RWTH Aachen University

Doctoral Thesis

Applications of fault-tolerant topological quantum error correction in near-term devices

Author:
Sascha Heinrich Heußen

Supervisor:
Prof. Dr. Markus Müller

Second examiner:
Prof. Dr. David P. DiVincenzo

Submitted to the Faculty of Mathematics, Computer Science and Natural Sciences at RWTH Aachen University

Institute for Quantum Information

Theoretical Quantum Technology Group

Fakultät für
Mathematik, Informatik
und Naturwissenschaften

RWTH AACHEN UNIVERSITY

Applications of fault-tolerant topological quantum error correction in near-term devices

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Sascha Heinrich Heußen, M. Sc.

aus

Viersen

Berichter: Univ.-Prof. Dr. rer. nat. Markus Müller
Univ.-Prof. Dr. David DiVincenzo

Tag der mündlichen Prüfung: 18. Januar 2024

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek verfügbar.

Abstract

Quantum computers can theoretically facilitate solving problems that classical supercomputers may only solve with resources – like computation time – exponentially growing with the size of the problem instance. Such devices are widely believed to require quantum error correction to reach the level of reliability that is needed to run useful large-scale quantum algorithms. A fault-tolerant system design can help to achieve practical computational advantages from quantum computers built from noisy components. In this thesis, we demonstrate how modern fault-tolerant quantum circuit designs can be used to systematically suppress noise, which is detrimental to the information stored and processed by the quantum computer. For this purpose, we employ small topological quantum error correcting codes that are suitable for experimental devices.

We develop a numerical simulation technique that is capable of efficiently simulating adaptive sequences of quantum circuits built from components with weak noise on a classical computer. Failure of the whole system in this case becomes a rare event, with high-fidelity physical operations, which may be attainable in practice due to the impressive improvements of experimental control capabilities. Then, using active quantum error correction can become a fruitful endeavor to suppress failure rates even further. The quantum computation can be sustained for in principle arbitrarily long times this way. We show that the first experimental realization of an encoded fault-tolerant universal quantum gate set in a trapped-ion quantum computer can be accurately modeled by a few-parameter noise model that is only informed by the infidelities of experimental operations. The effectiveness of fault-tolerant implementations is showcased by the superior performance of magic state preparation compared to non-fault-tolerant approaches. Subsequently, we study the effect of improved noise strengths on physical operations in order to gain an advantage of logical qubit operation over physical qubits. Additionally, we provide a new circuit for fault-tolerant unitary logical qubit initialization that eliminates the need for in-sequence measurements, which pose major roadblocks for experimental realizations of fault-tolerant protocols. Further developing these ideas, we propose and analyze a measurement-free quantum error correction scheme, which is fully fault-tolerant with respect to any type of noise on all components of the circuit. This may enable fault-tolerant quantum error correction in experimental architectures that struggle with implementations of fast and reliable in-sequence measurements and feed-forward corrections. We show via numerical simulations that the scheme can potentially achieve lower logical failure rates than a conventional fault-tolerant quantum error correction scheme when implemented in a neutral-atom quantum computer.

Our findings contribute to the advancement of practical fault-tolerant quantum computation. They serve to realize primitives of error-corrected universal quantum computation using state-of-the-art and near-term devices.

Kurzzusammenfassung

Quantencomputer besitzen theoretisch das Potential zur effizienten Lösung bestimmter Problemklassen, die klassische Supercomputer nur mit Ressourcen – wie beispielsweise Rechenzeit – lösen können, die exponentiell mit der Größe des Problems anwachsen. Gemäß breit vertretener Auffassung bedarf es Quantenfehlerkorrektur, um die zuverlässige Ausführung von nützlichen Quantenalgorithmen in großem Maßstab sicher zu stellen. Eine fehlertolerante Systemarchitektur kann dabei helfen, einen praktischen Rechenvorteil mit Quantencomputern zu erreichen, die aus rauschenden Komponenten bestehen. In dieser Dissertation zeigen wir, wie moderne fehlertolerante Quantenschaltkreise zur systematischen Reduktion von Rauschen verwendet werden können, das sich ansonsten verheerend auf die vom Quantencomputer gespeicherten und verarbeiteten Informationen auswirkt. Dazu verwenden wir topologische Quantenfehlerkorrektur-Codes, die für existierende experimentelle Aufbauten geeignet sind.

Wir entwickeln eine effiziente numerische Simulationstechnik für adaptive Sequenzen von Quantenschaltkreisen mit schwachem Rauschen. Ein Defekt des Gesamtsystems ist unter diesen Umständen ein seltenes Ereignis; physikalische Operationen mit hinreichend hoher Güte sind dank verbesserter experimenteller Kontrollmöglichkeiten möglicherweise in der Praxis erreichbar. Quantenfehlerkorrektur kann sodann dabei helfen, Fehlerraten weiter zu reduzieren und die Quantenberechnung prinzipiell beliebig lang aufrecht zu erhalten. Wir zeigen, dass die erste experimentelle Realisierung eines kodierten fehlertoleranten universellen Satzes von Quantengattern in einem Ionenfallen-Quantencomputer akkurat durch ein Fehlermodell mit wenigen Parametern modelliert werden kann, die einzig durch die experimentellen Fehlerraten bestimmt werden. Der Nutzen von fehlertoleranten Implementierungen wird am Beispiel einer verbesserten Ausführung sogenannter „magic state“-Präparation gegenüber einer nicht-fehlertoleranten Prozedur herausgestellt. Weiterhin untersuchen wir die Auswirkungen von weiteren Verbesserungen von physikalischen Operationen, um einen operationellen Vorteil von logischen Qubits über physikalische Qubits zu erreichen. Zusätzlich entwickeln wir einen neuartigen Quantenschaltkreis für die fehlertolerante unitäre Präparation eines logischen Qubits, der keiner Messung einzelner Qubits bedarf, da Messungen innerhalb eines Schaltkreises ein signifikantes Hindernis für die experimentelle Realisierung fehlertoleranter Protokolle darstellen. Darauf aufbauend entwerfen und analysieren wir ein Quantenfehlerkorrektur-Schema ohne Messungen, das vollständig fehlertolerant gegenüber eines allgemeinen Fehlermodells bzgl. aller Schaltkreiskomponenten ist. Dieses Schema ermöglicht prinzipiell fehlertolerante Quantenfehlerkorrektur in experimentellen Plattformen ohne schnelle und zuverlässige Messungen und ohne die darauf basierenden Anwendungen von Korrekturoperationen in Echtzeit. Wir demonstrieren mithilfe numerischer Simulationen, dass unser Schema in einem Neutral-Atom-Quantencomputer niedrigere Fehlerraten als ein konventionelles Protokoll erreichen kann.

Unsere Ergebnisse tragen zur Entwicklung fehlertoleranter Quantencomputer mit praktischem Nutzen bei. Sie dienen der Realisierung von Grundbausteinen für fehlerkorrigiertes universelles Quantenrechnen mit kurzfristig verfügbaren experimentellen Aufbauten.

List of publications and preprints

The material published in the following works is also partly contained in this thesis:

- L. Postler, S. Heußen, I. Pogorelov, M. Rispler, T. Feldker, M. Meth, C. D. Marciniak, R. Stricker, M. Ringbauer, R. Blatt, P. Schindler, M. Müller and T. Monz, **Demonstration of fault-tolerant universal quantum gate operations**, *Nature*, 605, 675 (2022), cited as [Pos+22], mostly contained in Sec. 4.1.2.
- S. Heußen, L. Postler, M. Rispler, I. Pogorelov, C. D. Marciniak, T. Monz, P. Schindler and M. Müller, **Strategies for a practical advantage of fault-tolerant circuit design in noisy trapped-ion quantum computers**, *Physical Review A*, 107, 042422 (2023), cited as [Heu+23a], mostly contained in Secs. 4.1.2 and 4.1.3.
- F. Butt, S. Heußen, M. Rispler and M. Müller, **Fault-tolerant code switching protocols for near-term quantum processors**, arXiv preprint 2306.17686 (2023), cited as [But+23], referenced in Sec. 2.5.1.3.
- S. Heußen, D. F. Locher and M. Müller, **Measurement-free fault-tolerant quantum error correction in near-term devices**, arXiv preprint 2307.13296 (2023), cited as [HLM23], mostly contained in Sec. 4.2.
- S. Heußen, D. Winter, M. Rispler and M. Müller, **Dynamical subset sampling of quantum error correcting protocols**, arXiv preprint 2309.12774 (2023), cited as [Heu+23b], mostly contained in Sec. 3.5.

My contributions to [Pos+22] and [Heu+23a] include carrying out all numerical simulations, analysis of the circuits used for the experiments as well as theory modeling and interpretation of the numerical and experimental data. I contributed to the supervision of the Master's project that resulted in [But+23]. For [HLM23], I developed the scheme, its compilations for the various native gate sets, performed all numerical simulations and contributed to theory modeling. Regarding [Heu+23b], I devised the scheme, performed all numerical simulations and supervised the development of the Python package `qsample` [WH23].

Acknowledgments

First and foremost, I wish to thank my supervisor Markus Müller for giving me the chance to dive into a new field and guiding me with his encouraging and enduring patience over the last years. Without his ingenious and inspiring enthusiasm as well as his constructive criticism I would not have been capable of getting the job done. I am deeply grateful for all the animating discussions that we had, no matter how large or small the issue, and especially during pandemic times where being in contact was even harder.

I am also indebted to David Amaro and Davide Vodola for strengthening my determination to join Markus' group. My sincere thanks extends to Kai P. Schmidt and Gil Refael who did not fail to advice me in finding my way.

It was a special experience to be part of a group growing (almost) from the beginning. I had the great pleasure of working with Pedro Parrado Rodríguez, Fernando Martínez García, Eliana Fiorelli, Thomas Botzung, Lorenzo Cardarelli, Manuel Rispler, Thiago Lucena de Macedo Guedes, Luis Colmenarez, Santiago Varona Angulo, Lukas Bödeker, David F. Locher, Friederike Butt, Josias Old, Katrin Bolsmann, Don Winter, Leon Bohnmann, Lars Esser, Luc Kusters, Dmitri Dobrynin, David Scheer and Roshni Singh and I very much appreciate any group meeting, journal club, thirsty thursday, whiteboard discussion, train ride or lunch time I was granted to share with them. I also want to appreciate the company of Fabian Hassler, Lisa Otten, Alexander Ziesen, Steven Kim, Anton Montag, Konstantinos Kontogeorgiou, David DiVincenzo, Martin Rymarz, Evangelos Varvelis, Mira Ramakant Sharma, Maarten Wegewijs, Valentin Bruch and Anand Sharma and everyone in Aachen who made my time there worthwhile. I would like to acknowledge all the assistance that I received from Hélène Barton, Luise Snyders and Magda Baer Radermacher.

Particularly helpful was the support of Pedro Parrado Rodríguez and Ciarán Ryan-Anderson in jumpstarting my numerical undertakings. I am very thankful to Thomas Botzung and Friederike Butt for showing me that I am not alone. Special thanks should also go my office mates Josias Old and Katrin Bolsmann for taking my half-baked blabber. I wish to express my genuine gratitude for Thomas Monz, Philipp Schindler, Lukas Postler, Ivan Pogorelov and Christian Marciniak for not only a great and fun collaboration but also for offering me another invaluable perspective on research that I could not have found anywhere else.

I am very grateful to everyone who contributed their expertise and provided helpful feedback for preliminary versions of this thesis: Valentin Bruch for abstract, chapters 1, 2 and parts of 3, Josias Old for abstract, chapters 1 and 5, Philipp Walk for abstract, chapter 1 and parts of 2, David Scheer for chapter 2, Nicholas Fazio for section 2.5.1.2, Manuel Rispler for parts of chapter 2, Friederike Butt for chapter 3, Thiago Lucena de Macedo Guedes for parts of chapter 3, Luis Colmenarez for chapter 4, David F. Locher for section 4.2, Lukas Postler for parts of chapter 4 and Christian Marciniak for abstract, chapters 1, 2, 4 and 5.

Lastly, I owe a debt of gratitude to my friends who have been bearing with me and supporting me throughout my effort; Philipp Zander, Pascal Breuer and Jule Wunderlich on a special note. I noticed quite late that my interest in science and computers might have been seeded by Richard Janissen to whom I am very thankful. Meinen Eltern danke ich für ihre Wertschätzung für Bildung und die Beharrlichkeit, die mir zuteilwurde.

Contents

Abstract **v**

Kurzzusammenfassung **vii**

1 Introduction **1**
1.1 Physical platforms for quantum computation . . . 2
1.2 Quantum error correction and fault tolerance . . . 3
1.3 Structure of this thesis . . . 5

2 Universal fault-tolerant quantum computation **7**
2.1 Qubit systems and universality . . . 7
2.2 Quantum error correction . . . 9
2.2.1 Stabilizer formalism . . . 9
2.2.2 Stabilizer codes . . . 10
2.2.3 Fidelity measures . . . 12
2.3 Fault tolerance and the threshold theorem . . . 13
2.4 Fault-tolerant circuit design . . . 14
2.4.1 Definitions . . . 14
2.4.2 Fault propagation . . . 15
2.4.3 Shor-type error correction . . . 16
2.4.4 Flag fault-tolerant error correction . . . 17
2.4.5 Advantage of fault-tolerant protocols . . . 19
2.5 Topological QEC codes . . . 19
2.5.1 Paths to universality . . . 22
2.5.1.1 Braiding in topological quantum computers . . . 22
2.5.1.2 Magic state injection . . . 23
2.5.1.3 Code deformation . . . 29
2.5.2 Low-distance color code . . . 31

3 Numerical simulation of noisy quantum computers **35**
3.1 Noisy quantum channels . . . 35
3.2 Classical simulation of quantum circuits . . . 37
3.3 Fault path counting . . . 39
3.4 Direct Monte Carlo simulation . . . 40
3.5 Subset sampling . . . 42
3.5.1 Single-circuit subset sampling . . . 43
3.5.1.1 Multi-parameter noise . . . 47
3.5.1.2 Adaptive sampling . . . 47
3.5.2 Dynamical subset sampling . . . 51
3.5.2.1 Uncertainty intervals . . . 57
3.5.2.2 Application to protocols with intermediate measurements . . . 66

3.5.2.3 Efficiency and run time considerations 71

4 Advances for fault-tolerant quantum computation 75
4.1 Contributions for experimental progress 75
4.1.1 Trapped-ion quantum computation 75
4.1.1.1 Physical architecture and native gate operations . . . 77
4.1.1.2 Sources of noise in the particular ion trap setup . . . 78
4.1.1.3 Noise modeling 79
4.1.2 Fault-tolerant universal gate set realization 86
4.1.2.1 Pauli state preparation and logical Clifford gates . . . 87
4.1.2.2 Logical CNOT gate 91
4.1.2.3 Magic state preparation and T-gate teleportation . . . 93
4.1.3 Extensions of logical state preparation 97
4.1.3.1 Reaching break-even with non-deterministic schemes . 97
4.1.3.2 Comparison to deterministic schemes 100
4.1.3.3 Multiple logical qubits 109
4.1.3.4 Crosstalk-resistant state preparation 110
4.2 Measurement-free fault-tolerant QEC 113
4.2.1 Steane-type error correction 114
4.2.2 Deterministic encoding circuit without measurements 115
4.2.3 Quantum error correction unit 117
4.2.3.1 Logical level 117
4.2.3.2 Application to the Steane code 117
4.2.3.3 Resources 120
4.2.4 Implementations 120
4.2.4.1 Decomposition into two-qubit gates 121
4.2.4.2 Native multi-qubit gates in neutral-atom platforms and ion traps 122
4.2.4.3 Measurement-free advantage 127
4.2.5 Scaling up 132

5 Conclusions & outlook 135

Bibliography 141

List of Figures

2.1 Quantum circuit to measure the n-qubit observable U with the help of an additional measurement qubit 11
2.2 Illustration of noise terminology 14
2.3 Propagation of X- and Z-faults through a CNOT gate 15
2.4 For some quantum codes, some logical gates can be implemented in a bitwise-fashion 16
2.5 Circuit to measure the stabilizer operator $U^{\otimes 4}$ in a QEC code with distance $d = 3$ fault-tolerantly 17
2.6 In a flag circuit for FT stabilizer readout, a dangerous fault not only propagates to the data qubits but also triggers the flag 18
2.7 Common topological QEC stabilizer codes 21
2.8 Circuits for single-qubit gate teleportation 23
2.9 The stabilizer octahedron lies within the Bloch sphere 24
2.10 Circuit blocks for the 5-to-1 magic state distillation protocol 27
2.11 A collection of code deformation techniques 30
2.12 The Steane code 32

3.1 The stabilizers $\{Z_1, Z_2\}$ of the initial state $|00\rangle$ are sequentially transformed into the stabilizers $\{X_1X_2, Z_1Z_2\}$ of the Bell state $|\Phi^+\rangle$ 38
3.2 Illustration of w-fault-subsets for single-circuit subset sampling 44
3.4 Box representation of different fault-weight-subsets for a single circuit with binomial factors A_w 48
3.5 Subset sampling for GHZ state preparation at $p_{\max} = 10^{-3}$ 50
3.6 The ERU criterion opens the next larger fault-weight-subset 50
3.7 Protocol illustrations for dynamical subset sampling (DSS) 52
3.8 Circuit sequences can be represented by nested boxes as in Fig. 3.7(c) 54
3.9 Subboxes of the A_1-box, labeled D and E, are weighted by the branching ratio q_i or $1 - q_i$ respectively 55
3.10 Visualization of the choice–measurement process from Fig. 3.7(b) as a circular flowchart 56
3.11 Example event trees with circuit nodes and subset nodes 58
3.12 Example tree from running the circuit c two consecutive times 59
3.13 We can recursively construct any event tree from circuit and subset node branchings 61
3.14 Example trees illustrate the cutoff error after a single shot of an FT protocol containing the circuits ENC and MEAS 62
3.15 Event tree examples that illustrate the cutoff error contributions from longer circuit sequences 63
3.16 The worst-case contribution of the orange 1-subset to the failure rate from Fig. 3.14(a) can be upper-bounded by $LA_1(1 - M_0)$ 64
3.17 An event tree of an FT protocol that illustrates the four cases of the FT cutoff error as given in the main text 65

3.18 Dynamical circuit sequences for adaptive FT state preparation of $|0\rangle_L$ in the Steane code . . . 67
3.19 Logical failure rate estimations for the adaptive FT state preparation protocol . . . 68
3.20 FT readout of the stabilizer plaquette K_1^X of the Steane code can be performed with a measurement auxiliary qubit and a flag qubit . . . 69
3.21 Longest possible circuit sequence that occurs when using the flag fault-tolerant protocol to encode the $|0\rangle_L$ state . . . 70
3.22 Logical failure rate estimations for the flag-FT stabilizer measurement state preparation protocol . . . 71

4.1 Sketch of the considered ion trap architecture and MS gate symbols . . 77
4.2 Multiple ions are illuminated by bichromatic laser light in order to perform the MS gate . . . 82
4.3 Illustration of all possible crosstalk faults on a single MS gate . . . 85
4.4 Logical Pauli state preparation . . . 88
4.5 Pauli state preparation circuit from Fig. 4.4(a) compiled into MS gates 88
4.6 Stabilizer expectation values are estimated numerically using depolarizing and incoherent overrotation noise . . . 90
4.7 Preparing two logical qubits and applying the logical CNOT gate . . . 92
4.8 Three-step FT magic state preparation . . . 93
4.9 The non-FT magic state preparation circuit from Fig. 4.8(a) is compiled into two-qubit MS gates . . . 94
4.10 The three blocks for FT magic state preparation from Fig. 4.8(a) are compiled into a single circuit containing only two-qubit MS gates . . . 95
4.11 Fault-tolerant T-gate realization . . . 96
4.12 Logical failure rates of state preparation circuits are obtained via numerical simulations where all physical fault rates are scaled uniformly . 99
4.13 Logical building blocks to deterministically and fault-tolerantly prepare Pauli and magic states . . . 102
4.14 Trade-off between fidelity and repetition overhead for state preparation protocols . . . 103
4.15 Flag-FT circuit for measuring the logical Hadamard operator . . . 104
4.16 Parallel stabilizer readout . . . 105
4.17 Sequential stabilizer readout . . . 105
4.18 Flowchart for the deterministic FT magic state preparation procedure 107
4.19 Preparation time advantage of non-deterministic FT logical state preparation schemes . . . 110
4.20 Decomposition of a CNOT gate into an MS gate $\mathrm{MS}_{0,0}(-\pi/2)$ and local rotations according to Eq. (4.6) . . . 111
4.21 Crosstalk-resistant circuit for FT Pauli state preparation . . . 111
4.22 Logical failure rates of FT Pauli state preparation variants . . . 112
4.23 Illustration of Steane-type EC on the logical level . . . 114
4.24 Circuit to fault-tolerantly prepare $|0\rangle_L$ in the Steane code without measurements . . . 116
4.25 Simplified decomposition of the Toffoli gate . . . 116
4.26 Measurement-free fault-tolerant quantum error correction cycle built from logical blocks . . . 117
4.27 Circuit on the level of physical operations for MF FT QEC in the Steane code . . . 118

4.28 The repetition overhead of the MFEC scheme can be transformed into a qubit overhead . . . 119
4.29 Simplified decomposition of the C_3NOT gate . . . 121
4.30 Comparison of logical failure rates for MFEC and flag EC . . . 122
4.31 Sketched initial configurations for implementations of MFEC . . . 123
4.32 Exact decomposition of the Toffoli gate into 3-qubit X-type MS gates and local rotations . . . 124
4.33 The expectation value of the stabilizer K_1^Z can be mapped to an auxiliary qubit by two 5-qubit X-type MS gates . . . 124
4.34 Decomposition of the C_3NOT gate into 4-qubit X-type MS gates and local rotations . . . 125
4.35 The cost function in Eq. (4.77) converges to the target value of -28 for the variational circuit in Fig. 4.34 . . . 125
4.36 Logical failure rates for the compiled MFEC scheme . . . 127
4.37 Regimes of advantageous use for MFEC and flag EC . . . 130

List of Tables

2.1 Matrix representations of two common universal gate sets 8
2.2 Common magic state distillation schemes and some of their properties 26
2.3 Look up table for the seven-qubit Steane code 33

4.1 Experimentally-informed fault rates for the depolarizing noise model used in our numerical simulations 87
4.2 Comparison of quantum state fidelity and code space population . . . 90
4.3 Input-output table for the logical CNOT gate 91
4.4 Experimentally-informed fault rates for the extended noise model . . . 98
4.5 Recovery operations based on the flag error set 101
4.6 Look up table for flag-FT measurement of the logical Hadamard operator in the deterministic scheme 108
4.7 Physical fault rates used in the multi-parameter depolarizing noise model to numerically simulate three different MFEC implementations 123
4.8 Angle parameters for the C_3NOT gate decomposition in Fig. 4.34 . . . 126
4.9 Realistic fault rates of physical operations in a neutral-atom platform . 131
4.10 Simulated logical failure rates for MFEC and flag EC in a neutral-atom platform 131

List of Abbreviations

CTR	**C**ross**t**alk-**r**esistant
DSS	**D**ynamical **s**ubset **s**ampling
EC	**E**rror **c**orrection
ED	**E**rror **d**etection
ERU	**E**xpected **r**eduction of **u**ncertainty
FES	**F**lag **e**rror **s**et
FT	**F**ault-**t**olerant
h.c.	**H**ermitian **c**onjugate
MC	**M**onte **C**arlo
MF	**M**easurement-**f**ree
MS	**M**ølmer-**S**ørensen
MSD	**M**agic **s**tate **d**istillation
MZM	**M**ajorana **z**ero **m**ode
NISQ	**N**oisy **i**ntermediate **s**cale **q**uantum
no.	**N**umber **o**f
PA	**P**hase-**a**veraged
PTA	**P**auli **t**wirling **a**pproximation
QEC	**Q**uantum **e**rror **c**orrection
QC	**Q**uantum **c**omputation
QCCD	**Q**uantum **c**harge-**c**oupled **d**evice
qLDPC	**Q**uantum **l**ow **d**ensity **p**arity **c**heck
QND	**Q**uantum **n**on-**d**emolition

Chapter 1

Introduction

Quantum mechanics has proven successful in providing the most accurate description of nature, applicable to the smallest length scales. In the realm of atoms, electrons, photons – quantum particles in short – definitive classical descriptions of, for instance, a particle's position and momentum are replaced by a quantum mechanical description in terms of operators. The theory makes probabilistic statements about measurement outcomes when one aims to observe a physical quantity. A growing interest on the interconnections of physics and information science [Whe90; Lan91] has manifested with practical real-world implications. The impressive road taken on by the development of classical computers has lead to the manufacturing of ever smaller devices that have eventually become so small that quantum mechanics must be consulted. Since nature, to the best of our knowledge and excluding gravitational effects, behaves fundamentally according to the laws of quantum mechanics, the storage and processing of information might as well be facilitated by following a quantum mechanical approach.

Feynman's idea of employing an original quantum system, sufficiently well under control, to simulate another quantum system of interest and study its properties has pushed open the door to the whole new field of quantum computation [Fey82]. Such computational devices are in vast demand for research. The number of degrees of freedom of a quantum system grows exponentially with the number of particles. Simulation of quantum systems quickly becomes practically intractable with classical computers even for systems with a moderate number of particles that can already be built in laboratories today. Analogously to a classical computer, built from physical components realizing binary digits (bits), a quantum computer can be built from quantum mechanical units that are commonly called *qubits* [Sch95; NC10][1]. A qubit-based quantum computer, however, need not be a single-purpose machine that simulates a specific, potentially large-scale, quantum system. It stores information in a register built from many qubits. Digital universal quantum computation can be implemented if one is able to perform some discrete native set of operations on the qubits by appropriate physical mechanisms in a programmable way. A digital universal quantum computer is a multi-purpose device. Nonetheless, during the writing of this thesis, we witness the first complex quantum simulations being carried out on large-enough (although not necessarily universal) quantum computers that actually allow the study of physical properties of exotic states of matter, which are hard to access otherwise [Sat$^+$21; Mi$^+$22; Goo23a; Xu$^+$23b; Iqb$^+$23].

The source of a potential computational advantage of quantum computers lies in the larger number of degrees of freedom compared to classical computers. This root

[1] Qudits – quantum mechanical d-level systems as units of quantum information processing – are out of scope for this thesis [Wan$^+$20b].

cause also grants *noise* more possibilities to corrupt the information stored in the computer. A lot more can go wrong in a quantum computer than just erroneously flipping a bit from 0 to 1 or from 1 to 0 in a classical computer. In fact, due to the prevalence of noise in current quantum computing devices hosting tens to thousands of physical qubits, the recent stage of development is commonly referred to as the Noisy Intermediate Scale Quantum (NISQ) era [Pre18]. NISQ devices are already capable of running *shallow* circuits on noisy physical qubits to solve problems such as finding the ground state of a small molecule by employing appropriate quantum algorithms [Bha$^+$22]. For a practical usage of quantum processors, it is imperative to keep the quantum computation coherent over a sufficiently long time despite all physical components being fundamentally noisy [Pre98]. Quantum error correcting codes aim to protect the computation by spreading the information non-locally over many qubits [DMN13]. The applications of such codes can be designed in a fault-tolerant way in order to minimize the deteriorating effect of noise [CDT09]. Provided that noise is weak enough to be tolerated by the particular code – a rule of thumb is an average operation fidelity of at least 99.9% –, encoding information into error correcting codes using multiple physical qubits can offer an advantage over using only individual physical qubits to store the information [Got09]. Consequentially, the coherence time of the encoded, logical qubits can be extended compared to single physical qubits. Then, quantum algorithms consisting of *deep* quantum circuits are envisioned to be performed on error-corrected logical qubits.

A low-noise error-corrected digital universal quantum computer holds the promise to run algorithms that asymptotically outperform the best classical supercomputers. The most famous examples of such algorithms include Shor's algorithm [Sho94] to find the prime factors of large numbers, which forms the basis for some modern encryption schemes, and Grover's algorithm [Gro96] for searching in an unsorted database. Other tasks, such as high-dimensional optimization problems, are believed to profit greatly from the availability of quantum computers [Mon16]. Fields of applications include communication [Zha09], metrology [Sim$^+$17], machine learning [SSP14; SSP15] and chemistry [McA$^+$20].

1.1 Physical platforms for quantum computation

In order to be deemed a quantum computer, any physical machine should comply with DiVincenzo's criteria [DL98]: It must be able to host a number of well-characterized qubits that can be initialized to a known state. The qubits should come with long coherence times and one must be able to perform a universal set of gates on them. Finally, it is necessary to be able to measure the state of the qubits. Ion traps and neutral-atom platforms can be identified as promising candidates for the realization of quantum computers and are in the focus of this thesis.

Ion traps are a well-established architecture that has demonstrated the capabilities as required for a quantum computing device (see [Bru$^+$19] and Refs. therein). Individual ions can be held in a trap for hours and the electronic state of the ion can be kept with a long coherence time. Single qubits, i.e. each ion, can be manipulated consistently via laser pulses that implement single-qubit gate operations. The creation of entanglement was proposed and realized with laser-based multi-qubit gate operations using a shared motional mode of the ions. Loading of up to 32 ions with individual, high-fidelity, universal control has been achieved in established ion trap quantum processing architectures [Mos$^+$23; Che$^+$23]. These platforms can reach coherence times up to the

order of seconds and achieve physical operation fidelities of 99.75% to 99.9975% for single-qubit operations and 97% to 99.8% for two-qubit operations [Pog$^+$21; Ega$^+$21; Mos$^+$23]. However, there is still room for improvement with demonstrated coherence times of more than an hour [Wan$^+$21] and even higher two-qubit operation fidelities in other systems [Gae$^+$16; Bru$^+$19] but it is challenging to uphold these values while scaling up to systems with more physical qubits.

Neutral-atom platforms on the other side are among the newest architectures entering the race of building a large-scale quantum computer (see [Hen$^+$20; Win$^+$23] and Refs. therein). They are catching up quickly [Saf16; Gra$^+$22; Ma$^+$22; Jen$^+$22; Bar$^+$22], with demonstrated capabilities to load a large number of atoms into an array of optical tweezers that can be held stable for a long time. By shining laser light onto the spatially well-separated atoms, gate operations can be performed in parallel. Especially entangling gates are possible to execute in parallel naturally this way on atoms that are in close proximity by exploiting the so-called Rydberg blockade that shifts the electronic energy levels due to a large but short-ranged atom-atom interaction [Lev$^+$19]. Hundreds of atoms can be loaded in these platforms [Win$^+$23] with coherence times on the order of seconds up to 10 s. For instance, the architecture in [Blu$^+$22; Eve$^+$23] has demonstrated primitives for quantum error correction and is capable of loading 60 atoms with a coherence time of 1 s. Typical fidelities of physical operations are currently 99.6% to 99.9% for single-qubit gates and 95.5% to 99.5% for two-qubit gates [Win$^+$23].

Other frontier architectures, which are by no means less promising, include superconducting transmons [Kja$^+$20; Bra$^+$22] as well as nitrogen-vacancy centers and quantum dots in solid-state systems [Aws$^+$18; Cha$^+$21]. Silicon-based systems with demonstrated high-fidelity operations could be particularly compatible with established industry manufacturing standards [BS19]. Quantum computers based on photons have also been established as viable platforms [Kok$^+$07; Arr$^+$21; Tab$^+$23]. Experiments have been conducted that claim a quantum advantage in the sense that no classical computer could have performed the particular task within a reasonable amount of space and time [Aru$^+$19; Zho$^+$20; Zho$^+$21; Wu$^+$21; Mad$^+$22; Kim$^+$23][2]. Nevertheless, we point out that quantum error correction can help to realize fault-tolerant quantum hardware that performs computations that are *practically useful*, i.e. quantum algorithms that solve practical problems.

1.2 Quantum error correction and fault tolerance

Quantum fault tolerance marks an important cornerstone within the framework of quantum error correction: All physical hardware should be designed in such a way that the effects of noise do not spread uncontrollably through the quantum computer [Sho96; Got98b; Got09; NC10]. In a similar way to DiVincenzo's criteria for physical qubits, *logical* qubits encoded in a quantum error correcting code must be operated fault-tolerantly [CTV17]. This means that one requires the ability to initialize the logical qubit state fault-tolerantly. Errors on the logical qubit must be diagnosed in a fault-tolerant manner, i.e. without introducing even more noise in the process. A universal set of fault-tolerant logical gate operations should be available. The final requirement is the capability to fault-tolerantly measure the state of the logical

[2]The claims of [Aru$^+$19] have been challenged by [IBMapa; Ped$^+$19; Liu$^+$21; PCZ22] and the claims of [Kim$^+$23] have been challenged by [Tin$^+$23; Lia$^+$23].

qubit. This last step means to retrieve the result of a quantum algorithm that is fault-tolerantly performed on the encoded, logical level. The realization of large-scale error-corrected quantum computation requires both theoretical and experimental research progress and in fact the two fields constantly propel each other.

The *theory* foundations involved in this work are commonly formalized within the quantum circuit model [NC10] where, inspired by electric circuit diagrams from classical electronics, qubits are drawn as horizontal wires and gate operations are represented by blocks acting on one or more qubits[3]. *Stabilizer* quantum error correction is the most prominent technique to promote classical error correction to the quantum level [Got98a; DMN13]. The measurement of a stabilizer operator that serves to diagnose potential errors on a logical qubit does *not* collapse the wavefunction of the ideal, error-free logical qubit. Errors, which may be coherent superpositions, *are* however collapsed by these measurements. In effect, continuous, unitary errors become correctable by discrete recovery operations in stark contrast to analog classical computation. Such intermediate measurements and in-sequence logic are required to apply this type of *active* quantum error correction: Errors are first detected by the measurements and then corrected by an appropriate recovery operation based on the classical measurement information.

We make use of *topological* quantum error correcting codes that encode information into global operators while the error diagnostic is a purely local process [LB13; Bom13]. This characteristic is advantageous for the implementation of fault tolerance requirements since we want to guarantee that noise, when unintentionally resulting from a local measurement, cannot have malicious effects on the global degrees of freedom. Most recently, the theory of *flag* fault tolerance [Got16a; YK17; CR18b; CR18c; CB18; CR20; Rei20] has pushed the chance of actual realization of these theoretical ideas much closer toward experimental capabilities. Compared to previous fault-tolerant schemes [Sho96; DA07], the necessary number of qubits and gate operations is significantly decreased. Let us remark that this theoretical framework for active quantum error correction should be carefully distinguished from other approaches of putting quantum computing hardware into practical use, such as adiabatic quantum computation [AL18], quantum error mitigation [Cai$^+$23], self-correction [Bom$^+$13; Ter15], autonomous quantum error correction [Ger$^+$21; HMM22] and bosonic quantum error correction [TCV20], which we do not consider further in this thesis.

Most *experiments* are currently occupied with realizing quantum memories [Ter15]. This means that they aim to encode a logical qubit and store it for some time, eventually extended with the help of quantum error correction counteracting the effect of noise, without performing a computational task using logical gates on the logical qubit. Encoding of logical qubits has been shown in ion traps [Nig$^+$14], superconducting transmons [Tak$^+$17], neutral-atom platforms [Blu$^+$22] and nitrogen-vacancy centers [Abo$^+$22]. Fault-tolerant error detection experiments include [Lin$^+$17] for ion traps and [Vui18; And$^+$20] for superconducting transmons. Fault-tolerant stabilizer measurements were demonstrated for ion traps [Hil$^+$22] and nitrogen-vacancy centers [Abo$^+$22]. Full quantum error correction cycles have been realized so far only in ion traps [RA$^+$21] and superconducting transmons [Goo21; Kri$^+$22; Zha$^+$22; Goo23b]. Most impressively, the experiment in [Goo23b] for the first time observed that increasing the number of noisy physical qubits that encode the logical qubit leads to further suppression of noise after performing quantum error correction. A major roadblock

[3] In this thesis, we are not concerned with measurement-based quantum computation [Bri$^+$09], which also allows for fault-tolerant quantum computation [ND05; RHG06; RHG07; Zha$^+$23a].

to realizing quantum error correction in most architectures is the persistent difficulty to perform mid-circuit measurements and condition the application of feedback operations on these measurement outcomes. This is because measurements are typically much slower than gate operations and might even reach the order of magnitude of the qubit coherence time [Bru+19; Win+23]. Although challenging, we see first experimental success in this direction for neutral-atom platforms [Gra+23; Nor+23; Lis+23; Hui+23; Sin+23] while fast-enough measurements do not seem to be a major issue of concern for the superconducting transmon experiment in [Goo23b] and the ion trap experiment in [RA+21], which also implements feed-forward corrections.

Fault-tolerant single-qubit operations can already be performed in most architectures [Ega+21; Mar+22; Abo+22] and logical two-qubit gates have been shown both non-fault-tolerantly [Erh+21] and fault-tolerantly [RA+22]. A fault-tolerant universal set of gate operations has up to this day only been demonstrated in the ion trap experiment of [Pos+22], however, without reaching the threshold for beneficial quantum error correction.

1.3 Structure of this thesis

Our focus is the advancement of fault-tolerant quantum error correction with the long-term goal in mind to realize error-corrected large-scale quantum computation. This thesis, for this purpose, is structured as follows. In Chapter 2, we introduce theoretical tools necessary to comprehend the recent development of the field of fault-tolerant quantum error correction. We discuss aspects of universal quantum computation and introduce the stabilizer formalism that constitutes the foundation for active quantum error correction as well as the threshold theorem of quantum fault tolerance. Then, we also outline the concept of flag fault tolerance, discuss its practical advantage and delineate how techniques for universal quantum computation can be realized in topological quantum error correcting codes. Chapter 3 comprises a discussion of techniques that classically simulate the effect of noise in a quantum computing device. Here, our main contribution is a new method for the efficient simulation of weak noise on quantum algorithms with in-sequence measurements where logical failure is a rare event. In Chapter 4, we present results that may serve as guidance to the development of new experiments and help to reach important milestones on the long road toward practical quantum computation. Firstly, sources of noise and the realization of a fault-tolerant universal gate set in a particular ion trap architecture are described followed by a thorough discussion of possible extensions. Secondly, we provide a novel measurement-free fault-tolerant quantum error correction scheme that might be particularly useful for neutral-atom platforms but also in ion traps. We discuss its applications, resource requirements, practical implementations and a possible advantage over other state-of-the-art error correction schemes. In Chapter 5, we draw conclusions and dare an outlook toward the potential future developments of our findings and the field of fault-tolerant quantum computation in general.

Chapter 2

Universal fault-tolerant quantum computation

It is natural to use a quantum system to represent binary units of information. In this chapter we review the fundamental theoretical framework necessary to construct an error-corrected quantum computer from noisy physical qubits. First, we outline how qubit systems can enable universal quantum computation (QC). With the help of quantum error correcting stabilizer codes, which we introduce in Sec. 2.2, the theory of fault tolerance can facilitate the operation of logical qubits with higher accuracy than bare physical qubits once certain threshold criteria are fulfilled, as we describe in Secs. 2.3 and 2.4. Finally, in Sec. 2.5, we show various ways of how topological systems permit the construction of universal fault-tolerant quantum computers and introduce a minimal model suitable for implementation in modern quantum computing hardware.

2.1 Qubit systems and universality

The most common basic unit to represent quantum information is the qubit [NC10]. Any physical two-level system that must be described by the laws of quantum mechanics is eligible to practically implement a qubit. The general pure qubit state

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{2.1}$$

with $\alpha, \beta \in \mathbb{C}, |\alpha|^2 + |\beta|^2 = 1$ is an arbitrary, normalized superposition of the computational basis states $|0\rangle$ and $|1\rangle$ in Dirac notation. These basis states are typically part of a more complex energy-level structure of a physical host system, for instance an individual atom or a superconducting circuit. When parametrized as

$$|\psi\rangle = \cos\left(\frac{\theta}{2}\right)|0\rangle + \mathrm{e}^{\mathrm{i}\phi}\sin\left(\frac{\theta}{2}\right)|1\rangle \tag{2.2}$$

with two angles $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi)$, the single-qubit state can be represented as a point on the surface of the three-dimensional Bloch sphere. The state may always be chosen up to an irrelevant global phase $\mathrm{e}^{\mathrm{i}\varphi}|\psi\rangle$. More generally, a statistical mixture of single-qubit states

$$\rho = \frac{1}{2}(I + \vec{\rho}\cdot\vec{\sigma}) \tag{2.3}$$

is represented by the Bloch vector $\vec{\rho} \in \mathbb{R}^3$ of length $|\vec{\rho}| \leq 1$, where I is the identity and $\vec{\sigma} = (X, Y, Z)$ is the vector of Pauli matrices and for pure states $\rho = |\psi\rangle\langle\psi|$.

For universal quantum computation, any n-qubit quantum state must be reachable with arbitrary precision starting from a generic product state $|0\rangle^{\otimes n}$. Two ingredients are sufficient to achieve universality. The first is the ability to perform rotations of single-qubit states by any angle θ about an arbitrary axis specified by a unit vector $\hat{n}$. These rotations are described by the rotation operators

$$R_{\hat{n}}(\theta) = \exp\left(-\mathrm{i}\frac{\theta}{2}\hat{n}\vec{\sigma}\right) = \cos\left(\frac{\theta}{2}\right)I - \mathrm{i}\sin\left(\frac{\theta}{2}\right)\hat{n}\vec{\sigma}. \tag{2.4}$$

If $\hat{n}$ is a Cartesian unit vector, we denote the rotation operator as $R_\sigma(\theta)$ with $\sigma \in \{X, Y, Z\}$. The second requirement is the ability to create entanglement between two states. The controlled-NOT (CNOT) gate uses the state of a control qubit (1) to condition its effect on a target qubit state (2) as

$$|x, y\rangle \overset{\mathrm{C_1NOT_2}}{\Longrightarrow} |x, x \oplus y\rangle \tag{2.5}$$

with binary variables $x, y \in \{0, 1\}$. Remarkably, it has been shown that a finite number of gate operations is sufficient for universal quantum computation, much like in classical computation the NAND gate alone is universal. For example, the CNOT gate can be combined with the Hadamard gate H and the gate $T = \exp\left(-\mathrm{i}\frac{\pi}{8}Z\right)$ to form the universal gate set $\{\mathrm{CNOT}, H, T\}$ [NC10], which we use extensively in Sec. 4.1.2 of this thesis. Another universal gate set is $\{\mathrm{Toffoli}, H\}$ [Shi03] and we summarize properties of both gate sets in Tab. 2.1.

