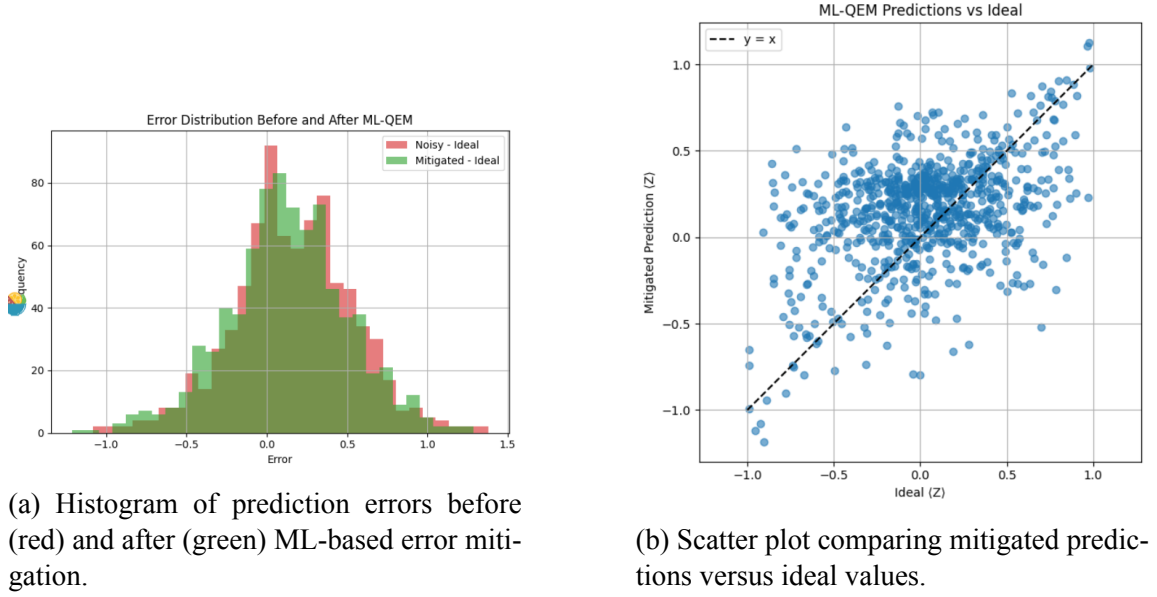


Figure 5.4: ML-QEM performance analysis

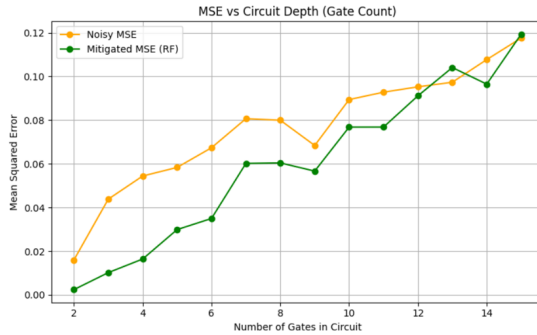
Table 5.1: Features across different patches

| Feature Description      | Patch (0,1) | Patch (1,2) | Patch (1,3) |
|--------------------------|-------------|-------------|-------------|
| T1 of qubit q0 (s)       | 0.00015     | 0.00009     | 0.00009     |
| T2 of qubit q0 (s)       | 0.00018     | 0.00016     | 0.00016     |
| Readout error of q0      | 0.0342      | 0.0307      | 0.0307      |
| T1 of qubit q1 (s)       | 0.00009     | 0.00007     | 0.00006     |
| T2 of qubit q1 (s)       | 0.00016     | 0.00009     | 0.00008     |
| Readout error of q1      | 0.0307      | 0.0558      | 0.0367      |
| CNOT gate error (q0, q1) | 0.0111      | 0.02774     | 0.01757     |
| CX direction validity    | 1           | 1           | 1           |