Universal gate sets		Matrix representation	Properties
Gate	Gate	of gate	
CNOT		$\mathrm{diag}(I,\, X)$	Clifford
T		$\mathrm{diag}(1,\, \mathrm{e}^{\mathrm{i}\frac{\pi}{4}})$	non-Clifford
H	H	$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$	Clifford
	Toffoli	$\mathrm{diag}(I,\, I,\, I,\, X)$	non-Clifford

Table 2.1: Matrix representations of two common universal gate sets. Each gate set must contain a non-Clifford gate because the Clifford group alone is not universal. Here, I is the two-dimensional identity matrix.

Using H and T, the Solovay-Kitaev theorem ensures that any single-qubit rotation can be approximated to in principle arbitrary precision ε using a finite sequence of $\mathcal{O}(\log^c(1/\varepsilon))$ steps, where c is a constant [Kit97; Sol00; NC10]. The T-gate plays a special role. It is a *non-Clifford* gate that extends the set of *Clifford* gates to universality. The latter is comprised by all gates that map Pauli matrices to Pauli matrices and thus belong to the n-qubit Clifford group

$$C_n = \{U \mid UPU^\dagger \in G_n \; \forall P \in G_n\} \tag{2.6}$$

with unitary matrices U and G_n representing the n-qubit Pauli group $\langle X, Y, Z\rangle^{\otimes n}$, generated by the Pauli matrices on each qubit individually. A circuit, which only contains Clifford gates, qubit initializations to Pauli bases and measurements in Pauli bases is called a Clifford circuit. The Clifford group, despite not being universal on its own, is crucial for fault-tolerant quantum error correction, as we lay out in in the following sections.

2.2 Quantum error correction

Multiple physical qubits can be composed to gain control over more degrees of freedom than just the two-dimensional subspace of a single physical qubit. These extra degrees of freedom, i.e. the states of additional qubits, are used in the context of quantum error correction (QEC) to diagnose errors that may happen to a qubit [DMN13]. For useful quantum computation to become a reality, two opposing goals must be balanced. On the one hand, the qubit state should be as isolated from the environment as possible so that coherence times are longer than the time we wish to operate the qubits. On the other hand, perfect isolation is not desirable since it would render it impossible to actually perform gates and thus computation on the qubits. Consequently, prevalence of some form of noise on physical qubits is inevitable. Active QEC provides a prescription on how to manipulate a quantum system in order to diagnose and correct errors, which we describe in this section. We first introduce the concept of stabilizers in order to subsequently explain how stabilizer codes are used for active QEC.

2.2.1 Stabilizer formalism

The stabilizer formalism is an efficient way to describe stabilizer states[1], alternative to the statevector notation, of n qubits [Got97]. In statevector notation, the n-qubit state is denoted as a vector with 2^n complex entries in a Hilbert space. An equivalent representation is to define the state via the set of 2^n operators to which the state is an eigenvector with eigenvalue $+1$:

$$\mathcal{S} = \{S_i \mid S_i|\psi\rangle = +1 \cdot |\psi\rangle, S_i \in G_n, \forall i = 1, \ldots, 2^n\}. \tag{2.7}$$

For stabilizer states, only n of these operators are linearly independent and are thus sufficient to generate the abelian stabilizer group $\mathcal{S}$. So it is a large simplification to denote the state by its stabilizer *generators*. For a single-qubit state for example, the states $|+\rangle$, $|+\mathrm{i}\rangle$ and $|0\rangle$ ($|-\rangle$, $|-\mathrm{i}\rangle$ and $|1\rangle$) are the $+1$ eigenstates of $(-)X$, $(-)Y$ and $(-)Z$ respectively, so we can write $\pm X$, $\pm Y$ and $\pm Z$ to uniquely identify the respective states $|\pm\rangle$, $|\pm\mathrm{i}\rangle$ and $|0/1\rangle$.

A gate operation U that is applied to a stabilizer state $|\psi\rangle$ to produce the state $U|\psi\rangle$ can just as well be written as an application to any stabilizer $S \in \mathcal{S}$ of the state like

$$U|\psi\rangle = US|\psi\rangle = USU^\dagger U|\psi\rangle \tag{2.8}$$

so that the stabilizer S is mapped to $USU^\dagger$, which stabilizes the state $U|\psi\rangle$. As an example, consider the Hadamard gate H applied to the state $|0\rangle$ which creates the equal superposition state $H|0\rangle = |+\rangle$. Equivalently, we can denote the transformation as

$$HZH = X \tag{2.9}$$

in terms of the single-qubit stabilizers $Z|0\rangle = |0\rangle$ and $X|+\rangle = |+\rangle$.

[1] Pure stabilizer states are all n-qubit states that can be generated from $|0\rangle^{\otimes n}$ using only Clifford operations.

2.2.2 Stabilizer codes

The stabilizer formalism unfolds its potential to design codes for quantum error correction [Got98a][2]. In a quantum code, we aim to encode logical information in the code word states $|0\rangle_L$ and $|1\rangle_L$ of a quantum system with more degrees of freedom than a single physical qubit with states $|0\rangle$ and $|1\rangle$. Stabilizer codes provide an elegant way to fulfill the quantum error correcting conditions [KL97], which state that for two correctable errors E_i and E_j acting on two code words $|\psi_k\rangle$ and $|\psi_l\rangle$ the overlap

$$\langle\psi_k|E_i^\dagger E_j|\psi_l\rangle = \alpha_{ij}\delta_{kl} \tag{2.10}$$

must vanish for different code words or be a complex number for the same code word. This complex number, encoded in the Hermitian matrix α, is the same for any two errors acting on any code word $|\psi_k\rangle$ since we may not obtain any logical information by measuring $\langle\psi_k|E_i^\dagger E_j|\psi_k\rangle$. The code must be designed in such a way that correctable errors E_i and E_j are mapped to orthogonal subspaces $E_i|\psi_k\rangle$ and $E_j|\psi_l\rangle$ regardless of the underlying code words, assuming $k \neq l$. A crucial finding is that correcting Pauli errors is sufficient for a code to correct any type of noise that has a representation in terms of Kraus operators[3] since any set of Kraus operators can be expanded in the basis of Pauli matrices [NC10].

The *code space* of a stabilizer code using n qubits with parameters $[[n, k, d]]$ is generated by a number of $n-k$ linearly independent stabilizer generators $g_1, \ldots, g_{n-k}$. All 2^{n-k} stabilizer operators $S_i, S_j \in \mathcal{S} = \langle g_1, \ldots, g_{n-k}\rangle$ mutually commute $[S_i, S_j] = 0$, $\forall i, j$. The code space C is the simultaneous $+1$ eigenspace of all stabilizers

$$C = \{|\psi\rangle \mid S|\psi\rangle = +1 \cdot |\psi\rangle \ \forall S \in \mathcal{S}\}, \tag{2.11}$$

which is spanned by the 2^k logical states $|\psi_1, \psi_2, \ldots, \psi_k\rangle_L$ where each $\psi_i \in \{0, 1\}$, $\forall i = 1, \ldots, k$. There are k pairs of logical operators $X_{L,1}, Z_{L,1}, \ldots, X_{L,k}, Z_{L,k}$ that act on their respective subspace of the logical state as the single-qubit Pauli matrices X and Z act on the computational basis states $|0\rangle$ and $|1\rangle$. At the same time, the logical operators leave the code space invariant, i.e. they map stabilizers to stabilizers. Thus, we say that the $[[n, k, d]]$ code encodes k logical qubits. The logical operators commute with all stabilizer operators but mutually anticommute $\{X_{L,i}, Z_{L,i}\} = 0$, $\forall i = 1, \ldots, k$. The minimal number of non-trivial single-qubit Pauli operators that can jointly form a logical operator is defined as the *distance* d of the code. Two operators A and B are called "stabilizer-equivalent" if multiplication of A by some stabilizer $S_i \in \mathcal{S}$ yields $B = S_i A$.

All operators in the n-qubit Pauli group G_n either commute or anticommute with any element in $\mathcal{S}$. A Pauli error $E \in G_n$ on a logical state $|\psi\rangle_L$ can be detected by application of a stabilizer S_i like

$$E|\psi\rangle_L = S_i E|\psi\rangle_L = (-1)^m E S_i|\psi\rangle_L = (-1)^m E|\psi\rangle_L \tag{2.12}$$

if $m = 1$, i.e. the stabilizer S_i anticommutes with E. If the error commutes with *all* stabilizer operators and is thus undetectable, it is either itself a stabilizer or a logical operator. The group formed by these operators is called the centralizer of $\mathcal{S}$.

[2] Aspects of efficient classical simulation will be discussed in Sec. 3.2. Other applications that make use of stabilizer states, besides quantum error correction, include quantum communication, for instance quantum key distribution [NC10], and certain variants of randomized benchmarking [Dan+09; MGE12; Hel+22], which are out of scope for this thesis.

[3] More detail will be given in Chapter 3.

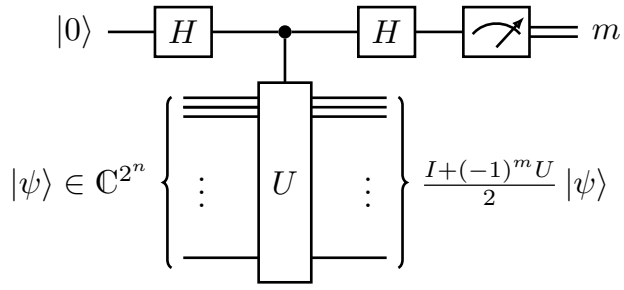


FIGURE 2.1: Quantum circuit to measure the n-qubit observable U with the help of an additional measurement qubit. The n-qubit state $|\psi\rangle$ is projected onto the respective $(-1)^m$ eigenstate of U by the measurement. We assume that the reader is familiar with the standard quantum circuit notation [NC10]. Here m can be either 0 or 1. We refer to the value of the measurement outcome as $(-1)^m$.

We see from Eq. (2.12) that the eigenvalue of the stabilizer that detects the error E is flipped to -1 compared to the $+1$ eigenvalue on the ideal code word $|\psi\rangle_L$. Stabilizer operators can be measured with the help of an auxiliary qubit via the circuit shown in Fig. 2.1. These are the two fundamental building blocks to stabilizer error correction: Reproducing the single-qubit Pauli algebra on the logical level formed by n qubits and detecting errors via measurements of stabilizer operators.

In general, we may measure $n - k$ linearly independent stabilizer operators. This set of measurement outcomes is called the *syndrome*. In practice, errors cannot be observed directly without measuring the quantum state. We can only measure the syndrome and must deduce from the syndrome what is the most likely error that has probably occurred to cause the given syndrome. However, there is no guarantee that this process of *decoding* is always correct. Two different errors E_i and E_j might cause the same syndrome. If we confuse one error for the other and apply a recovery operation $R = E_j$, assuming $E_j = E_j^{-1}$, such that the product $R \cdot E_i$ is equal to a logical operator, we can inflict a failure on the logical qubit.

Note that the recovery R need not necessarily be applied as a physical gate as long as R is a Pauli operator. Any Pauli operator P is mapped to another Pauli operator P' by subsequent Clifford operations (see Eq. (2.6)) so one can keep track in a classical register, called the Pauli frame, how to decode, i.e. interpret, any measurement result of physical qubits [Kni05; Rie+17]. This has the advantage that one does not introduce more noise by the extra physical recovery operation. Only before applying a non-Clifford gate, the Pauli corrections have to be applied on the actual quantum state because a Pauli operator P is mapped to an operator outside the Pauli group by the non-Clifford gate in general.

It is common to perform decoding on classical measurement data after the quantum circuit since this can be done noise-free (assuming perfect classical computation) instead of running a final round of QEC with noisy quantum hardware in order to check whether a noisy quantum algorithm has terminated with failure. As a simple example, consider the bitstring $m = 001$, which may be the result of the error X_3 when using a noisy circuit to encode a logical state in the three-qubit repetition code with code words $|0\rangle_L = |000\rangle$ and $|1\rangle_L = |111\rangle$ and then measure all three qubits in the Z-basis. We can represent the Z-stabilizers of the code $S_1 = Z_1Z_2$ and $S_2 = Z_2Z_3$ as binary strings $s_1 = 110$ and $s_2 = 011$ respectively, obtain the classical syndrome

as the overlap of m with s_1 and s_2 as $(0, 1)$ and conclude, without performing additional quantum measurements, that the bitflip has likely happened to the third bit and interpret the measured bitstring m as logical zero.

2.2.3 Fidelity measures

We use two different figures of merit to assess the quality of logical quantum states. The *logical failure rate* p_L of a QEC protocol is the average infidelity $1 - F$ of output states ρ, where the logical fidelity

$$F = \langle P_{\pm O} \rangle \tag{2.13}$$

is the expectation value of the projector

$$P_{\pm O_L} = \frac{I \pm O_L}{2} \tag{2.14}$$

onto the ± 1 eigenstate of a logical operator O_L. The logical fidelity describes the overlap of the logical Bloch vector of the respective logical state ρ_i, which would be the output of an ideal or noiseless protocol, with the logical Bloch vector of the actual noisy output state ρ. For instance, if we ideally expect the $|0\rangle_L$ state, then the logical fidelity of a noisy state ρ with respect to $|0\rangle_L$ is given by $\langle P_{Z_L} \rangle = (1 + \langle Z_L \rangle)/2$. Since the logical fidelity is only defined in the code space, we always perform one round of ideal EC to obtain a final measurement result as described above. Then the logical fidelity reflects the probability to successfully recover the state after performing a noisy circuit.

On the other hand, the *quantum state fidelity* not only measures how well a state is prepared in the logical subspace but also contains information about the full state such as quality of the stabilizers that prepare the code space in the first place. The quantum state fidelity is the expectation value

$$\mathcal{F}(\rho_i, \rho) = \langle \rho_i \rangle = \mathrm{tr}(\rho_i \rho) \tag{2.15}$$

of the ideal quantum state ρ_i with respect to the noisy state ρ. Any quantum state is characterized by the quantum state fidelity independent of any QEC framework. For a general n-qubit state the quantum state fidelity can be expressed as

$$\mathcal{F}(\rho_i, \rho) = \frac{1}{4^n} \mathrm{tr} \left(\sum_{j=1}^{4^n} [\mathrm{tr}(S_j \rho_i) S_j] \rho \right) \tag{2.16}$$

where we expand the general n-qubit state ρ_i in the operator basis that is spanned by all 4^n possible n-qubit Pauli operators S_j [Nig+14]. For stabilizer states only the 2^n stabilizer elements $S_j \in \mathcal{S}$ give a non-zero coefficient $\mathrm{tr}(S_j \rho_i) = \pm 1$ so that

$$\mathcal{F}(\rho_i, \rho) = \frac{1}{2^n} \sum_{j=1}^{2^n} \langle S_j \rangle. \tag{2.17}$$

We can write the ideal state ρ_i as a product of projectors P_{CS} onto the code space – the $+1$ eigenspace of the stabilizer generators g_j – and its logical operator $\rho_i = P_{\pm O_L} P_{\mathrm{CS}}$ with

$$P_{\mathrm{CS}} = \prod_{j=1}^{n-k} \frac{I + g_j}{2} \tag{2.18}$$

and the sum in Eq. (2.17) originates from multiplying out the products of projectors that define the state ρ_i.

2.3 Fault tolerance and the threshold theorem

At the heart of fault-tolerant quantum computation lies the threshold theorem, which guarantees that a quantum computation can be kept under control as long as errors are corrected faster than they occur [KLZ96; AGP06; ABO08]. This puts an upper bound on the inaccuracy that a single step of the quantum computation must abide by: A quantum algorithm that consists of g gates should return the desired result with constant error $\mathcal{O}(1)$, meaning that – on average – any gate should contribute an error of at most $\mathcal{O}(1/g)$. Although for practical experiments the threshold is in general a complicated hypersurface in the high-dimensional space of error parameters, which severely depends on the quantum code and noise processes in a given architecture, it is remarkable that the notion of a threshold can be derived from more general considerations.

A code capable of correcting t errors that randomly and independently occur with probability p will fail with probability $p_L \sim cp^{t+1}$ as $p \to 0$ with a constant c. If we concatenate j layers of encoding, this logical failure rate can be suppressed like

$$p_L^{(j)} \sim c^{\sum_{k=0}^{j-1}(t+1)^k} p^{(t+1)^j} = c^{-1/t}(c^{1/t}p)^{(t+1)^j} \tag{2.19}$$

as $p \to 0$ and in the limit of $j \to \infty$ we expect $p_L \equiv 0$ provided that $p < 1/\sqrt[t]{c}$. For many useful quantum codes and noise models, it can be shown that the constant c indeed has a finite value so that a threshold exists [Got09; DMN13; Rof19]. While the number of physical qubits needed to concatenate j code layers increases exponentially with j, we see that $p_L^{(j)}$ is suppressed double-exponentially. This means that the *larger* number of noisy qubits can lead to *smaller* logical failure rates compared to systems with fewer noisy qubits. Finite values of j can already suffice to suppress $p_L^{(j)}$ to a desired level. The crossing point of the $p_L^{(j)}$ with different finite j marks the *pseudothreshold*.

In general, the pseudothreshold is the crossing point p_{pt} where the logical failure rate

$$p_L(d > 1, p = p_{\mathrm{pt}}) = p_L(d = 1, p = p_{\mathrm{pt}}) \tag{2.20}$$

of QEC codes of the same code family with finite distance d are equal to that of an unencoded qubit, given a specific noise model [Ter15]. The code's threshold p_{t} then results from taking the thermodynamic limit of an infinitely large code and is defined via

$$\lim_{d\to\infty} p_L(d, p = p_{\mathrm{t}}) = p_L(d = 1, p = p_{\mathrm{t}}). \tag{2.21}$$

Codes that do not feature a threshold may still be useful to suppress noise in practice using finite-distance members of such a code family [NP12; BP13]. Values for thresholds can change with additional assumptions about the form of the noise [BNB19].

In the above consideration we only included errors that occur on a quantum code directly. In the following we review modern approaches to fault tolerance and incorporate the noisy circuitry needed to practically implement QEC in a real noisy hardware into the framework of fault tolerance.

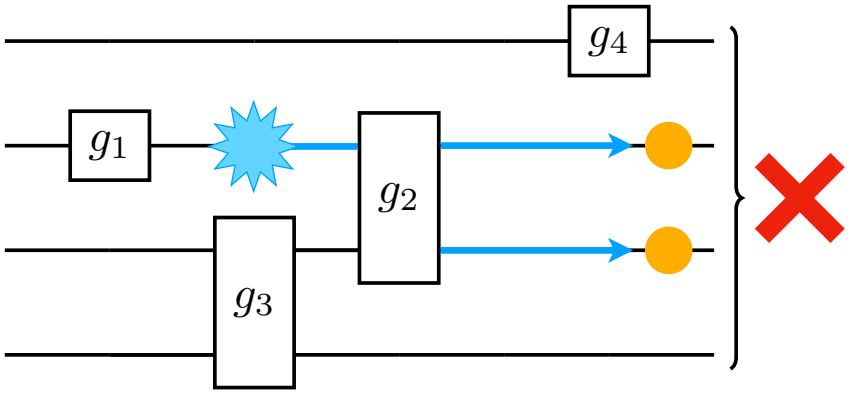


FIGURE 2.2: Generic quantum circuit with qubits (wires) and gates (boxes). Illustration of a fault (blue star) that arises from the faulty gate g_1 and propagates through the subsequent gate g_2. The propagated error (orange circles) can result in failure after decoding (red cross). The gates g_3 and g_4 are unaffected by the fault.

2.4 Fault-tolerant circuit design

The general notion of fault tolerance must be translated into practical prescriptions for *fault-tolerant circuit design* for actual applications [Pre98; CTV17]. Merely knowing that a threshold exists does not yet help in devising concrete quantum circuits to encode logical qubits and execute logical operations on them. In this section, we review the tools for fault-tolerant circuit design that enable the advantageous operation of logical over physical qubits in a regime of sufficiently weak noise.

2.4.1 Definitions

Within the framework of stabilizer codes, noise is discretized as Pauli operators. Although noise on physical operations can in principle be unitary, it is known that QEC decoheres noise over time through encoding and stabilizer measurements [Bea$^+$18; IP20]. For fault-tolerant (FT) circuit design, we always consider circuit-level noise, i.e. all components of a quantum circuit must be modeled as prone to noise as they would be in an actual experiment [RH07]. Other noise types such as phenomenological noise [Den$^+$02; Ohn$^+$04] or code capacity noise [KBMD09] do not take the microscopic details of the noisy circuit components into account (a summary of various noise models is given, for example, in [LAR11]). They rather deal with code characteristics instead or assume simplifications or limiting cases such as, for instance, perfect Clifford gates.

When considering noisy quantum circuits, we distinguish three types of noise events that are sketched in Fig. 2.2: faults, errors and failures. A *fault* is an instance of noise on a *circuit location*, e.g. a single Pauli operator F applied after an ideal circuit operation U, which could be any gate operation or qubit initialization, or before a measurement. When we assume single faults to happen stochastically and independently with a probability p, a particular configuration of m faults can occur with probability p^m in leading order. The *order* m is also called the *fault weight*. An *error* is the resulting operator on the logical qubit after executing the noisy circuit, e.g. a Pauli fault caused an error E as in Eq. (2.12). Here, the *Pauli weight* is the number of qubits an operator acts on with non-trivial Pauli operators. For example, the error

$E = IXIXIXI = X_2X_4X_6$ acting on a seven-qubit state has weight $\mathrm{Wt}(E) = 3$. Note that the order m is in general different from the Pauli weight since, for example, the single fault operator F can have $\mathrm{Wt}(F) \leq q$ if U is a q-qubit gate. We define the *error weight* of an error E in a code with stabilizer group $\mathcal{S}$ as the minimal Pauli weight $\mathrm{wt}(E) = \min_{S\in\mathcal{S}} \mathrm{Wt}(SE)$ where the minimum is taken with regards to all operators that are stabilizer-equivalent to E. A *failure* is a wrongly decoded state as the result of an error of weight $w > t = \lfloor \frac{d-1}{2} \rfloor$ in an $[[n, k, d]]$ QEC code, which uses n physical qubits to encode k logical qubits with distance d, i.e. the minimal Pauli weight of a logical operator O_L is $\mathrm{wt}(O_L) = d$. The three types of events are sketched in Fig. 2.2. A fault is called *dangerous* when it can lead to failure and the location of a dangerous fault is referred to as a *bad* location [Cha18]. There will always exist some faults of order $t + 1$ that cause failure, i.e. they lead to uncorrectable errors of weight $t + 1$ and thus application of an unintended logical operator when performing QEC.

A quantum circuit is deemed "fault-tolerant at level t" (FT-t) when all possible fault configurations up to order t on all circuit locations never lead to failure. The logical failure rate p_L of an FT implementation of a distance d QEC code thus scales as $p_L \sim p^{t+1}$ in the limit of low physical fault rate $p \to 0$. Crucially, the weight of the *error* determines whether it is correctable or causes failure and the probability of occurrence for the microscopic *fault* configuration determines the order in which it contributes to the polynomial for p_L in case of failure. However, in case multiple faults can occur in a correlated fashion between different circuit locations, such as for crosstalk on entangling gates that we discuss in detail in Chapter 4, the leading-order exponent $t + 1$ for the polynomial scaling behavior of logical failure rates can be reduced and we say that fault tolerance is not respected by these types of faults.

2.4.2 Fault propagation

The effect of individual Pauli faults in Clifford circuits can be examined via fault propagation. Since Clifford gates map Pauli operators to Pauli operators, swapping the order of Pauli faults and Clifford gates can be used to determine the error at the end of the circuit resulting from a fault within the circuit. For example, the Hadamard gate transforms between different Pauli types since $HZ = XH$. More importantly, faults can proliferate through the circuit via CNOT gates, as Fig. 2.3 illustrates, since

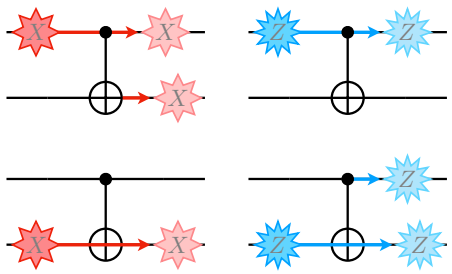


FIGURE 2.3: Propagation of X- and Z-faults through a CNOT gate.

$$X_1X_2\,\mathrm{C_1NOT_2} = \mathrm{C_1NOT_2}\,X_1 \quad \text{and} \tag{2.22}$$
$$Z_1Z_2\,\mathrm{C_1NOT_2} = \mathrm{C_1NOT_2}\,Z_2. \tag{2.23}$$

A fault of low Pauli weight can propagate to an uncorrectable error so that even if the fault is of order $m \leq t$ we can end up with an error of weight $w > t$ if the circuit is designed in an unfavorable way. Up to t faults can in principle be prevented from propagating to cause errors of weight larger than t by advantageous circuit design.

It is however challenging to find such circuits in practice and a significant effort is made to automate fault-tolerant circuit compilation (for a summary, see [Pal+17]). Two

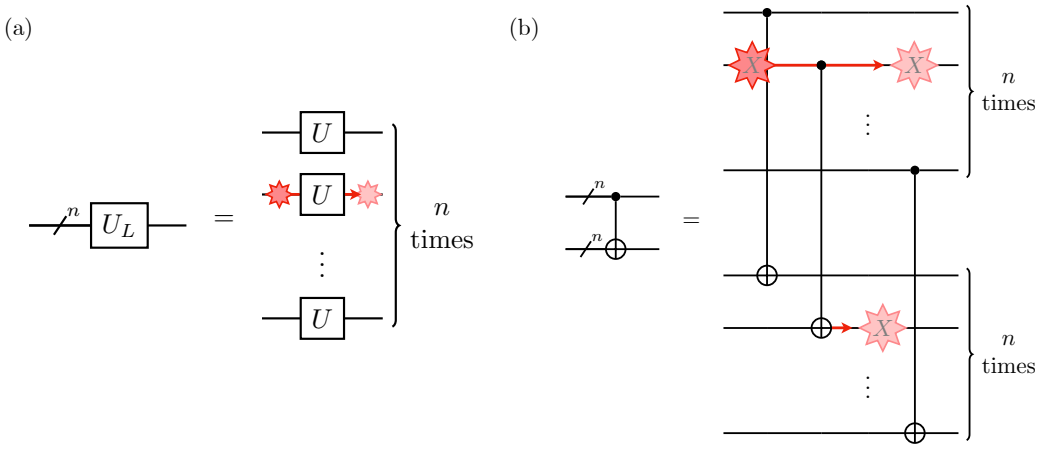


FIGURE 2.4: For some quantum codes, some logical gates can be implemented in a bit-wise fashion. Transversal single-qubit gates, as in (a), and transversal two-qubit gates, as the transversal CNOT gate in (b) are inherently fault-tolerant because qubits are never coupled within the same logical block. Consequentially, faults (red stars) cannot spread uncontrollably throughout any block. Note that, in general, not all physical operations U need to be the same. Also, the operations on the physical level may differ from the logical one.

guiding principles have emerged that can facilitate the design of FT circuits: repetition and transversality [Pre98]. The repeated execution of circuit blocks, for example to measure the eigenvalue of an operator, adds redundancy and can help to protect against concluding a wrong measurement result in case of errors. Transversal gates prohibit the spread of faults within single code blocks and are thus FT by definition: In Fig. 2.4, logical gates are realized by bitwise application of physical gates so that no coupling between the physical qubits that form the same logical qubit is required. Both principles are central ingredients to Shor-type error correction [Sho96]. As one of the first proposals for FT QEC, we briefly discuss Shor-type error correction as an illustrative example in the following. It is outperformed by more recent developments such as the flag-qubit paradigm, which enables fault tolerance with very little qubit overhead. A connection between the two is described subsequently.

2.4.3 Shor-type error correction

Shor-type error correction makes use of both repetition and transversality in order to achieve fault tolerance [Sho96]. Two other techniques that were developed early on but presumably need more physical qubits are Steane-type error correction [Ste97], which requires encoded auxiliary qubits and will be reviewed in Sec. 4.2.1, and Knill-type error correction [Kni04; Kni05], which relies on teleportation. Using a bare auxiliary qubit to measure a stabilizer operator makes the measurement non-FT since single faults can cause both the measurement to yield a flipped result and corrupt the logical information on the logical data qubit. In Shor-type EC, the bare auxiliary qubit is replaced with a pre-verified Greenberger–Horne–Zeilinger (GHZ) state

$$|\mathrm{GHZ}\rangle = \frac{|00...0\rangle + |11...1\rangle}{\sqrt{2}} \tag{2.24}$$

as illustrated in Fig. 2.5 [GHZ89]. This multi-qubit state is coupled transversally to the data qubit block so that the stabilizer value is encoded in the parity of the GHZ state, i.e. the ± 1 eigenvalue is represented by the state $(|00...0\rangle \pm |11...1\rangle)/\sqrt{2}$.

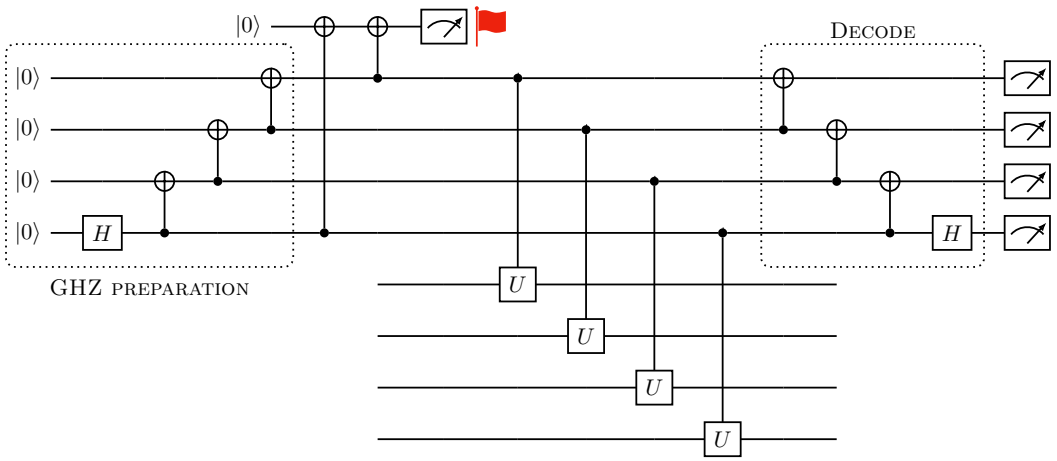


FIGURE 2.5: Circuit to measure the stabilizer operator $U^{\otimes 4}$ in a QEC code with distance $d = 3$ fault-tolerantly, reproduced from [DMN13]. The flag qubit verifies FT preparation of the auxiliary four-qubit GHZ state. The preparation circuit and the flag measurement must be repeated until the flag is clear. The parity of the GHZ state encodes the stabilizer eigenvalue. For an FT measurement, the circuit must be applied two or three times until the measurement result after decoding can be determined via majority voting.

Therefore the information is protected against X-errors by the size of the GHZ state. Verification of the GHZ state can be done, for the four-qubit example in Fig. 2.5, via an extra qubit called a "flag", which heralds dangerous faults when it is *triggered*, i.e. measured as -1. The GHZ state preparation must be repeated until the flag is *clear*, i.e. measured as $+1$. Then it is guaranteed that no single dangerous fault can propagate from the GHZ state to the data qubits. Only then it can be used for the stabilizer readout. This procedure alone does not protect against measurement faults or Z-errors on the GHZ state. In order to be fully FT, the GHZ preparation and transversal stabilizer readout must be repeated until the stabilizer value can be determined via majority voting from the individual measurements. For a distance d code, at least $t + 1$ and at most $(t + 1)^2$ repetitions are necessary to perform the majority vote [Cha18].

Later on, it has been suggested by DiVincenzo and Aliferis that the extra flag qubit and the repetition overhead needed for GHZ verification can be saved by employing an alternative decoding block on the GHZ state *after* the stabilizer readout [DA07]. In this scheme the number of required auxiliary qubits is determined by the Pauli weight of the measured stabilizer. The flag qubit paradigm has reduced the qubit overhead for FT measurements further (and a historical overview is given, for instance, in [PR21]).

2.4.4 Flag fault-tolerant error correction

Instead of preparing the GHZ state *before* coupling to the logical data qubit, the physical CNOT gates that encode and decode a GHZ state can be interspersed between the CNOTs of the transversal gate. In the fault-free case, we still prepare a GHZ state this way, but all individual faults that can occur in the circuit can be propagated in a unique way so that the flag pattern, i.e. the measured auxiliary qubits, can be used to correct all errors up to the fault tolerance level t.

Flag qubits act as heralds for dangerous faults [CR18c]. Flag circuits are designed in a way that, when all flag qubits are clear, no dangerous fault of order t or smaller

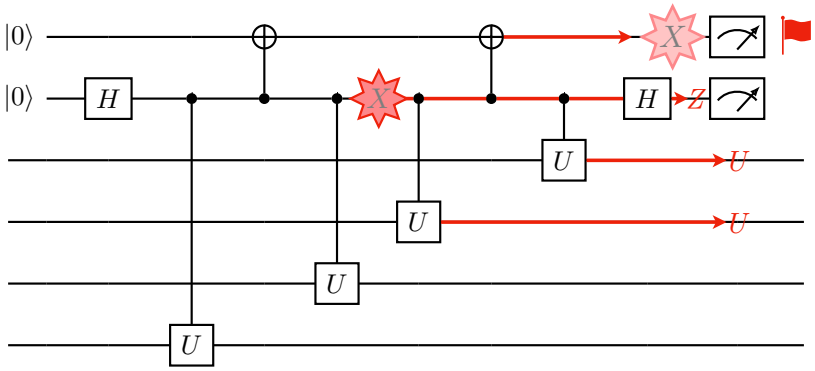


FIGURE 2.6: In a flag circuit for FT stabilizer readout, a dangerous fault not only propagates to the data qubits but also triggers the flag. The appropriate two-qubit recovery operation can be applied, which corrects the weight-2 error, instead of the non-FT look up table decoder that would lead to failure in this case.

can have occurred. When a flag is triggered, it indicates that a dangerous fault may have occurred. As a result, one can decide to abort a protocol and start a new trial or adapt the decoding strategy in accordance with the additional flag information. With the help of flags, which are added to a non-FT circuit, the decoding strategy can be extended to include all bad locations of the non-FT circuit.

For example, a distance-3 code enables the correction of all weight-1 errors with a standard look up table decoder. While a non-FT stabilizer readout circuit can cause uncorrectable weight-2 errors from bad locations, the extra degrees of freedom provided by adding a flag qubit can suffice to also correct the weight-2 errors. Note that *not all* weight-2 errors must be corrected but only the *specific* ones that can actually follow from the bad locations in the circuit. The circuit in Fig. 2.6 contains a bad location where the otherwise dangerous fault is propagated to the flag qubit, which triggers the respective two-qubit recovery operation. Only two auxiliary qubits are required instead of the four-qubit GHZ state in Shor-type EC.

This is the paradigm-shift that came with the invention of flag circuits [Got16a; YK17; CR18b; CR18c; CB18; CR20; Rei20; PR21]: Instead of protecting the stabilizer value by the size of the auxiliary state it is sufficient to map all uncorrectable errors that can occur with probability up to order $m \leq t$ to a unique flag-measurement pattern.

The expected improvements for practical implementations are large. Not only is the qubit overhead for flag fault tolerance reduced compared to Shor-type EC. It has been shown that the repetition overhead for stabilizer measurements can be reduced from $(t+1)^2$ to $\binom{t+2}{2}$ [Cha18]. Also, fewer CNOT gates are required to couple the flag qubits to the measurement qubit. On devices with limited qubit connectivity, such as superconducting transmon systems, flag-bridge circuits can enable flag-FT QEC by redistributing the connections of the physical CNOT gates while maintaining the fault-tolerance property of the circuit design [LA20].

2.4.5 Advantage of fault-tolerant protocols

It is clear that FT protocols will outperform their non-FT counterparts eventually at small enough physical fault rates p. The polynomial suppression of failures in an FT protocol $p_L^{(\mathrm{FT})} \sim c_{\mathrm{FT}} \cdot p^{t+1}$ makes it advantageous to use over a non-FT protocol where $p_L^{(\text{non-FT})} \sim c_{\text{non-FT}} \cdot p^m$ with $m \leq t$ in the limit $p \to 0$. The constants $c_{(\text{non-})\mathrm{FT}}$ are determined by the number of bad locations in the respective circuits. FT implementations typically require more qubits and gate operations and thus have more locations than their non-FT counterparts so we expect $c_{\mathrm{FT}} > c_{\text{non-FT}}$. It is imperative to keep c_{FT} as small as possible, i.e. the overhead to achieve fault tolerance should be minimized, in order to keep the noise strength p for which the use of FT circuits is advantageous as large as possible [Got14].

As long as an unencoded physical qubit, which fails with $p_L^{(\mathrm{phys})} \sim p$ (corresponding to the special case of $m = 1$ above), preserves logical information better than an encoded, logical qubit, it is not desirable to perform the encoding in the first place. For a true practical advantage of FT implementations over physical qubits, Gottesman suggested to compare the failure rates of the respective encoded and unencoded operations as achievable each in the best way within the same quantum computing architecture under consideration [Got16b]. One may also opt to compare only the dominant physical fault rates in a specific experiment, which we elaborate further in Sec. 4.1.3.1. Note that these advantage criteria are different from the (pseudo-)threshold, which is concerned about varying the distance of the underlying QEC code and eventually taking the thermodynamic limit $d \to \infty$. To judge upon an FT advantage we compare different implementations of logical operations for codes with the *same* distance and to physical operations.

In the following section, we give an overview of topological QEC codes and various methods to achieve universal fault-tolerant quantum computing using surface codes and color codes as prominent representatives.

2.5 Topological QEC codes

The theoretical development and discovery of topological phases of matter and topological order marked a paradigm shift in condensed matter physics beyond the Landau theory of phase transitions [Ber13; Zen$^+$19]. Contrary to purely local order parameters like magnetization that take on a finite value in a particular phase but vanish in another, it was conceived that different matter phases can be distinguished by (non-local) topological invariants. For example, topological insulators are characterized by a finite topological invariant and show a finite gap in their (bulk) electronic band structure, which prohibits adiabatic deformation to the atomic limit of the topologically trivial band insulator. Topologically ordered phases typically show ground state degeneracies and long-range entanglement.

In the two-dimensional toric code model, which is sketched in Fig. 2.7(a), the ground state is fourfold degenerate with zero, one or two non-contractible loops wrapping around the two independent directions of the underlying topologically non-trivial manifold, the torus [Kit97; Sat$^+$21]. Elementary excitations in the toric code are local quasi-particles with anyonic exchange statistics. Many of these local excitations are required to change the expectation value of a global operator spanning across a non-contractible loop and thus take the system to another ground state. This is the general working principle for topological QEC codes. The logical information

is encoded into global operators, which are protected by the system size, i.e. the code distance, while the error diagnostic is done via local stabilizer operators. A two-dimensional random-bond Ising model with quenched disorder on a square lattice with periodic boundary conditions can be mapped to a noisy toric code where stabilizer eigenvalues are randomly flipped [Den$^+$02; Vod$^+$22]. By solving the Ising model one can obtain a noise threshold for when the QEC code will fail to protect the logical state because a logical operator is inferred, which corresponds to a percolating string of spin flips in the Ising model.

More common than the toric code, due to its experimentally challenging periodic boundary conditions, are surface codes [BK98]. These flat cousins of the toric code have the same bulk structure but come with open boundary conditions, as shown in Fig. 2.7(b), which offer additional possibilities to manipulate logical information. Due to its high threshold on the order of $p \approx 0.8\%$ under circuit-level depolarizing noise, the family of surface codes ranges among the most prominent candidates for the realization of FTQC [Ste14]. Another variant of topological QEC codes are two-dimensional color codes, which are constructed by tiling a manifold with a three-colorable lattice (see Fig. 2.7(c)) [BMD06]. On each plaquette, one places distinct X- and Z-type stabilizers in a symmetric way. Logical operators of a color code are stringnets that span across the whole lattice, which can have both periodic and open boundary conditions. An example for the latter are triangular color codes [BMD06]. The threshold of the color code family under circuit-level depolarizing noise has been numerically estimated to $p \approx 0.3\%$ [Kub18]. Approximate values for the surface and color code thresholds have been reproduced in [Cha$^+$20][4].

Both surface and triangular color codes each encode a single logical qubit ($k = 1$) into their n physical qubits with a distance $d \sim \sqrt{n}$ in two spatial dimensions. They are examples of the more general class of quantum low-density parity check (qLDPC) codes, where the number of physical qubits involved in each stabilizer and the number of stabilizers acting on each physical qubit are bound by a constant for all members of a particular code family [Got14; BE21]. The low Pauli weight of stabilizer operators that form such codes allows one to measure them in practice with a small number of, potentially long-range, connections to the data qubits (see Fig. 2.6) in order to practically perform QEC. This more general framework of qLDPC codes is currently under investigation to realize codes that preserve a finite ratio of logical per physical qubits in the thermodynamic limit $n \to \infty$ while keeping the advantages, for instance a high threshold, of surface or color codes [Bra$^+$23; Xu$^+$23a].

Some gate operations can be performed relatively easy in surface codes or color codes. For example, logical X- or Z-operators are implemented by applying physical operations in a bitwise fashion along appropriate edges of the system. Triangular color codes are particularly neat in two spatial dimensions because all Clifford gates can be realized transversally here. However, it is known by the Eastin-Knill theorem that no QEC code can have a transversal universal gate set [EK09]. Thus, for any code, additional techniques are required to reach universality with fault-tolerant quantum computers.

[4]Note that numerical threshold values highly depend on the concrete circuits, the microscopic noise model and the specific decoding strategy.