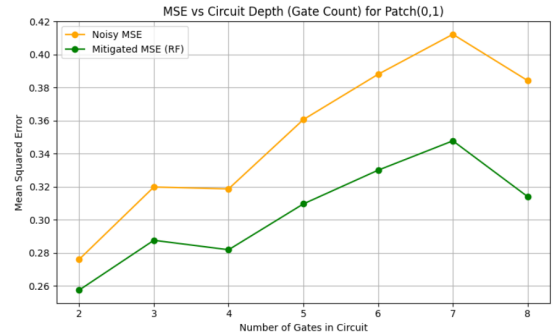
role in guiding the model’s learning process. Table 5.1 provides a comparison of these features across the three evaluated 2-qubit patches: (0,1), (1,2), and (1,3). From the table, we observe that Patch (1,2) has the highest CNOT gate error and readout error among the three, while Patch (0,1) has the longest coherence times for qubit 0. Patch (1,3), on the other hand, shows moderate values across most metrics. These physical characteristics reflect the diversity of noise profiles encountered during training and motivate the use of patch-specific models to account for such variations.

Rather than assuming uniform behavior across all qubit pairs, our framework tailors the model to each patch, enabling it to learn localized noise-response patterns. The resulting  $R^2$  scores—0.8600 for Patch (0,1), 0.9186 for Patch (1,2), and 0.9219 for Patch (1,3)—demonstrate that the model is capable of capturing these feature-conditioned relationships and achieving accurate prediction within each noise context.

To further evaluate the scalability and generalization of our ML-QEM framework, we conducted two additional simulations targeting different inference regimes. The first test used 2-qubit circuits for both training and evaluation, with circuit depths ranging from 2 to 11. In this case, the training circuits and test circuits shared the same patch configuration and noise environment. As shown in Fig. 5.5a, the model consistently reduced MSE across all depths compared to the unmitigated results. Even as circuit depth increased beyond the training range (e.g., depth 9–11), the model still demonstrated partial transferability, capturing observable trends in noise-affected measurements.



(a) native 2-qubit circuits aligned with the training distribution.



(b) patch-wise inference in larger entangled circuits.

**Figure 5.5: Comparison of ML error mitigation MSE across different gate depth:**

(a) 2-qubit circuits used during training and testing. The regression model is trained on circuits confined to single 2-qubit patches.

(b) Testing on 4-qubit circuits with  $\langle Z_0 Z_1 \rangle$  as the observable of interest. Although the model is trained only on patch(0,1), interference from patch(1,2) and patch(1,3) introduces additional noise.

In the second experiment, we introduced more complex scenarios by testing on 4-qubit circuits with entangled interactions. Specifically, training was still conducted using circuits confined to patch (0,1), but the test circuits included additional entanglement from patch (1,2) and patch (1,3), simulating a more realistic and noisy local environment. This setting mirrors a common structure in larger superconducting layouts, where localized gates may still be affected by crosstalk or nearby operations. As seen in Fig.5.5b, although the added interference degraded overall performance—as expected—the model was still able to reduce the MSE on patch (0,1), validating its ability to generalize to entangled subcircuits.

These simulations serve to complement the full 4-qubit ML results presented earlier, where mitigation improvements were modest. By isolating specific circuit patches and evaluating their behavior under both clean and entangled conditions, we provide additional evidence that patch-based ML-QEM can scale with circuit complexity and retain localized effectiveness in larger systems.

Future work may explore deeper architectures (e.g., graph neural networks), adaptive training strategies, or integration with hardware calibration data. Additionally, robustness under distributional shift and improved tail performance could be addressed via outlier-aware objectives or transfer learning, further enhancing the model’s deployment viability on near-term quantum devices.

## 5.5 Comparison and Discussion of QEM Strategies

Next, we will compare the three quantum error mitigation (QEM) strategies by combining the simulated results discussed before, highlighting their respective strengths, limitations, and suitability for near-term quantum devices.

ZNE operates by amplifying the existing noise and extrapolating the observed measurements back to the zero-noise limit. On the one hand, our simulation results confirm that ZNE is effective in reducing MSE and improving fidelity across a range of noise models and gate types, especially in circuits with limited depth. This makes it a promising tool for near-term quantum applications that rely on shallow variational or measurement-based protocols. On the other hand, the method’s performance degrades significantly as circuit depth increases. This limitation stems from both the intrinsic instability of extrapolation at large noise amplification factors and the exponential increase in sampling cost (Temme et al., 2017). In addition, the method assumes a smooth and well-behaved response of observables to noise amplification, which may not hold under strongly non-Markovian or

hardware-specific noise interactions.