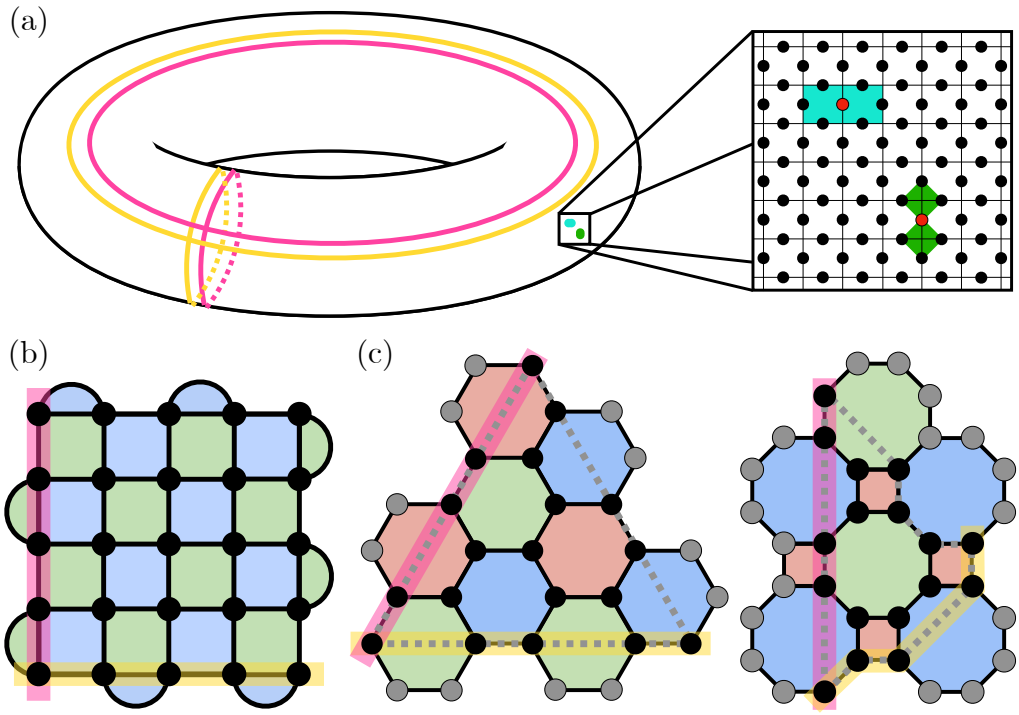


FIGURE 2.7: Common topological QEC stabilizer codes. All of them have local stabilizers and global logical operators. Since the stabilizers have a constant Pauli weight and every physical qubit takes part in a constant number of stabilizers, the shown codes belong to the class of qLDPC codes (see main text). (a) Torus with two non-contractible loops in each direction. Two quasi-particles are shown as green and blue dots. Inset: Qubits (dots) are placed on edges of a square lattice. Single-qubit errors (red dots) can excite stabilizers acting as star (green) and plaquette (blue) operators on four qubits each. Errors along a non-contractible loop around the torus correspond to applying a logical operator. The toric code encodes two logical qubits by the two X- and Z-type strings wrapping around the two directions on the torus (yellow and pink). (b) The toric code is deformed to open boundary conditions and rotated by 45° to realize a finite size surface code patch, which encodes a single logical qubit. Logical operators (yellow and pink) connect opposite boundaries. (c) Tilings of the 2D plane in, for instance, a 6.6.6 or 4.8.8 structure form three-colorable lattices that can be used to define stabilizers of color codes with qubits represented by vertices of the lattice. The qubits enclosed by the gray dashed lines belong to triangular color codes with examples of logical operators shown along the boundaries of each code. For these codes, some plaquettes are reduced from containing 6 or 8 qubits to only 4 qubits.

2.5.1 Paths to universality

Different strategies to realize universal gate sets in fault-tolerant quantum computing platforms have been proposed. In the following, we discuss the three most commonly pursued paths to universality: Realization of a physical system that hosts non-abelian anyons. Universal quantum computation is aimed to be performed through the exchange statistics of these exotic quasi-particles. Another route employs preparing so-called magic states, which act as a resource that is consumed during the execution of circuits that effectively realize non-Clifford gates. Lastly, we describe code deformation techniques. Here, instead of realizing gates in a complicated way, the goal is to change the underlying code so that the desired gate is easy to execute.

It is worth noting that these different approaches are not mutually exclusive. Ideas from one section can usually also be used to facilitate one of the other approaches. These connections will be pointed out in the respective subsections.

2.5.1.1 Braiding in topological quantum computers

In a topological quantum computer, as suggested by Kitaev [Kit97; Kit03], logical qubits are non-local degrees of freedom that are protected by topological properties of the underlying physical system. For example, the Kitaev chain has a topological insulator phase where edge states on each end of the one-dimensional chain can only be destroyed by fluctuations that span across the whole system from one edge to the other [Kit01]. These edge states are Majorana zero modes (MZMs) [SFN15]. MZMs have non-abelian exchange statistics. Only the composite particle of two MZMs behaves like a fermion. Information can be encoded and manipulated by moving anyons around each other, i.e. by braiding their world-lines, and thus imprint non-trivial phase factors onto the wavefunction of the system, which can be probed by measurements. For quantum computation with anyons, the exchange matrix R and the fusion matrix F represent the fundamental operations ("gates") on logical qubit states [Nay$^+$08].

MZMs belong to the class of Ising anyons, which by themselves can implement Clifford operations. For Ising anyons, the non-trivial F-matrix is a representation of the Hadamard gate and the non-trivial R-matrix is equivalent to the phase gate. In a quantum computer relying on Ising anyons, universality must be achieved by enhancing the set of possible gates by a non-Clifford gate, e.g. via magic states (see Sec. 2.5.1.2).

Another flavor of non-abelian anyons are Fibonacci anyons. They may be used to realize a universal gate set in their braiding properties, since their braiding operators are dense in SU(2) [FLW02]. Explicit circuit constructions have been given to realize a Fibonacci code [BD12].

Evidence has been reported on first experimental hardware realizations of physical systems that may host MZMs [Agh$^+$23; Dvi$^+$23]. In an experiment with superconducting qubits, braiding of Ising anyons has been demonstrated with the help of twist defects in 2D surface code patches [Goo23a; Xu$^+$23b]. Additionally, non-abelian quasi-particles have been "realized" as a quantum simulation with trapped-ion qubit architectures [Iqb$^+$23]. These states are also candidate explanations for the 5/2 state of the fractional quantum Hall effect [SFN15]. It is suspected that the 12/5 fractional quantum Hall state contains a Fibonacci sector but there is no experimental realization of topological universal quantum computation as of today [GSF21].

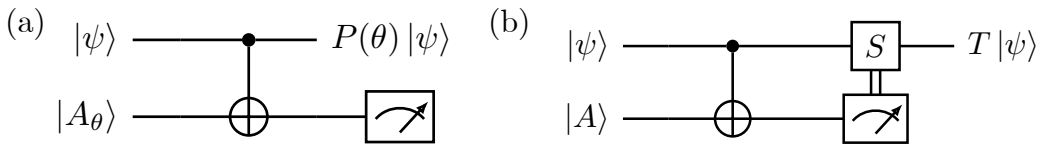


FIGURE 2.8: Circuits for single-qubit gate teleportation. (a) The gate $P(\theta) = \mathrm{diag}(1, \mathrm{e}^{\mathrm{i}\theta})$ is applied to an arbitrary input state $|\psi\rangle$ when the measurement result is $+1$. One must post-select non-deterministically on the desired measurement outcome or repeat the teleportation circuit with a shifted angle $\theta \to 2\theta$ when the measurement yields the -1 outcome. (b) By applying an extra S-gate conditioned on the non-trivial measurement result, the T-gate can be realized via magic state injection deterministically. The S-gate can, for instance, be implemented transversally in two-dimensional color codes. In surface codes, code deformation techniques (see Sec. 2.5.1.3) or other auxiliary-qubit-assisted circuits (see [AP08; Fow+12] for example) can be used.

2.5.1.2 Magic state injection

Any gate can be realized with the help of appropriate teleportation circuits, given an auxiliary input state that mediates the desired gate [GC99]. For example, single-qubit gates can be performed using the circuits in Fig. 2.8. In Fig. 2.8(a), an arbitrary input state $|\psi\rangle$ is entangled to the state $|A_\theta\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle + \mathrm{e}^{\mathrm{i}\theta}|1\rangle\big)$ via the CNOT gate [Pre98]. Then the qubit state is collapsed by the measurement in order to apply the gate $P(\theta) = \mathrm{diag}(1, \mathrm{e}^{\mathrm{i}\theta})$ and yield the output state $P(\theta)|\psi\rangle$ in case the measurement result is $+1$. Because the measurement yields the outcomes ± 1 randomly, this teleportation circuit is *non-deterministic*.

Given that a feedback operation can be conditioned on the measurement outcome, the teleportation can be made *deterministic*. The teleportation circuit in Fig. 2.8(b) realizes the non-Clifford T-gate by injecting an appropriate magic state $|A\rangle \equiv \big|A_{\pi/4}\big\rangle$ into the Clifford circuit [NC10]. For a -1 measurement outcome, an additional phase gate $S = \mathrm{diag}(1, \mathrm{i})$ corrects the wrongly projected state. The circuit then always outputs the state $T|\psi\rangle$ for any input state $|\psi\rangle$ at the price of consuming the resource magic state $|A\rangle$.

Teleportation on the logical level requires that a logical entangling gate can be performed. Recall that for surface codes and color codes, the logical CNOT is implemented transversally (and thus fault-tolerantly) by bitwise application of physical CNOT gates between the physical qubits of two code patches. If the transversal CNOT gate is not available in a given quantum computing architecture, other techniques such as lattice surgery or defect braiding (see Sec. 2.5.1.3) can be employed to realize the logical CNOT gate.

With the help of magic state injection, performing high-fidelity non-Clifford gates reduces to preparing high-fidelity magic states. It is commonly assumed that the Clifford gates involved in magic state preparation and injection can be performed without noise. We stress again that the magic state is a consumable resource in the sense that it is a necessary input to the teleportation circuit and the magic state on the input qubit is "gone" after the measurement when the non-Clifford gate is performed.

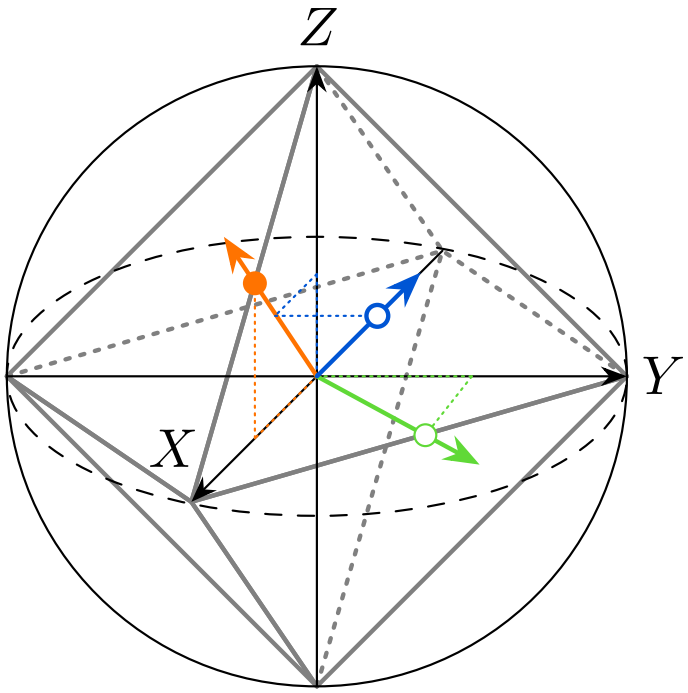


FIGURE 2.9: The stabilizer octahedron (gray, thick lines) lies within the Bloch sphere. Magic states correspond to intersections of symmetry axes (colored arrows) of the stabilizer octahedron with the Bloch sphere. Intersection points with the octahedron are indicated for the T-type magic state ρ_T (blue, thick dot, Eq. (2.27)) and the two H-type magic states ρ_A (green, thin dot, Eq. (2.28)) and ρ_H (orange, filled dot, Eq. (2.29)). Symmetric magic states can be obtained from the corresponding intersections with all edges and faces of the octahedron.

Magic state distillation

In what follows, we briefly review a class of procedures to produce high-fidelity magic states. Later on, we discuss aspects of fault tolerance and practical infidelity requirements for non-Clifford gates for large-scale quantum computation, for which we consider Shor's factoring algorithm as an illustrative example.

T-type and H-type magic states were discussed by [Rei05; BK05][5]. These states lie on points of the Bloch sphere that correspond to certain symmetry axes of the stabilizer octahedron[6], which is parametrized by the components of the Bloch vector $\vec{\rho}$ fulfilling $|\rho_x| + |\rho_y| + |\rho_z| \leq 1$, as illustrated in Fig. 2.9. Eight symmetric T-type magic states lie on those points on the Bloch sphere that are determined by the axes that pierce through the centers of any of the eight octahedron faces. Twelve symmetric H-type magic states lie on the corresponding Bloch sphere surface's points given by the axes cutting any octahedron edge in half.

The infidelity of magic states is typically assumed as stemming from twirled noise. For instance for the magic state $|H\rangle = \cos\left(\frac{\pi}{8}\right)|0\rangle + \sin\left(\frac{\pi}{8}\right)|1\rangle$, which is the $+1$ eigenstate

[5]Another type of magic state, the $|CZ\rangle$ state, has recently been prepared fault-tolerantly with higher fidelity than the corresponding unencoded magic state in an experiment with superconducting transmons [Gup+23].

[6]These are all states that can be generated, starting from $|0\rangle$, by noisy Clifford operations.

of the Hadamard operator H, the superoperator

$$\mathcal{H}(\rho) = \frac{1}{2}\rho + \frac{1}{2}H\rho H^\dagger \tag{2.25}$$

maps any state ρ to a mixture of the form

$$\rho(p) = (1-p)|H\rangle\langle H| + p|{-H}\rangle\langle{-H}| \tag{2.26}$$

with $0 \leq p \leq 1$ where $|{-H}\rangle$ is the state $Y|H\rangle$ orthogonal to $|H\rangle$ [MEK12]. It is called a twirling operator since, by stochastically applying the identity operation or the Hadamard gate H with equal chance, it decoheres any state into the eigenbasis of H.

The initial proposal to yield high-fidelity magic states by a purification routine of mixtures like Eq. (2.26) became known as *magic state distillation* [BK05]. Here, the idea is to obtain a single magic state with high fidelity from many magic states with lower fidelity using an appropriate distillation circuit. Bravyi and Kitaev have shown for the state

$$\rho_T = |T\rangle\langle T| = \frac{1}{2}\left(I + \frac{X+Y+Z}{\sqrt{3}}\right) \tag{2.27}$$

that distillation can be achieved by measuring the stabilizer generators of the five-qubit code [Laf$^+$96; Ben$^+$96] that has a transversal gate $\mathrm{e}^{\mathrm{i}\pi/4}SH$, to which ρ_T is an eigenstate. In the scheme, one post-selects on the trivial syndrome and all other states are discarded. Then, applying a decoding circuit of the code to an accepted state yields the distilled magic state. A similar routine uses the $[[15,1,3]]$ code for the H-type magic state

$$\rho_A = |A\rangle\langle A| = \frac{1}{2}\left(I + \frac{X+Y}{\sqrt{2}}\right) \tag{2.28}$$

that is related to $|H\rangle$ via a Clifford gate as $|A\rangle = \mathrm{e}^{\mathrm{i}\pi/8}HS^\dagger|H\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle + \mathrm{e}^{\mathrm{i}\pi/4}|1\rangle\big)$. In this code, the logical X- and T-gates are transversal and ρ_A is the $+1$ eigenstate to the operator $TXT^\dagger$. Their distillation routines, called "5-to-1" and "15-to-1", can be applied repeatedly to distill a single magic state of any desired fidelity, given enough low-fidelity magic states. Later, Meier, Eastin and Knill proposed a different scheme, called "10-to-2", based on an error detecting $[[4,2,2]]$ code with distance $d = 2$ [MEK12]. Here one encodes two magic states of the form

$$\rho_H = |H\rangle\langle H| = \frac{1}{2}\left(I + \frac{X+Z}{\sqrt{2}}\right), \tag{2.29}$$

then measures the transversal Hadamard operator[7] to project onto the logical magic state with the help of eight more magic states and lastly applies a decoding block to obtain two output magic states with higher fidelity than the input magic states. Note that while both distillation methods rely on post-selection, not all achieve an acceptance rate R of 1 in the limit of noiseless magic states. Some characteristics are shown for a selection of protocols for magic state distillation in Tab. 2.2.

Further protocol examples

Consider the 5-to-1 protocol as sketched in Fig. 2.10, which works on an input product state $\rho_T^{\otimes 5}$ with twirled noise on each ideal magic state. For an initial noise strength

[7] Measurement of the logical Hadamard operator was already suggested in [Kni04] using the seven-qubit Steane code.

Scheme (n_{in}-to-n_{out})	5-to-1	15-to-1	7-to-1	10-to-2	$3k$+8-to-k	49-to-1
Magic state	$\|T\rangle$	$\|A\rangle$	$\|H\rangle$	$\|H\rangle$	$\|A\rangle$	$\|A\rangle$
Code	$[[5,1,3]]$	$[[15,1,3]]$	$[[7,1,3]]$	$[[4,2,2]]$	$[[3k+8,k,2]]$	$[[49,1,5]]$
Error rate	$5p^2$	$35p^3$	$7/9p$	$9p^2$	$(1+3k)p^2$	$1411p^5$
Acc. rate R	$1/6$	$1-15p$	$9/128$	$1-10p$	$1-(8+3k)p$	-
Threshold p_{th}	0.173	0.141	0.146	0.089	$(3k+1)^{-1}$	0.1366
Cost exponent γ	2.32	2.46	-	2.32	$\log_2 \frac{3k+8}{k}$	2.42
Reference	[BK05]	[BK05]	[Rei05]	[MEK12]	[BH12]	[BH12]

TABLE 2.2: Common magic state distillation schemes and some of their properties for the states $|T\rangle$, $|H\rangle$ and $|A\rangle$. Note that $|A\rangle$ is related to $|H\rangle$ via Clifford gates. The underlying quantum code is given and the error rate and acceptance rate refer to a single distillation round the low-p limit. The threshold p_{th} denotes the error rate below which the scheme improves the fidelity of the output magic state(s). The cost exponent (see main text) can be defined as $\gamma = \log_s(\frac{n_{\text{in}}}{n_{\text{out}}R})$ with the number of input and output states of a distillation protocol, its average acceptance rate R and its error rate scaling exponent s, which we both consider in the low-p limit. Due to the linear scaling of the output error rate for "7-to-1", the cost does not have a polylog scaling for this protocol. Further distillation protocols using the [[23,1,7]] Golay code and the [[7,1,3]] Steane code can be found in [Rei05] and [Fow$^+$12] respectively.

of $p_0 = 0.1$, the output error rate $p_1 = 5p_0^2$ after one round will be $p_1 = 0.05$; an improvement factor of 2. For n distillation rounds the output error rate $p_n = \frac{1}{5}(5p_0)^{2^n}$ is double-exponentially suppressed. For instance, four rounds yield an improvement factor of $p_0/p_4 = 32768$ using $(5 \times 6)^4 = 8.1 \times 10^5$ input magic states on average in the low-p limit due to the finite acceptance rate $R = 1/6$. The closer p_0 is to the threshold, the more distillation rounds n are needed to reach a desired high fidelity and $n(p_0 \to p_{\text{th}}) \to \infty$ diverges logarithmically at the threshold value p_{th}.

For the application of any protocol one must trade-off their (dis-)advantages and there is no conclusive answer, which protocol is the "best". For example, although the 5-to-1 protocol has the largest threshold, i.e. it can distill high-fidelity magic states from the most-noisy input states up to a noise strength of $p_{\text{th}} = 0.173$, in practice it might be more feasible to use a protocol with a large acceptance rate, such as 10-to-2, when the input states can be prepared with less noise in the first place so that a high threshold is not needed. Bravyi and Haah combine two ideas from both methods, namely projecting into a code with a transversal gate from [BK05] and using distance-2 codes from [MEK12], to devise the "$3k+8$-to-k" protocol from a code construction using triorthogonal matrices, which distills the state $|A\rangle$ [BH12]. They introduce the distillation cost C as the number of raw, noisy input magic states N divided by the number of output magic states M that are distilled to a lower error rate ε after several distillation rounds, which scales like

$$C = \mathcal{O}(\log^\gamma(1/\varepsilon)). \tag{2.30}$$

The cost exponent γ is defined as

$$\gamma = \log_s\left(\frac{n_{\text{in}}}{n_{\text{out}}R}\right) \tag{2.31}$$

with the number of input states n_{in} and output states n_{out} of a single round of the distillation protocol, its average acceptance rate R and its error rate scaling exponent s in the low-p limit. The protocol from [BH12] requires fewer qubits in the large-k limit, where the cost exponent $\gamma \approx 1.6$ (see Tab. 2.2). However, when $k \to \infty$,

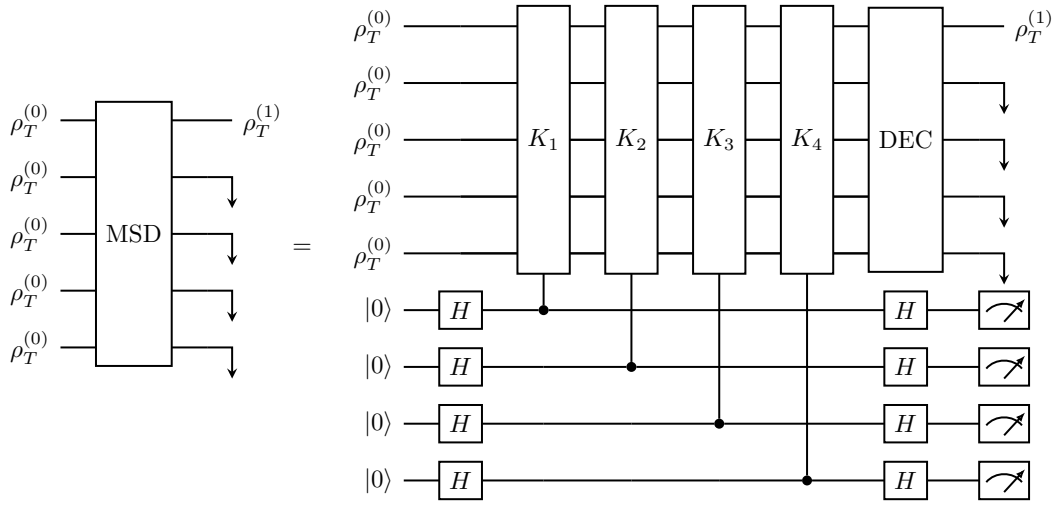


FIGURE 2.10: Circuit blocks for the 5-to-1 magic state distillation (MSD) protocol (see Tab. 2.2). The 5-to-1 protocol distills five noisy magic states of the form $\rho_T(p_0)$ to a single magic state $\rho_T(p_1)$ with $p_1 < p_0$ if p_0 is smaller than a distillation threshold p_{th}. The MSD block consists of measuring the stabilizers K_1, K_2, K_3, K_4 of the five-qubit code followed by a decoding block (DEC). Here, we may, for instance, use the decoding circuit of the five-qubit code from [DP16] and an additional single-qubit Clifford operation to obtain $\rho_T^{(1)}$. The output state can be accepted if the trivial syndrome is measured and discarded otherwise.

the protocol does not have a finite threshold. This conceptional problem is purely theoretical and does not cause practical problems at finite k. Using larger codes, the error rate suppression per round of distillation can be increased at the cost of requiring more physical qubits, for instance with a "49-to-1" protocol based on a 49-qubit code with a transversal T-gate [BH12].

To conclude our review of magic state distillation, we briefly mention various recent works on the matter. Numerical studies show that sequential application of different protocols instead of repeatedly applying the same protocol can yield better distillation results [MEK12; BH12]. Intuitively it is sensible to use high-threshold protocols for initial rounds and high-acceptance protocols for later rounds, which agrees with the numerical findings of [BH12]. It has been found that using entangled states instead of single-qubit states as magic states can improve the distillation outcome [BCS22]. A large body of research is considered with developing a resource theory of magic and classifying the set of distillable states [Vei$^+$14; HC17; OC17; CH17; CB09; CB10]. For example, the authors of [OC17] state that the cost of magic state distillation can be much lower than the cost to perform error correction with a surface code. In Ref. [Lit19], the size of surface code patches is systematically reduced to a minimal qubit overhead at the price of larger circuit depths. The encoding rate $1/n$ of the surface code with n physical qubits is identified as the major cost factor instead of magic state distillation. We point out that multi-qubit gates, such as the Toffoli or CCZ gate, can also be realized with appropriate magic states [Eas13; HH18a; GF19]. Surprisingly, codes with distillation cost exponents $\gamma \leq 1$ have been found [Haa$^+$17; HH18b]. Multi-level distillation deals with combining magic states of different levels of purification [Jon13].

Let us now discuss aspects of fault tolerance for the preparation of high-fidelity magic states.

Fault tolerance

Recall that the standard assumption for magic state distillation is that the Clifford circuits used for the distillation can be executed *noise-free* [JO$^+$12; Bro13][8]. Since the actual noisy distillation circuits are not FT, their circuit failure rate puts an upper bound on the fidelity that can be achieved in practice by performing multiple rounds of distillation. Suppose that after one round of the 5-to-1 protocol the output error rate p_1 from noisy input states with error rate p_0 is modified by the noisy distillation circuit to become

$$p_1' = 5p_0^2 + cp_d. \tag{2.32}$$

Here $c > 1$ is a constant and p_d is the noise strength for gates in the non-FT distillation circuit. Applying the distillation twice will yield

$$p_2' = 5p_1'^2 + cp_g = 5(5p_0^2 + cp_d)^2 + cp_d \tag{2.33}$$

and for an arbitrary number of n distillation rounds the resulting failure rate

$$p_n' = 5p_{n-1}'^2 + cp_d \tag{2.34}$$

cannot become smaller than cp_d. The apparent solution to this problem is to use magic state distillation only on the logical level, i.e. on logical qubits that are already encoded in a suitable QEC code, so that – despite of the decoding block of the distillation routine – the "inner" QEC code will still protect against faults in a way that preserves general fault tolerance. For example using the seven-qubit Steane code for additional protection, a single round of the 5-to-1 protocol would require at least 5 logical data qubits and 4 logical auxiliary qubits, each encoded by 7 physical qubits, leading to a total of 63 physical qubits and at least 262 entangling gates[9]. While these numbers may still be feasible for current or near-term experimental devices, performing multiple rounds of distillation leads to an exponential increase in the number of qubits.

An alternative to magic state distillation is to prepare the logical magic state in a fault-tolerant manner directly [Got16a; CC19]. An FT magic state comes with a failure rate $\sim p^{t+1}$ that is suppressed by the number of correctable errors t in the respective code compared to the small-enough fault rate of physical components p (see Sec. 2.4.1). The FT Clifford circuit for logical magic state injection preserves fault tolerance of the overall output state $T_L|\psi\rangle_L$ where the logical T-gate is applied to a logical input state $|\psi\rangle_L$. FT magic state preparation can be achieved by performing a quantum non-demolition (QND) measurement of an operator that will project the input state onto an eigenstate of the respective operator. Since the corresponding operators are known for the most common magic states, the problem reduces to fault-tolerantly measuring the respective operator. This can be done efficiently, for example, with the

[8] The assumption of twirled noise on the input magic states can be lifted for an appropriate choice of codes [HH21a].

[9] We introduce the Steane code in Sec. 2.5.2. 7 physical CNOTs are needed for a single logical CNOT gate. 4 logical CNOTs are required for each of the four stabilizer readouts in the protocol. 9 entangling gates are used for the decoding of the five-qubit code using the circuit from [DP16]. The circuits to encode logical Pauli and magic states from [Got16a] contain 8 and 11 CNOTs respectively. An additional large number of CNOT gates would be required by QEC units, which we do not quantify further.

help of flag qubits (see Fig. 2.6). When all building blocks of magic state injection are FT, the overall failure rate $\sim p^{t+1}$ will eventually outperform the distillation failure rate p_d at low physical fault rates. In Sec. 4.1.2, we use this approach for the smallest two-dimensional topological color code, the Steane code, to fault-tolerantly prepare the logical $|H\rangle$ state. In Ref. [CC19], the authors find that this method can lead to substantial reductions in qubit count and CNOT gate count compared to the 10-to-2 protocol. Circuits for larger-distance FT magic state preparation in color codes are provided in [CN20]. A magic state prepared in a color code can be teleported to a surface code via lattice surgery as was shown in [SC22] (also see Sec. 2.5.1.3).

Large-scale quantum computation

Let us now consider, as an example, the accuracy requirements of a large-scale quantum algorithm such as Shor's factoring algorithm [Sho94] and the implications on the noisy hardware that is used to practically realize the quantum computation. Shor's algorithm requires run time $\mathcal{O}(L^3)$ to factor an L-bit integer. The number of necessary non-Clifford gates also scales like L^3. Thus we need on the order of 10^{10} non-Clifford gates to factor a 2048-bit number, such as used in RSA encryption, using Shor's algorithm (also see the estimation in [Fow$^+$12]). To perform the algorithm with constant error, an error rate of at most $\varepsilon = 10^{-10}$ per non-Clifford gate would be required. Using FT magic state preparation, a code of distance d will yield a logical failure rate proportional to $p^{\lfloor \frac{d+1}{2} \rfloor}$ in the low-p limit. So, given that physical operations fail with rates $p \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-6}\}$, we need at least a code of distance $d \in \{9, 7, 5, 3\}$, correcting $t \in \{4, 3, 2, 1\}$ errors, to obtain a magic state failure rate lower than ε respectively. The numerical study in [CN20] however suggests that the low-p limit is hard to reach. For example, a physical fault rate of $p = 10^{-4}$ in a distance-5 code is shown not to be sufficient for their FT magic state preparation protocol to reach ε in practice. To suppress error rates to ε via magic state distillation, the authors of [BH12] claim that the best protocol involves three distillation rounds with different codes using 180 qubits in total with magic state noise strength $p = 10^{-2}$ and noise-free Clifford gates. With both methods, ε is out of reach for state-of-the-art quantum computing platforms. FT magic state preparation however is more suitable for devices with $\lesssim 100$ qubits since it avoids the concatenation overhead of magic state distillation and can still be used to demonstrate crucial primitives toward the goal of fault-tolerant universal quantum computation.

2.5.1.3 Code deformation

Code deformation techniques aim to circumvent the need to perform complicated physical operations in order to realize a corresponding logical gate operation. This is achieved by changing the QEC code such that the logical gate can be done in a relatively simple way. Fault tolerance can be established if deforming the code is possible in an FT fashion and the logical gate has an FT, for example a transversal, implementation in the deformed code.

Consider a surface code patch with punctures that create internal boundaries and thus encode additional logical qubits as shown in Fig. 2.11(a). These punctures, or defects, can be moved through the surface code lattice by measuring adjacent stabilizers [Fow$^+$12]. Braiding one puncture around another can be performed to implement the operator map for a logical CNOT gate. *Defect braiding* can be relatively easy if stabilizer measurements are possible to perform in a given architecture anyway. A

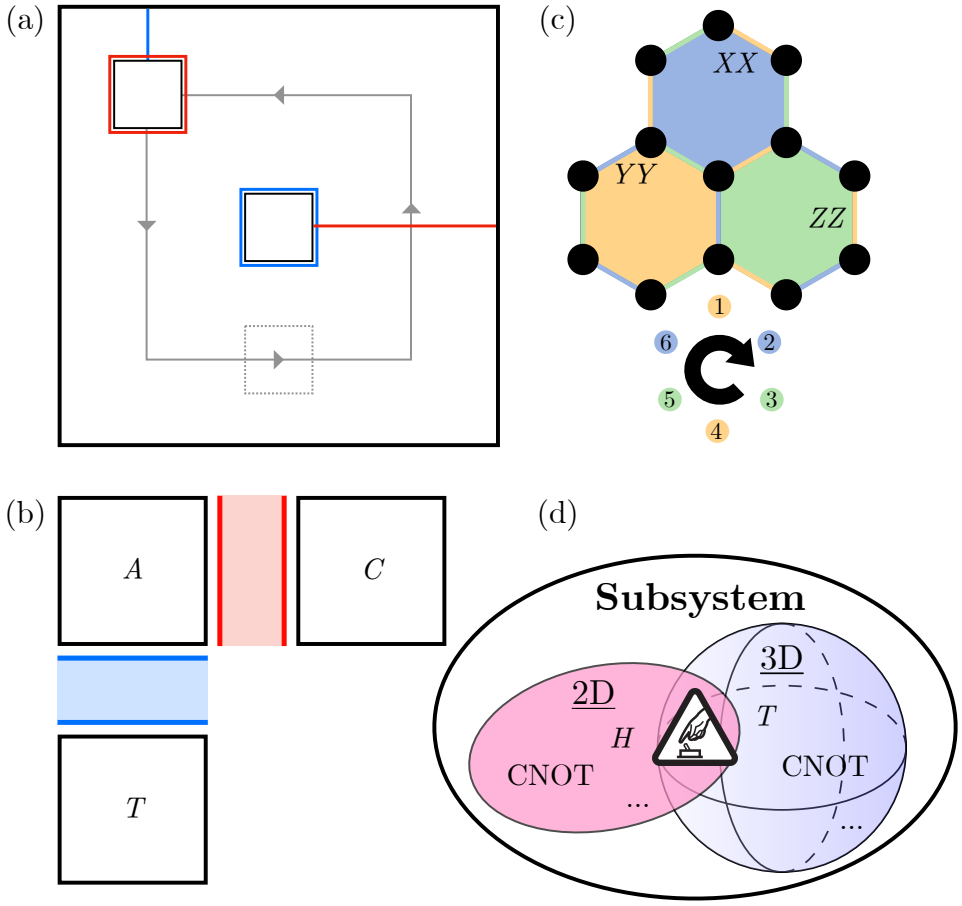


FIGURE 2.11: A collection of code deformation techniques. (a) Logical operators (colored) along the internal boundaries and to the external boundary of a surface code patch with two punctures as described in [Fow+12]. (b) Lattice surgery implementation of a logical CNOT gate using three logical qubits functioning as control (C), auxiliary (A) and target (T) qubit. For surface code or color code patches, logical operators can be measured on common boundaries of X- and Z-type. (c) Smallest unit for the honeycomb code; a Floquet code, taken from [Pae+23]. Weight-2 operators of all three Pauli types along edges with the same orientation (XX on all $\diagdown$, YY on all $\diagup$, ZZ on all vertical edges) are measured in a six-step cycle (yellow edges in step 1, blue edges in step 2 and so on as indicated). (d) Illustration of a subsystem code that contains two stabilizer codes. Together the two codes possess a transversal universal gate set. FT code switching allows one to realize all of its gates fault-tolerantly.

major drawback, however, is that large surface code patches, i.e. many physical qubits, are needed to create multiple logical qubits with a desired distance with punctures and boundaries. Performing such stabilizer measurements in an FT way can be done by including flag qubits into the planar embedding of the surface code as, for instance, suggested in [Cha$^+$20].

Using a logical auxiliary qubit, the technique of *lattice surgery* instead enables one to perform a logical CNOT gate between two logical qubits encoded in code patches by only measuring operators along a common boundary as illustrated in Fig. 2.11(b) [Hor$^+$12]. This is especially useful if physical access to the qubits along the boundary is easier than to bulk qubits. By the measurement, one temporarily creates a larger code between the control and the auxiliary qubit and then between the auxiliary and the target qubit. At the end of the procedure, the code space for each logical qubit is restored. Since the stabilizer measurements in these two techniques can be done fault-tolerantly, they enable an FT logical CNOT. We remark that these techniques for logical CNOT gates can be used for the teleportation circuits that realize magic state injection.

More recently, *Floquet codes* have been proposed where the code is periodically deformed by performing cyclic sequences of operator measurements [HH21b; Vui21; Gid$^+$21; GNM22; Pae$^+$23; Fah$^+$23]. This has the effect that one can only define an *instantaneous* stabilizer group during each step of the cycle. The advantage of these codes is that they can offer lower stabilizer weights than their static counterparts making the code less prone to noise by reducing the gate count. An example of weight-2 stabilizers in the honeycomb code is sketched in Fig. 2.11(c).

Over the course of this thesis, we have contributed to a theoretical proposal to perform *code switching* [Bom15a; KB15] in a fault-tolerant way. Instead of realizing, for instance, a T-gate via magic state injection, we combine a two- and a three-dimensional color code into a larger subsystem code. Each of the two stabilizer codes can be reobtained by choosing a specific gauge of the subsystem code. The transversal gate sets of both codes *together* are universal; in the two-dimensional color code all Clifford gates are transversal and in the three-dimensional color code, the T-gate is transversal. These relationships are depicted in Fig. 2.11(d). Code switching can be made FT by using flag circuits for stabilizer measurements and "fixing" the gauge when switching from one code to the other. By employing flag circuits it can be assured that no dangerous faults can propagate to the data qubits in order $m \leq t$. We have shown this for the smallest instances with distance $d = 3$ of the two code families that correct $t = 1$ error in [But$^+$23]. Code switching thus avoids the qubit overhead that would result from concatenating these two codes [CJOL17].

For the remaining chapters of this thesis, our work mostly focuses on the smallest representative of the family of two-dimensional topological color codes, which is suitable for near-term devices – the $[[7, 1, 3]]$ Steane code.

2.5.2 Low-distance color code

Using $n = 7$ physical qubits, the Steane code allows one to encode $k = 1$ logical qubit with a distance $d = 3$ [Ste96a]. Thus, the Steane code can correct $t = \lfloor \frac{d-1}{2} \rfloor = 1$ arbitrary Pauli error but the occurrence of $t + 1 = 2$ or more errors in general leads to logical failure. The $n - k = 6$ stabilizer generators of the code can be defined as

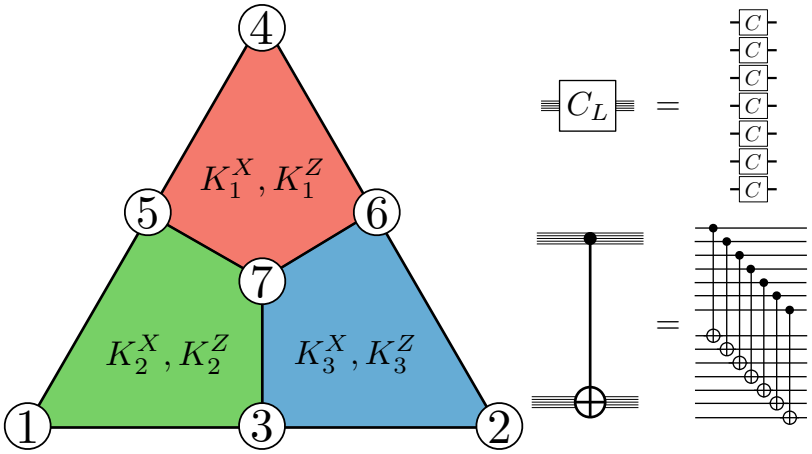


FIGURE 2.12: The $[[7,1,3]]$ Steane code uses $n = 7$ physical qubits to encode $k = 1$ logical qubit with distance $d = 3$. The colored plaquettes of the code graph symbolize the three X- and three Z-type operators K_i^σ that generate the stabilizer group. All logical Clifford gates can be realized by bitwise application of physical gates. The CNOT gate is transversal because the Steane code belongs to the class of CSS codes, which have distinct X- and Z-type generators. Since X- and Z-type generators are symmetric, the Hadamard gate is transversal.

$$\begin{aligned} K_1^X &= X_4X_5X_6X_7, & K_1^Z &= Z_4Z_5Z_6Z_7, \\ K_2^X &= X_1X_3X_5X_7, & K_2^Z &= Z_1Z_3Z_5Z_7, \\ K_3^X &= X_2X_3X_6X_7, & K_3^Z &= Z_2Z_3Z_6Z_7 \end{aligned} \tag{2.35}$$

and can be interpreted as three plaquettes for both X- and Z-type generators formed by four adjacent physical qubits. Figure 2.12 illustrates that each physical qubit either belongs to one, two or three of the plaquettes. Their eigenvalues can uniquely identify single-qubit errors. Three plaquettes of one Pauli type can each take two binary values in order to associate $2^3 - 1 = 7$ non-trivial three-bit syndromes with the appropriate single-qubit recovery operation in the look up table 2.3.

The logical operators can be chosen as $X_L = X^{\otimes 7}$ and $Z_L = Z^{\otimes 7}$. They are stabilizer-equivalent to weight-3 operators, for example, along the boundaries of the code graph. The logical zero code word of the Steane code

$$\begin{aligned} |0\rangle_L = \frac{1}{\sqrt{8}}(&|0000000\rangle + |1010101\rangle + |0110011\rangle + |1100110\rangle \\ &+|0001111\rangle + |1011010\rangle + |0111100\rangle + |1101001\rangle) \end{aligned} \tag{2.36}$$

is the simultaneous $+1$ eigenstate to all six code generators and Z_L. All Clifford gates can be implemented transversally in the Steane code. The existence of a transversal CNOT gate is a direct consequence of the fact that the Steane code is a Calderbank-Shor-Steane (CSS) code: There are distinct X- and Z-type stabilizer generators and the two sectors do not mix [CS96; Ste96b]. Additionally, the transversality of the Hadamard gate follows from the symmetry of the K_i^X and K_i^Z.

Due to its symmetries and low qubit count, the Steane code is a candidate for realization of single or multiple logical qubits in near-term devices that can host on

K_1^Z, K_2^Z, K_3^Z	recovery R
$+++$	I
$++-$	X_2
$+-+$	X_1
$+--$	X_3
$-++$	X_4
$-+-$	X_6
$--+$	X_5
$---$	X_7

TABLE 2.3: Look up table for the seven-qubit Steane code as shown in Fig. 2.12. + and − indicate a positive and negative expectation value of the respective stabilizer generator. All six of them form the error syndrome $(K_1^X, K_2^X, K_3^X, K_1^Z, K_2^Z, K_3^Z)$. Only the Z-type syndromes and corresponding X-type recoveries are shown. Since the Steane code is symmetric under exchange of X and Z, the Z-type recoveries from X-syndrome measurements can be applied analogously. The two three-bit syndromes (K_1^X, K_2^X, K_3^X) and (K_1^Z, K_2^Z, K_3^Z) are sufficient to correct all single Pauli errors.

the order of ten to 100 physical qubits. In experiments, the Steane code has been used to prepare a logical qubit [Nig+14; Blu+22; Abo+22]. Also, repetitive QEC cycles have been shown experimentally on the Steane code [RA+21]. With two logical qubits prepared, a logical CNOT has been experimentally realized and ultimately, using FT magic state preparation and magic state injection with the Steane code, we have contributed to the experimental demonstration of an FT universal gate set (see Sec. 4.1.2) [RA+22; Pos+22].

Chapter 3

Numerical simulation of noisy quantum computers

Qubits are never perfectly isolated from their environment since the ability to manipulate logical information requires some sort of access to the system. This coupling between the system and its surroundings also makes it prone to unintended interactions that we classify as noise. In this chapter we discuss methods to describe noisy quantum systems and techniques to simulate them on classical computers. In Sec 3.1, we provide representative examples of coherent and incoherent quantum channels that are used to describe imperfect quantum gate operations. Efficient simulation of stabilizer states on classical computers is subsequently explained in Sec. 3.2. We then review two commonly used standard techniques to simulate the effect of incoherent noise on quantum circuits, namely fault path counting and direct Monte Carlo simulation, in Secs. 3.3 and 3.4 respectively. Lastly in Sec. 3.5, we describe subset sampling of quantum error correcting circuits and propose its extension to sequences of noisy circuits, which may include mid-circuit measurements, as an efficient importance sampling technique at low noise strengths as in experimental hardware platforms available today[1].

3.1 Noisy quantum channels

A reasonably general formalism to describe the evolution of a quantum state ρ is a Kraus map

$$\mathcal{E}(\rho) = \sum_i K_i \rho K_i^\dagger, \tag{3.1}$$

which transforms the state ρ by the operation elements (or Kraus operators) K_i [NC10]. They form a complete set such that $\sum_i K_i^\dagger K_i = I$ and can be used to describe unitary evolution, measurements or open-system dynamics. An ideal single-qubit rotation $R_X(\theta)$ of angle θ about the X-axis, for example, changes the state ρ_0 to

$$\rho = R_X(\theta)\rho_0 R_X^\dagger(\theta) = \exp\left(-\mathrm{i}\frac{\theta}{2}X\right)\rho_0 \exp\left(\mathrm{i}\frac{\theta}{2}X\right). \tag{3.2}$$

Noise on the rotation angle θ could be of coherent nature such that, instead of performing the intended rotation of θ, a miscalibrated rotation about the angle $\theta+\xi$ with

[1] Numerical simulations in this chapter are performed using the Python package `qsample` [WH23]. Uncertainties on sampled quantities are given as 68% confidence intervals in this chapter unless indicated otherwise.

small ξ is applied. The noise channel $\mathcal{N}$ for this unitary error transforms the ideally rotated state ρ to

$$\mathcal{N}_\xi(\rho) = \exp\left(-\mathrm{i}\frac{\xi}{2}X\right)\rho\exp\left(\mathrm{i}\frac{\xi}{2}X\right) \tag{3.3}$$

$$= \cos^2\left(\frac{\xi}{2}\right)\rho + \sin^2\left(\frac{\xi}{2}\right)X\rho X + \frac{\mathrm{i}}{2}\sin(\xi)(\rho X - X\rho). \tag{3.4}$$

The Kraus operators can be expanded in any complete operator basis set as $K_i = \sum_j a_{ij}A_j$ so that the process matrix

$$\chi_{ij} = \sum_k a_{ki}a_{kj}^* \tag{3.5}$$

parametrizes the Kraus map in the basis formed by the A_j. The channel in Eq. (3.3) can then be written as

$$\mathcal{N}_\xi(\rho) = \sum_{ij}\chi_{ij}A_i\rho A_j^\dagger. \tag{3.6}$$

We see from Eq. (3.4) that the process matrix for this channel is

$$\chi = \frac{1}{2}\begin{pmatrix} 1+\cos(\xi) & \mathrm{i}\sin(\xi) & 0 & 0 \\ -\mathrm{i}\sin(\xi) & 1-\cos(\xi) & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.7}$$

in the basis of the Pauli matrices $A = (I, X, Y, Z)$. The off-diagonal entries are of order $\mathcal{O}(\xi)$ linear in the angle ξ and the leading diagonal entry[2] is of order $\mathcal{O}(1)$. So we can expect that neglecting the off-diagonal entries induces an error of order $\mathcal{O}(\xi)$, for instance when calculating an expectation value $\mathrm{Tr}(O\mathcal{N}_\xi(\rho))$ of an observable O. This motivates the *Pauli twirling approximation* (PTA), which is commonly used to approximate a coherent noise channel by a stochastic or incoherent noise channel [Ben+96; Eme+07; Sil+08; Dan+09; GZ13]. The Pauli-twirled channel $\tilde{\mathcal{N}}$ is obtained from $\mathcal{N}$ as

$$\tilde{\mathcal{N}}_\xi(\rho) = \frac{1}{4}\sum_{P\in\Lambda}P\mathcal{N}_\xi(P\rho P)P \tag{3.8}$$

$$= \cos^2\left(\frac{\xi}{2}\right)\rho + \sin^2\left(\frac{\xi}{2}\right)X\rho X \tag{3.9}$$

$$\equiv (1-p)\rho + pX\rho X \tag{3.10}$$

with $\Lambda = \{I, X, Y, Z\}$. The channel is called stochastic because the parameter p can be interpreted as a probability. At noise strength p, with probability $1-p$ no fault operator is applied to the state ρ by the channel $\tilde{\mathcal{N}}$ and with probability p the fault operator X is applied. Since Pauli twirling amounts to neglecting the off-diagonal entries of the process matrix χ, it can lead to underestimation of failure rates. It is known that the most drastic underestimation[3] of errors can occur for large coherent

[2] In Eq. (3.7), the matrix element χ_{00} is of order $\mathcal{O}(1)$ and the matrix element χ_{11} is of order $\mathcal{O}(\xi^2)$.

[3] The term "accuracy" is commonly used to describe how small the approximation error is and the term "honesty" refers to only overestimating but not underestimating the approximated logical failure rates in the literature.

gate errors [Gut+16; DP17; PR+21][4]. Still, the PTA provides a method to efficiently simulate the leading-order effect of noisy quantum channels in a stabilizer simulation without tracking all components of the statevector under the evolution described by such channels (see Sec. 3.2).

However, incoherent noise channels must not only be understood as approximations. Some physical noise processes are inherently incoherent, for example dephasing. The density matrix ρ at time t of a single qubit subject to pure dephasing is of the form

$$\rho(t) = \begin{pmatrix} \rho_{00} & \rho_{01}e^{-t/T_2} \\ \rho_{10}e^{-t/T_2} & \rho_{11} \end{pmatrix} \tag{3.11}$$

with the dephasing or coherence time T_2. In the limit $t \to \infty$, the off-diagonal elements in Eq. (3.11) decay to zero. The incoherent channel

$$\mathcal{E}_p(\rho) = (1-p)\rho + pZ\rho Z \tag{3.12}$$

describes this time evolution exactly if we identify $e^{-t/T_2} = 1 - 2p$.

The above channels model an underlying physical process within the Kraus formalism. From a fundamental point of view, we might ask how a noisy circuit transforms a quantum state without considering the specific hardware implementation. From a fault tolerance perspective, the most general noise channel on an n-qubit state ρ is the depolarizing channel

$$\mathcal{E}_p(\rho) = (1-p)\rho + \frac{p}{4^n-1}\sum_{\sigma\in\Lambda'} \sigma\,\rho\,\sigma \quad \text{with} \quad \Lambda' = \Lambda^{\otimes n} \setminus I^{\otimes n}, \tag{3.13}$$

which applies one out of all possible n-qubit Pauli operators with probability $\frac{p}{4^n-1}$ and leaves the state unchanged with probability $1-p$. These Pauli operators form a complete basis, in which we can expand any fault operator. This is especially notable in the context of fault tolerance: If a circuit is FT under application of Eq. (3.13) to all circuit locations, it is FT under any noise channel on physical operations and idling locations that acts on the computational subspace of the physical qubits[5].

3.2 Classical simulation of quantum circuits

In order to store a pure n-qubit quantum state $|\psi\rangle$ in a classical computer, in general one must keep track of the 2^n complex entries of the statevector that describes the quantum state[6]. For a mixture ρ, this number increases to $4^n - 1$ independent (real) entries of the density matrix. Gate operations are represented as multiplications of the state with unitary matrices of size $2^n \times 2^n$. The memory requirement to store the state and the time to calculate the effect of gate operations on the state here both grow exponentially with the number of qubits n.

For the broad class of stabilizer states, the evolution of a pure n-qubit quantum state $|\psi\rangle$ in a Clifford circuit can be described completely by only tracking its n stabilizer

[4] Other numerical studies find overestimated logical failure rates when performing the PTA with amplitude or phase damping [TS14; KG15]. For the Steane code, the PTA was shown to be accurate when simulating combined amplitude damping and dephasing [GB15]. Different regimes of the PTA are discussed in [Puz+14; Kat17].

[5] This excludes leakage, which we are not concerned with in this thesis [DMN13; BVT21; Mia+23].

[6] For a general normalized n-qubit quantum state where the global phase is irrelevant, there are $2^{n+1} - 2$ independent real entries.

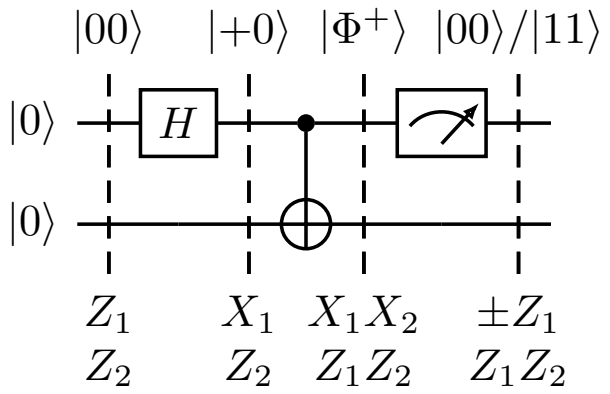


FIGURE 3.1: The stabilizers $\{Z_1, Z_2\}$ of the initial state $|00\rangle$ are sequentially transformed into the stabilizers $\{X_1X_2, Z_1Z_2\}$ of the Bell state $|\Phi^+\rangle$ by the gates H_1 and $\mathrm{C_1NOT_2}$ of the circuit. Measurement of the first qubit collapses the Bell state to either $|00\rangle$ or $|11\rangle$.

generators [Got98a; AG04][7]. Circuits that only contain Clifford gates, qubit initializations to Pauli states and measurements in Pauli bases then transform one set of n-qubit Pauli operators to another one for every time step of the circuit. Updating the stabilizers of the state can be performed by the classical computer within a computation time of $\mathcal{O}(n^2)$ [AG04].

As an example, consider preparing the Bell state

$$\left|\Phi^+\right\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}} \tag{3.14}$$

starting from two qubits initialized in $|00\rangle$ using the circuit in Fig. 3.1. A set of stabilizer generators for the initial state is $\{ZI, IZ\} = \{Z_1, Z_2\}$. The index of the Pauli matrix refers to the qubit it acts on. A single Hadamard gate on the first qubit H_1 changes the generator set to $\{XI, IZ\} = \{X_1, Z_2\}$ (see Eq. (2.9)). Then, the entangling $\mathrm{C_1NOT_2}$ gate transforms the set of generators to $\{XX, ZZ\} = \{X_1X_2, Z_1Z_2\}$, which indeed stabilizes the Bell state $|\Phi^+\rangle$ (see Fig. 2.3).

The measurement outcome for a single qubit, e.g. in the Z-basis, can be determined in a classical simulation by its stabilizer. If the stabilizer on the said qubit is $\pm Z$, the measurement will deterministically produce the respective value ± 1 and leave the state $|0/1\rangle$ unchanged. But if the stabilizer is X, i.e. the state's stabilizer and the measurement operator anticommute, then the state $|+\rangle$ is randomly projected onto the Z-eigenstates $|0/1\rangle$ and the Z-measurement will produce random outcomes ± 1, which can be drawn from a uniform distribution as part of the classical computation. Measuring single qubits of an n-qubit state collapses the stabilizer state according to the measurement outcome. For example, measuring the first qubit of the Bell state in Eq. (3.14) in the Z-basis, produces an outcome of ± 1 because Z_1 anticommutes with the Bell state stabilizer X_1X_2. Thus, the post-measurement state is stabilized by $\{\pm Z_1, Z_1Z_2\}$, correctly describing the state $|00\rangle$ when the measurement outcome is $+1$ or the state $|11\rangle$ when the measurement outcome is -1. The anticommuting operator X_1X_2 is replaced with the measurement operator $\pm Z_1$ and the commuting operator Z_1Z_2 remains in the stabilizer.

[7] A faster simulation method of stabilizer states than [AG04] based on so-called graph states has been given in [AB06]. Also see [Gid21] for more recent optimizations.

Estimating failure rates on the logical level via stabilizer simulation is a common method to evaluate the performance of stabilizer QEC codes. It can be applied for incoherent Clifford noise channels by adding fault operators from the Clifford group to the ideal circuit in the stabilizer simulation. However, there exist more advanced methods that can efficiently simulate/approximate non-Clifford circuits under additional assumptions, for instance that the number of T-gates in the circuit is small compared to the number of Clifford gates [Bra$^+$19; Sed$^+$21] or by extending the stabilizer formalism so that additional phases can also be tracked [NBN15; WBB22].

In the following, we discuss methods to classically simulate noise on quantum circuits. We always assume that the classical simulation of the actual quantum circuit can be done either via stabilizer or statevector simulation.

3.3 Fault path counting

The conceptually most straightforward method to obtain the failure rate of a noisy quantum circuit is to exhaustively assess all microscopic fault events, which can occur with the noise model under consideration [ABO08; RAK18]. For each fault, one evaluates whether or not its propagation results in an uncorrectable error and thus failure of the circuit. The circuit failure rate p_C can then be reconstructed from the occurrence probabilities p_F of all individual dangerous faults that lead to failure like

$$p_C = \sum_F p_F + \sum_{F,F'} p_{F,F'} + \cdots . \tag{3.15}$$

Here, the probability for a specific single fault p_F out of N_F possible faults is given by a probability p_f that this fault occurs multiplied by the probability $(1-p_f)^{N_F-1}$ that all other possible faults do not occur. In the second sum, all possible combinations of two faults that each occur with a joint probability $p_{F,F'}$ would also be evaluated. Additional terms would contain the fault configurations of order $m \geq 3$ (see Sec. 2.4.1).

When employing a common noise channel to several circuit locations, e.g. all gate operations, many individual faults occur with equal probabilities. For example, under depolarizing noise of strength p, any single-qubit gate fault X, Y or Z occurs each with probability $p/3$. Thus, we group together faults of a common type α and rewrite Eq. (3.15) as

$$\begin{aligned} p_C = &\sum_\alpha N_{\{\alpha\}} p_\alpha (1-p_\alpha)^{N_\alpha - 1} + N_{\{\alpha,\alpha\}} p_\alpha^2 (1-p_\alpha)^{N_\alpha - 2} \\ &+ \sum_{\alpha \neq \beta} N_{\{\alpha,\beta\}} p_\alpha (1-p_\alpha)^{N_\alpha - 1} p_\beta (1-p_\beta)^{N_\beta - 1} + \cdots \end{aligned} \tag{3.16}$$

where $N_{\{\alpha\}}$ is the number of dangerous faults of order $m = 1$ that happen with leading-order probability p_α. The total number of faults from locations of type α is denoted by N_α and $N_{\alpha,\beta}$ ($N_{\{\alpha,\beta\}}$) is the number of (dangerous) faults from two locations, one of type α and one of type β. The latter are the fault configurations of order $m = 2$ that together cause failure. One fault type has fault probability p_α and the other fault happens with probability p_β and so on for higher orders in p. This could for example be the combined effect of a particular single-qubit gate fault with probability $p/3$ and a specific two-qubit gate fault with probability $p/15$ under depolarizing noise of strength p, which, in conjunction, cause failure.

In practice, it may not be feasible to exhaustively place *all* fault events for circuits even with a moderate number of locations due to the exponentially growing number of m fault combinations. However, at small fault rates $p_\alpha \to 0$ it may be sufficient to evaluate a small number of terms of Eq. (3.16). Let us consider the case where we only exhaustively place all single faults, i.e. determine the $N_{\{\alpha\}}$ exactly. The error κ that we make by neglecting failures that are caused by two or more faults for the circuit failure rate p_C can be estimated as

$$\kappa = \sum_\alpha \binom{N_\alpha}{2} p_\alpha^2 (1-p_\alpha)^{N_\alpha - 2} + \sum_{\alpha \neq \beta} N_\alpha N_\beta p_\alpha (1-p_\alpha)^{N_\alpha - 1} p_\beta (1-p_\beta)^{N_\beta - 1} + \cdots \tag{3.17}$$

since the numbers of dangerous faults $N_{\{\alpha,\alpha\}} \leq \binom{N_\alpha}{2}$ and $N_{\{\alpha,\beta\}} \leq N_\alpha N_\beta$ are bounded by the respective total number of combinations of two faults.

Fault path counting yields practical estimations of circuit failure rates in the limit where all fault rates $p_\alpha \to 0$. Assuming only a single location type α and assuming that the total number of faults N_α is equal to the number of dangerous faults $N_{\{\alpha\}}$, the estimation of the error κ from faults of weight 2 is smaller than the single-fault estimation to p_C when

$$\binom{N_\alpha}{2} p_\alpha^2 (1-p_\alpha)^{N_\alpha - 2} \leq N_\alpha p_\alpha (1-p_\alpha)^{N_\alpha - 1} \Rightarrow p_\alpha \leq \frac{2}{N_\alpha + 1}. \tag{3.18}$$

It is clear that this bound on the maximally permitted value of p_α only becomes smaller when more terms are added to the left-hand-side of the inequality (3.18), i.e. faults of higher weights are also considered. This illustrates that fault path counting is only practical at small fault rates.

Additionally, for an FT circuit of fault tolerance level t with a large number of locations N_c, fault path counting might require a prohibitively large number of circuit evaluations $\mathcal{O}(N_c^{t+1})$ to obtain even the lowest non-vanishing order estimation of $p_C \sim p^{t+1}$. In this context, stochastic methods like direct Monte Carlo simulation might be more feasible, which we describe in the following section.

3.4 Direct Monte Carlo simulation

Incoherent noise channels as described in Sec. 3.1 have a probabilistic interpretation that can be used to estimate failure rates by sampling random noisy circuit instances in a Monte Carlo (MC) type simulation [Ste03; RHG07]. Here, every noisy location of a circuit is represented by its ideal operation U combined with a fault operator F which is drawn randomly from a distribution that represents the particular noise model for the respective operation. For a gate operation (this may include the identity gate), the faulty unitary U_{faulty} acts on a qubit state $|\psi\rangle$ like

$$U_{\text{faulty}}|\psi\rangle = F U_{\text{ideal}}|\psi\rangle \tag{3.19}$$

so first the ideal operation U_{ideal} is applied and it is followed by application of the fault operator F. Ideal qubit initialization to the state $|\psi\rangle$ is followed by the fault operator and a fault on a measurement operation is placed before applying the ideal measurement operator. The noisy circuit is then simulated by traversing all locations,

drawing a random number according to the respective noise channel of the location and probabilistically applying the relevant fault operators to the ideal operation. Any instance of a noisy circuit realization – usually called a "sample" or "shot" – then leads to a failure or not. The failure rate can be estimated as

$$\hat{p}_C = \frac{\text{number of failures}}{\text{number of MC shots}}. \tag{3.20}$$

Since a single shot is a Bernoulli trial, the standard error[8] of $\hat{p}_C$ is estimated in the limit of a large number of MC shots $N_{\text{MC}} \to \infty$ as

$$\hat{\sigma}_C \sim \sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{N_{\text{MC}}}}. \tag{3.21}$$

The true failure rate p_C likely lies within the interval $[\hat{p}_C - z_{\alpha/2}\hat{\sigma}_C, \hat{p}_C + z_{\alpha/2}\hat{\sigma}_C]$, where the confidence level $1-\alpha$ determines the value of the quantile function $z_{\alpha/2}$. This confidence interval is called the Wald interval when Eq. (3.21) determines $\hat{\sigma}_C$ [BCD01]. We assume that

$$z_{\alpha/2} = \sqrt{2}\,\text{erf}^{-1}(1-\alpha), \tag{3.22}$$

where $\text{erf}^{-1}(.)$ is the inverse error function, is the quantile function of the normal distribution so that confidence levels of approximately 68% and 95% correspond to $z \approx 1$ and $z \approx 1.96$. It is known that the Wald interval can fail to provide an accurate description of the confidence interval for certain numerical values [Dev15]. One obvious example is that it yields $\hat{\sigma}_C = 0$ for $\hat{p}_C \in \{0, 1\}$. However, Eq. (3.21) is only an estimation to the true standard error of the circuit failure rate p_C, which reads $\sigma_C = \sqrt{\frac{p_C(1-p_C)}{N_{\text{MC}}}}$ [Dev15; Soc23]. The true confidence interval is bounded by

$$p_C^{\text{up/low}} = \hat{p}_C \pm z_{\alpha/2}\sqrt{\frac{p_C(1-p_C)}{N_{\text{MC}}}}. \tag{3.23}$$

By transforming Eq. (3.23) into a quadratic equation in p_C, it can be solved exactly. The confidence interval set by the upper and lower bounds

$$p_C^{\text{up/low}} = \frac{1}{1+\frac{z_{\alpha/2}^2}{N_{\text{MC}}}}\left(\hat{p}_C + \frac{z_{\alpha/2}^2}{2N_{\text{MC}}} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_C(1-\hat{p}_C)}{N_{\text{MC}}} + \frac{z_{\alpha/2}^2}{4N_{\text{MC}}^2}}\right) \tag{3.24}$$

is called the Wilson interval [Wil27].

The failure rates we estimate are usually small, due to the low physical fault rates that we assume for circuit operations and because the leading-order term for the failure rate of an FT-t circuit is proportional to p^{t+1}. The small value of p is suppressed even further by the fault tolerance level t in the exponent, pushing p_C close to zero, where the Wald interval is unreliable. Thus, we provide confidence intervals as Wilson intervals obtained by evaluation of Eq. (3.24) in a symmetrized form $\left[\hat{p}_C - \frac{p_C^{\text{up}} - p_C^{\text{low}}}{2}, \hat{p}_C + \frac{p_C^{\text{up}} - p_C^{\text{low}}}{2}\right]$.

Still, direct MC requires more shots the lower p_C to achieve a given accuracy, i.e. relative error, on the estimator $\hat{p}_C$. This property of the estimator is known as *robustness*

[8] We use the conventional terminology "standard error" here since there is no risk of confusion with an error operator E.

in the literature [RT09]. For example, the necessary number of shots to estimate p_C with a relative error of at most ϵ can be obtained via

$$\epsilon\hat{p}_C \geq z_{\alpha/2}\sqrt{\frac{p_C(1-p_C)}{N_{\mathrm{MC}}}} \Leftrightarrow N_{\mathrm{MC}} \geq z_{\alpha/2}^2 \frac{p_C(1-p_C)}{\epsilon^2 \hat{p}_C^2}. \tag{3.25}$$

When $\hat{p}_C \approx p_C$ for a sufficiently large number of shots and $z_{\alpha/2} = 1$, a failure rate of $p_C = 10^{-3}$ would require $N_{\mathrm{MC}} \approx 4 \times 10^5$ shots to be estimated with a relative error of $\epsilon = 5\%$.

This number of shots is large compared to the number of evaluations that we considered previously in the context of fault path counting in the following sense: If we spent the $N_{\mathrm{MC}} = 4 \times 10^5$ shots on exhaustive placement of fault operators for fault path counting instead, we could determine the circuit failure rate in $\mathcal{O}(p^2)$ for an FT-1 circuit with $\sqrt{N_{\mathrm{MC}}} \approx 630$ possible faults. It must be decided from the context which method is preferential.

From the previous two sections, it becomes clear however that both methods have caveats. In the following we describe subset sampling as a technique to circumvent both problems of large required MC shot number and fault path counting a large number of location combinations.

3.5 Subset sampling

We can estimate the effect of noise at low physical fault rates, instead of exhaustively placing all possible fault events as in fault path counting, by employing an MC-type procedure to sample faults that belong to distinct subsets selectively. Failure rates can then be reconstructed approximately with a finite sampling uncertainty, which, as long as physical fault rates are low, will be smaller than with direct MC.

Consider a circuit with $g = 100$ gates at a physical fault rate of $p = 5 \times 10^{-4}$. When using direct MC, no fault will be realized in the sampling procedure for $(1-p)^g \approx 95\%$ of the shots. If said circuit is FT at level 1, this problem even becomes more severe since it is clear that $(1-p)^g + gp(1-p)^{g-1} \approx 99.9\%$ of the shots will never cause failure. These instances of direct MC are practically useless because they only reveal what is already known: that the fault-free circuit (or the circuit with a single fault if FT-1 holds) only terminates with success and never with failure. Subset sampling avoids this inefficiency by allowing one to exclude these cases from the sampling procedure.

In the following, we review the concept of subset sampling and discuss possible extensions. It can be applied to any quantum circuit of a fixed size [Li$^+$17; Tro$^+$18; GMB19; Li20; Pae$^+$23; Heu$^+$23a]. Subsequently, we then build on these previous works to develop a generalization of subset sampling that we coin "dynamical subset sampling", which allows one to also model quantum protocols as sequences of noisy quantum circuits with intermediate measurements and protocol branching options that depend on these measurement results within the framework of subset sampling [Heu$^+$23b]. A major advantage of this scheme is that a well-defined confidence interval on the failure rate estimator $\hat{p}_C$ is maintained at all times. This distinguishes the method, for instance, from Metropolis-type techniques such as [BV13] where convergence cannot be guaranteed.

3.5.1 Single-circuit subset sampling

Let us consider a single quantum circuit of a fixed size with a varying number of faults w, i.e. fault configurations of order $m = w$. The fault weight w can be used to uniquely label distinct *subsets* by which the space of all sampled noisy circuits can be partitioned and we refer to any specific subset as the w-fault-subset. If p is low enough, the fault-free subset, with $w = 0$, is the "largest" subset. This means that it is most likely that a particular instance of direct MC will realize the fault-free subset, as the above example illustrates. For any w-fault-subset, the probability that a specific w-fault, i.e. a fault of weight w, occurs is given by the probability to have exactly w faulty circuit locations, p^w, times the probability to have exactly $N_c - w$ fault-free circuit locations, $(1-p)^{N_c-w}$, where N_c is the total number of circuit locations for a circuit c. Then, there are $\binom{N_c}{w}$ different possibilities to choose a w-fault. Thus, we denote the size of the w-fault-subset by its *binomial factor* given by

$$A_w^c(p) = \binom{N_c}{w} p^w (1-p)^{N_c-w} \tag{3.26}$$

and the 0-fault-subset is the largest subset when the physical fault rate p fulfills

$$A_0^c(p) > A_1^c(p) \Leftrightarrow p < \frac{1}{N_c + 1}. \tag{3.27}$$

The notion of subsets is illustrated in Fig. 3.2. The circuit failure rate p_C can be reconstructed from individual subset failure rates $p_{\text{fail}}^{(w)}$ as

$$p_C = \sum_{w=0}^{N_c} A_w^c(p) p_{\text{fail}}^{(w)}. \tag{3.28}$$

We may now estimate any subset failure rate $p_{\text{fail}}^{(w)}$ in an MC-type fashion by numerically drawing w-faults randomly a number of $N_{\text{SS}}^{(w)}$ times according to a particular noise model of interest. The 0-fault-subset is explicitly excluded since its failure rate $p_{\text{fail}}^{(0)} \equiv 0$ is known already. The central ingredient to subset sampling is a particular feature of Eq. (3.28): The subset failure rates $p_{\text{fail}}^{(w)}$ are completely independent from the physical fault rate p, which is contained within the binomial factors that are known analytically. In practice, this means that the subset failure rates need only be determined once for a fixed numerical value $p = p_{\max}$ and the functional behavior $p_C(p)$ can be reconstructed after the sampling procedure. The physical fault rate is naturally separated from the sampling in this formalism. Therefore, scalings can be easily extracted this way. This advantage of subset sampling becomes even more useful when we consider FT-t circuits where, by definition, $p_{\text{fail}}^{(w)} \equiv 0 \ \forall w \leq t$. Note that any type of incoherent fault operators can be chosen and that, since we did not make any assumption about the quantum circuit, subset sampling is suitable not only for QEC but any type of quantum circuit with circuit-level noise.

As for fault path counting, it would be impractical to evaluate the sum in Eq. (3.28) for *all* subsets up to the N_c-fault-subset and practically it will be sufficient to only evaluate a finite number of terms at sufficiently low p. By introducing a maximum fault subset $w_{\max}$ as a cutoff, we underestimate the true circuit failure rate as the lower bound p_l reads

$$p_l \equiv \sum_{w=0}^{w_{\max}} A_w p_{\text{fail}}^{(w)} \leq p_C, \tag{3.29}$$

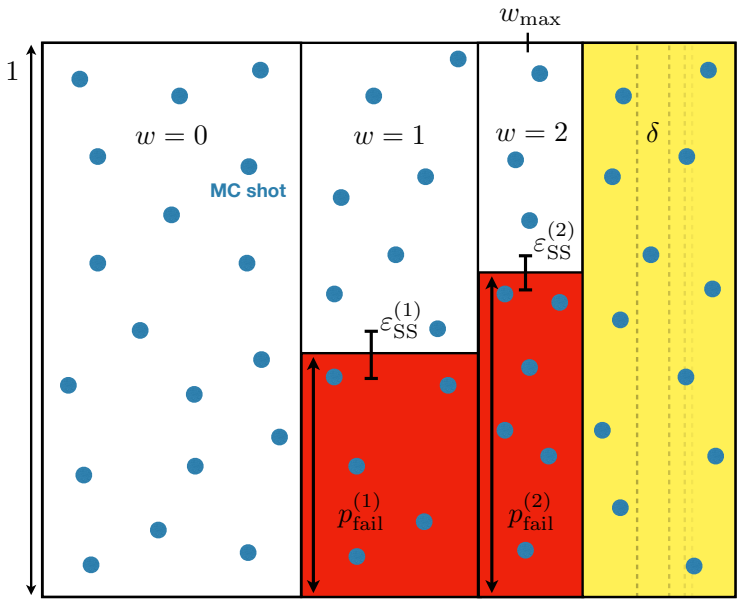


FIGURE 3.2: Illustration of w-fault-subsets for single-circuit subset sampling. Essential to subset sampling, each subset (box) is sampled distinctly and the fault-free subset, labeled $w = 0$, is omitted by design. Direct MC would sample uniformly in the whole w-space (blue dots) such that most shots are placed in the fault-free subset, which is assumed to be the largest box. Subset failure rates $p_{\mathrm{fail}}^{(w)}$ (ratio of vertically extended box colored red) are estimated with respective uncertainties $\varepsilon_{\mathrm{SS}}^{(w)}$ up to a weight cutoff $w_{\max}$. The cut off subsets (yellow) collectively amount to the cutoff error δ.

which is illustrated in Fig. 3.2. We omit the index c and argument p for the binomial factor from now on where the circuit c and the physical fault rate p are fixed quantities in the discussion. On the other hand, we can add an upper bound p_u to Eq. (3.29) by assuming that all neglected subsets contribute the maximum value $p_{\mathrm{fail}}^{(\bar{w})} = 1 \; \forall \bar{w} > w_{\max}$ to the failure rate, so that

$$p_C \leq \sum_{w=0}^{w_{\max}} A_w p_{\mathrm{fail}}^{(w)} + \sum_{w=w_{\max}+1}^{N_c} A_w \equiv p_u. \tag{3.30}$$

This allows us to define the *cutoff error* δ as the sum over binomial factors of all neglected subsets

$$\delta = \delta(p) = \sum_{w=w_{\max}+1}^{N_c} A_w = 1 - \sum_{w=0}^{w_{\max}} \binom{N_c}{w} p^w (1-p)^{N_c - w}, \tag{3.31}$$

which is marked in yellow in Fig. 3.2 and can be expressed as the complement to the size of all subsets that were already sampled so no further circuit evaluations are required to determine the cutoff error. The bounds posed by Eqs. (3.29) and (3.30) tighten as $p \to 0$, i.e. the cutoff error δ vanishes in this limit; the size of the fault-free subset $(1-p)^{N_c} \to 1$.

Thus, we define the subset sampling estimator

$$\hat{p}_C \equiv \frac{p_l + p_u}{2} = p_l + \frac{\delta}{2} \tag{3.32}$$

as the average value between upper and lower bound. The estimator is unbiased [Dev15; Soc23] in the sense that

$$\mathrm{Bias}(\hat{p}_C, p_C) = \sum_{w=1}^{w_{\max}} A_w(p)\hat{p}_{\mathrm{fail}}^{(w)} + \frac{1}{2}\left(1 - \sum_{w=0}^{w_{\max}} A_w(p)\right) - p_C \to 0 \tag{3.33}$$

the bias vanishes in the limit of $p \to 0$ or, equivalently, when we take all subsets into account $w_{\max} \to N_c$. Convergence of the subset failure rate estimators $\hat{p}_{\mathrm{fail}}^{(w)}$ to their true values is assured by the central limit theorem and the law of large numbers when a large amount of shots is taken in every subset.

The advantage of subset sampling over direct MC is larger at lower p. For larger p, one must choose a larger cutoff $w_{\max}$ to keep the cutoff error $\delta(p)$ limited to a desired numerical value at a fixed $p = p_{\max}$ of interest. If $w_{\max}$ becomes too large, so that one has to sample a large number of subsets, direct MC in turn becomes more efficient because there is no practical value in distinguishing all subset failure rates $p_{\mathrm{fail}}^{(w)}$ but the overall quantity of interest is the circuit failure rate p_C.

The sampling uncertainty for any subset failure rate, i.e. the uncertainty about the height of a red box in Fig. 3.2, analogously to Eq. (3.21), reads

$$\varepsilon_{\mathrm{SS}}^{(w)} \sim \sqrt{\frac{p_{\mathrm{fail}}^{(w)}\left(1 - p_{\mathrm{fail}}^{(w)}\right)}{N_{\mathrm{SS}}^{(w)}}} \qquad (\text{as } N_{\mathrm{SS}}^{(w)} \to \infty) \tag{3.34}$$

when $N_{\mathrm{SS}}^{(w)}$ samples are used to estimate a single subset failure rate $p_{\mathrm{fail}}^{(w)}$. For the total sampling error on the failure rate

$$\varepsilon_{\mathrm{SS}} = \sqrt{\sum_{w=1}^{w_{\max}} \left[A_w^c(p)\varepsilon_{\mathrm{SS}}^{(w)}\right]^2} \tag{3.35}$$

the individual subset sampling errors $\varepsilon_{\mathrm{SS}}^{(w)}$ are suppressed by the according binomial factors. Especially, there is no uncertainty about the fault-free subset, so $\varepsilon_{\mathrm{SS}}^{(0)} \equiv 0$. The uncertainty interval of the subset sampling estimator $\hat{p}_C$ is thus defined as

$$[p_l - \varepsilon_{\mathrm{SS}}, p_u + \varepsilon_{\mathrm{SS}}] = [\hat{p}_C - \varepsilon_{\mathrm{SS}} - \delta/2, \hat{p}_C + \varepsilon_{\mathrm{SS}} + \delta/2]. \tag{3.36}$$

so that the total length of the interval is

$$\eta \equiv 2\varepsilon_{\mathrm{SS}} + \delta. \tag{3.37}$$

Since the binomial factors $A_w^c(p)$ in Eq. (3.35) are small for low p, much fewer samples $\sum_w N_{\mathrm{SS}}^{(w)} \ll N_{\mathrm{MC}}$ are required to estimate the circuit failure rate p_C to a given accuracy via subset sampling compared to direct MC. Both types of error, sampling error and cutoff error, are assured to be small if p is low enough and, notably, they are independent from each other for single-circuit subset sampling, which ceases to be the case for dynamical subset sampling (see Sec. 3.5.2).

Moreover, in case of a small number of circuit locations N_c, it might be practical to not sample, say, the 1-fault-subset but instead use fault path counting to also determine $p_{\mathrm{fail}}^{(1)}$ exactly such that $\varepsilon_{\mathrm{SS}}^{(1)} = 0$. In order to do so, fault path counting would need to evaluate $a \times N_c$ faults, where the constant a is determined by the circuit and the noise

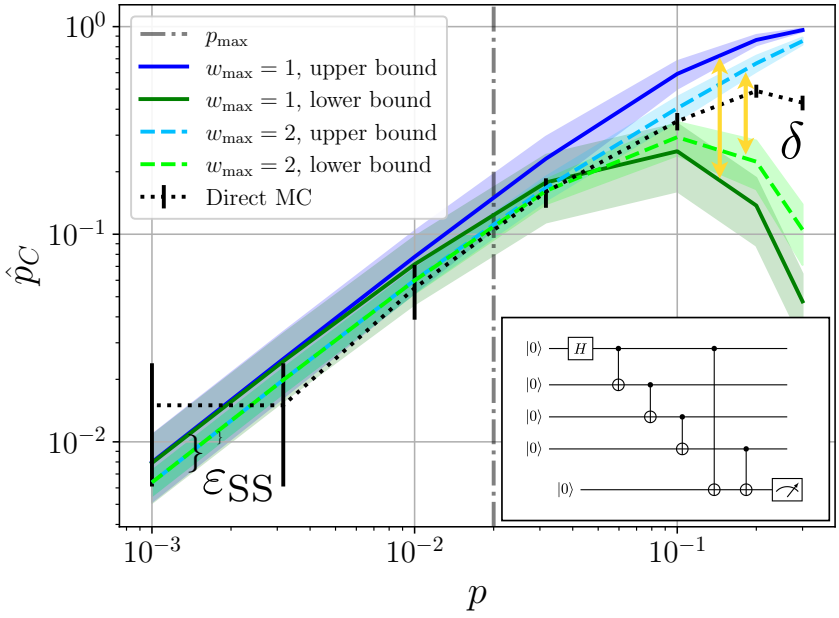


FIGURE 3.3: Upper and lower bounds to the true circuit failure rate p_C of the FT four-qubit GHZ preparation circuit (inset) are shown, dependent on the physical fault rate p, for subset sampling with two different weight cutoffs $w_{\max} = 1$ (solid, 20 shots) and $w_{\max} = 2$ (dashed, 200 shots). The cutoff error $\delta(p)$ is the difference between upper and lower bound (yellow arrows) and approaches zero as $p \to 0$. The respective total sampling errors $\varepsilon_{\mathrm{SS}}$ (shaded areas around bounds, marked with curly brackets) are independent of p. For comparison, data points obtained via direct Monte Carlo (MC, dotted) with at most 200 shots are shown, where uncertainty intervals grow larger for lower p.

model. If, for example, the circuit consisted of only single-qubit (two-qubit) gates and we had depolarizing noise (see Eq. (3.13)), we would have $a = 3$ (15) different faults F of Pauli weight $\mathrm{Wt}(F) = 1$ (or 2). If the number of samples $N_{\mathrm{SS}}^{(1)}$ needed to determine the contribution $A_1^c(p)\varepsilon_{\mathrm{SS}}^{(1)}$ of the sampling uncertainty of the 1-fault-subset to the total sampling uncertainty $\varepsilon_{\mathrm{SS}}$ up to a desired accuracy, say 5%, is smaller than the number of circuit evaluations required for fault path counting $a \times N_c$, then one should use the sampling approach in practice. This will be the case for large circuits.

We apply subset sampling, as an example, for a circuit that prepares the four-qubit GHZ state fault-tolerantly under depolarizing noise. The bounds to the true failure rate p_C according to Eqs. (3.29) and (3.30) are shown in Fig. 3.3 for the circuit we encountered earlier as part of Shor-type EC in Fig. 2.5, which is also reproduced in the inset of Fig. 3.3. A failure is recorded when the flag is triggered for this example[9]. The upper and lower bounds tighten when lowering p and can barely be distinguished by eye for $p \lesssim 10^{-2}$. With a small $w_{\max} = 1$, the bounds are narrowing more slowly than with $w_{\max} = 2$. It is equivalent to say that at $p = p_{\max} = 2 \times 10^{-2}$ the cutoff error δ, i.e. the difference between the upper and lower bound, is larger for $w_{\max} = 1$ than for $w_{\max} = 2$. Clearly, the sampling errors $\varepsilon_{\mathrm{SS}}$ for both examples do not depend on the physical fault rate p but only on the maximum fault weight $w_{\max}$ and the number of samples N_{SS}. We also show in Fig. 3.3 that the direct MC sampling error

[9]Note that this is different from counting erroneous state preparation as failure. Such an event, where the flag is not triggered but still the wrong state is prepared, can only occur with faults of at least weight 2 in the given circuit.

grows larger as p is lowered while we keep the number of samples for each data point N_{MC} fixed. The relative error is not constant but $\varepsilon_{\mathrm{MC}}/\hat{p}_C \sim \hat{p}_C^{-1/2}$ as $\hat{p}_C \to 0$.

We limited the above discussion of subset sampling to a single noise parameter p. The approach has been extended to multi-parameter noise models that distinguish different physical fault rates for several circuit components and thus allow for a more realistic description of faulty hardware, for instance in [Li+17; GMB19].

3.5.1.1 Multi-parameter noise

To employ subset sampling with a multi-parameter noise model that is parametrized by a vector $\vec{p}$, the integer-valued subset labels $w \in \mathbb{N}_0$ are replaced by vectors $\vec{w} \in \mathbb{N}_0^M$ with integer-valued entries. Here, M is the number of noise parameters for various types of circuit locations such as single-qubit gates, qubit initializations, measurements and two-qubit gates. The index w of the binomial factors A_w is replaced accordingly with a vector-valued subset index $\vec{w}$ so that

$$A_{\vec{w}}(\vec{p}) = \prod_{\mu=1}^{M} \binom{N_c^{(\mu)}}{w_\mu} p_\mu^{w_\mu} (1-p_\mu)^{N_c^{(\mu)} - w_\mu} \tag{3.38}$$

where $N_c^{(\mu)}$ is the number of circuit locations of type μ. For example, let us consider a two-parameter noise model with $\vec{p} = (p_1, p_2)$ for the GHZ preparation circuit we discussed before. We denote by p_1 the fault rate of single-qubit operations, namely qubit initialization, single-qubit gates and measurements and by p_2 the two-qubit gate fault rate. Then, for the above circuit, $\vec{N}_c = (7, 5)$ describes the respective number of circuit locations (see inset of Fig. 3.3).

A complication arises from the fact that we cannot, in this case, directly put the subset labels $\vec{w}$ in a natural order as for the scalar case, where, for sufficiently low p, we have $A_w(p) > A_{w+1}(p)\ \forall w \geq 0$. For instance, for a particular value of $\vec{p}$, we must determine if $A_{(0,1)}(\vec{p})$ is smaller or larger than $A_{(1,0)}(\vec{p})$, i.e. we have to assign an ordering to all $\vec{w}$-fault-subsets for a given $\vec{p}$.

3.5.1.2 Adaptive sampling

For both single- and multi-parameter noise models, the question arises how we should distribute our shots over the various subsets. While one may uniformly sample from all subsets, it is desirable that only the most "relevant" subsets are sampled. In Ref. [Pae+23], a pre-sampling step is performed for this reason to roughly estimate how many shots should be used for each subset. Due to the two different sources of uncertainty, either the sampling error of a single subset failure rate or the cutoff error from neglecting certain subsets, we must decide for any given shot whether to spend it on decreasing any $\varepsilon_{\mathrm{SS}}^{(w)}$ or decreasing δ.