Probabilistic Error Cancellation (PEC) is theoretically capable of exactly inverting noise channel and can recover the ideal output by representing noisy operations as linear combinations of ideal ones. However, this technique relies on precise knowledge of the noise model and suffers from an exponentially growing sampling overhead.

To understand this, consider the depolarizing noise channel as an example. When a quantum gate suffers from depolarizing noise of strength  $p$ , the noisy operation can be described as

$$\mathcal{E}_{\text{dep}}(\rho) = (1 - p)U\rho U^\dagger + p \cdot \frac{I}{2},$$

To cancel this noise, PEC constructs the inverse channel:

$$\mathcal{E}^{-1} = \frac{1}{1 - p}\mathcal{E}_{\text{dep}} - \frac{p}{3(1 - p)}(X\rho X + Y\rho Y + Z\rho Z) \quad (31)$$

While this method can in principle restore the ideal expectation values, it introduces several critical challenges. First, PEC demands precise knowledge of the underlying noise model—typically obtained through full noise tomography, which scales exponentially with qubit count. For a 20-qubit system, this requires on the order of  $4^{20} \approx 10^{12}$  measurements, posing a significant calibration burden. Second, PEC suffers from a rapidly increasing sampling cost. In order to cancel out noise, each noisy gate must be probabilistically replaced by a mixture of ideal gates with positive and negative weights. This means that a single run of the original circuit is no longer enough—you need to run many randomized versions of it and combine their results. Even for modest depolarizing noise levels ( $p = 0.1$ ), the overhead per gate is approximately  $\gamma \approx 1 + \frac{4p}{3} = 1.13$ . For 10 gates per qubit - the total number of gates is  $N=200$  -, the total sampling overhead becomes  $\gamma^{200} \sim 10^{10}$ , making it intractable on medium-scale implementation.

ML-QEM offers a data-driven alternative that avoids the need for repeated noise amplification or explicit noise inversion. In our framework, supervised learning models—trained on noisy and ideal results of 2-qubit circuits—are used to infer error-mitigated results in larger 4-qubit circuits. This approach is inherently scalable to larger circuits as it does not require additional quantum resources during inference. Despite this, our ML-QEM framework still faces limitations in scalability and generality. In contrast to techniques like ZNE, which can be applied to circuits of varying structure with minimal modification, the ML model requires retraining for circuits with different depths or gate compositions. Moreover, while mitigation accuracy improves, the absolute gain in fidelity remains modest—suggesting the need for more expressive models or richer features to unlock further benefits.

## CHAPTER 6

### CONCLUSION

Quantum error mitigation (QEM) is a critical research direction in the NISQ era, where hardware limitations preclude the use of full-scale quantum error correction. In this work, we systematically explored and compared three representative QEM methods—Zero-Noise Extrapolation (ZNE), Probabilistic Error Cancellation (PEC), and machine learning-based QEM (ML-QEM)—using a combination of theoretical formulation and numerical simulation on realistic noise models.

Our results demonstrate that ZNE can effectively improve measurement accuracy in shallow quantum circuits, particularly under single- and mixed-noise models. However, its effectiveness diminishes significantly with increasing circuit depth or gate count, due to numerical instability, extrapolation divergence, and the onset of complete decoherence. While PEC offers a theoretically optimal solution for mitigating Markovian noise, it is ultimately impractical for medium-scale systems such as our 20-qubit target architecture, owing to its exponential cost in circuit sampling and calibration—rendering it unsuitable for circuits beyond a few qubits.

In contrast, our machine learning-based approach—trained on small 2-qubit subcircuits using random forest models—shows clear improvement when applied to new 4-qubit circuits. Although the increase in fidelity is modest (from 0.8335 to 0.8409), the method has several practical advantages. Most importantly, it does not require ideal outputs during testing, which means it can be used even when simulating large circuits is no longer possible. This makes our approach more scalable and suitable for future quantum devices where traditional simulation becomes infeasible. In addition, our model is simple to train, runs quickly, and does not need detailed noise models or changes to the quantum circuits themselves. These features make it easy to use in real-world experiments. While this is only a basic version of the method, the results already show a measurable benefit, and there is clear potential for further improvement by using more advanced models or better features.