We propose to maximize the *expected reduction of uncertainty* (ERU) in order to spend any single shot on a subset such that, on average, the total uncertainty η will be reduced the most. The *expected* uncertainty after any next shot in the w-fault-subset can be expressed as the average over the two possible outcomes "fail" or "no fail". We assume that the total uncertainty after the next shot changes to either $\eta\big|_{w\to\text{fail}}$ or $\eta\big|_{w\to\text{no fail}}$ depending on the assumed outcome. Then, the expected uncertainty reads

$$p_{\text{fail}}^{(w)}\, \eta\big|_{w\to\text{fail}} + (1 - p_{\text{fail}}^{(w)})\, \eta\big|_{w\to\text{no fail}}. \tag{3.39}$$

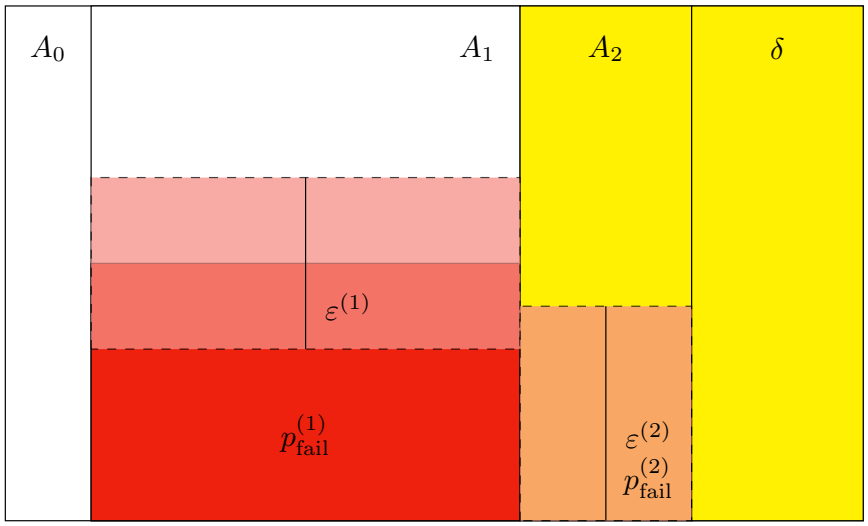


FIGURE 3.4: Box representation of different fault-weight-subsets for a single circuit with binomial factors A_w. The subset failure rate $p^{(1)}_{\text{fail}}$ has a sampling uncertainty $\varepsilon^{(1)}$. Opening the 2-fault-subset, i.e. sampling it for the first time, reduces the cutoff error δ by the binomial factor A_2 at the price of increasing the overall sampling uncertainty because an initial estimate from few shots of $p^{(2)}_{\text{fail}}$ comes with a large uncertainty $\varepsilon^{(2)}$. (The 0-fault-subset is drawn on the side for better visibility of the higher-weight subsets but it is still assumed to be the largest subset as in Fig. 3.2.)

We define the expected reduction of uncertainty $\Delta(w)$ for any w-fault-subset as

$$\Delta(w) \equiv \eta - \left[p^{(w)}_{\text{fail}} \, \eta\big|_{w\to\text{fail}} + (1 - p^{(w)}_{\text{fail}}) \, \eta\big|_{w\to\text{no fail}} \right] \tag{3.40}$$

and choose the subset

$$w^* = \underset{w}{\operatorname{argmax}} \, \Delta(w) \tag{3.41}$$

such that the ERU is maximized when choosing w^* out of all possible w ("ERU criterion"). The maximization can be performed in practice over all w-fault-subsets that were already sampled and the largest subset yet unsampled $w \in [1, w_{\max} + 1]$ because sampling in this subset will decrease the cutoff error more than any other unsampled subset, assuming that $A_w > A_{w+1} \, \forall w \geq 0$, which is fulfilled for small enough p. For the multi-parameter case, this space of choices is understood to refer analogously to the subset ordering defined in Sec. 3.5.1.1.

The two expected uncertainties $\eta\big|_{w\to(\text{no})\,\text{fail}}$ are calculated from the respective subset failure rate that changes with the two assumed next shot's outcomes as follows. When $p^{(w)}_{\text{fail}}$ is determined as

$$p^{(w)}_{\text{fail}} = \frac{m^{(w)}}{N^{(w)}_{\text{SS}}}, \tag{3.42}$$

where $m^{(w)}$ denotes the number of fails and $N^{(w)}_{\text{SS}}$ is the number of samples in the w-fault-subset, we can define the subset failure rate after an assumed positive/negative

outcome of the *next* shot, i.e. the $N_{\mathrm{SS}}^{(w)} + 1$st shot, as

$$p_{\mathrm{fail}}^{(w)}\big|_{+} \equiv \frac{m^{(w)}+1}{N_{\mathrm{SS}}^{(w)}+1} = \frac{m^{(w)}+1}{N_{\mathrm{SS}}^{(w)}} \frac{N_{\mathrm{SS}}^{(w)}}{N_{\mathrm{SS}}^{(w)}+1} = \left(p_{\mathrm{fail}}^{(w)} + \frac{1}{N_{\mathrm{SS}}^{(w)}}\right) \frac{N_{\mathrm{SS}}^{(w)}}{N_{\mathrm{SS}}^{(w)}+1} \tag{3.43}$$

$$p_{\mathrm{fail}}^{(w)}\big|_{-} \equiv \frac{m^{(w)}}{N_{\mathrm{SS}}^{(w)}+1} = p_{\mathrm{fail}}^{(w)} \frac{N_{\mathrm{SS}}^{(w)}}{N_{\mathrm{SS}}^{(w)}+1} \tag{3.44}$$

and use $p_{\mathrm{fail}}^{(w)}\big|_{\pm}$ to calculate $\eta\big|_{w\to(\mathrm{no})\,\mathrm{fail}}$ respectively by Eq. (3.37).

An illustration of the two different uncertainties after a finite amount of shots is depicted in Fig. 3.4. In this example, we assume a currently estimated $p_{\mathrm{fail}}^{(1)}$ with a sampling error $\varepsilon^{(1)}$ for the 1-fault-subset and all other subsets are still contained in the cutoff error $\delta = 1 - A_0 - A_1$, which explicitly includes A_2. The total length of the uncertainty interval η on $\hat{p}_C$ in this case is $2\varepsilon^{(1)} + \delta$. If we spent the next shot in the 2-fault-subset, the total uncertainty would change to $2\varepsilon^{(1)} + 2\varepsilon^{(2)} + \delta - A_2$. The previous cutoff error δ is reduced by A_2 but the sampling error is increased by $\varepsilon^{(2)}$. If, instead, we spent the next shot in the 1-fault-subset, we would decrease $\varepsilon^{(1)}$ but leave δ unchanged. For this case, we can calculate the ERU values for both subsets as

$$\begin{aligned}\Delta(1) &= \left(2\varepsilon^{(1)} + \delta\right) - \left[p_{\mathrm{fail}}^{(1)} \cdot \left(2\varepsilon^{(1)}|_{+} + \delta\right) + (1 - p_{\mathrm{fail}}^{(1)}) \cdot \left(2\varepsilon^{(1)}|_{-} + \delta\right)\right] \\ &= 2\varepsilon^{(1)} - \left[p_{\mathrm{fail}}^{(1)} \cdot 2\varepsilon^{(1)}|_{+} + (1 - p_{\mathrm{fail}}^{(1)}) \cdot 2\varepsilon^{(1)}|_{-}\right]\end{aligned} \tag{3.45}$$

and

$$\begin{aligned}\Delta(2) = &\left(2\varepsilon^{(1)} + \delta\right) \\ &- \left[p_{\mathrm{fail}}^{(2)} \cdot \left(2(\varepsilon^{(1)} + \varepsilon^{(2)}|_{+}) + \delta - A_2\right) + (1 - p_{\mathrm{fail}}^{(2)}) \cdot \left(2(\varepsilon^{(1)} + \varepsilon^{(2)}|_{-}) + \delta - A_2\right)\right] \\ &= A_2 - 2\varepsilon^{(2)}|_{+}\end{aligned} \tag{3.46}$$

The last equality holds, since $\varepsilon^{(2)}|_{+} = \varepsilon^{(2)}|_{-}$ in case that $N_{\mathrm{SS}}^{(2)} = 0$, so that $p_{\mathrm{fail}}^{(2)}$, which at this point would be undetermined because no samples have been run in the 2-fault-subset so far, naturally drops out of Eq. (3.46). The length of the Wilson interval (3.24), in general, is symmetric under exchange of $\hat{p} \leftrightarrow 1 - \hat{p}$, so, specifically, it does not matter for the sampling error if the first shot in a subset yields a failure or not.

Let us now apply the ERU criterion to the four-qubit GHZ preparation circuit. The flag rate of the circuit, obtained via subset sampling with ERU, is shown in Fig. 3.5(a). We only require 100 shots at $p_{\max} = 10^{-3}$ to obtain the bounds, which are reasonably tight over the whole interval of fault rates $p \in [10^{-4}, 10^{-1}]$. Also, the sampling result from direct MC is shown: With this method we need 10^4 shots to achieve a comparable accuracy at $p_{\max} = 10^{-3}$. This can be seen more clearly from Fig. 3.5(b). For lower values of p, the relative uncertainties for direct MC are even larger. The values of the ERU $\Delta(w^*)$ for the chosen subset in each shot are given for this example in Fig. 3.6(a). The fault-free subset is chosen exactly once in the very first circuit run. No improvement can be expected from further sampling the ideal circuit. Afterwards, the 1-fault-subset is sampled most of the time, except in the 21st shot, where the 2-fault-subset is chosen. At this point, the sampling uncertainty of the 1-fault-subset has become so small that there is more gain to the overall accuracy to reduce the cutoff error by A_2. In Fig. 3.6(b), we illustrate a more diverse choice of subsets that occurs

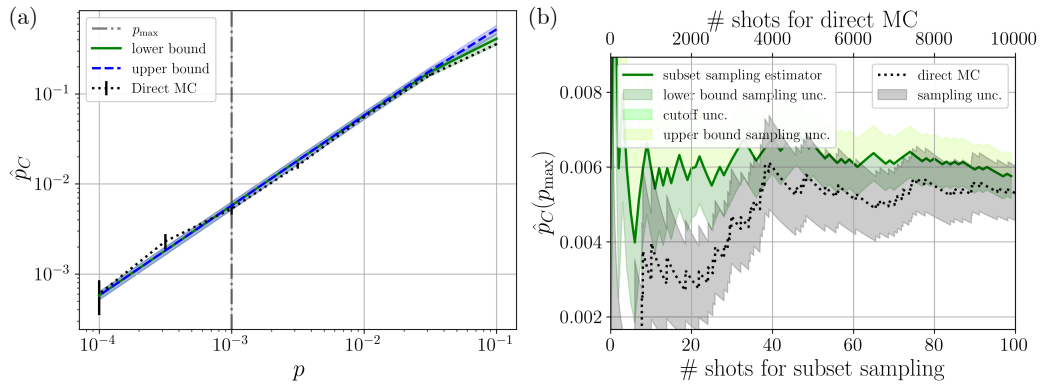


FIGURE 3.5: Subset sampling for GHZ state preparation at $p_{\max} = 10^{-3}$. (a) Upper and lower bound on the flag rate p_C are tight. While the uncertainty intervals of the direct MC estimation grow larger for smaller p, all rates for smaller p can be extracted analytically in subset sampling and the relative error stays constant. (b) Subset sampling (green solid), after 10^2 shots, reaches a similar accuracy as direct MC (black dotted) after 10^4 shots. Note that two different horizontal axes are used here. The advantage of DSS over direct MC is here two orders of magnitude in the number of samples for reaching a comparable total estimation uncertainty (unc.).

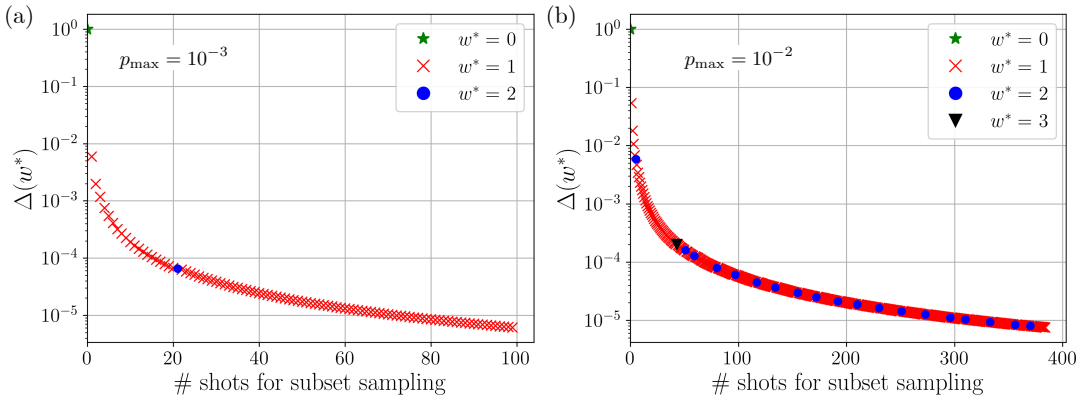


FIGURE 3.6: The ERU criterion opens the next larger fault-weight-subset when decreasing the sampling uncertainty ε_{SS} does not improve the total uncertainty η more than decreasing the cutoff error δ. The 1-fault-subset is chosen most of the times. (a) At $p_{\max} = 10^{-3}$, the 2-fault-subset is chosen once. (b) At $p_{\max} = 10^{-2}$, the 2-fault subset is selected various times and the 3-fault-subset is sampled once. Since the binomial factors A_2 and A_3 are larger at $p_{\max} = 10^{-2}$ than at $p_{\max} = 10^{-3}$, the larger-weight subsets are more relevant as compared to (a).

at a larger value of $p_{\max} = 10^{-2}$ since the higher-weight subsets are more relevant the larger $p_{\max}$ is set. In this case, the 2-fault-subset is sampled more often because its contribution to the failure rate and the total uncertainty η is now larger than at lower $p_{\max}$.

3.5.2 Dynamical subset sampling

In this section, we generalize single-circuit subset sampling to the case of efficiently estimating failure rates of a quantum algorithm, or protocol, that consists of executing *non-deterministic* sequences of noisy quantum circuits [Heu+23b]. A particular sequence of circuits is not fixed a priori but may only be determined at run time of the protocol. This is due to the stochastic nature of in-sequence measurements that after each circuit lead to a decision which circuit is executed next. Common FT QEC protocols are non-deterministic in this sense since they feed-forward intermediate measurement information: For instance in Shor-type EC, the GHZ preparation is repeated until the flags are clear and the actual transversal stabilizer measurements are repeated a varying number of times that depends on the actual measurement outcomes. Circuit sequences executed for flag-FT stabilizer readout or other flag schemes can as well be classified as non-deterministic [Got16a; CB18; CC19; Heu+23a]. Other scenarios include the dynamical execution of state preparation circuits in [FF+23] or the FT QEC cycles of [RA+21].

Low physical fault rates lead us to employing subset sampling as an importance sampling technique since failure of a circuit is rare. In the context of FT QEC, protocol failures become even more rare since a true protocol failure rate $p^* \sim p^{t+1}$ is suppressed with the number of errors t that the underlying QEC code can correct in the limit $p \to 0$. Because the useful regime of operation for an FT QEC protocol is below the break-even point where the logical failure rate becomes lower than the failure rate of the unencoded qubit, a practical numerical simulation technique should be able to handle the potentially low values of these break-even points. Thresholds for common QEC codes – if existent – can range over several orders of magnitude, from 10^{-9} to 10^{-2} faults per operation depending on the code and noise model under consideration (see [DMN13] for an overview).

Dynamical subset sampling (DSS) maintains the advantages that we previously identified for subset sampling: Any circuit-level noise model of incoherent fault operators can be used with any number of noise parameters. Scalings can be extracted analytically since the p-dependence lies entirely in the binomial factors. Only the most relevant subsets need to be sampled and well-defined error bars are provided at any time in the process. Also, the ERU criterion carries over to this more general setting.

In the following, we first introduce the general framework of dynamical subset sampling to appropriately model quantum protocols. We then discuss how the sampling uncertainties and cutoff error become interleaved in this new setting and define an uncertainty interval for the DSS estimator. Next, we demonstrate the DSS procedure using two different protocols for FT logical state preparation in the Steane code. Our discussion is concluded by remarks on the efficiency of DSS and run time considerations.

A protocol $\mathcal{P}$, which non-deterministically realizes sequences of quantum circuits $\mathcal{C}$ that suffer a varying number of faults, is illustrated as an event tree in Fig. 3.7(a).

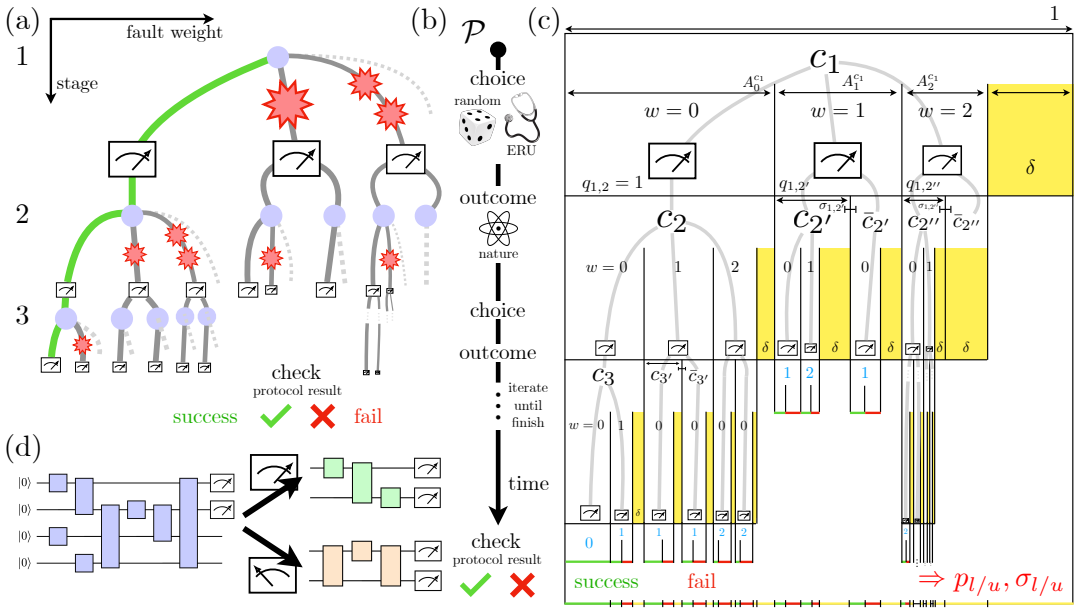


FIGURE 3.7: Protocol illustrations for dynamical subset sampling (DSS). (a) Sketch of a DSS event tree. Purple nodes symbolize individual quantum circuits. For each circuit a fault with weight $w \in \mathbb{N}_0$ (red stars), which increases from left to right, can be applied. Subsequent to circuit nodes, measurement outcomes determine which circuit is run as the next part of a protocol (vertical direction). For simplicity, it is assumed that any measurement leads to a binary decision which circuit is run next. We check whether a failure has occurred after termination of the protocol. The fault-free path is highlighted in green. We assume there is only one fault-free path. (b) Protocol sampling consists of two alternating types of steps. For the given circuit, a subset must be chosen ("choice"). At the end of that faulty circuit, the measurement ("outcome") determines the next circuit. These steps are iterated until the protocol finishes. The choice can be taken at random or by using a diagnostic criterion such as maximizing the expected reduction of uncertainty (ERU). Measurement results are stochastically determined by "nature". (c) A box representation of the tree in (a) contains circuits c and measurements that cause branchings. The horizontal width of any w-box symbolizes its binomial factor A_w^c. The failure rate lower bound p_l is the sum over all individual path failure rates (red portion of lower horizontal edge). The true failure rate is underestimated by at most all boxes that were not explored (yellow), i.e. the cutoff error δ, which allows calculation of an upper bound to the failure rate p_u (red + yellow portions of lower horizontal edge). Both bounds have respective sampling uncertainties $\sigma_{l/u}$ that are caused by the sampling uncertainties σ_i of the branching ratios q_i. The fault-free path never leads to failure. (d) Generic dynamical sequence of two circuits. Measurement of the first two qubits in the first circuit (purple) determines whether either the upper (green) or the lower (orange) circuit is run on the last two qubits.

Any *stage* $i \in \mathbb{N}$ of a sequence $\mathcal{C}$ is specified by a tuple

$$\mathcal{C}_i = (c_i, w_i) \tag{3.47}$$

that denotes the ideal circuit c_i and the w_i-fault-subset. Paths from the root to any leaf of the tree represent alternating patterns of circuits c_i and subsets w_i (see Fig. 3.7(b)). We define the *transition rate* from any faulty circuit $\mathcal{C}_i$ in the sequence to any successor at stage $i+1$ as

$$q_i \equiv \mathbb{P}(\mathcal{C}_i \to c_{i+1} |\, \mathcal{C}_{<i}) \tag{3.48}$$

given all previous faulty circuits $\mathcal{C}_{<i}$, which can be estimated from a number of $N_{\mathcal{C}_i}$ samples. Sequential branchings into distinct subsets of size $A^{c_i}_{w_i}$ for each circuit c_i and branching options for the next circuit after any w_i with branching ratios[10] q_i and $1-q_i$ are shown[11] with more detail in Fig. 3.7(c). The sampling uncertainty of q_i is given by the Wilson interval (3.24) and denoted as

$$\sigma_i = \sqrt{\mathrm{Var}(q_i)}. \tag{3.49}$$

As a crucial difference to single-circuit subset sampling, it is impossible to determine locally, i.e. at stage i, by qubit measurements whether or not the execution of a faulty circuit $\mathcal{C}_i$ will lead to failure after realizing the full sequence $\mathcal{C}$. Not only because the particular fault w_i is drawn randomly but essentially because the effect of any fault on subsequent circuits may not be determined a priori at stage i. We can, however, assume an expected value for a protocol failure obtained from previous runs of the whole protocol and update this expectation value continuously when more measurements of $\mathcal{C}_i$ are sampled. Only a posteriori, after termination of the whole protocol, we can determine whether a protocol failure has occurred as a result of a specific faulty circuit sequence that was realized in this particular shot, which is symbolized in Fig. 3.7(c) by a red outcome at the end of a tree path. The product of all factors (A and q) along a path determine its total contribution to the overall failure rate.

We define the DSS estimator, as in Eq. (3.32), as the mean value $\hat{p} \equiv \frac{p_l+p_u}{2}$ between a lower and an upper bound, which can be determined from the tree structure as

$$p_l = \sum_{P_{\mathrm{fail}}} \prod_{i \in P_{\mathrm{fail}}} A_{\mathcal{C}_i} q_i \tag{3.50}$$

$$p_u = p_l + \delta. \tag{3.51}$$

The lower bound p_l is the weighted sum over all paths P_{fail} that lead to failure and the last branching ratio of any length-l path $q_l \equiv \mathbb{P}(\mathcal{C}_{\leq l} \to \mathrm{fail})$ is the path failure rate. The cutoff error δ, which determines the upper bound p_u, is quantified, analogously to Eq. (3.31), as the worst-case error made by neglecting subsets and assuming they always produce failures. Therefore, the cutoff error $\delta_{\mathcal{C}_i}$ of any circuit node $\mathcal{C}_i$ of the tree is the complement of the binomial factors of all non-neglected subsets

$$\delta_{\mathcal{C}_i} = 1 - \sum_{w_i=0}^{w_{i,\max}} A_{\mathcal{C}_i}, \tag{3.52}$$

[10] We use the terms "branching ratio" and "transition rate" interchangeably.

[11] We assume that any measurement can only lead to a binary decision which one of two possible circuits follows next.

illustrated by any yellow box in Fig. 3.7(c). Then, the individual circuit cutoff errors $\delta_{\mathcal{C}_i}$ are aggregated along the respective paths in the event tree to the total cutoff error

$$\delta = \sum_{\mathcal{C}_i} \left(\prod_{k=1}^{i} A_{\mathcal{C}_k} q_k \right) \delta_{\mathcal{C}_i} \tag{3.53}$$

where one must take into account the path product that leads from the root node to any $\delta_{\mathcal{C}_i}$ in the tree (see Fig. 3.7(c)). Note that, therefore, δ now contains the transition rates q, which are sampled quantities and thus the two types of uncertainty, sampling uncertainty σ and cutoff error δ, are no longer well-separated as was the case in single-circuit subset sampling. The uncertainties of the upper and lower bounds $p_{l/u}$,

$$\sigma_{l/u} = \sqrt{\mathrm{Var}\big(p_{l/u}\big)}, \tag{3.54}$$

determine the uncertainty of the DSS estimator $[p_l - \sigma_l,\, p_u + \sigma_u] = [\hat{p} - \sigma_l - \delta/2,\, \hat{p} + \sigma_u + \delta/2]$ where, due to the dependence of δ on the branching ratios, we now have $\sigma_l \neq \sigma_u$ and we denote the total estimation uncertainty as

$$\eta \equiv \sigma_l + \sigma_u + \delta. \tag{3.55}$$

A detailed derivation for the calculation of $\mathrm{Var}\big(p_{l/u}\big)$ is given in Sec. 3.5.2.1.

Let us complete our specification of the DSS procedure by discussing suitable choice criteria for w_i-fault subsets. The ERU criterion translates from the single-circuit case to the dynamical case. It can be applied to any branching ratio q_i, which is affected by the assumed additional positive/negative measurement outcome, analogously to Eqs. (3.43) and (3.44), as

$$q_i \to q_i^{\pm} = \begin{cases} \left(q_i + \frac{1}{N_{\mathcal{C}_i}}\right) \frac{N_{\mathcal{C}_i}}{N_{\mathcal{C}_i}+1}, & + \\ q_i \frac{N_{\mathcal{C}_i}}{N_{\mathcal{C}_i}+1}. & - \end{cases} \tag{3.56}$$

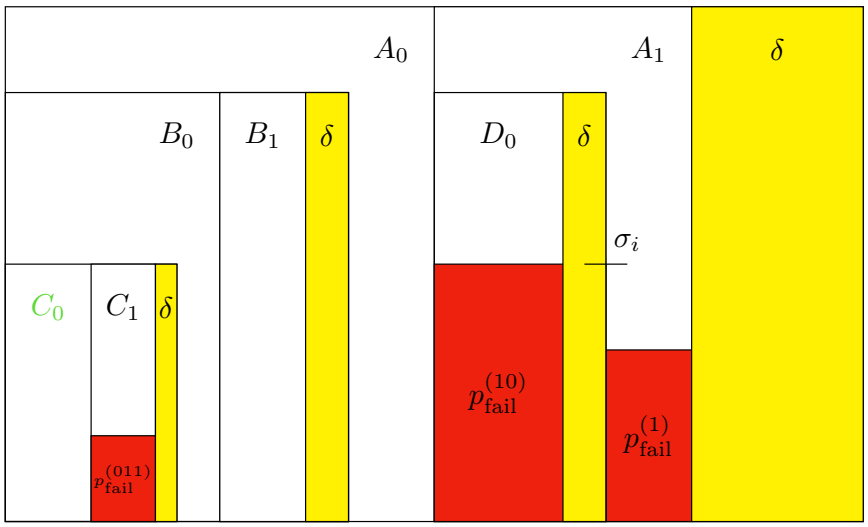


FIGURE 3.8: Circuit sequences can be represented by nested boxes as in Fig. 3.7(c). A subbox can take up at most all the space provided by the enclosing box. The fault-free path has the total binomial factor $A_0 B_0 C_0$ here. Within the subbox labelled by the green C_0, no failure can occur (compare to the green path in Fig. 3.7(a)).

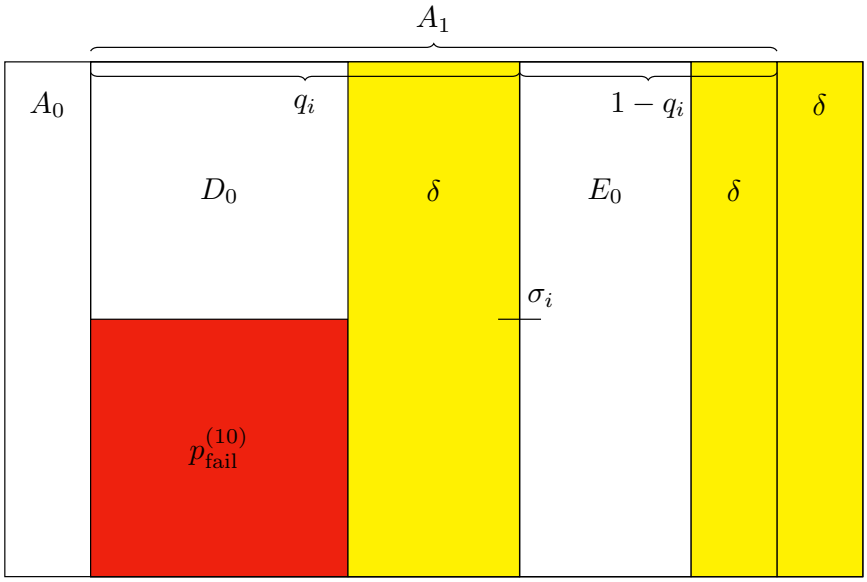


FIGURE 3.9: Subboxes of the A_1-box, labeled D and E, are weighted by the branching ratio q_i or $1-q_i$ respectively (cf. Fig. 3.7(c)). Since the branching ratio estimate q_i has an uncertainty σ_i, the size of any subbox is not determined with total confidence.

Then, the expected total uncertainties $\eta|_{w_i\to\pm}$ can be calculated from Eq. (3.56) and the ERU in Eq. (3.40) changes to

$$\Delta(w_i) = \eta - [q_i\eta|_{w_i\to+} + (1-q_i)\eta|_{w_i\to-}]. \tag{3.57}$$

The ERU is maximized by choosing the subset $w_i^* = \text{argmax}_{w_i}\,\Delta(w_i)$. Calculating the expected reduction of uncertainty $\Delta(w_i)$ for every shot and every circuit for all $w_{\max}+1$ possible choices of subsets might be infeasible and some computational aspects are further considered in Sec. 3.5.2.3.

Some additional remarks on Eq. (3.57) are in order. Note that when choosing a fault-weight-subset in any stage of an event tree, the total uncertainty of the whole subtree from this stage on and the cutoff error in the given stage must be considered. This is illustrated more clearly by the nested boxes in Fig. 3.8. When choosing the 1-subset for instance, the total uncertainty for all subboxes of A_1 must be taken into account. This especially includes the size of the subboxes, which are only known with the sampling uncertainty σ_i of the branching ratio q_i, which here is the transition rate from A_1 to another circuit with binomial factors D_w. For the fault-free path, in the leftmost part of Fig. 3.8, this uncertainty does not exist since we assume that there can be no branchings off from a sequence of ideal circuits. We "zoom in" to the A_1-subbox in Fig. 3.9 to further illustrate its internal structure. It not only contains the D_w partition but also the other branching option that is taken with the complementary ratio $1-q_i$ and is labeled by E_w. The contribution of failures in the D_0 subbox varies with q_i and in this example its relative contribution is reduced when, due to additional samples, the E_0 box, which does not contain failures, is weighted more strongly, i.e. $1-q_i$ becomes larger.

When we probe to open a new subset with the ERU criterion, it is sensible to assume, before actually running a shot, that this subset will lead to failure with a finite probability, e.g. 1/2 or the current average value of all sampled subset failure rates $\frac{1}{N_w}\sum_w p_{\text{fail}}^{(w)}$. Assuming a subset failure rate of 1 is impractical because this way the cutoff error δ does not change (see Sec. 3.5.2.1). Compared to Eq. (3.46) we do not get

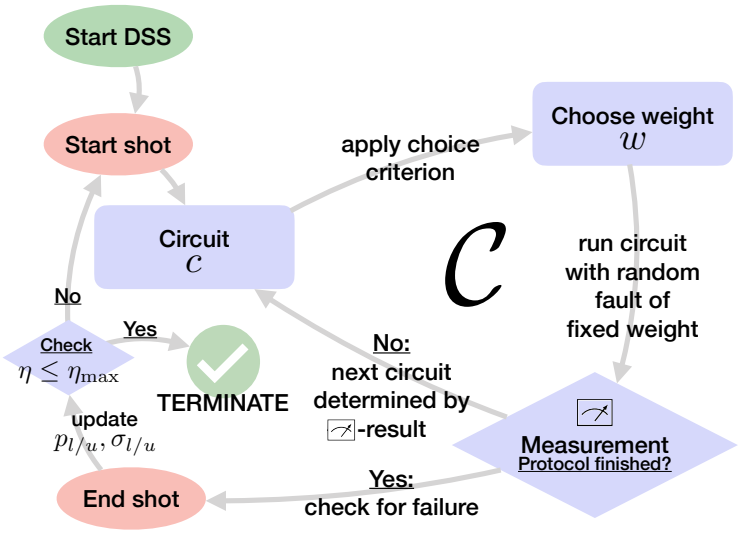


FIGURE 3.10: Visualization of the choice–measurement process from Fig. 3.7(b) as a circular flowchart. Any single shot of DSS consists of alternating between choosing the next circuit c_i to sample and choosing the fault-weight-subset w_i for this circuit. The underlying protocol $\mathcal{P}$ and the stochastic intermediate measurement results determine the circuit sequence $\mathcal{C}$. Fault-weight-subsets can be chosen according to the binomial factor choice criterion or the ERU criterion. Random fault instances within each subset are drawn uniformly. Repeated execution of the inner cycle generates circuit sequences $\mathcal{C}$, see Eq. (3.47), and the event tree in Fig. 3.7. If the previous shot does not lead to a resulting η – the total estimation uncertainty – below the target uncertainty $\eta_{\max}$ or the maximum number of shots $N_{\max}$ is not reached yet, a new shot is run.

immediate feedback about failure of the circuit because the protocol does, in general, not terminate after our choice of a subset. Both possible subsequent branchings must be taken into account and both are assumed to yield failure with probability 1. For example, this would be the case if, in Fig. 3.9, all the D- and E-subboxes were filled red (or yellow). Also, assuming an initial failure probability of 0 is unrealistic for a non-FT path, i.e. a path that could potentially result in failure.

A simpler method for choosing subsets is to draw random fault weights w_i according to the distribution of binomial factors $A_{\mathcal{C}_i}$ of the current circuit c_i. This strategy, that we call *binomial factor choice*, will produce small total fault weights $w_{\text{tot}} = \sum_{i \in P} w_i$ in a single shot if p is not too large and the circuit sequence is sufficiently long. If p is too small or the circuit sequence is too short, w_{tot} might not be different from zero and thus the trivial, ideal protocol will be sampled. In the single-circuit case or for repeated runs of the same circuit until the result is accepted where faults cannot propagate from one circuit to the next (as for flag verification for instance, see Sec. 2.4.3), one could just exclude the 0-fault-subset. In principle, any distribution can be used to draw fault operators, so the pre-sampling strategy from [Pae+23] could be employed here as well.

The full dynamical subset sampling algorithm is shown in the flowchart in Fig. 3.10 and expressed in pseudocode in Alg. 1. Its output are numerical values of $p_{l/u}$ and $\sigma_{l/u}$, i.e. the bounds on the true failure rate p^* and their uncertainties. The regime of low physical fault rates $\hat{p}(p < p_{\max})$ can then be inferred analytically since the dependency on p lies entirely in the binomial factors $A_{\mathcal{C}_i}$, which do not depend on the sampled branching ratios q_i.

Algorithm 1 Dynamical subset sampling algorithm to estimate the failure rate p^* of a non-deterministic QEC protocol $\mathcal{P}$ including in-sequence measurements using the ERU criterion

Input: maximum number of samples $N_{\max}$ or target uncertainty $\eta_{\max}$, noise model $\mathcal{E}$, maximum noise strength $p_{\max}$
Output: failure rate bounds $p_{l/u}$ and uncertainties $\sigma_{l/u}$

while $N < N_{\max}$ and $\eta > \eta_{\max}$ **do**
  calculate $\Delta(w_i)$ for $w_i = 0, ..., w_{i,\max} + 1$
  set $w_i \leftarrow \text{argmax}_{w_i} \Delta(w_i)$
  run $\mathcal{C}_i$ with w_i-fault drawn from $\mathcal{E}$
  determine c_{i+1} from measurement outcome of $\mathcal{C}_i$ according to $\mathcal{P}$
  update branching ratio q_i
  if $c_{i+1} \neq$ "finish" **then**
    $i \leftarrow i + 1$ go to next circuit
  else
    check for failure
    update p_l, p_u, σ_l and σ_u
    $N \leftarrow N + 1$ go to next shot
  end if
end while

3.5.2.1 Uncertainty intervals

In this section, we provide a rigorous derivation how to practically calculate the variances of the upper and lower bounds $\text{Var}(p_{l/u})$ in Eq. (3.54) and thus determine the uncertainty interval of the DSS estimator. In general, the failure rate bounds $p_{l/u}$ take the form of sums of path products (see Eq. (3.50)) and the origin of the variance are the transition rate variances $V_i \equiv \text{Var}(q_i)$ while the binomial factors have no uncertainty. Any transition rate q_i is the expectation value of a Bernoulli random variable and all binomial factors A_w are constants in this regard. However, we interchangeably use q_i as both the random variable and its expectation value since the concrete meaning is always clear from the context.

The variance of a sum of n variables q_i with constants A_i is

$$\text{Var}\left(\sum_{i=1}^{n} A_i q_i\right) = \sum_{i=1}^{n} A_i^2 V_i + 2 \sum_{j=1}^{n} \sum_{i=1}^{j-1} A_i A_j \,\text{Cov}(q_i, q_j) \tag{3.58}$$

and the variance of a product of n variables q_i with constants A_i is

$$\text{Var}\left(\prod_{i=1}^{n} A_i q_i\right) = \prod_{i=1}^{n} A_i^2 (V_i + q_i^2) - \prod_{i=1}^{n} A_i^2 q_i^2 \tag{3.59}$$

as long as the q_i and q_i^2 are independent [Goo60; Goo62].

Indeed, two subsequent transition rates q_j and q_k, which belong to the same path in a tree, have a vanishing covariance $\text{Cov}(q_j, q_k) = 0$. To see this, assume two Bernoulli random variables X and Y. The event described by Y follows conditioned on X taking on the value 1 out of the two possible values 0 and 1. Then, from the definition of the covariance

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y), \tag{3.60}$$

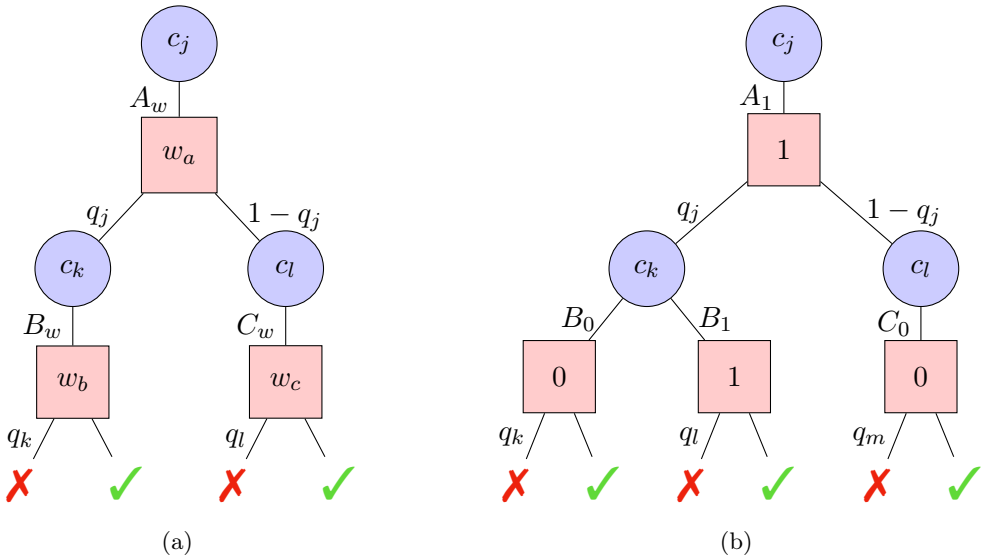


FIGURE 3.11: (a) Example event tree where circuit nodes (blue circles) are labeled c_j, c_k and c_l. Subset nodes (red squares) are labeled w_a, w_b and w_c. The branching ratios q_j and q_k as well as q_j and q_l (and also q_k and q_l) are uncorrelated. (b) Event tree with non-zero covariance of q_k- and q_l-paths, which share the random variable q_j. Indexing of the subset nodes is reduced to labels 0 and 1 for better readability.

where $\mathbb{E}(X)$ denotes the expectation value of X, it is clear that the covariance vanishes if X always takes on the value 1, which is the case for the transition rates q_j and q_k along a given path. However, non-zero covariance contributions in general arise from overlapping paths that split at some node in the tree and we give two examples for calculating the lower bound variance $\mathrm{Var}(p_l)$ with a branching from a subset node and branching from a circuit node in the following. Then follows a third example that illustrates calculation of the upper bound variance $\mathrm{Var}(p_u)$.

Subset node branching

The tree in Fig. 3.11(a) splits at the *subset node* w_a and it has two paths that lead to failure, namely the path along q_j and q_k and the path along $1-q_j$ and q_l. Binomial factors of the circuits c_k and c_l are denoted B_w and C_w respectively. The variance of the sum of these two paths reads

$$\begin{aligned}
&\mathrm{Var}(A_w(B_w q_j q_k + C_w(1-q_j)q_l)) \\
&= A_w^2\left[B_w^2\,\mathrm{Var}(q_j q_k) + C_w^2\,\mathrm{Var}((1-q_j)q_l) + 2B_wC_w\,\mathrm{Cov}(q_j q_k, (1-q_j)q_l)\right] \\
&= A_w^2\left[B_w^2(q_j^2V_k + q_k^2V_j + V_jV_k) + C_w^2(q_l^2V_j + (1-q_j)^2V_l + V_jV_l) - 2B_wC_wq_kq_lV_j\right]
\end{aligned} \tag{3.61}$$

using Eqs. (3.58) and (3.59) and the fact that

$$\mathrm{Cov}(q_j q_k, (1-q_j)q_l) = -q_k q_l \mathrm{Var}(q_j) \tag{3.62}$$

since $\mathrm{Cov}(q_k, q_l) = 0$.

Circuit node branching

The tree in Fig. 3.11(b) splits at the *circuit node* c_k and it has two paths that thereafter lead to fail, namely the path along q_j and q_k and the path along q_j and q_l (and it also

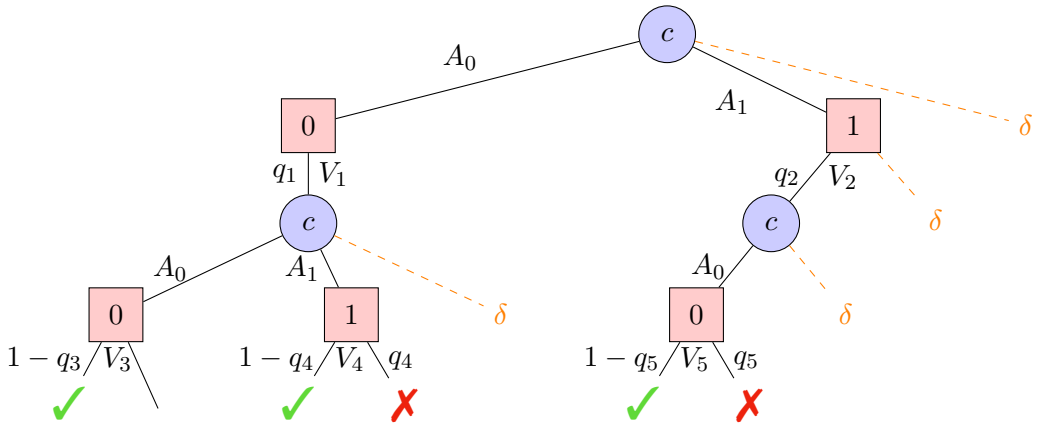


FIGURE 3.12: Example tree from running the circuit c two consecutive times. At the end of the fault-free path, no failure can happen so $q_3 = V_3 = 0$. Dashed orange lines, branching from both circuit and subset nodes, indicate yet unexplored branches but could potentially lead to failures and thus contribute to the cutoff error δ (δ-node). If the protocol for this tree is FT-1, then $q_4 = q_5 = 0$ and $V_4 = V_5 = 0$ because no failures can result from paths that only contain a single fault. Since we assume no branching off the fault-free path, $V_1 = 0$.

splits at the upper subset node 1 and the path along $1 - q_j$ and q_m can also lead to fail). The variance of the lower bound p_l for this tree reads

$$\begin{aligned}
&\mathrm{Var}(A_1(q_j(B_0q_k + B_1q_l) + (1-q_j)C_0q_m)) \\
&= A_1^2\mathrm{Var}(q_j(B_0q_k + B_1q_l) + (1-q_j)C_0q_m) \\
&= A_1^2[\mathrm{Var}(q_j(B_0q_k + B_1q_l)) + \mathrm{Var}((1-q_j)C_0q_m) \\
&\qquad\qquad + 2\,\mathrm{Cov}(q_j(B_0q_k + B_1q_l), (1-q_j)C_0q_m)] \\
&= A_1^2\big[B_0^2\,\mathrm{Var}(q_jq_k) + B_1^2\,\mathrm{Var}(q_jq_l) + 2B_0B_1\,\mathrm{Cov}(q_jq_k, q_jq_l) + C_0^2\mathrm{Var}((1-q_j)q_m) \\
&\qquad\qquad + 2\,\mathrm{Cov}(q_j(B_0q_k + B_1q_l), (1-q_j)C_0q_m)\big] \qquad (3.63)
\end{aligned}$$

and could be expanded further by employing Eq. (3.59) for the product terms as well as Eq. (3.62) and $\mathrm{Cov}(q_jq_k, q_jq_l) = q_kq_l\,\mathrm{Var}(q_j)$.

Upper bound variance

When calculating the upper bound variance $\mathrm{Var}(p_u)$ we must take into account the cutoff error δ, i.e. we must consider all possible but yet unsampled branchings in any event tree that could *potentially* lead to failure when running more shots. The incomplete tree in Fig. 3.12 illustrates this as an example. Here we assume a protocol that repeatedly runs the circuit c. Also, another circuit $\bar{c}$ could potentially be run after the first execution of c. All unsampled circuit nodes contribute to the cutoff error, as marked by the δ-nodes that follow from the purple circuit nodes in Fig. 3.12. However, there is also a contribution to $\mathrm{Var}(p_u)$ from the red subset nodes, in particular the "incomplete" branching ratio q_2. In this example, so far only the branching to the left has been realized so that $q_2 = 1$ but, since $V_2 > 0$, in principle in any of the next shots one could also realize a branching to the circuit $\bar{c}$ thus decreasing q_2.

We can write the complete cutoff error δ from this tree, according to Eq. (3.53), by adding the term $1 - q_2 = 0$ as

$$\delta = 1 - A_0 - A_1 + A_0q_1(1 - A_0 - A_1) + A_1(q_2(1 - A_0) + (1 - q_2))$$

so that the upper bound $p_u = p_l + \delta$ can be expressed as

$$p_u = 1 - A_0 - A_1 + A_0 q_1 (1 - A_0 - A_1(1 - q_4)) + A_1(q_2(1 - A_0(1 - q_5)) + (1 - q_2)).$$

The variance of p_u then reads

$$\mathrm{Var}(1 - A_0 - A_1 + A_0 q_1 (1 - A_0 - A_1(1 - q_4)) + A_1(q_2(1 - A_0(1 - q_5)) + (1 - q_2))) \tag{3.64}$$

and can be split into two terms since the we can treat the subtrees in the left and right halves of Fig. 3.12 distinctly. The constant terms in the beginning, which do not contain any q, can be left out. The first term in Eq. (3.64) then is

$$\begin{aligned} &A_0^2 \,\mathrm{Var}(q_1(1 - A_0 - A_1(1 - q_4))) \\ &= A_0^2((1 - A_0)^2 V_1 + A_1^2 \,\mathrm{Var}(q_1(1 - q_4)) - 2(1 - A_0)A_1(1 - q_4)V_1) \end{aligned} \tag{3.65}$$

and the second term is

$$\begin{aligned} &A_1^2 \,\mathrm{Var}(q_2(1 - A_0(1 - q_5)) + (1 - q_2)) \\ &= A_1^2(\mathrm{Var}(q_2(1 - A_0(1 - q_5))) + V_2 - 2(1 - A_0(1 - q_5))V_2), \end{aligned} \tag{3.66}$$

which contains the contribution $V_2 > 0$. Both terms are in leading order quadratic in p, as required.

However, we see from these three examples that explicitly calculating the variance quickly becomes cumbersome, although straightforward.

Recursive calculation

As a more practical procedure to calculate the upper and lower bound and their respective variances, we note that one may exploit the recursive structure of any event tree. Our quantities of interest can be "propagated" up from the leaf nodes until its root node in the following way. The two components depicted in Fig. 3.13 can be used to build up any event tree.

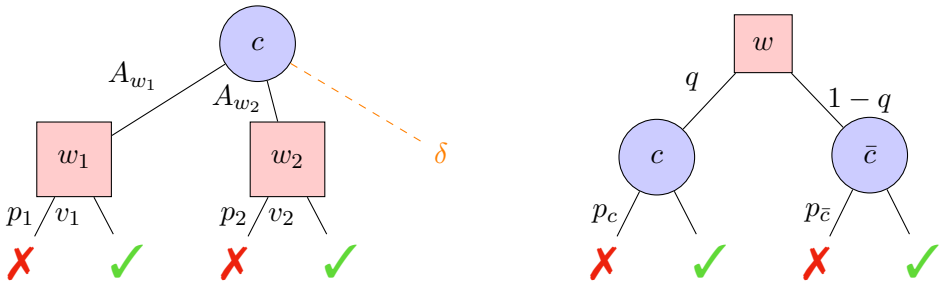
For this purpose, we introduce a slightly different notation. Any subset node w has a node failure rate $w.p$, which may be interpreted as p being a property or an attribute of w. This node failure rate can be calculated from the transition rate $w.q$ from w to its subsequent circuit node c (and $1 - w.q$ is the transition rate to the other subsequent circuit node $\bar{c}$) as

$$w.p = w.q \times c.p + (1 - w.q) \times \bar{c}.p. \tag{3.67}$$

If w is a leaf node, i.e. there is no subsequent circuit, $w.q$ is the failure rate. The node failure rate of a circuit node

$$c.p = \sum_{w \in c} A_w \, w.p \tag{3.68}$$

is defined analogously as the total failure rate of all subtrees below c where w are all subset nodes directly subsequent to c. The variance of a circuit node is given by $c.v = \sum_{w \in c} A_w^2 \, w.v$ and the variance of any subset node w can be calculated by



(a) $c.p = A_{w_1} p_1 + A_{w_2} p_2 \,(+\,\delta)$ (b) $w.p = q \times p_c + (1-q) \times p_{\bar{c}}$

FIGURE 3.13: We can recursively construct any event tree by branchings from (a) circuit nodes c to subset nodes $w_1, w_2, ...$ and (b) subset nodes w to circuit nodes c and $\bar{c}$. The indicated failures correspond to coarse-grained subtree failure rates at higher levels, e.g. if c is the root node. The base case of recursion is reached if any w does not have a circuit node successor but is followed by termination of the protocol. Failure rates and variances of upper bounds are calculated by including the cutoff error δ to a node's failure rate. Its value must be determined according to the rules given at the end this section.

employing the general formulas from Eqs. (3.58) and (3.59) as

$$\begin{aligned} w.v &= \mathrm{Var}(w.p) = \mathrm{Var}(w.q \times c.p + (1 - w.q) \times \bar{c}.p) \\ &= \mathrm{Var}(w.q \times c.p) + \mathrm{Var}((1 - w.q) \times \bar{c}.p) - 2\, c.p \times \bar{c}.p\, \mathrm{Var}(w.q) \\ &= w.q^2 \times c.v + \mathrm{Var}(w.q)\, c.p^2 + \mathrm{Var}(w.q)\, c.v \\ &+ (1 - w.q)^2\, \bar{c}.v + \mathrm{Var}(w.q)\, \bar{c}.p^2 + \mathrm{Var}(w.q)\, \bar{c}.v - 2\, c.p \times \bar{c}.p\, \mathrm{Var}(w.q) \\ &= \mathrm{Var}(w.q)\big[c.p^2 + c.v + \bar{c}.p^2 + \bar{c}.v - 2\, c.p \times \bar{c}.p\big] + w.q^2 \times c.v + (1 - w.q)^2\, \bar{c}.v \\ &= \mathrm{Var}(w.q)\big[(c.p - \bar{c}.p)^2 + c.v + \bar{c}.v\big] + w.q^2 \times c.v + (1 - w.q)^2\, \bar{c}.v \end{aligned} \tag{3.69}$$

if w is not a leaf node and $w.v = \mathrm{Var}(w.q)$ if w is a leaf node.

When calculating an upper bound property, we only need to include the circuit cutoff error $c.\delta = 1 - \sum_{w=0}^{w_{\max}} A_w^c$ such that Eq. (3.68) is extended to $c.p = \sum_{w \in c} A_w\, w.p + c.\delta$.

FT cutoff error

The cutoff error as discussed above is not sensitive to potential fault tolerance properties that a protocol might have. For instance, we saw that the variance of the upper bound in Eq. (3.64) scales quadratically in p so that the uncertainty σ_u scales linearly. Now assume that the protocol that we modeled with the tree in Fig. 3.12 is FT-1. We would have a true failure rate $p^* \sim p^2$ as $p \to 0$ but the upper bound uncertainty $\sigma_u \sim p$ so that the relative error $\sigma_u / p^* \sim p^{-1}$ increases as p is lowered, which destroys the constant error guarantee of subset sampling. This upper bound would also be overly pessimistic since we know beforehand that the scaling will be quadratic for an FT-1 protocol. The challenge of combining the fault tolerance guarantee with the worst-case bound given by the cutoff error is illustrated in Fig. 3.14(a) for a protocol that consists of two circuit nodes ENC and MEAS. While naively we could calculate the circuit cutoff error as $1 - A_0 \sim p$, we know beforehand that there can be no failure in order p due to FT-1. But assuming a circuit cutoff error $1 - A_0 - A_1 \sim p^2$ is not a worst-case bound since in fact there could follow failures below the A_1 subset when additional faulty circuits are run.

In the following we derive how the cutoff error δ can be calculated such that the fault tolerance property is incorporated and the constant relative error guarantee for

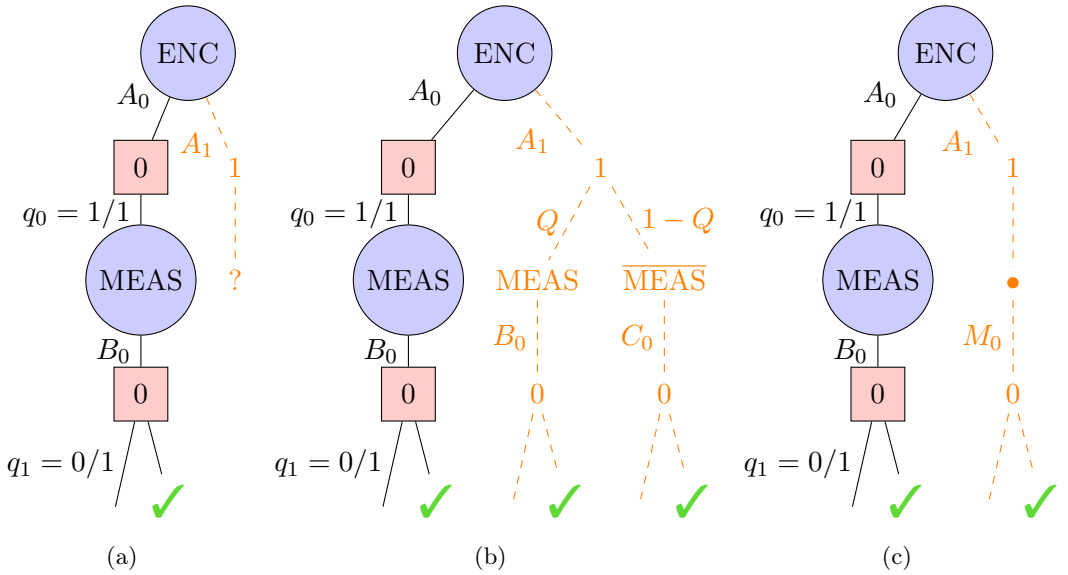


FIGURE 3.14: Example trees illustrate the cutoff error (orange, no frames) after a single shot of an FT protocol containing the circuits ENC and MEAS (black frames). We assume here that the protocol consists of two stages which is not the case in general. (a) What is the worst-case contribution of the orange 1-subset to the failure rate? (b) Two possible circuit branchings within the unsampled subtree have an FT path. (c) The FT path contributions can be bounded by a single node $\bullet$ with binomial factors $A_w^\bullet = M_w$.

subset sampling is restored. We show in two steps – for a QEC code of distance $d = 3$ – how one can choose the upper bound such that the relative error stays constant when scaling p, i.e. the scaling of the upper bound and its sampling error are of the same order in p as the failure rate. The first step is an estimation for the branching ratios at a single stage of the tree. The second step then considers the worst-case FT tree depth or, equivalently, the maximum possible length of a circuit sequence in case only tolerable fault configurations happen, which we call the maximum FT path length L. This length is always known in practice since the protocol has been actively constructed in order to comply with fault tolerance.

Consider two possible paths that could follow after ENC-1 in Fig. 3.14(b) and for the moment assume that the protocol only executes one additional circuit, which is either MEAS or a circuit other than MEAS (called $\overline{\text{MEAS}}$). The path products for these two paths are $A_1 Q B_0$ and $A_1(1-Q)C_0$. They contain an unknown branching ratio Q. The total contribution to the cutoff error from the orange subtree is given by $A_1(Q(1-B_0) + (1-Q)(1-C_0))$. Assuming, without loss of generality, that one of the two binomial factors is larger than the other one, for instance $B_0 > C_0$, we can bound the cutoff error contribution from above as

$$A_1(Q(1-B_0) + (1-Q)(1-C_0)) \leq A_1(Q(1-C_0) + (1-Q)(1-C_0)) = A_1(1-C_0), \tag{3.70}$$

which removes the unknown branching ratio Q from the expression. In practice, we know all possible circuits when we model a protocol and we know their binomial factors independently from the sampling process. So, as Fig. 3.14(c) illustrates, we can replace the branching with transition rate Q by a single node with the minimal binomial factor

$$M_0 = \min_{c \in \mathcal{C} \in \mathcal{P}} (A_0^c) \tag{3.71}$$

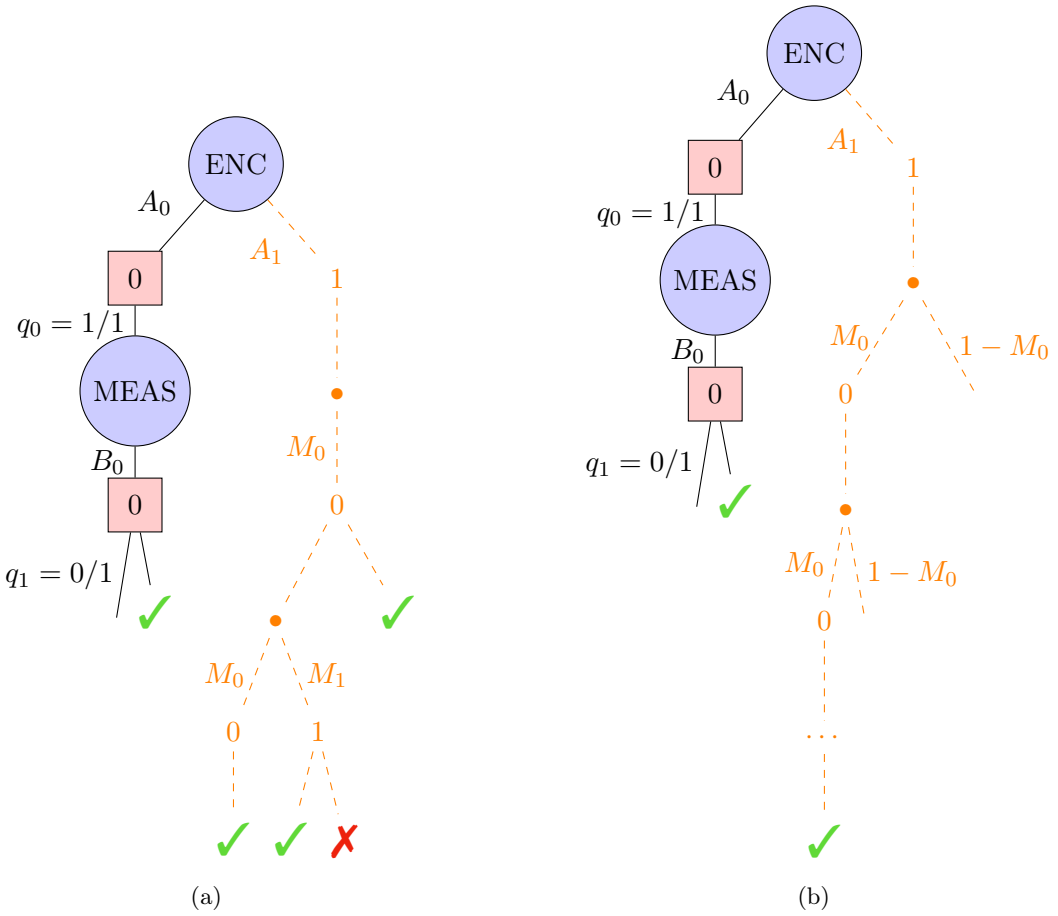


FIGURE 3.15: Event tree examples that illustrate the cutoff error contributions from longer circuit sequences. (a) A contribution to the failure rate can come from a deeper stage within the orange subtree (1-0-1), further down than another FT path (1-0). (b) All cutoff error contributions are of the form $1 - M_0$ within the orange subtree illustrated by the two branchings to the right. The actual circuit sequence is unimportant and the cutoff error can be bounded by all possible $1 - M_0$ branchings.

out of all circuits that are part of the protocol $\mathcal{P}$ so that $1 - M_0$ can be used as an upper bound in the sense of Eq. (3.70).

The second step is to allow for longer circuit sequences, which are however bounded by the maximum FT sequence length L that is known from the design of the protocol. The situation of a later fault causing failure in conjunction with the earlier 1-fault-subset is depicted in Fig. 3.15(a). While the weight sequence 1-0 can never lead to failure, in a longer sequence, the weight sequence 1-0-1 is not FT anymore but can indeed lead to failure and must be considered in the FT cutoff error.

This particular example is contained in the path along 1-0-$(1 - M_0)$ in Fig. 3.15(b). From this tree, we can calculate the total contribution from all possible circuit cutoff errors contained in the orange subtree as

$$\begin{aligned} &A_1((1 - M_0) + M_0(1 - M_0) + M_0^2(1 - M_0) + \cdots + M_0^{L-1}(1 - M_0)) \\ &= A_1(1 - M_0) \sum_{k=0}^{L-1} M_0^k \leq L A_1(1 - M_0) \sim p^2 \end{aligned} \tag{3.72}$$

where we upper-bound the factors M_0 by 1. This is necessary because otherwise

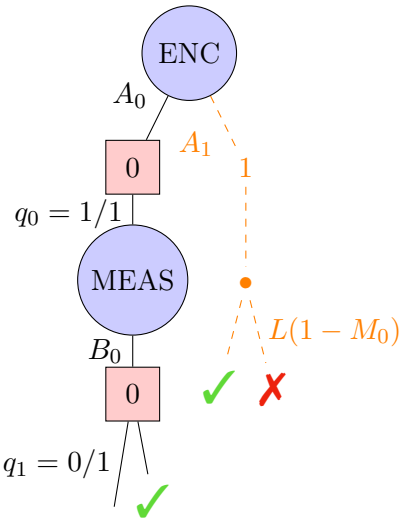


FIGURE 3.16: The worst-case contribution of the orange 1-subset to the failure rate from Fig. 3.14(a) can be upper-bounded by $LA_1(1 - M_0)$.

we may underestimate the path contribution from longer sequences, i.e. the left-branchings denoted M_0 in Fig. 3.15(b) can actually have binomial factors larger than M_0 in general. The final result of the estimation in Eq. (3.72), which we illustrate in Fig. 3.16, now scales quadratically and is a worst-case bound, as desired.

For a more general tree, we now illustrate different cases of how the (FT) cutoff error can be chosen in Fig. 3.17, which correspond to the following rules:

(I) If the total path weight up to a subset node is 0, add no additional branching circuit because we assume that there is only one fault-free path.

(II) Only for an FT protocol: If the total path weight up to a subset node is 1 and the branching ratio q of the subset node is equal to 1 and there is only one (circuit) node below the subset node but the other (circuit) node is still unknown, then add a complementary branching with factor $1 - q$ and a δ-node with value $L(1-M_0)$. While this node does not contribute directly to the cutoff error, it contributes to the variance of the upper bound via $\mathrm{Var}(1 - q) \neq 0$.

(III) If the total path weight up to a subset node is 2 (for an FT protocol) / 1 (for a non-FT protocol) or larger and the branching ratio q of the subset node is equal to 1 and there is only one circuit node below the subset node but the other one is still unknown, then add a complementary branching with factor $1 - q$ and a δ-node with value 1 below the complementary branching with $1 - q$.

(IV) Only for an FT protocol: If the total path weight up to a circuit node is 0 and if the circuit node does not have a 1-subset, add a virtual one. The virtual 1-subset has a δ-node with value $L(1 - M_0)$. The δ-node of the circuit node then has the usual value according to Eq. (3.52).

We end our discussion of the uncertainty interval of the DSS estimator by some concluding remarks. First note that for a "repeat until success" protocol we could count its maximum FT sequence length from the last qubit reset on. For instance, assume that the circuit ENC is repeated until success and then MEAS follows. Since all qubits are reset in the beginning of ENC, faults cannot propagate from one execution to the next and the maximum FT sequence length can practically be set to $L = 2 + t$

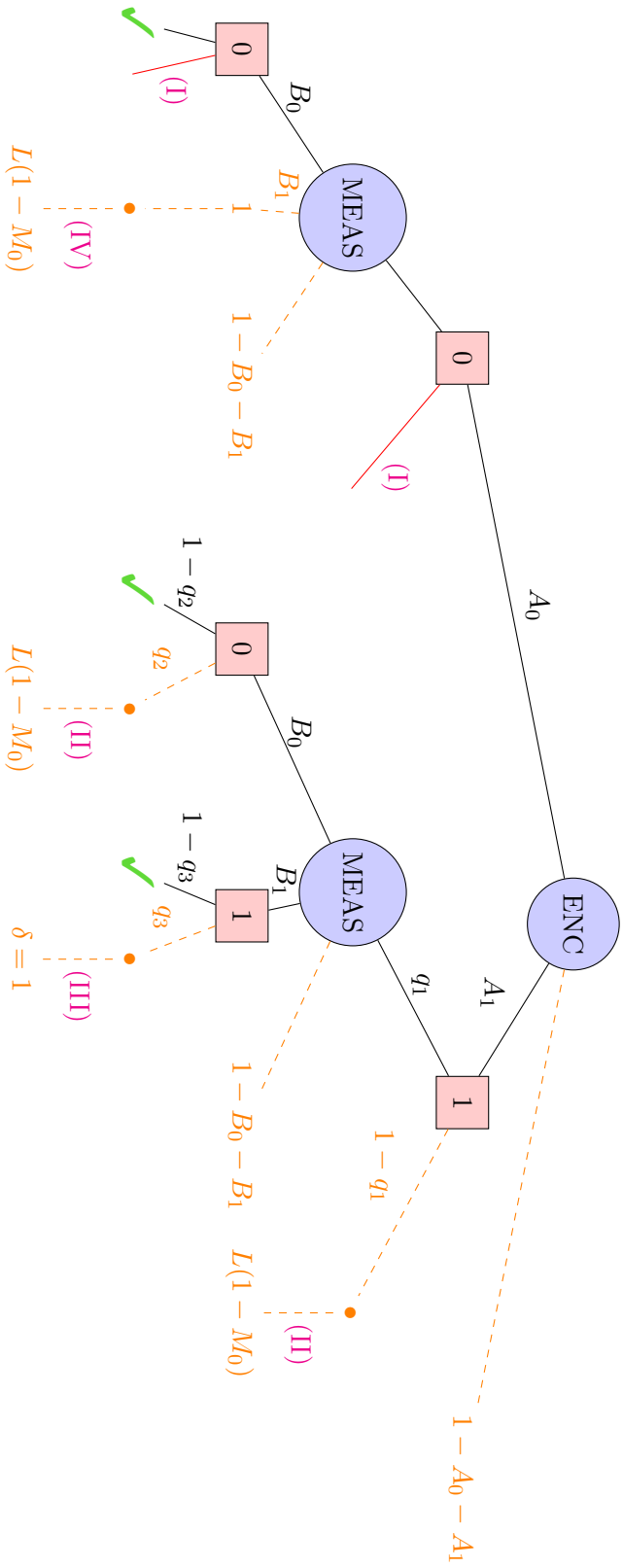


FIGURE 3.17: An event tree of an FT protocol that illustrates the four cases of the FT cutoff error as given in the main text. If the protocol were non-FT, we could replace the cases (II) and (IV) with instances of case (III).

with $t = 1$. A path ENC-1-ENC-0-MEAS-0, for instance, can be realized in practice. While a path ENC-1-ENC-1-ENC-0-MEAS-0 can also be realized, its contribution is of order p^2 and is contained in the cutoff error contribution $A_1(1-M_0)$ in Fig. 3.15(b). Another aspect is that one does not necessarily always have to use the factor L but, since the number of circuits that were already run i is always known in practice, L may be replaced by a maximum possible FT "rest of sequence" of length $l = L - i$.

A generalization of the FT cutoff error to QEC codes with arbitrary distance d should be possible by straightforward combinatorics, i.e. counting the worst-case number of subsequent circuit node cutoff contributions and upper bounding the binomial factors of their path products, but has not been conducted in this work.

3.5.2.2 Application to protocols with intermediate measurements

In this section, we apply dynamical subset sampling to two state-of-the-art protocols to fault-tolerantly prepare the $|0\rangle_L$ state of the Steane code. The first is an extension of the flag circuit of [Got16a] that always terminates with the correctly prepared state after running at most three circuits [Heu+23a]. Although the inner workings of this protocol are interesting in their own regard and thus are explained in Sec. 4.1.3.2, here we focus on the noise simulation. The second protocol is FT state preparation by means of sequential stabilizer measurements using flag circuits, according to [CB18]. Here, eight circuits can be run in total.

For all simulations, we apply the standard circuit-level depolarizing noise model specified by the channel in Eq. (3.13), which is applied to single-qubit gates, two-qubit gates, physical qubit initializations and measurements, with noise parameters $\vec{p} = (p_1, p_2)$. This means that we apply the fault-operators

$$F_1 \in \{\sigma_i, \forall i \in [1,2,3]\} \tag{3.73}$$

$$F_2 \in \{\sigma_i \otimes \sigma_j, \forall i,j \in [0,1,2,3]\} \setminus I \otimes I \tag{3.74}$$

to any ideal operation. Single-qubit operations are executed ideally with probability $1 - p_1$ and a fault operator F_1 is applied uniformly with probability p_1. Two-qubit operations are executed ideally with probability $1 - p_2$ and a fault operator F_2 is applied uniformly with probability p_2. As a consequence

1. any single-qubit gate is followed by a Pauli fault F_1 drawn uniformly and independently from $\{X, Y, Z\}$ with probability $p_1/3$,
2. any two-qubit gate is followed by a Pauli fault F_2 drawn uniformly and independently from $\{I, X, Y, Z\}^{\otimes 2} \setminus I \otimes I$ with probability $p_2/15$,
3. any qubit is initialized to a flipped state (for example $|0\rangle \to |1\rangle$ or $|+\rangle \to |-\rangle$) with probability $2p_1/3$ and
4. any qubit measurement reports a flipped result ($\pm 1 \to \mp 1$) with probability $2p_1/3$.

We now move forward to the actual protocol examples.

Adaptive FT state preparation

A protocol that uses adaptive noisy circuits to fault-tolerantly prepare the $|0\rangle_L$ state of the Steane code with a finite sequence is illustrated in Fig. 3.18 [Heu+23a]. The first step is non-FT state preparation and measurement of the logical operator $Z_L = Z_3 Z_5 Z_6$

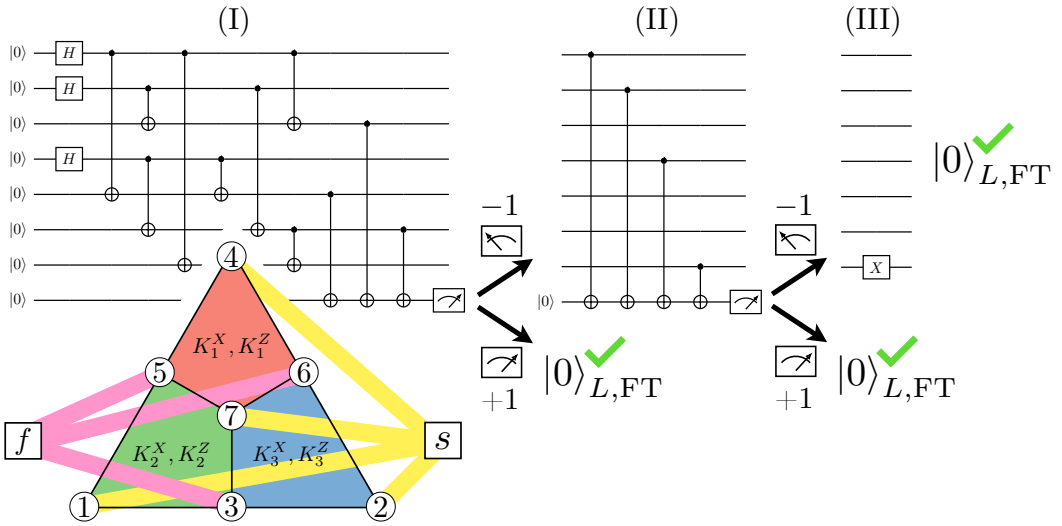


FIGURE 3.18: Dynamical circuit sequences for adaptive FT state preparation of $|0\rangle_L$ in the Steane code. When the flag (pink connections to f) is triggered in step (I), the stabilizer $K_1^Z K_2^Z K_3^Z = Z_1 Z_2 Z_4 Z_7$ is measured (yellow connections to s) in step (II). Then, if the next measurement also yields -1, the correction X_7 is applied in step (III). Otherwise the protocol terminates and $|0\rangle_L$ is prepared fault-tolerantly at any possible point of termination.

with a single flag qubit, as suggested in [Got16a]. If the flag is clear, one accepts the state. If instead the flag is triggered, the complementary operator – the product of all three stabilizer generators – $K_1^Z K_2^Z K_3^Z = Z_1 Z_2 Z_4 Z_7$ is measured subsequently. This second step can again be performed with a single auxiliary qubit. If the measurement yields $+1$, the state is accepted. If the measurement yields -1, a recovery operation X_7 is applied in a third step and then the protocol terminates. The result for any of the three possible protocol terminations is that the state $|0\rangle_L$ is prepared correctly up to a correctable weight-1 error[12]. In our simulations, we then measure all seven data qubits in the Z-basis, apply a final round of ideal EC and evaluate in software whether or not a logical failure has occurred. The maximum FT sequence length of this protocol is $L = 4$.

The binomial factors for our two-parameter noise model read

$$A^c_{\vec{w}}(p_1, p_2) = \binom{N_c^{(1)}}{w_1} \binom{N_c^{(2)}}{w_2} p_1^{w_1} p_2^{w_2} (1-p_1)^{N_c^{(1)} - w_1} (1-p_2)^{N_c^{(2)} - w_2} \tag{3.75}$$

where $N_c^{(1)}$ and $N_c^{(2)}$ are the number of single- and two-qubit gates in circuit c respectively. We fix our maximal physical fault rates, for this example, to the values $\vec{p}_{\max} = (p_{\max,1}, p_{\max,2}) = (10^{-3}, 10^{-2})$ and employ the binomial factor choice criterion to randomly choose fault-weight-subsets $\vec{w} = (w_1, w_2)$.

We scale both noise parameters p_1 and p_2 together uniformly to obtain the scaling in Fig. 3.19(a), which, as expected for an FT-1 protocol, is of quadratic form $\hat{p} \sim p^2$ as $\vec{p} \to \vec{0}$. The bounds estimated by DSS are tight in the relevant range $p_1 < p_{\max,1}$ and $p_2 < p_{\max,2}$ of the two parameters. Due to the quadratic scaling, the uncertainty

[12]The protocol is explained in more detail in Sec. 4.1.3.2 and reappears with a slight variation in Sec. 4.2.2.

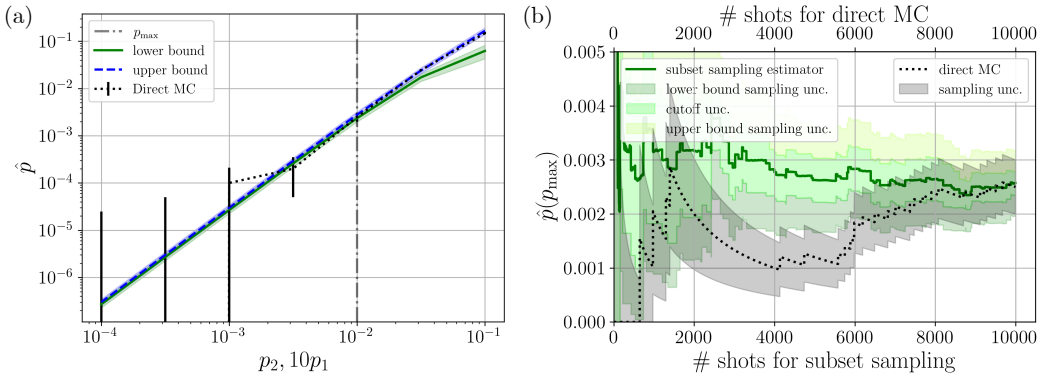


FIGURE 3.19: Logical failure rate estimations for the adaptive FT state preparation protocol. (a) Direct MC achieves a reasonable estimation for $p_2 \geq 10^{-2}$ (and $p_1 \geq 10^{-3}$). When diverging at growing $\vec{p}$, the upper and lower bounds estimated via subset sampling enclose the direct MC estimator. At low $\vec{p}$, however, direct MC records few or no failures after $N_{\mathrm{MC}} = 10^4$ shots. The analytical scaling of subset sampling yields tight uncertainty intervals around the estimated failure rate bounds in this regime. (b) The uncertainty interval η of the subset sampling estimator to the logical failure rate $\hat{p}$ (shades of green, solid) is composed of the upper and lower bound sampling uncertainties $\sigma_{l/u}$ and the cut-off uncertainty δ. After 10^4 shots at $\vec{p}_{\mathrm{max}} = (10^{-3}, 10^{-2})$, dynamical subset sampling yields η similar to the direct MC estimation uncertainty (black dotted with shaded area). We drew the faults weights randomly according to the circuits' binomial factors for this example.

intervals of the direct MC estimators, shown for comparison, quickly become unreasonably large for a fixed number of $N_{\mathrm{MC}} = 10^4$ samples as $\vec{p}$ is lowered. At small $\vec{p}$, very few or even no logical failures are registered anymore by direct MC because the failure rate is too low.

Figure 3.19(b) reveals that the two uncertainty intervals for both DSS and direct MC are of comparable size after 10^4 shots for our choice of $\vec{p}_{\mathrm{max}}$. However, many more shots would be needed with direct MC to reliably estimate failure rates at lower $\vec{p}$ values whereas the DSS estimator can be extracted analytically without running any additional shots.

Flag-FT stabilizer measurements

The $|0\rangle_L$ state of the Steane code may as well be initialized by sequentially measuring all three X-type stabilizer generators. All three Z-type stabilizer eigenvalues are already fixed to $+1$ by initializing all data qubits to the $|0\rangle$ state. The state initialization can be rendered FT if the stabilizer measurements are FT, which can be established with the help of flag circuits as introduced in [CB18] and recently demonstrated experimentally [RA$^+$21; Hil$^+$22]. By extending the Steane look up table 2.3 by additional weight-2 correction operators that are applied if flags are triggered, FT-1 can be achieved. For instance, the flag circuit that fault-tolerantly reads out the stabilizer K_1^X is depicted in Fig. 3.20. If the measurement of K_1^X yields -1 and the other two X-type stabilizers yield $+1$ measurement outcomes, we would naively apply Z_4 as a correction by the Steane look up table 2.3. However, X-errors of weight-2 can propagate from bad locations of the flag circuit, as illustrated in Fig. 3.20, though not without triggering the flag. Note that only X-faults can propagate to the data qubits through the subsequent CNOT gates. Those faults are accounted for by repeating

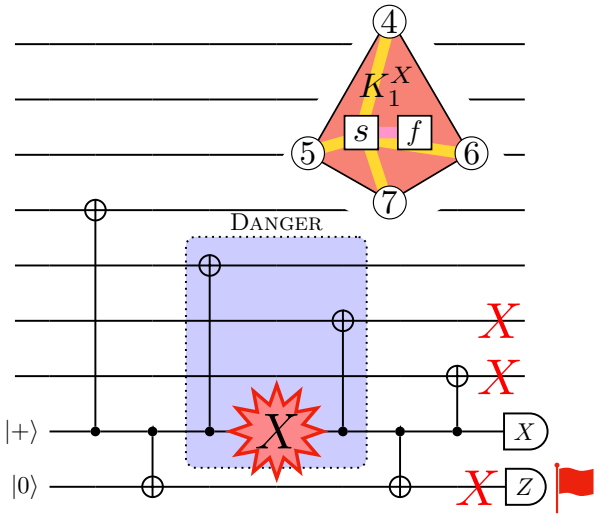


FIGURE 3.20: FT readout of the stabilizer plaquette $K_1^X = X_4X_5X_6X_7$ of the Steane code can be performed with a measurement auxiliary qubit and a flag qubit (measured in the X- and Z-basis respectively, see Fig. 2.6). The two highlighted middle CNOT gates are bad locations and can cause dangerous faults. An example X-fault is marked, which propagates to trigger the flag for causing a weight-2 Pauli error on the data qubits. In the subsequent non-flagged stabilizer readout this weight-2 error results in the syndrome $+ - +$ and, according to case 3(a) of the protocol, the correction X_6X_7 is (correctly) applied. The syndrome readout qubit s, which is measured in the X-basis, is connected to the four data qubits (yellow) and the flag qubit f, measured in the Z-basis, is only connected to s (pink).

stabilizer measurements until one of the following three cases is fulfilled:

1. The same syndrome is measured with all flags being clear in two consecutive rounds a and b. Then, apply the Z-correction according to the Steane look up table 2.3.

2. Two different syndromes are measured but all flags are clear in two consecutive rounds a and b. Then, run a third round of non-flagged syndrome readout and apply the Z-correction using that last syndrome according to the Steane look up table 2.3.

3. Any flag is triggered. Then, abort the flag circuit sequence and directly continue with a round of non-flagged syndrome readout:

   (a) If the measured syndrome corresponds to an X-error that can be caused by a fault that triggers the flag, apply this X-error as a correction. (This type of error makes up the so-called *flag error set* [CB18].)

   (b) If no such X-error exists, apply the Z-correction according to the Steane look up table 2.3.