Recent ML-based error mitigation methods—such as those by Liao and Srushti (Liao et al., 2024; Patil et al., 2025)—achieve good performance, but often rely on deep learning models trained on large and complex circuits. These methods usually need either other error mitigation results (like ZNE) as labels, or customized training for each circuit, which makes them harder to apply generally. By contrast, our method trains on simple, fully

simulated small circuits and works directly on new circuits without retraining. This makes it much easier to scale up and apply in practice.

## 6.1 Recommendation for Future Work

While our results validate the feasibility of patch-based ML-QEM, several directions remain open for future improvement:

**Adopt more expressive machine learning models:** The current framework uses random forests for interpretability and ease of training, but advanced models—such as graph neural networks (GNNs), transformers, or hybrid physics-informed networks—may capture more complex noise behaviors and offer better generalization to unseen circuit structures.

**Evaluate scalability beyond 4-qubit circuits:** Our modular patch-based method is inherently extensible to larger circuits. Future experiments should explore applying the model to 7-, 10-, or even 20-qubit circuits using the same patch inference strategy, and evaluate how well mitigation quality holds up.

**Combine with existing QEM techniques:** Hybrid schemes—e.g., using ML to estimate ZNE fit curves or filter PEC sample weights—may allow combining the best of both worlds: the interpretability and theoretical guarantees of classical methods, with the adaptability and low overhead of machine learning.

**Deploy on real quantum hardware:** As the final step toward practical deployment, future work should test this framework directly on IBM QPUs or similar superconducting platforms. Since our model does not require ideal outputs at inference time, it is well-suited for real-device implementation. Nevertheless, a key challenge in real-device deployment is the lack of ideal reference values, which complicates the assessment of whether mitigation is truly effective.

Moving forward, the integration of physics-informed machine learning with partial noise knowledge and device-specific topology offers a promising pathway for robust, hardware-adaptive error mitigation. As quantum hardware continues to scale, such hybrid strategies may play a key role in bridging the gap between near-term noisy devices and future fault-tolerant quantum computation.

## REFERENCES

- Cai, Z., Babbush, R., Benjamin, S. C., Endo, S., Huggins, W. J., Li, Y., McClean, J. R., & O'Brien, T. E. (2023). Quantum error mitigation. *Reviews of Modern Physics*, 95(4), 045005. <https://doi.org/10.1103/RevModPhys.95.045005>
- Chia, A., Tan, K. C., Pawela, Ł., Kurzyński, P., Paterek, T., & Kaszlikowski, D. (2016). Coherent chemical kinetics as quantum walks. i. reaction operators for radical pairs. *Phys. Rev. E*, 93, 032407. <https://doi.org/10.1103/PhysRevE.93.032407>
- Clarke, J., & Wilhelm, F. K. (2008). Superconducting quantum bits. *Nature*, 453(7198), 1031–1042. <https://www.nature.com/articles/nature07128#citeas>
- Czarnik, P., Arrasmith, A., Coles, P. J., & Cincio, L. (2021). Error mitigation with Clifford quantum-circuit data. *Quantum*, 5, 592. <https://doi.org/10.22331/q-2021-11-26-592>
- de Oliveira, A., Gomes, R., Brasil, V., Rubiano da Silva, N., Céleri, L., & Souto Ribeiro, P. (2020). Full thermalization of a photonic qubit. *Physics Letters A*, 384(36), 126933. <https://doi.org/10.1016/j.physleta.2020.126933>
- Devitt, S. J., Munro, W. J., & Nemoto, K. (2013). Quantum error correction for beginners. *Reports on Progress in Physics*, 76(7), 076001. <https://doi.org/10.1088/0034-4885/76/7/076001>
- Fowler, A. G., Mariantoni, M., Martinis, J. M., & Cleland, A. N. (2012). Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(3). <https://doi.org/10.1103/physreva.86.032324>