At most eight circuits can be run in sequence for this protocol: Six flag circuits, one noisy non-flagged syndrome readout circuit and one circuit that contains the noisy final measurements of all data qubits. This longest possible circuit sequence is shown, without the final data qubit measurements, in Fig. 3.21. Some instances of the noisy protocol can be much shorter, for example, when skipping ahead as in case 3 of the protocol specification. Branching events of the protocol, that are modeled in the event

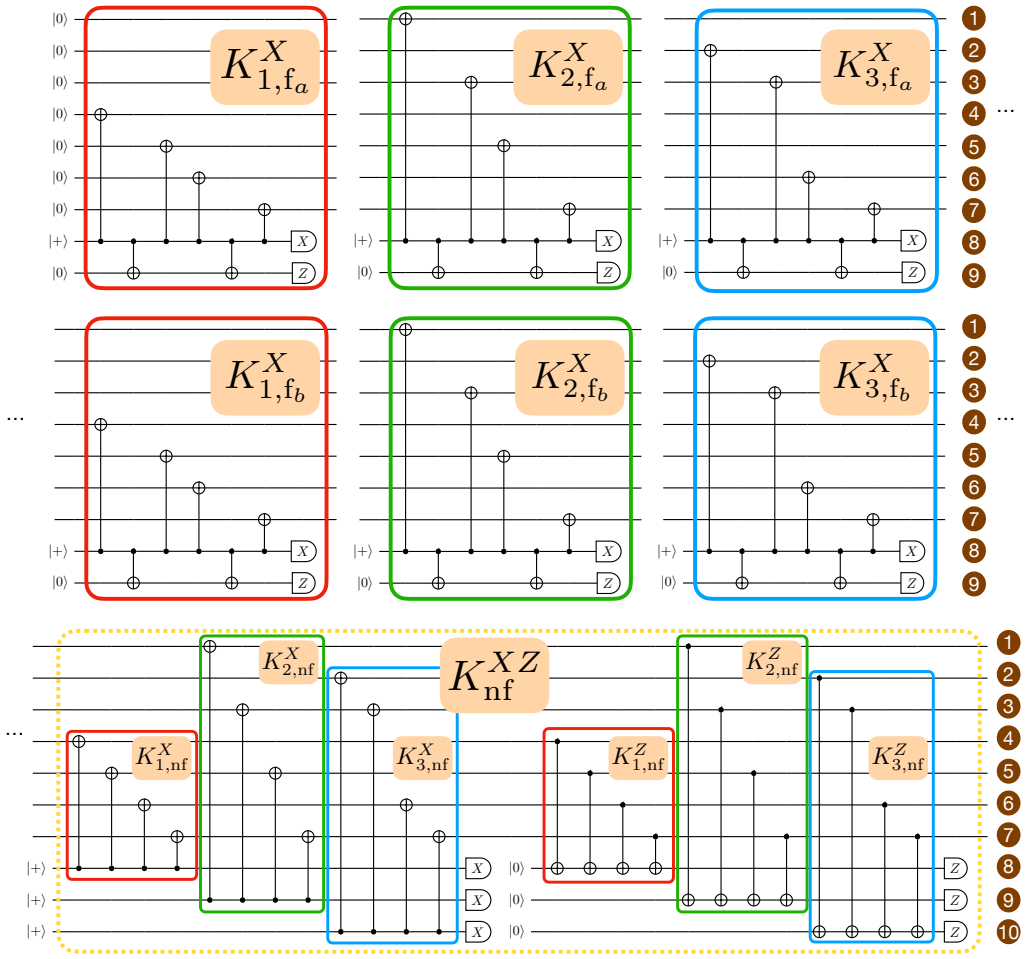


FIGURE 3.21: Longest possible circuit sequence that occurs when using the flag fault-tolerant protocol to encode the $|0\rangle_L$ state. The respective readout of the Steane code plaquette K_i^σ with a flag circuit ($.\mathrm{f}$) or a non-flag circuit ($.\mathrm{nf}$) is marked by RGB-colored boxes. If the flag qubit (qubit number 9, protocol case 3) is measured as -1 or if two consecutive flagged measurements of the same stabilizers $K^\sigma_{i,\mathrm{f}_a}$ and $K^\sigma_{i,\mathrm{f}_b}$ disagree (qubit number 8, protocol case 2), the circuit run is interrupted and continues at the yellow boxed circuit. The non-flagged syndrome measurement (dashed yellow box "K^{XZ}_{nf}") in the end is only omitted if all six previous stabilizer readouts agree and no flag has been triggered (protocol case 1).

tree, are determined by the measurement outcomes of the current flag circuit: Either we go to the next flag circuit or we jump ahead and run the non-flagged syndrome readout circuit ("K^{XZ}_{nf}"). The large number of branching possibilities renders DSS suitable to effectively explore the most relevant subsets. We again use the binomial factor choice criterion to preferentially sample low-weight subsets.

Failure rate estimations with DSS and direct MC are shown in Fig. 3.22 for the depolarizing noise model with single parameter[13] $p = p_1 = p_2$. Again, all estimated quantities scale $\hat{p} \sim p^2$ because of level-1 fault tolerance. Direct MC does not produce reliable estimations for the data points at low p for $N_{\mathrm{MC}} = 10^4$ shots. We use DSS at $p_{\mathrm{max}} = 10^{-3}$ in Fig. 3.22(a) with the same number of shots and obtain upper and lower bounds to the true protocol failure rate that enclose the direct MC estimators at larger

[13]The protocol contains of mostly two-qubit gates except the Hadamard gates that are performed for initialization to $|+\rangle$ and measurement in the X-basis.

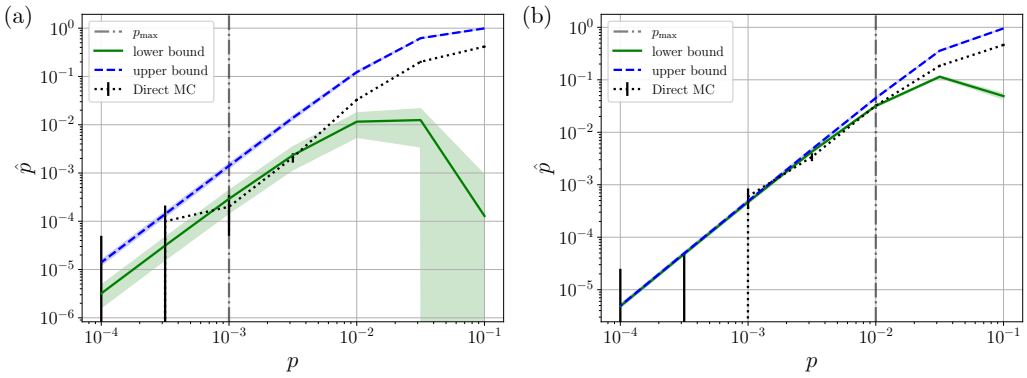


FIGURE 3.22: Logical failure rate estimations for the flag-FT stabilizer measurement state preparation protocol. For the limited number of $N_{\mathrm{MC}} = 10^4$ shots, the direct MC estimators are practically useless in the low-p regime ($p \lesssim 10^{-3}$) because the logical failure rate is too low. (a) A relatively large cutoff error remains at $p_{\max} = 10^{-3}$ (gray vertical dash-dotted line) but the total uncertainty on $\hat{p}$ stays bounded as $p \to 0$. (b) The bounds on the true failure rate can be improved by sampling at $p_{\max} = 10^{-2}$. The cutoff error vanishes in the limit $p \to 0$ because, here, all relevant subsets are explored by binomial factor choice.

p. The DSS bounds are not as tight as for the previous protocols. Some relevant paths with a relatively large product of binomial factors are not realized with binomial factor choice and are thus contained in the cutoff error δ. An additional effect is that our assumed maximum FT path length $L = 8$ may lead to an overly pessimistic estimation of the upper bound if the typical circuit sequence length at the given $p_{\max}$ is shorter. This issue could be solved by employing the ERU criterion instead and improving the estimation of the FT cutoff error. By sampling at larger $p_{\max} = 10^{-2}$, shown in Fig. 3.22(b), the cutoff error is improved upon without changing the DSS algorithm. The increase of $p_{\max}$ has the effect that more different fault-weight-subset are chosen by the binomial factor choice criterion. The bounds to the true protocol failure rate tighten when lowering p since the space of possible paths is more thoroughly explored.

3.5.2.3 Efficiency and run time considerations

We add a couple of general remarks on the efficiency and run time of the DSS algorithm.

As with single-circuit subset sampling, we consider DSS efficient when, by an appropriate subset choice criterion, it can be ensured that no large fraction of shots is spent on irrelevant subsets, for instance the fault-free path. These subsets should be large for a given $\vec{p}_{\max}$, in the sense analogous to Eq. (3.27), so that the protocol failure rate can be efficiently estimated with a narrow uncertainty interval from not too many non-trivial paths.

It may take a relatively long time to calculate total estimation uncertainties η in order to apply the ERU criterion for a large event tree compared to just running a single circuit as part of a protocol. We point out that these procedures can be parallelized. On the one hand, one may continue running further samples while, on the other hand, the expected reduction of uncertainty $\Delta(w_i)$ is evaluated numerically. A previously taken ERU choice or the binomial factor choice can be employed for these shots. Also, the evaluation for $\Delta(w_i)$ itself can be parallelized for all possible w-subsets that are probed and the assumed positive/negative outcome for the expected total uncertainty

can as well be calculated in parallel. In practice, it is important that the maximum subset weights $w_{i,\max}$ are kept small enough anyway.

As a direction of future research it would be interesting to study if the sampling efficiency can be improved by other subset choice criteria. We imagine, for example, that machine learning techniques could be used to choose fault weights in such a way that trivial paths are avoided but low-weight paths that can actually cause failure are encouraged. The application of dynamical subset sampling may also be extended to more complex situations, for instance, circuits that use multiple logical qubits such as FT teleportation protocols.

Chapter 4

Advances for fault-tolerant quantum computation

Practical progress for the actual experimental realization of fully-fledged error-corrected quantum information processing devices with large numbers of qubits consists of several milestones regarding fault-tolerant encoding, operation and interconnection of logical qubits. In this chapter we contribute theoretically-motivated avenues for the experimental advancement toward universal fault-tolerant quantum computation. In Sec. 4.1, we particularly focus on latest achievements with ion trap devices. Numerical simulations of noisy FT circuits[1], using the techniques of Chapters 2 and 3, are shown to accurately predict experimental findings and enable an outlook toward future experiments. In Sec. 4.2, we devise a novel scheme to perform a fault-tolerant quantum error correction cycle without measurements of physical qubits. The scheme is especially useful in near-term devices that cannot reliably execute physical in-sequence measurements fast enough or even at all, which we demonstrate via numerical simulations. A special focus is put on neutral-atom and ion trap devices when discussing practical implementations of the scheme.

4.1 Contributions for experimental progress

After describing the specific ion trap setup under consideration, we introduce the noise models employed in this study. Then follows a key step on the long road toward error-corrected universal quantum computation: Accompanying the first experimental realization of a fault-tolerant universal gate set [Pos$^+$22], we demonstrate that a simple but experimentally-informed depolarizing noise model accurately describes the experimental data. We simulate logical state preparation, logical Clifford operations, a logical CNOT gate and a logical T-gate implementation using magic state injection on the logical level. Lastly, we analyze how to reach the break-even point for advantageous use of FT state preparations, discuss aspects of scaling up to multiple logical qubits and provide additional insights on the influence of crosstalk in entangling gates.

4.1.1 Trapped-ion quantum computation

Ion traps are among the leading platforms for FT quantum information processors [Lei$^+$03; HRB08; Oze11; Bru$^+$19]. Types of ion traps that are in use today include Penning traps, which confine ions with the help of static electric and magnetic fields,

[1]Numerical simulations in this chapter are performed using a modified version of the Python package `PECOS` [RA18; RA19]. Uncertainties on sampled quantities are given as 95% confidence intervals in this chapter unless indicated otherwise.

as well as Paul traps, which operate with time-dependent electric fields [Gho95]. Both types have been miniaturized into planar variants [Sta$^+$05; Chi$^+$05]. In a linear trap, the trapped ions, interacting via the Coulomb force, form a one-dimensional ion crystal [Win$^+$98].

One typically refers to optical, Zeeman or fine-structure or hyperfine qubits depending on the electronic states of the ions, which are used to represent the computational basis states $|0\rangle$ and $|1\rangle$. Single-qubit gate operations are implemented by directly driving the transition between the two computational basis states, for instance with a laser of suitable wavelength, or via Raman transitions that require two appropriate laser beams and an auxiliary state [Oze11; Bal17]. While lasers that operate in the optical range are used to drive optical qubits, microwave gates can be used for the smaller energy separations of the basis states of other qubit types. Microwave-based architectures may be preferential for guiding light to the ions on an integrated chip via on-chip waveguides, although laser light can be focused more tightly [MW01; Lek$^+$17].

After physically loading ions into the trap by ionization of a neutral atom gas, initialization of the qubits into the $|0\rangle$ state is typically done by cooling the trapped ion and optically pumping into the long-lived electronic ground state [Roo00; Esc$^+$03]. To measure the physical qubit state, one can resonantly drive a transition from one of the basis states to a short-lived auxiliary state [Sch$^+$13b]. Photons emitted due to state-dependent fluorescence can be detected by an appropriate camera so that the two basis states can be discriminated by the photon detection. Scattered photons can heat up the ion crystal in the process. To continue coherent operation of the quantum processor after the measurement of single qubits, lengthy re-cooling procedures would be required and efficient techniques for mid-circuit measurements and in-sequence feedback operations are not yet available in the system described in [Pog$^+$21], which is in focus for this work.

Modern schemes of entangling gates in ion traps rely on geometric phases that are imprinted upon a two-qubit state while closed trajectories in phase space are performed [Bal17]. Although the ion-ion interaction is mediated by motional modes, the ion's internal and motional states are separable before and after performing the gate (assuming the gate is performed ideally). A commonly used entangling gate in ion trap platforms is the Mølmer-Sørensen (MS) gate [MS99; SM00] described by the q-qubit unitary operator

$$\mathrm{MS}(\theta, \vec{\varphi}, q) = \exp\left(-\mathrm{i}\theta S_{\vec{\varphi}}^2\right) \tag{4.1}$$

where

$$S_{\vec{\varphi}} = \frac{1}{2}\sum_{j=1}^{q}(X_j \cos(\varphi_j) + Y_j \sin(\varphi_j)) \tag{4.2}$$

is the rotated collective spin operator in the equatorial plane of the Bloch sphere for each physical qubit. One big advantage of the MS gate is that it can be used without cooling ions to the motional ground state. Bichromatic laser beams that are slightly detuned from the nearest sideband transition can be used to realize this gate (more detail will be given in Sec. 4.1.1.3).

Several linear traps may be coupled together in a modular fashion to enable scaling up to large numbers of physical qubits. Within the quantum charge-coupled device (QCCD) architecture, individual ions are physically shuttled in a two-dimensional

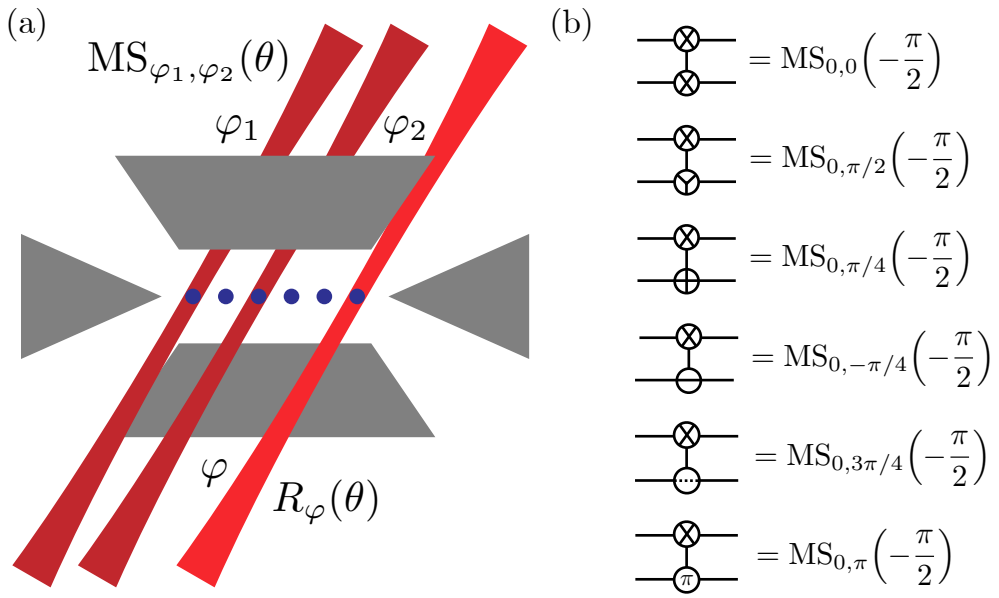


FIGURE 4.1: (a) In the considered ion trap architecture, the ions (blue dots) are suspended in a macroscopic linear Paul trap. Single-qubit rotations are facilitated by tightly focused laser beams (lighter-red). Entangling gates can be performed on arbitrary ion pairs (darker-red beams). The phase φ of any laser beam can be adjusted arbitrarily. The rotation angle θ is determined by the beam intensity and pulse duration. (b) Circuit symbols for two-qubit MS gates with six different phases and fixed angle $\theta = -\pi/2$ according to Eq. (4.4).

setup between dedicated storage and interaction zones, which each host a small one-dimensional ion crystal [KMW02; Pin$^+$21]. Alternatively, photonic links can be used to transfer individual qubit states between modules [Mon$^+$14; Drm$^+$23]. Additional aspects and more recent developments of scaling up are discussed in Chapter 5.

4.1.1.1 Physical architecture and native gate operations

The experimental system that was used to perform the experiments related to the content of this section is sketched in Fig. 4.1(a). It hosts a static ion string of up to 16 trapped $^{40}\mathrm{Ca}^+$ ions with all-to-all connectivity[2] and is described in detail in [Pog$^+$21; Heu$^+$23a]. The qubit is encoded in the Zeeman sublevels $4S_{1/2,m_j=-1/2} = |0\rangle$ and $3D_{5/2,m_j=-1/2} = |1\rangle$ of the ground state and a metastable excited state with a lifetime of $T_1 \approx 1.2\,\mathrm{s}$ [Bar$^+$00]. Single-qubit Pauli gates of X- and Y-type can be performed using laser pulses resonant with the qubit transition. The laser phase φ determines the rotation axis in the rotation operator

$$R_\varphi^{(i)}(\theta) = \exp\left(-\mathrm{i}\frac{\theta}{2}\sigma_\varphi^{(i)}\right) \quad \text{with} \quad \sigma_\varphi^{(i)} = X_i\cos(\varphi_i) + Y_i\sin(\varphi_i) \tag{4.3}$$

that targets the i-th qubit. The rotation angle θ is determined by the duration and intensity of the laser pulse. The setup allows one to implement two-qubit maximally-entangling MS gates with arbitrary phases $\vec{\varphi} = (\varphi_1, \varphi_2)$ that are described by the

[2]The larger the number of ions in the one-dimensional crystal, the closer the growing number of vibrational modes of the ion string become in terms of frequency. Undesired couplings to vibrational modes due to mode crowding need to be prevented with suitable counter-strategies in order to operate two-qubit MS gates in a long ion chain (see [LB18; Lan$^+$19] for instance).

unitary

$$\mathrm{MS}_{\varphi_1,\varphi_2}(\theta) = \exp\left(-\mathrm{i}\theta S^2_{\varphi_1,\varphi_2}\right) \tag{4.4}$$

with the spin operator

$$S_{\varphi_1,\varphi_2} = \frac{1}{2}(X_1\cos\varphi_1 + X_2\cos\varphi_2 + Y_1\sin\varphi_1 + Y_2\sin\varphi_2) \tag{4.5}$$

and with a fixed rotation angle $\theta = -\pi/2$. Circuit symbols for some $\mathrm{MS}_{\varphi_1,\varphi_2}(-\pi/2)$ gates are shown in Fig. 4.1(b). The two-qubit unitary $\mathrm{MS} \equiv \mathrm{MS}_{0,0}(-\pi/2)$ is equivalent to a CNOT gate up to local rotations by the identity

$$\mathrm{C}_i\mathrm{NOT}_j = R_Y^{(i)}(-\pi/2)R_X^{(i)}(\pi/2)R_X^{(j)}(\pi/2)\,\mathrm{MS}_{0,0}^{(ij)}(-\pi/2)\,R_Y^{(i)}(\pi/2) \tag{4.6}$$

taken from [Mas17], which we frequently use to effectively execute the CNOT operation.

Single-qubit rotations about the Z-axis need not be implemented physically but are tracked in software. The phases of operations subsequent to a Z-rotation of angle α are shifted according to the identities

$$\mathrm{MS}_{\varphi_1,\varphi_2}(-\pi/2)R_Z^{(1)}(\alpha) = R_Z^{(1)}(\alpha)\mathrm{MS}_{\varphi_1-\alpha,\varphi_2}(-\pi/2) \tag{4.7}$$

$$R_\varphi(\theta)R_Z(\alpha) = R_Z(\alpha)R_{\varphi-\alpha}(\theta). \tag{4.8}$$

All gate operations are executed sequentially and qubit initializations and measurements are executed in parallel for the simulations performed and experiments considered in this work.

4.1.1.2 Sources of noise in the particular ion trap setup

We now describe the dominant noise processes of the aforementioned hardware architecture. The dephasing time of single qubits can be estimated as $T_2 \approx 100\,\mathrm{ms}$, which is much shorter than the (longitudinal) relaxation time T_1 [Pog$^+$21]. While this relaxation leads to amplitude damping and leakage out of the computational subspace, *dephasing* is the predominant process affecting idling qubits due to fluctuations in the laser frequency or magnetic field [Heu$^+$23a]. When amplitude damping is negligible for the idling times under consideration $t \ll T_1$, we assume that the time-dependent density matrix in Eq. (3.11) for pure dephasing is an appropriate approximation of the single-qubit idling dynamics. The dephasing channel in Eq. (3.12) can thus be used to describe uncorrelated idling noise with rate

$$p_{\mathrm{idle}} = \frac{1}{2}\left(1 - \exp\left(-\frac{t}{T_2}\right)\right) \tag{4.9}$$

on locations where single qubits are idling during the execution of an operation on other qubits, which takes a time t. In previous investigations, effects of correlated dephasing were also reported [RM15; Pos$^+$18; Pal$^+$22].

For single- and two-qubit gates, laser fluctuations or miscalibrations can lead to undesired *overrotations* of a small angle ξ on top of the correct angle θ. These overrotations have been identified previously as a dominant source of error in ion traps [Deb$^+$20; Zha21]. For systematic miscalibrations, each intended rotation angle is shifted by a small amount $\theta \to \theta + \xi$ (see Eq. (3.4)). Also, fluctuations with finite variance ξ^2

around a mean value θ can occur. For instance, the coherent channel corresponding to systematic, coherent overrotations on a two-qubit state ρ reads

$$\mathcal{E}^{(2)}_{\text{cor}}(\rho) = \exp\left(-\mathrm{i}\frac{\xi}{2}X_1X_2\right)\rho\,\exp\left(\mathrm{i}\frac{\xi}{2}X_1X_2\right) \tag{4.10}$$

for X-type MS gates, i.e. $\vec{\varphi} = \vec{0}$. The physical fault rates that we assume for single- and two-qubit gates are experimentally informed by performing single-qubit randomized benchmarking and from the fidelity of a 16-qubit GHZ state respectively, as described in [Heu$^+$23a]. The experimental single-qubit gate fault rate is $p_1 = 5 \times 10^{-3}$ and the two-qubit gate fault rate is $p_2 = 2.5 \times 10^{-2}$ for all simulations presented in this section.

Physical qubit initializations to $|1\rangle$ instead of $|0\rangle$ happen with probability $p_i = 3\times10^{-3}$ and the result of a single qubit measurement is flipped at a rate $p_m = 3 \times 10^{-3}$ (see [Heu$^+$23a] and Refs. therein).

When laser light not only shines on targeted ions but also on their nearest-neighbor ions due to leakage light from the focused laser beam, *crosstalk* occurs [Sar$^+$20; Deb$^+$20; PR$^+$21]. The Rabi frequency Ω of the targeted ion relates the rotation angle $\theta = \Omega t$ to the time t the laser light couples to the ion [GK05]. The ratio $\varepsilon = \Omega_n/\Omega$ of the Rabi frequencies at a neighboring ion (Ω_n) and the targeted ion describes the strength of the crosstalk, which averages to a mean value of $\varepsilon_{\text{mean}} = 1\%$ over all ions in the trap. While the crosstalk amplitude for MS gates takes a maximum value of $\varepsilon_{\text{max}} = 1.6\%$, the crosstalk phase φ_c of the leakage light can essentially take on any value between 0 and 2π depending on the target ion. Experimental data in [Heu$^+$23a] shows that the value of this phase for a fixed target ion is constant over a long time period. Techniques to suppress crosstalk during single- and two-qubit gates by optimizing or applying additional pulses are known [VKL99; BHC04; Nig$^+$14; Mer$^+$14; Pil$^+$14; Fig$^+$19; Fan$^+$22; Xie$^+$22]. However, these techniques were not applied in the considered experiments. Crosstalk on single-qubit gates is small anyway and we show in Sec. 4.1.3 that the experiments are currently not limited by crosstalk on entangling gates to reach the break-even point of advantageous operation of FT circuits.

We now move on to describe different ways to model the above-mentioned noise in a way suitable for numerical simulations.

4.1.1.3 Noise modeling

In what follows, we describe our noise models for the description of the ion trap experiments partly published in [Pos$^+$22; Heu$^+$23a]. We first describe the relevance of the depolarizing noise channel. Subsequently a model of fluctuating overrotation noise is derived. Finally, we lay out a detailed derivation of incoherent phase-averaged crosstalk channels for single- and two-qubit gates.

Depolarizing noise

We mostly model noise on gate operations as depolarizing noise, using the channel in Eq. (3.13), for simplicity. Microscopic noise models that consider amplitude fluctuations or gate miscalibrations [Ber$^+$19] as well as thermal errors or motional heating [Bal17] and overrotations [Li$^+$17], have been devised in other works. However, our arguments for advantageous FT quantum computation primarily regard the appropriate FT circuit design. Recall that it can be assured that a circuit is FT toward

any type of computational noise on operations and idling locations if it is FT toward depolarizing noise.

Incoherent overrotations

In Sec. 4.1.2.1, we compare the accuracy of the depolarizing noise channel to overrotations. For this purpose, note that the incoherent approximation of the coherent overrotation channel in Eq. (4.10) can be obtained by Pauli twirling. The incoherent noise channel that we employ for simulations of X-overrotations in MS gates is

$$\mathcal{E}_{\text{ior}}^{(2)}(\rho) = (1-p_2)\rho + p_2 X_1 X_2 \rho X_1 X_2 \tag{4.11}$$

with $p_2 = \sin^2 \xi/2$.

Let us inspect fluctuations on the rotation angle of an X-type MS gate (uppermost gate in Fig. 4.1(b)) in more detail. The unitary $\text{MS}_{0,0}(\theta + x) = \text{MS}_{0,0}(x)\text{MS}_{0,0}(\theta)$ factorizes into an intended rotation about the angle θ and an unintended rotation about the angle x. We now consider fluctuations of the unintended rotation angle further. A rotation $\text{MS}_{0,0}(x)$ where x follows a Gaussian distribution with mean value 0 and standard deviation ξ takes the two-qubit state

$$\rho \to \int_{-\infty}^{\infty} \mathrm{d}x \exp\Big(-\mathrm{i}\frac{x}{2} X_1 X_2\Big) \rho \exp\Big(\mathrm{i}\frac{x}{2} X_1 X_2\Big) P(x,\xi) \tag{4.12}$$

with probability distribution

$$P(x,\xi) = \frac{1}{\sqrt{2\pi}\xi} \exp\left(-\frac{x^2}{2\xi^2}\right). \tag{4.13}$$

Pauli twirling the channel commutes with the integral so that we can write the channel as in Eq. (4.11) with

$$p_2 = \int_{-\infty}^{\infty} \mathrm{d}x\, P(x,\xi) \sin^2\Big(\frac{x}{2}\Big). \tag{4.14}$$

This integral evaluates to

$$\frac{1}{\sqrt{2\pi}\xi} \int_{-\infty}^{\infty} \mathrm{d}x \exp\left(-\frac{x^2}{2\xi^2}\right) \sin^2\Big(\frac{x}{2}\Big) = \frac{1}{2}\left(1 - \exp\left(-\frac{\xi^2}{2}\right)\right) = \left(\frac{\xi}{2}\right)^2 + \mathcal{O}(\xi^4). \tag{4.15}$$

Concluding, if we assume small Gaussian angle fluctuations of width ξ we can identify $p_2 = (\xi/2)^2$ for the incoherent channel in Eq. (4.11).

Single-qubit gate crosstalk

For an ion that is the direct neighbor of a target ion, where a rotation $R_\varphi(\theta)$ shall be performed, residual laser light causes the *single-qubit crosstalk* rotation

$$R_{\varphi_c}(\varepsilon\theta) = \exp\left(-\mathrm{i}\frac{\varepsilon\theta}{2}\sigma_{\varphi_c}\right) \tag{4.16}$$

of angle $\varepsilon\theta$ and crosstalk phase φ_c. The corresponding channel on a single-qubit density matrix ρ reads

$$\begin{aligned}\mathcal{E}(\rho) &= R_{\varphi_c}(\varepsilon\theta)\,\rho\,R^\dagger_{\varphi_c}(\varepsilon\theta)\\ &= \cos^2\frac{\varepsilon\theta}{2}\,\rho + \sin^2\frac{\varepsilon\theta}{2}\,(\sigma_{\varphi_c}\,\rho\,\sigma_{\varphi_c}) + \frac{\mathrm{i}}{2}\sin\varepsilon\theta\,[\rho,\sigma_{\varphi_c}]\\ &= \cos^2\frac{\varepsilon\theta}{2}\,\rho + \sin^2\frac{\varepsilon\theta}{2}\big(\cos^2\varphi_c\,X\rho X + \sin^2\varphi_c\,Y\rho Y + \cos\varphi_c\sin\varphi_c(X\rho Y + Y\rho X)\big)\\ &\qquad + \frac{\mathrm{i}}{2}\sin\,\varepsilon\theta\,(\rho(X\cos\varphi_c + Y\sin\varphi_c) - (X\cos\varphi_c + Y\sin\varphi_c)\rho). \quad (4.17)\end{aligned}$$

Pauli twirling this channel yields the incoherent approximation

$$\begin{aligned}\tilde{\mathcal{E}}(\rho) &= \cos^2\frac{\varepsilon\theta}{2}\,\rho + \sin^2\frac{\varepsilon\theta}{2}\big(\cos^2\varphi_c\,X\rho X + \sin^2\varphi_c\,Y\rho Y\big) &(4.18)\\ &\equiv (1-p_{\mathrm{c}_1})\rho + p_{\mathrm{c}_1}(r_x X\rho X + (1-r_x)Y\rho Y) &(4.19)\end{aligned}$$

where we define the physical fault rate as

$$p_{\mathrm{c}_1} = \sin^2\left(\frac{\varepsilon\theta}{2}\right) \quad (4.20)$$

and introduce the *phase ratios* $r_x = \cos^2\varphi_c$ and $r_y = 1 - r_x = \sin^2\varphi_c$. All terms in the commutator and the off-diagonal terms in the $\sin^2$-term cancel. This directly corresponds to taking only the diagonal terms of the process matrix χ parametrizing the coherent channel of Eq. (4.17) in the Pauli basis

$$\chi = \frac{1}{2}\begin{pmatrix} 2\cos^2(\varepsilon\theta/2) & \mathrm{i}\sin\varepsilon\theta\cos\varphi_c & \mathrm{i}\sin\varepsilon\theta\sin\varphi_c & 0\\ -\mathrm{i}\sin\varepsilon\theta\cos\varphi_c & 2\sin^2(\varepsilon\theta/2)\cos^2\varphi_c & 2\sin^2(\varepsilon\theta/2)\cos\varphi_c\sin\varphi_c & 0\\ -\mathrm{i}\sin\varepsilon\theta\sin\varphi_c & 2\sin^2(\varepsilon\theta/2)\cos\varphi_c\sin\varphi_c & 2\sin^2(\varepsilon\theta/2)\sin^2\varphi_c & 0\\ 0 & 0 & 0 & 0\end{pmatrix}. \quad (4.21)$$

For crosstalk phases that are distributed across the whole interval of all possible values $\varphi_c \in [0, 2\pi)$, as observed experimentally in [Heu$^+$23a], we use the identities $\int_0^{2\pi}\mathrm{d}\varphi_c\cos^2\varphi_c = \int_0^{2\pi}\mathrm{d}\varphi_c\sin^2\varphi_c = \pi$ to average over the crosstalk phase φ_c in Eq. (4.18) and obtain the phase-averaged crosstalk channel

$$\langle\tilde{\mathcal{E}}\rangle_{\varphi_c}(\rho) = (1-p_{\mathrm{c}_1})\rho + \frac{p_{\mathrm{c}_1}}{2\pi}\int_0^{2\pi}\mathrm{d}\varphi_c\,\big(\cos^2\varphi_c\,X\rho X + \sin^2\varphi_c\,Y\rho Y\big). \quad (4.22)$$

From this, we define the incoherent noise channel for single-qubit crosstalk

$$\mathcal{E}_{\mathrm{c}_1}(\rho) = (1-p_{\mathrm{c}_1})\rho + \frac{p_{\mathrm{c}_1}}{2}(X\rho X + Y\rho Y), \quad (4.23)$$

which we use in our numerical simulations. On any single-qubit crosstalk location a fault operator is placed with probability p_{c_1} and we randomly choose this operator to be X or Y with equal chance for simplicity. Note that one could as well take different but fixed crosstalk phases for all individual ions according to the experimental data from [Heu$^+$23a]. If one had control over the crosstalk phase, for instance if the phases were the same on target and neighbor ions, an $X(Y)$-fault could be placed at crosstalk locations that belong to intended $X(Y)$-rotations described by Eq. (4.18) with $\varphi_c = 0\,(\pi/2)$. For this channel, the physical crosstalk fault rate $p_{\mathrm{c}_1} = p_{\mathrm{c}_1}(\theta)$ depends on the rotation angle of the gate. For example, rotation angles $\theta \in \{\pi, \pi/2, \pi/4\}$

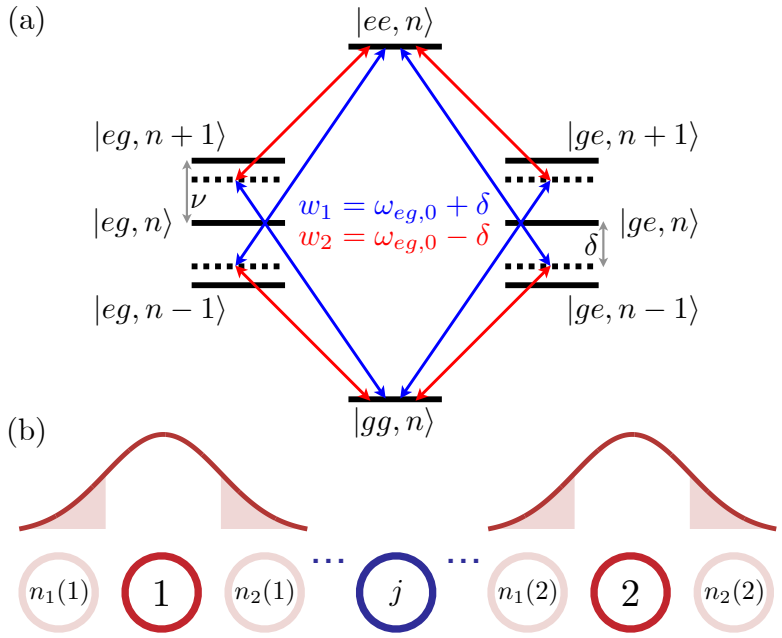


FIGURE 4.2: Multiple ions are illuminated by bichromatic laser light in order to perform the MS gate. (a) Level scheme for applying an MS gate to two ions via a shared motional mode. The states are labeled by the ion's internal electronic states g and e and the phonon number n of the vibrational mode of the chain. The laser frequencies $\omega_{1/2}$ are slightly detuned from the qubit frequency $\omega_{eg,0}$ by an amount δ. (b) The ion chain (circles) is partially illuminated by two laser beams with Gaussian beam profile. Not only the target ions 1 and 2 receive laser light but also their nearest neighbor ions $n_1(1), n_2(1), n_1(2)$ and $n_2(2)$ are illuminated unintentionally.

correspond to approximate probabilities $p_{c_1} \in \{2.5 \times 10^{-4}, 6.2 \times 10^{-5}, 1.5 \times 10^{-5}\}$ respectively according to Eq. (4.20) with $\varepsilon = 1\%$. For this incoherent channel, both neighbor ions n_1 and n_2 are independent single-qubit crosstalk locations.

Crosstalk on MS gates

We now derive our noise model for *crosstalk on MS gates* based on the Hamiltonian of the MS gate. The time-dependent MS gate Hamiltonian[3] reads

$$H(t) = H_0 + H_{\text{int}}(t) \quad \text{with} \tag{4.24}$$

$$H_0 = \sum_{j=1}^{q} \frac{\omega_{eg,0}}{2}\sigma_{z,j} + \nu\left(a^\dagger a + \frac{1}{2}\right) \quad \text{and} \tag{4.25}$$

$$H_{\text{int}}(t) = \sum_{j=1}^{q} \frac{\Omega_j(t)}{2}\left(e^{\mathrm{i}\left(\vec{k}_1\vec{x}_j - (\omega_{\text{eg},0}+\delta)t - \varphi_j\right)} + e^{\mathrm{i}\left(\vec{k}_2\vec{x}_j - (\omega_{\text{eg},0}-\delta)t - \varphi_j\right)} + \text{h.c.}\right)\left(\sigma_j^+ + \sigma_j^-\right) \tag{4.26}$$

where $\sigma_j^\pm = (X_j \pm \mathrm{i}Y_j)/2$ are the atomic raising and lowering operators and h.c. denotes Hermitian conjugation [MS99; SM00]. Here q is the number of ions that are illuminated by the lasers with wavevectors $\vec{k}_1$ and $\vec{k}_2$ and the Ω_j are their respective Rabi frequencies. The bosonic ladder operators a and $a^\dagger$ correspond to a single vibrational mode of the ion string with frequency ν and we assume that no other

[3] We work in units where $\hbar = 1$.

vibrational modes are excited by the laser beams. We express $\vec{k}_i\vec{x}_j = \eta_{ij}(a^\dagger + a)$ via the Lamb-Dicke parameters η_{ij}, which quantify the ratio of the ion's excursions to the laser wavelengths. We then operate in a regime where the laser detuning $\delta \ll \omega_{eg,0}$ is much smaller than the qubit frequency. Then the Lamb-Dicke parameters η_{ij} all approximately take on the same value η for all ions and we operate in the Lamb-Dicke regime where $\eta \ll 1$ [GK05]. A level scheme that illustrates the MS gate for two ions is shown in Fig. 4.2(a). For a larger ion chain, the sum over the q ions in Eq. (4.26) can be split into the ions $t' \in \{1, 2\}$ that are targeted by a two-qubit MS gate and their nearest-neighbor ions $n' \in \{n_1(1), n_2(1), n_1(2), n_2(2)\}$, which also receive laser light unintentionally. This situation is depicted in Fig. 4.2(b). The Rabi frequencies of the target ions $\Omega_{t'}(t) = \Omega$ and the neighbor ions $\Omega_{n'} = \varepsilon\Omega$ are taken to be constant over time so we can write the Hamiltonian as

$$H_{\text{int}}(t) = -\eta\Omega\Big(ae^{-\mathrm{i}\epsilon t} + a^\dagger e^{\mathrm{i}\epsilon t}\Big)\left(\sum_{t'}\frac{1}{2}\sigma_{\varphi_{t'}} + \sum_{n'}\frac{\varepsilon}{2}\sigma_{\varphi_{n'}}\right) \tag{4.27}$$

where $\epsilon = \nu - \delta$ and the rotating wave approximation has been applied [GK05]. The final form of the Hamiltonian can now be expressed with the collective spin operator $S_{\vec{\varphi}}$ as

$$H_{\text{int}}(t) = -\eta\Omega\Big(ae^{-\mathrm{i}\epsilon t} + a^\dagger e^{\mathrm{i}\epsilon t}\Big)S_{\vec{\varphi}} \tag{4.28}$$

where $\vec{\varphi} = (\varphi_1, \varphi_2, \varphi_{n_1(1)}, \varphi_{n_2(1)}, \varphi_{n_1(2)}, \varphi_{n_2(2)})$ contains all target and neighbor ion phases. The unitary time evolution operator for this Hamiltonian follows as

$$U(t) = D(\Gamma(t)S_{\vec{\varphi}})\exp\big(\mathrm{i}\theta(t)S_{\vec{\varphi}}^2\big) \tag{4.29}$$

with $\Gamma(t) = \int_0^t \gamma(t')\mathrm{d}t'$ and $\theta(t) = \operatorname{Im}\int_0^t \gamma(t')\mathrm{d}t' \int_0^{t'} \gamma^*(t'')\mathrm{d}t''$ where $\gamma(t) = \mathrm{i}\eta\Omega e^{\mathrm{i}\epsilon t}$ and the displacement operator $D(\alpha) = \exp\big(\alpha a^\dagger - \alpha^* a\big) \sim 1 + \big(\alpha a^\dagger - \alpha^* a\big)$, which fulfills the relation $D(\alpha)D(\beta) = D(\alpha + \beta)\exp(\mathrm{i}\operatorname{Im}(\alpha\beta^*))$. Adjustment of the gate parameters $\Gamma(t)$ and $\theta(t)$ to experimentally realize the MS gate is described further in [MG+22].

We can split the collective spin operator $S_{\vec{\varphi}}$ into a term that describes the target ions and another term that describes the nearest-neighbor ions as

$$S_{\vec{\varphi}} = S_{\vec{\varphi}}\Big|_{\text{targets}} + S_{\vec{\varphi}}\Big|_{\text{neighbors}} \tag{4.30}$$

$$= \frac{1}{2}\Big(\sigma_{\varphi_1} + \sigma_{\varphi_2} + \varepsilon\Big(\sigma_{\varphi_{n_1(1)}} + \sigma_{\varphi_{n_2(1)}} + \sigma_{\varphi_{n_1(2)}} + \sigma_{\varphi_{n_2(2)}}\Big)\Big) \tag{4.31}$$

where the Rabi frequencies of the neighbor ions are suppressed by the crosstalk ratio ε. By taking the square of $S_{\vec{\varphi}}$ we – in linear order of ε – create all combinations of target and neighbor ion operators. This can as well be expressed as two terms

$$S_{\vec{\varphi}}^2 = S_{\vec{\varphi}}^2\Big|_{\text{intended}} + S_{\vec{\varphi}}^2\Big|_{\text{crosstalk}}. \tag{4.32}$$

For the MS gate, which transforms the state ρ as

$$\mathcal{E}(\rho) = \exp\big(-\mathrm{i}\theta S_{\vec{\varphi}}^2\big)\,\rho\,\exp\big(\mathrm{i}\theta S_{\vec{\varphi}}^2\big), \tag{4.33}$$

the intended term in Eq. (4.32) realizes the MS gate rotation on the target ions. The MS interaction originates from the square of the intended spin operator $S_{\varphi_1,\varphi_2} = \frac{1}{2}(\sigma_{\varphi_1} + \sigma_{\varphi_2})$ as

$$S_{\varphi_1,\varphi_2}^2 \overset{\simeq}{=} \frac{1}{2}\sigma_{\varphi_1}\sigma_{\varphi_2} \tag{4.34}$$

where $\tilde{=}$ indicates that we omit irrelevant terms: Since the Pauli operators anticommute on the same qubit, some terms sum to zero. Squares of Pauli operators yield the identity and thus only contribute an irrelevant global phase. For the case $\varphi_1 = \varphi_2 = 0$ we find the usual X-type MS gate.