- Havel, T. F. (2003). Robust procedures for converting among lindblad, kraus and matrix representations of quantum dynamical semigroups. *Journal of Mathematical Physics*, 44(2), 534–557. <https://doi.org/10.1063/1.1518555>
- Khan, M. U., Kamran, M. A., Khan, W. R., Ibrahim, M. M., Ali, M. U., & Lee, S. W. (2024). Error Mitigation in the NISQ Era: Applying Measurement Error Mitigation Techniques to Enhance Quantum Circuit Performance. *Mathematics*, 12(14), 2235. <https://doi.org/10.3390/math12142235>
- Kim, J., Oh, B., Chong, Y., Hwang, E., & Park, D. K. (2022). Quantum readout error mitigation via deep learning. *New Journal of Physics*, 24(7), 073009. <https://doi.org/10.1088/1367-2630/ac7b3d>
- Kjaergaard, M., Schwartz, M. E., Braumüller, J., Krantz, P., Wang, J. I.-J., Gustavsson, S., & Oliver, W. D. (2020). Superconducting qubits: Current state of play. *Annual Review of Condensed Matter Physics*, 11(1), 369–395. <https://doi.org/10.1146/annurev-conmatphys-031119-050605>
- Koch, J., Yu, T. M., Gambetta, J., Houck, A. A., Schuster, D. I., Majer, J., Blais, A., Devoret, M. H., Girvin, S. M., & Schoelkopf, R. J. (2007). Charge-insensitive qubit design derived from the cooper pair box. *Physical Review A*, 76(4). <https://doi.org/10.1103/physreva.76.042319>
- Lee, C., & Park, D. K. (2023). Scalable quantum measurement error mitigation via conditional independence and transfer learning. *Machine Learning: Science and Technology*, 4(4), 045051. <https://doi.org/10.1088/2632-2153/ad1007>

- Liao, H., Wang, D. S., Sitdikov, I., Salcedo, C., Seif, A., & Mineev, Z. K. (2024). Machine learning for practical quantum error mitigation. *Nature Machine Intelligence*. <https://doi.org/10.1038/s42256-024-00927-2>
- Murali, P., Linke, N. M., Martonosi, M., Abhari, A. J., Nguyen, N. H., & Alderete, C. H. (2019, June). Full-Stack, Real-System Quantum Computer Studies: Architectural Comparisons and Design Insights. <https://doi.org/10.48550/arXiv.1905.11349>
- Nielsen, M. A., & Chuang, I. L. (2010). *Quantum computation and quantum information: 10th anniversary edition*. Cambridge University Press.
- OpenCourseWare, M. (2012). Quantum theory of radiation interactions: Chapter 8 [Accessed: 2024-12-17]. [https://ocw.mit.edu/courses/22-51-quantum-theory-of-radiation-interactions-fall-2012/b24652d7f4f647f05174797b957bd3bf\\_MIT22\\_51F12\\_Ch8.pdf](https://ocw.mit.edu/courses/22-51-quantum-theory-of-radiation-interactions-fall-2012/b24652d7f4f647f05174797b957bd3bf_MIT22_51F12_Ch8.pdf)
- Patil, S., Mondal, D., & Maitra, R. (2025, April). Machine Learning Approach towards Quantum Error Mitigation for Accurate Molecular Energetics. <https://doi.org/10.48550/arXiv.2504.07077>
- Preskill, J. (2018). Quantum computing in the nisq era and beyond. *Quantum*, 2, 79. <https://doi.org/10.22331/q-2018-08-06-79>
- Qin, D., Chen, Y., & Li, Y. (2023). Error statistics and scalability of quantum error mitigation formulas. *npj Quantum Information*, 9(1), 35. <https://doi.org/10.1038/s41534-023-00707-7>

- Song, C., Cui, J., Wang, H., Hao, J., Feng, H., & Li, Y. (2019). Quantum computation with universal error mitigation on a superconducting quantum processor. *Science Advances*, 5(9), eaaw5686. <https://doi.org/10.1126/sciadv.aaw5686>
- Takagi, R., Endo, S., Minagawa, S., & Gu, M. (2022). Fundamental limits of quantum error mitigation. *npj Quantum Information*, 8(1), 114. <https://doi.org/10.1038/s41534-022-00618-z>
- Temme, K., Bravyi, S., & Gambetta, J. M. (2017). Error mitigation for short-depth quantum circuits. *Physical Review Letters*, 119(18). <https://doi.org/10.1103/physrevlett.119.180509>
- Tsubouchi, K., Sagawa, T., & Yoshioka, N. (2023). Universal Cost Bound of Quantum Error Mitigation Based on Quantum Estimation Theory. *Physical Review Letters*, 131(21), 210601. <https://doi.org/10.1103/PhysRevLett.131.210601>