The crosstalk term in Eq. (4.32) consists of all combinations of two single-qubit operators on a target and a neighbor ion in order ε. These crosstalk locations are shown for an MS gate in Fig. 4.3. Higher orders of ε are neglected. Then each crosstalk location can be described independently by a coherent two-qubit rotation. A single location 1, $n_1(1)$ denoted by a target ion 1 and a neighbor ion $n_1(1)$ is contained in the squared spin operator as

$$S_{\vec{\varphi}}^2\Big|_{\text{crosstalk}} \supset \frac{\varepsilon}{2}\sigma_{\varphi_1}\sigma_{\varphi_{n_1(1)}}. \tag{4.35}$$

It generates the rotation

$$R_{1,n_1(1)} = \exp\left(-\mathrm{i}\frac{\varepsilon\theta}{2}\sigma_{\varphi_1}\sigma_{\varphi_{n_1(1)}}\right). \tag{4.36}$$

For any single crosstalk location t, n the coherent channel

$$\mathcal{E}(\rho) = R_{t,n}(\varepsilon\theta)\,\rho\,R_{t,n}^\dagger(\varepsilon\theta) = \exp\left(-\mathrm{i}\frac{\varepsilon\theta}{2}\sigma_{\varphi_t}\sigma_{\varphi_n}\right)\rho\,\exp\left(\mathrm{i}\frac{\varepsilon\theta}{2}\sigma_{\varphi_t}\sigma_{\varphi_n}\right) \tag{4.37}$$

describes the coherent evolution of the state ρ analogously to Eq. (4.17). Pauli twirling is performed with the 16 two-qubit Pauli operators $P_2 \in \{I, X, Y, Z\} \otimes \{I, X, Y, Z\}$ on this coherent channel analogously to single-qubit crosstalk. The twirled, incoherent channel for one MS gate crosstalk location reads

$$\begin{aligned}\tilde{\mathcal{E}}(\rho) = \cos^2\frac{\varepsilon\theta}{2}\,\rho + \sin^2\frac{\varepsilon\theta}{2}\big(&\cos^2\varphi_t\cos^2\varphi_n\,X_tX_n\rho X_tX_n + \sin^2\varphi_t\sin^2\varphi_n\,Y_tY_n\rho Y_tY_n\\ &+ \cos^2\varphi_t\sin^2\varphi_n\,X_tY_n\rho X_tY_n + \sin^2\varphi_t\cos^2\varphi_n\,Y_tX_n\rho Y_tX_n\big).\end{aligned} \tag{4.38}$$

We define the incoherent noise channel, identifying the probability $p_{c_2} = \sin^2(\varepsilon\theta/2)$, as

$$\begin{aligned}\tilde{\mathcal{E}}(\rho) = (1-p_{c_2})\rho + p_{c_2}(&r_{xx}X_tX_n\rho X_tX_n + r_{xy}X_tY_n\rho X_tY_n\\ &+ r_{yx}Y_tX_n\rho Y_tX_n + r_{yy}Y_tY_n\rho Y_tY_n)\end{aligned} \tag{4.39}$$

with phase ratios

$$\begin{aligned} r_{xx} &= \cos^2\varphi_t\cos^2\varphi_n,\ r_{xy} = \cos^2\varphi_t\sin^2\varphi_n\\ r_{yx} &= \sin^2\varphi_t\cos^2\varphi_n,\ r_{yy} = \sin^2\varphi_t\sin^2\varphi_n.\end{aligned} \tag{4.40}$$

In the same way as in Eq. (4.22), we average over phases of neighbor ions φ_n to obtain the incoherent noise channel

$$\begin{aligned}\langle\tilde{\mathcal{E}}\rangle_{\varphi_n}(\rho) = (1-p_{c_2})\rho + \frac{p_{c_2}}{2}\big(&\cos^2\varphi_t\,(X_tX_n\,\rho\,X_tX_n + X_tY_n\,\rho\,X_tY_n)\\ &+ \sin^2\varphi_t\,(Y_tX_n\,\rho\,Y_tX_n + Y_tY_n\,\rho\,Y_tY_n)\big).\end{aligned} \tag{4.41}$$

In order to formulate a simple noise model, which does not need to take into account the microscopic nature of the gate, as for the depolarizing channel, we also average

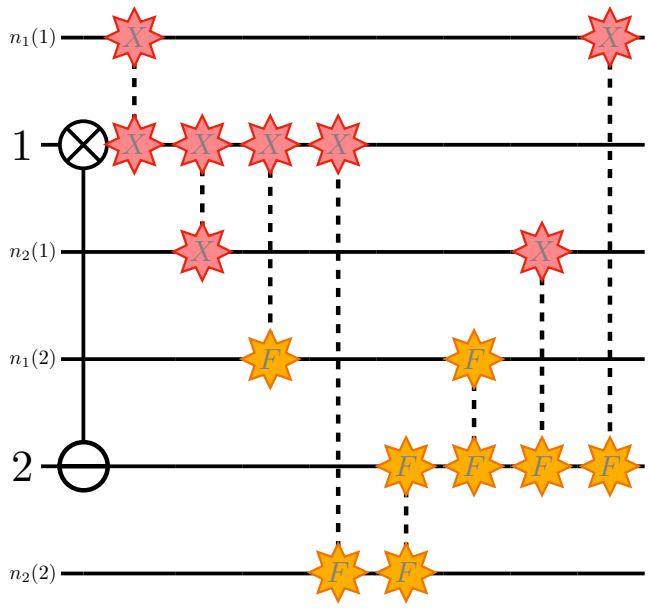


FIGURE 4.3: The gate $\mathrm{MS}_{0,-\pi/4}(-\pi/2)$ is shown (black, circles) followed by all possible crosstalk faults (stars with vertical dashed line) as an example of a crosstalk location. A Pauli-X fault is applied for the target ion 1 with $\varphi_1 = 0$ and its nearest neighbors and the operator $F = \frac{X-Y}{\sqrt{2}}$ is applied for the target ion 2 with $\varphi_2 = -\pi/4$ and its nearest neighbors, assuming that the respective crosstalk phase is the same as the associated target ion phase.

over the target ion phases φ_t. Our final result is the channel for phase-averaged MS gate crosstalk

$$\begin{aligned}\mathcal{E}_{\mathrm{c}_2}(\rho) = (1-p_{\mathrm{c}_2})\rho + \frac{p_{\mathrm{c}_2}}{4}(&X_tX_n\,\rho\,X_tX_n + X_tY_n\,\rho\,X_tY_n \\ &+ Y_tX_n\,\rho\,Y_tX_n + Y_tY_n\,\rho\,Y_tY_n),\end{aligned} \tag{4.42}$$

which we use in our numerical simulations.

Let us now inspect *special cases* that can appear for some configurations of MS gates. To begin, we examine the case where the two target ions share a common neighbor, i.e. that $n_2(1) = n_1(2)$. This common neighbor ion receives residual laser light from both target ions. If the laser phase on target and neighbor ions are the same, $\varphi_{n_2(1)} = \varphi_{n_1(2)}$, the spin operator in Eq. (4.31) now reads

$$S_{\vec{\varphi}} = \frac{1}{2}\Big(\sigma_{\varphi_1} + \sigma_{\varphi_2} + \varepsilon\Big(\sigma_{\varphi_{n_1(1)}} + 2\sigma_{\varphi_{n_2(1)}} + \sigma_{\varphi_{n_2(2)}}\Big)\Big) \tag{4.43}$$

so the angle of the coherent rotation is doubled $\varepsilon\theta \to 2\varepsilon\theta$ on the common neighbor. For the incoherent model this means that the fault probability shifts like $p_{\mathrm{c}_2} \to 4p_{\mathrm{c}_2}$ since $\sin^2(\varepsilon\theta) = 4\sin^2(\varepsilon\theta/2)\cos^2(\varepsilon\theta/2)$. We furthermore consider crosstalk where we assume that the phases of all MS gates and neighbor ions can be set to $\vec{\varphi} = \vec{0}$. The shift in angle must be performed on common-neighbor locations in this case. The unitary evolution operators from the coherent channel

$$\mathcal{E}_{\mathrm{cct}}(\rho) = \exp\Big(-\mathrm{i}\frac{\varepsilon}{2}\theta X_tX_n\Big)\,\rho\,\exp\Big(\mathrm{i}\frac{\varepsilon}{2}\theta X_tX_n\Big) \tag{4.44}$$

(see Eq. (4.36)) translate via the PTA to the incoherent channel

$$\mathcal{E}_{\mathrm{xct}}(\rho) = (1-p_{\mathrm{c}_2})\rho + p_{\mathrm{c}_2}X_tX_n\,\rho\,X_tX_n \tag{4.45}$$

with $p_{c_2}(\theta) = \sin^2 \varepsilon\theta/2$ on non-common neighbor locations ("X-type crosstalk"). For locations with common neighbor ions, the shifts $\varepsilon\theta \to 2\varepsilon\theta$ and $p_{c_2} \to 4p_{c_2}$ are taken into account in numerical simulations respectively.

As *another special case*, let us, sticking to the common neighbor case $n_2(1) = n_1(2)$, suppose that the neighbor ion phases were the same as their associated target ion's phase $\varphi_1 = \varphi_{n_1(1)} = \varphi_{n_2(1)}$ and $\varphi_2 = \varphi_{n_1(2)} = \varphi_{n_2(2)}$ but the target ion phases $\varphi_1 \neq \varphi_2$ are not the same. As an example, we take $\varphi_1 = 0$ and $\varphi_2 = -\pi/4$ as for the gate in Fig. 4.3 where the operator $F \equiv \frac{X-Y}{\sqrt{2}}$ is defined. Now, we find all operator combinations that contribute to crosstalk from

$$S_{0,-\pi/4} = \frac{1}{2}\big(X_1 + F_2 + \varepsilon\big(X_{n_1(1)} + X_{n_2(1)} + F_{n_1(2)} + F_{n_2(2)}\big)\big) \tag{4.46}$$

for the squared spin operator

$$\begin{aligned} S^2_{0,-\pi/4} \supset \frac{1}{4}\big(2\varepsilon\big(&X_1X_{n_1(1)} + X_1X_{n_2(1)} + X_1F_{n_1(2)} + X_1F_{n_2(2)} \\ &+ X_{n_1(1)}F_2 + X_{n_2(1)}F_2 + F_{n_1(2)}F_2 + F_{n_2(2)}F_2\big)\big). \end{aligned} \tag{4.47}$$

Note that four terms $X_1X_{n_2(1)}, X_1F_{n_1(2)}, X_{n_2(1)}F_2$ and $F_{n_1(2)}F_2$ occur in Eq. (4.47) so there is no angle doubling on the common neighbor qubit $n_2(1) = n_1(2)$ since $\varphi_1 \neq \varphi_2$ but all four terms contribute equally either as a coherent rotation or an independent fault operator in the incoherent channel.

As a final remark, if the phases of target ions can be adjusted as $\varphi_1 = 0$ and $\varphi_2 = \pi$ for X-type MS gates or as $\varphi_1 = \pi/2$ and $\varphi_2 = 3\pi/2$ for Y-type MS gates, the effect of the crosstalk on a common neighbor ion can be canceled completely with the above assumption since $\sigma_0 = -\sigma_\pi$ and $\sigma_{\pi/2} = -\sigma_{3\pi/2}$.

To conclude our discussion of crosstalk on entangling gates, let us point out the deteriorating effect of crosstalk on FT circuit designs. Since this type of crosstalk, as sketched in Fig. 4.3, induces correlated faults of Pauli weight 2 between an MS gate location and neighboring qubits, it must in general be expected that fault tolerance is broken by the crosstalk noise channels described above. As a consequence, logical failure rates, even if the respective circuits are otherwise FT, contain a linear scaling term $p_L \sim cp_{c_2}$ as $p_{c_2} \to 0$ where the proportionality constant c is determined by the number of bad crosstalk locations.

4.1.2 Fault-tolerant universal gate set realization

Using the ion trap architecture from [Pog+21], the first demonstration of a fault-tolerant universal gate set on the logical level was achieved experimentally in a recent work [Pos+22]. We contributed to this milestone – the realization of the gate set $\{\mathrm{CNOT}, H, T\}$ – by numerically performing stabilizer and statevector simulations that accurately predict the experimentally measured logical fidelities. The crucial finding is that the FT magic state preparation, needed to realize the non-Clifford gate, achieves a superior performance than its non-FT counterpart although built from a larger number of noisy components. The physical fault rates of the trapped-ion components given in Sec. 4.1.1.2 are listed as depolarizing noise model parameters in Tab. 4.1. The experimentally-informed depolarizing noise model that we use is sufficient in order to model the considered experiment without requiring additional microscopic modeling of physical noise processes.

In the following, we describe the FT universal gate set realization in three steps. First, the FT logical qubit initialization in the Steane code and FT single-qubit Clifford gates

Operation	Physical fault rate	Numerical value
Single-qubit gates	p_1	5×10^{-3}
Two-qubit gates	p_2	2.5×10^{-2}
Initializations	p_i	4.5×10^{-3}
Measurements	p_m	4.5×10^{-3}

TABLE 4.1: Experimentally-informed fault rates for the depolarizing noise model based on Eq. (3.13) used in our numerical simulations.

are performed. Here we also compare our experimentally-informed depolarizing noise model to a more microscopically-motivated overrotation noise model and find that microscopic modeling does not improve the simulation accuracy of the noisy logical qubit state. Then follows the demonstration of an FT logical CNOT gate. As a last step, logical magic state preparation and logical gate teleportation is performed fault-tolerantly to practically realize the FT T-gate. As a minor difference to Chapter 2, in the remainder of this section, we refer to the T-gate as a $\pi/4$-rotation in the Y-basis instead of the Z-basis, which only differ by a Clifford operation. Recall that Clifford operations are transversal in the Steane code and thus straightforward to implement fault-tolerantly. Meanwhile, FT magic state preparation has also been performed in a superconducting transmon device [Gup+23] and an FT non-Clifford gate has been demonstrated using the $[[8, 3, 2]]$ code [MRV23].

4.1.2.1 Pauli state preparation and logical Clifford gates

All six logical Pauli states in the Steane code can be prepared using the circuit shown in Fig. 4.4(a), which was originally devised in [Got16a]. The first block uses eight CNOT gates to prepare the $|0\rangle_L$ state non-fault-tolerantly on seven physical qubits. The number of CNOT gates is minimized by preparing the three plaquettes of the Steane code in an interleaved way; CNOT gates 1, 4 and 7 initialize the green plaquette K_2 (see Fig. 2.12), CNOT gates 2, 6 and 8 are needed for preparation of the blue plaquette K_3 and CNOT gates 3, 5 and 8 are involved in preparing the red plaquette K_1 (counting left to right and top to bottom). An additional physical flag qubit verifies correct state preparation via three extra CNOT gates in the second block. The flag qubit is used to measure the logical operator $Z_L = Z_3Z_5Z_6 = K_1^Z K_2^Z K_3^Z Z^{\otimes 7}$, which yields $+1$ in the absence of faults. All dangerous faults that may occur in the first block are propagated to trigger the flag and an example of such an X-fault is illustrated in Fig. 4.4(a). The propagated error X_1X_3 will lead to failure of the Steane code because it causes the X-syndrome $++-$. Thus, the recovery X_2 would be applied by the look up table 2.3. The resulting operator $X_L = X_1X_2X_3$ flips the logical state $|0\rangle_L$ to $|1\rangle_L$ so that a failure is inflicted. Whenever the flag is triggered, the encoded state is discarded and a new preparation trial is made. When the flag is clear, we accept the state since it is guaranteed that no single fault in the encoding circuit can have caused an uncorrectable error in this case. As a result, $|0\rangle_L$ is prepared fault-tolerantly. Since all Clifford gates are transversal in the Steane code, we can subsequently apply bitwise single-qubit gates $R_X(\pi)$, $R_Y(\pi/2)$, $R_Y(-\pi/2)$, $R_X(-\pi/2)$ or $R_X(\pi/2)$ to rotate $|0\rangle_L$ to any of the five other logical Pauli states $|1\rangle_L$, $|+\rangle_L$, $|-\rangle_L$, $|+\mathrm{i}\rangle_L$ or $|-\mathrm{i}\rangle_L$ respectively while obeying fault tolerance.

For usage in the considered ion trap device, we compile the encoding circuit into MS gates and local rotations using Eq. (4.6) and the actual circuit that was used in the experiments is given in Fig. 4.5 alongside with propagation rules of Pauli faults

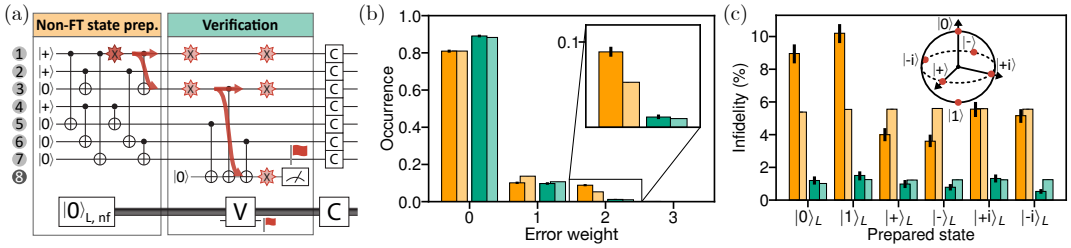


FIGURE 4.4: Logical Pauli state preparation simulation and experiment. (a) Fault-tolerant state preparation is performed in three steps: Non-FT preparation is followed by a verification step (V) that renders the overall procedure FT. Afterwards, the prepared $|0\rangle_L$ state can be transformed to any Pauli state fault-tolerantly by applying any logical Clifford gate (C), which is implemented transversally in the Steane code. A single fault X_1 (red star) propagates to an uncorrectable weight-2 error but also triggers the flag. (b) Percentage of output states $E|0\rangle_L$ that carry errors of different minimal Pauli weight wt(E) for non-FT (orange) and FT (green) state preparation. Lighter bars on the right show numerical simulation data and darker bars on the left are experimental data. (c) Logical fidelities for the six logical Pauli eigenstates, marked on the inset logical Bloch sphere, with one additional round of ideal in-software error correction performed after state preparation. All MC simulations are done with 10^6 samples. Error bars indicate the 68% confidence interval in the figure and in the main text.

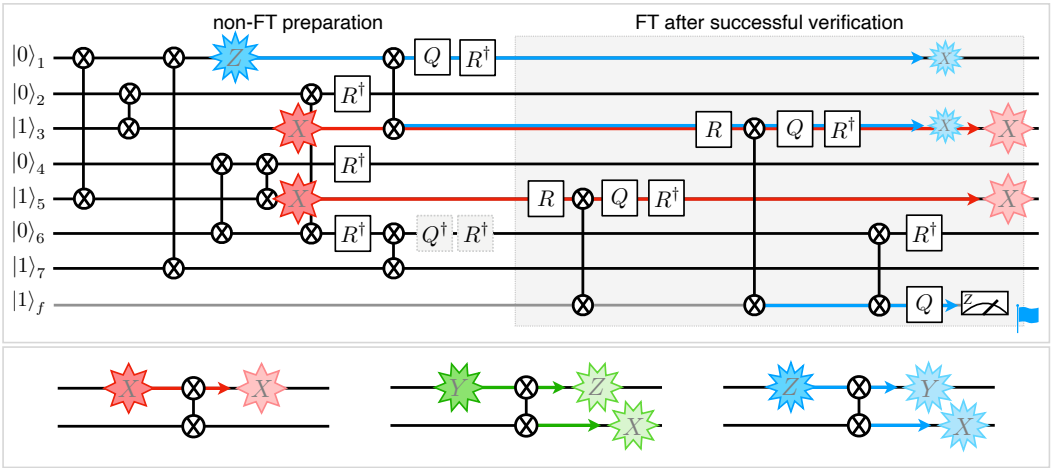


FIGURE 4.5: The Pauli state preparation circuit from Fig. 4.4(a) is compiled into MS gates (see Eq. (4.4) and Fig. 4.1(b)) using Eq. (4.6) and local rotations are optimized (upper panel). The first block prepares $|0\rangle_L$ non-fault-tolerantly. The two gates shaded gray can be omitted when the second block is also run to prepare $|0\rangle_L$ fault-tolerantly in case the flag is clear. The lower panel illustrates propagation rules for Pauli faults through MS gates. As an example, the Z_1 fault (12-cornered blue star) propagates to X_1X_3 and triggers the flag. The crosstalk fault X_3X_5 (red 8-cornered stars) breaks fault tolerance because it leads to a weight-2 error although the flag is clear. Further discussion of crosstalk is provided in Sec. 4.1.3.1.

through MS gates.

The quality of the prepared $|0\rangle_L$ state is analyzed by categorizing X-errors that end up on the state after running the encoding circuit. The bitstrings that label all eight constituent states of

$$|0\rangle_L = \frac{1}{\sqrt{8}}(|0000000\rangle + |1010101\rangle + |0110011\rangle + |1100110\rangle + |0001111\rangle + |1011010\rangle + |0111100\rangle + |1101001\rangle) \tag{4.48}$$

are the possible classical measurement outcomes when measuring the seven physical qubits that make up the logical qubit in the Z-basis (assuming absence of faults). In post-processing, we can determine the minimal Hamming distance D_H of any measured bitstring m as

$$\min(D_H(m, 0000000), D_H(m, 1010101) \ldots, D_H(m, 1101001)) \tag{4.49}$$

to classify whether $|0\rangle_L$ has been prepared with an X-error of weight $0, 1, 2$ or 3 by the noisy circuit. While X-errors of weight 0 or 1 are considered correctable, those of weight 2 are uncorrectable and weight-3 errors correspond to logical operators directly. Note that here, for instance the operator $X_5X_6X_7$, which is equivalent to X_4 by application of K_1^X, is counted as a weight-1 error. The occurrence rates of each error category is shown as simulated via stabilizer simulation and measured in the experiment in Fig. 4.4(b) for both the non-FT and FT $|0\rangle_L$ state preparation. Errors that lead to failure occur less frequently by approximately one order of magnitude when the flag verification technique is used in both simulation and experiment.

The logical infidelity $1 - F_0$ of the $|0\rangle_L$ state, with

$$F_0 = \frac{1 + \langle Z_L \rangle}{2}, \tag{4.50}$$

i.e. the overlap of the reconstructed Bloch vector with the desired logical state, is suppressed from 5.38(2)% to 1.01(1)% in simulation (9.0(6)% to 1.2(2)% in the experiment) as can be seen in Fig. 4.4(c). While in simulation the logical infidelity is almost the same for all six states – note that the logical Clifford rotation slightly increases infidelity for all states except $|0\rangle_L$ – the experimental values, taken on different days, fluctuate more strongly due to miscalibrations of the apparatus. The average infidelity for all six FT state preparations in the simulations (experiment) is 1.203(4)% (1.1(1)%) where, for the five other states, the respective logical fidelities are determined as

$$F_1 = \frac{1 - \langle Z_L \rangle}{2}, \quad F_\pm = \frac{1 \pm \langle X_L \rangle}{2} \quad \text{and} \quad F_{\pm \mathrm{i}} = \frac{1 \pm \langle Y_L \rangle}{2} \tag{4.51}$$

according to Eq. (2.13). It was also verified experimentally that the code space was prepared correctly by also measuring the six stabilizer generators for all Pauli states [Pos+22]. The average expectation value of the X-type (Z-type) stabilizers is 0.826(3) (0.760(3)) for the FT and 0.842(3) (0.790(3)) for the non-FT preparation scheme. The acceptance rate of the flag verification step in the preparation of $|0\rangle_L$ is obtained as 84.42(4)% in the numerical simulations. The experimentally observed acceptance rate is 78.9(5)%.

As a more general measure to assess the quantum state of the logical qubit, we determine its quantum state fidelity

$$\mathcal{F} = \frac{1}{128} \sum_{i=1}^{128} \langle S_i \rangle, \tag{4.52}$$

noise	$\mathcal{F}$	p_{CS}
depolarizing	82.63(3)%	82.62(4)%
overrotation	86.18(3)%	86.20(4)%
experiment	82.7(11)%	83.1(15)%

TABLE 4.2: Comparison of quantum state fidelity $\mathcal{F}$ and code space population p_{CS}, as predicted by the two noise models, to experimentally measured values.

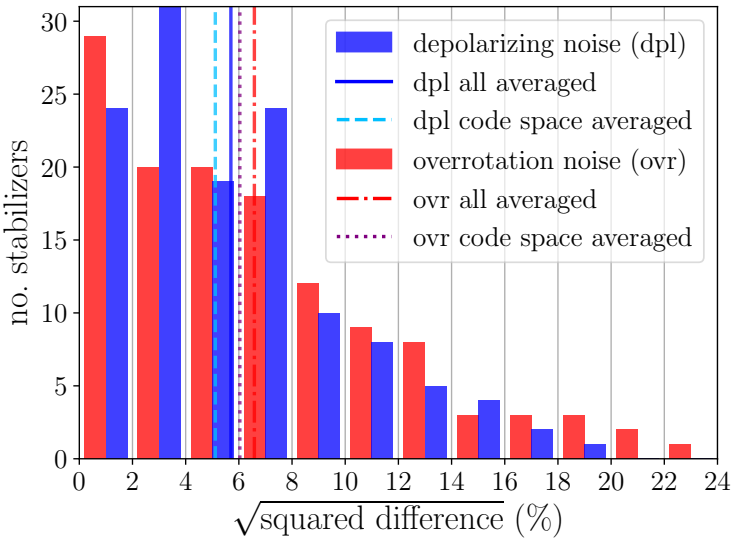


FIGURE 4.6: Stabilizer expectation values are estimated numerically using depolarizing (dpl) and incoherent overrotation (ovr) noise. The deviation $\sqrt{(\langle S_i\rangle_{\mathrm{exp}} - \langle S_i\rangle_{\mathrm{sim}})^2}$ to experimentally measured expectation values is shown for all 128 stabilizer operators. Average deviations for the quantum state fidelity and code space population of approximately 6% are indicated by vertical lines for both noise models. For the simulation we run 10^5 MC shots and make 100 measurements in the experiment.

which is the mean of expectation values of all $2^7 = 128$ stabilizer operators $S_i \in \mathcal{S}$. Here the mean is calculated with respect to the noisy state ρ and the S_i are the stabilizer elements $\langle K_1^X, K_2^X, K_3^X, K_1^Z, K_2^Z, K_3^Z, Z_L\rangle$ of the desired logical state $|0\rangle_L$. Analogously, the code space population p_{CS} is defined by averaging over all code space stabilizer expectation values as

$$p_{\mathrm{CS}} = \frac{1}{64}\sum_{i=1}^{64}\langle S_i\rangle, \tag{4.53}$$

i.e. without logical operators.

Since MS gates have the largest physical fault rate in the considered setup, we expect that they are the dominant source of noise on the quantum state fidelity and, as a consequence, also on the logical failure rates. We thus compare quantum state fidelities and code space populations as predicted by both the experimentally-informed depolarizing noise model and the presumably more microscopically-appropriate incoherent overrotation noise model[4] to the experimentally measured values given in Tab. 4.2. We observe that the depolarizing model captures these *average* values of the experiment more closely than the overrotation noise model.

[4]We also compare to predictions of a coherent overrotation noise model for another study that we present in Sec. 4.1.3.1.

Now, for *each* stabilizer operator of the logical qubit in the $|0\rangle_L$ state, we also determine the deviation $\sqrt{(\langle S_i\rangle_{\text{exp}} - \langle S_i\rangle_{\text{sim}})^2}$ of the simulated expectation value $\langle S_i\rangle_{\text{sim}}$ with both noise models to the experimentally measured expectation value $\langle S_i\rangle_{\text{exp}}$. We plot the deviation from experimental stabilizer expectation values for both simulated noise models in the histogram in Fig. 4.6. The distributions of deviations are very similar for both noise models. The largest fraction of stabilizers deviates little from the experimental values and few stabilizers can deviate up to 20% (24%) for the depolarizing (overrotation) noise model. Averaged deviations for the stabilizers that make up the quantum state fidelity and code space population are not significantly different with 5.7(11)% and 5.1(15)% (6.6(11)% and 6.0(15)%) respectively for depolarizing (overrotation) noise. Evidently, for the circuits that were used in the considered experiments, incoherent overrotation noise does not provide a more accurate description of the effect of noisy MS gates than depolarizing noise.

We conclude that from studying the quantum state fidelity, in addition to the logical fidelity, we do not gain additional information about the QEC procedure but logical fidelity is sufficient to compare experimental data to noise simulations.

4.1.2.2 Logical CNOT gate

The FT logical CNOT gate is realized between two logical qubits that are each prepared fault-tolerantly with the previously described procedure with a total of 29 entangling gates on 16 qubits. After both logical qubits are successfully initialized using 11 MS gates each, 7 physical entangling gates are applied pairwise to all 7 physical qubit pairs of the two registers as illustrated by the circuit in Fig. 4.7(a). The intended effect of the ideal logical CNOT gate on various logical product states is shown in the table on the right. Application of the logical CNOT to the state $|+,0\rangle_L$ creates the maximally-entangled Bell state $|\Phi^+\rangle_L$ on the logical level with a logical fidelity of 75.4(13)% in the experiment. The logical density matrix of the ideal Bell state is

input	output
$\|0,0\rangle_L$	$\|0,0\rangle_L$
$\|0,1\rangle_L$	$\|0,1\rangle_L$
$\|1,0\rangle_L$	$\|1,1\rangle_L$
$\|1,1\rangle_L$	$\|1,0\rangle_L$
$\|+,0\rangle_L$	$\|\Phi^+\rangle_L$
$\|+\mathrm{i},0\rangle_L$	$\|\Phi^{+\mathrm{i}}\rangle_L$

TABLE 4.3: Input-output table for the logical CNOT gate.

$$\rho_{\Phi^+} = \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \tag{4.54}$$

in the logical computational basis, which we show in Fig. 4.7(b) alongside with experimental logical state tomography data. The rotated Bell state $|\Phi^{+\mathrm{i}}\rangle = \frac{1}{\sqrt{2}}(|0,0\rangle + \mathrm{i}|1,1\rangle)$ is related to $|\Phi^+\rangle$ via a single phase gate $|\Phi^{+\mathrm{i}}\rangle = S_1|\Phi^+\rangle$ so that its stabilizer generators are obtained by

$$S_1 Z_1 Z_2 S_1^\dagger = Z_1 Z_2 \qquad \text{and} \qquad S_1 X_1 X_2 S_1^\dagger = Y_1 X_2. \tag{4.55}$$

For our six different logical input states, the logical infidelities of the output states after the logical CNOT are

$$F_{00} = \frac{1}{4}\big(1 + \langle Z_L^1\rangle + \langle Z_L^2\rangle + \langle Z_L^1 Z_L^2\rangle\big) \tag{4.56}$$

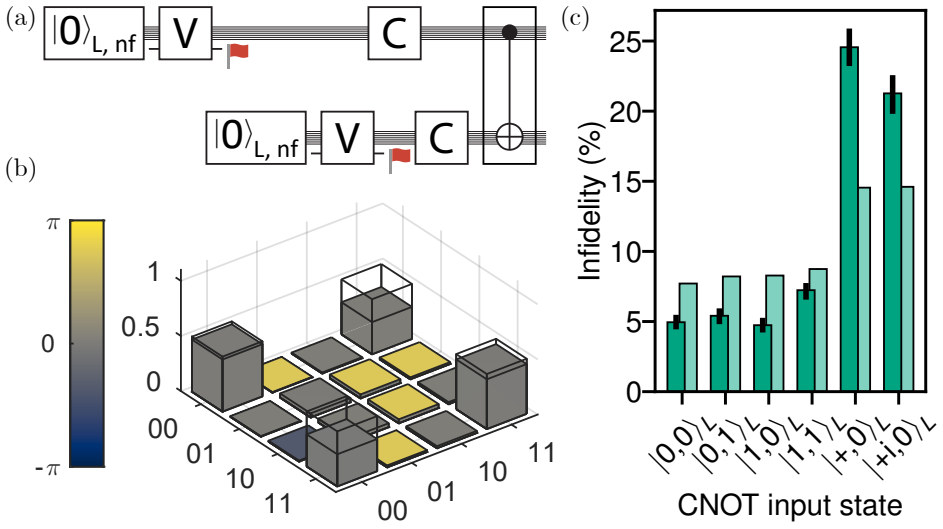


FIGURE 4.7: (a) FT circuit on the logical level for preparing two logical qubits and applying the logical CNOT gate. (b) Logical state tomography of the experimentally prepared noisy logical Bell state $|\Phi^+\rangle_L$ resembling the ideal density matrix in Eq. (4.54) depicted by the wireframes. Color represents the phase of the complex amplitudes. (c) Logical infidelities for six different logical input states to the logical CNOT gate obtained by numerical simulation (lighter, right, 10^6 samples) and in the experiment (darker, left). Error bars indicate the 68% confidence interval in the figure and in the main text.

$$F_{01} = \frac{1}{4}\left(1 + \langle Z_L^1\rangle - \langle Z_L^2\rangle - \langle Z_L^1 Z_L^2\rangle\right) \tag{4.57}$$

$$F_{11} = \frac{1}{4}\left(1 - \langle Z_L^1\rangle - \langle Z_L^2\rangle + \langle Z_L^1 Z_L^2\rangle\right) \tag{4.58}$$

$$F_{10} = \frac{1}{4}\left(1 - \langle Z_L^1\rangle + \langle Z_L^2\rangle - \langle Z_L^1 Z_L^2\rangle\right) \tag{4.59}$$

$$F_{\Phi^+} = \frac{1}{4}\left(1 + \langle X_L^1 X_L^2\rangle - \langle Y_L^1 Y_L^2\rangle + \langle Z_L^1 Z_L^2\rangle\right) \tag{4.60}$$

$$F_{\Phi^{+i}} = \frac{1}{4}\left(1 + \langle Z_L^1 Z_L^2\rangle + \langle X_L^1 Y_L^2\rangle + \langle Y_L^1 X_L^2\rangle\right) \tag{4.61}$$

as can be calculated from the projectors onto the respective two-qubit states

$$P_{ij} = \frac{1}{2}\left(I + (-1)^i Z_L^{(1)}\right)\frac{1}{2}\left(I + (-1)^j Z_L^{(2)}\right) \quad \text{with} \quad i,j \in \{0,1\}, \tag{4.62}$$

$$P_{\Phi^+} = \frac{1}{2}\left(I + X_L^1 X_L^2\right)\frac{1}{2}\left(I + Z_L^1 Z_L^2\right) \quad \text{and} \quad P_{\Phi^{+i}} = \frac{1}{2}\left(I + Y_L^1 X_L^2\right)\frac{1}{2}\left(I + Z_L^1 Z_L^2\right) \tag{4.63}$$

using Eq. (2.13). Their simulated and experimental values are shown in Fig. 4.7(c). The infidelities of output states is much higher when the control qubit is in one of the two superposition states $|+,0\rangle_L$ or $|+\mathrm{i},0\rangle_L$, thus causing an entangled output state, compared to the four product output states $|0/1,0/1\rangle_L$. The averaged logical infidelities over all six logical two-qubit states is 10.35(1)% in the simulations (11.0^{+3}_{-4}% in the experiment).

Conducting further studies on how the transversal CNOT implementation spreads error distributions from the input logical qubits could assist in developing logical entangling gate protocols with higher fidelities. A comprehensive characterization of the noisy logical CNOT gate has yet to be conducted (for instance, using the methods described in [Com+17; HFW20; FW20; CD+23]).

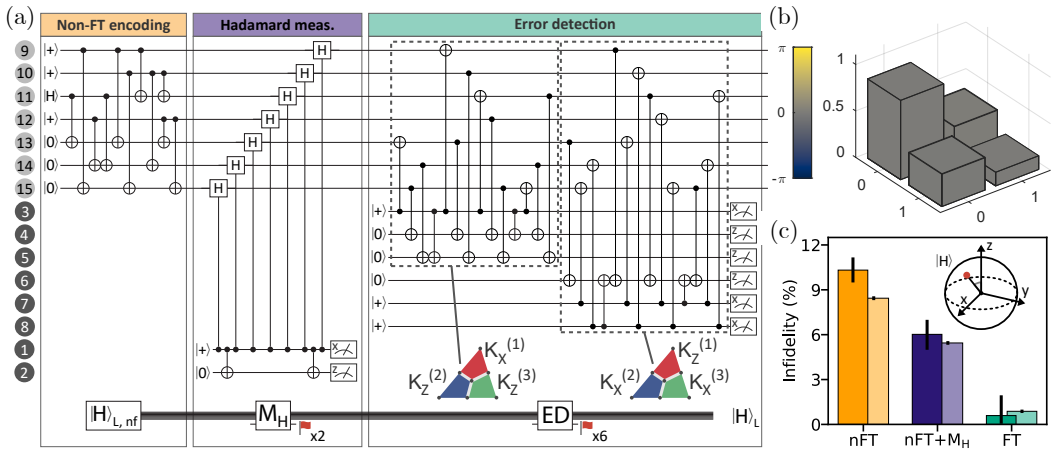


FIGURE 4.8: Three-step FT magic state preparation suggested by [Got16a; CC19]. (a) The first step prepares the state non-fault-tolerantly using the ion indexing as given on the left. In the second step, for an FT measurement of H_L two flag qubits indicate that the $+1$ eigenstate is prepared and that no dangerous fault has occurred. The third step is an FT error detection block, first given in [Rei20], where three X- and Z-type stabilizers are measured in an intertwined way so that the auxiliary qubits act as flag qubits. A preparation trial is aborted and repeated when any flag qubit is triggered and accepted when all flags are clear. (b) Logical state tomography data of the experimentally prepared logical magic state deviate from the ideal density matrix in Eq. (4.65) no more than 50 mrad for the complex phase factors and 0.7% for the amplitudes. (c) Infidelity of the logical magic state decreases with every preparation step both in numerical simulations (lighter bars on the right, 10^5 samples) and in the experiment (darker bars on the left). Error bars indicate the 68% confidence interval in the figure and in the main text.

4.1.2.3 Magic state preparation and T-gate teleportation

In order to fault-tolerantly implement the logical T-gate that performs the unitary $T_L = \mathrm{e}^{-\mathrm{i}\frac{\pi}{8}Y_L}$, we use the magic state injection technique, analogous to the one described in Sec. 2.5.1.2 and depicted in Fig. 2.8, on the logical level. The required logical magic state

$$|H\rangle_L = \cos\left(\frac{\pi}{8}\right)|0\rangle_L + \sin\left(\frac{\pi}{8}\right)|1\rangle_L \tag{4.64}$$

with density matrix

$$|H\rangle\langle H|_L = \frac{1}{2}\begin{pmatrix} 1+\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1-\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{2}\left(I + \frac{X_L + Z_L}{\sqrt{2}}\right) \tag{4.65}$$

(see Eq. (2.29)) is the $+1$ eigenstate of the logical Hadamard operator H_L. Our implementation follows the proposal in [Got16a; CC19], which is given in Fig. 4.8(a) and consists of three steps. Firstly, the magic state $|H\rangle_L$ is prepared with a non-FT circuit, similar to the Pauli state preparation circuit: Starting from a single physical magic state $|H\rangle$, the corresponding logical state $|H\rangle_L$ is encoded into the seven qubits that make up the Steane code. As a second step, the logical Hadamard operator H_L is measured fault-tolerantly using an additional flag qubit. This step projects the logical state onto the ±1 eigenspace spanned by $|H\rangle_L$ and the orthogonal state $|-H\rangle_L = Y_L|H\rangle_L$ and a measurement qubit, which is treated like a flag, indicates onto which of the two states the logical data qubit has been projected. The $|-H\rangle_L$ state could have been the result of a single fault in the non-FT encoding block, for instance

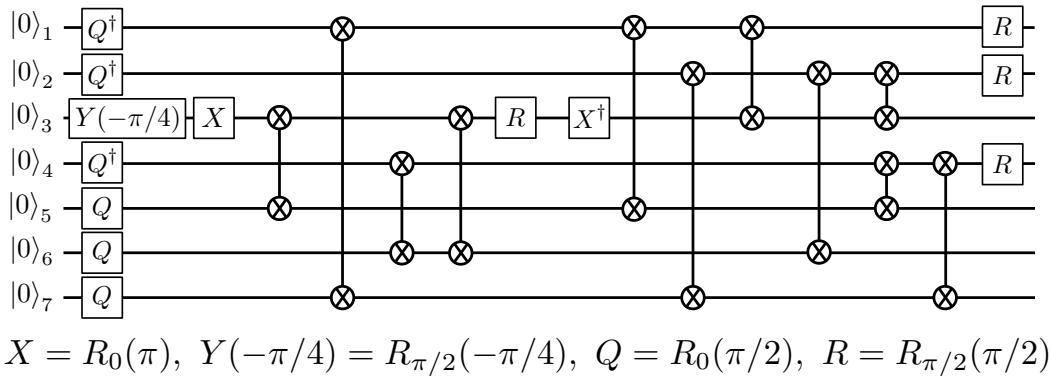


FIGURE 4.9: The non-FT magic state preparation circuit from Fig. 4.8(a) is compiled into two-qubit MS gates $\mathrm{MS}_{0,0}(-\pi/2)$ and local rotations in the XY-plane.

a Y-fault directly after preparing the physical magic state on the third physical qubit. During the second step, dangerous faults may propagate from the measurement qubit back to the physical data qubits while triggering the flag. However, a weight-2 error at the end of the first block may not cause a non-trivial measurement result in the second block. An additional last step is necessary that detects errors that may otherwise break fault tolerance, using a circuit given in [Rei20]. This error detection circuit is an efficient way to measure all six stabilizers with only 28 CNOT gates. The first part measures $K_X^{(1)}$, $K_Z^{(2)}$ and $K_Z^{(3)}$ and then $K_Z^{(1)}$, $K_X^{(2)}$ and $K_X^{(3)}$ are measured with one auxiliary qubit each. Due to their additional CNOT interconnections, the measurements can not only indicate non-trivial stabilizer eigenvalues in case the circuit block is executed without faults, but they also act as flags in case dangerous faults happen within the error detection block. If all three steps are executed and all auxiliary qubits are measured as $+1$, it is guaranteed that no single dangerous fault can have corrupted the state and as a consequence $|H\rangle_L$ is prepared correctly up to a weight-1 error that can be corrected in a subsequent round of EC or in software after all physical qubits have been measured.

We again compile both the non-FT encoding circuit and the three-step FT preparation circuit to only use the native gate set available in the considered ion trap architecture consisting of two-qubit MS gates and local rotations of arbitrary angle in the XY-plane as shown in Figs. 4.9 and 4.10. Note that, contrary to [MEK12], we do not need physical magic states for the Hadamard measurement step because the physical $\pi/4$-rotations that are required to perform the controlled-H operation can be done directly within the ion trap quantum computer instead of performing only Clifford gates (see the compilation in the right part of Fig. 4.10(a)). All measurements are deferred to the end of the FT preparation circuit.

The reconstructed logical density matrix of the logical magic state, which ideally would be given by Eq. (4.65), is shown in Fig. 4.8(b). The logical infidelities $1 - F_H$ of the logical magic state, where

$$F_H = \frac{1}{2}\left(1 + \frac{\langle X_L\rangle + \langle Z_L\rangle}{\sqrt{2}}\right), \tag{4.66}$$

are shown for statevector simulations and experiments after all three preparation steps in Fig. 4.8(c). Note that for the logical magic state, all types of logical Pauli operators, X_L, Y_L and Z_L, that may occur at the end of noisy state preparation

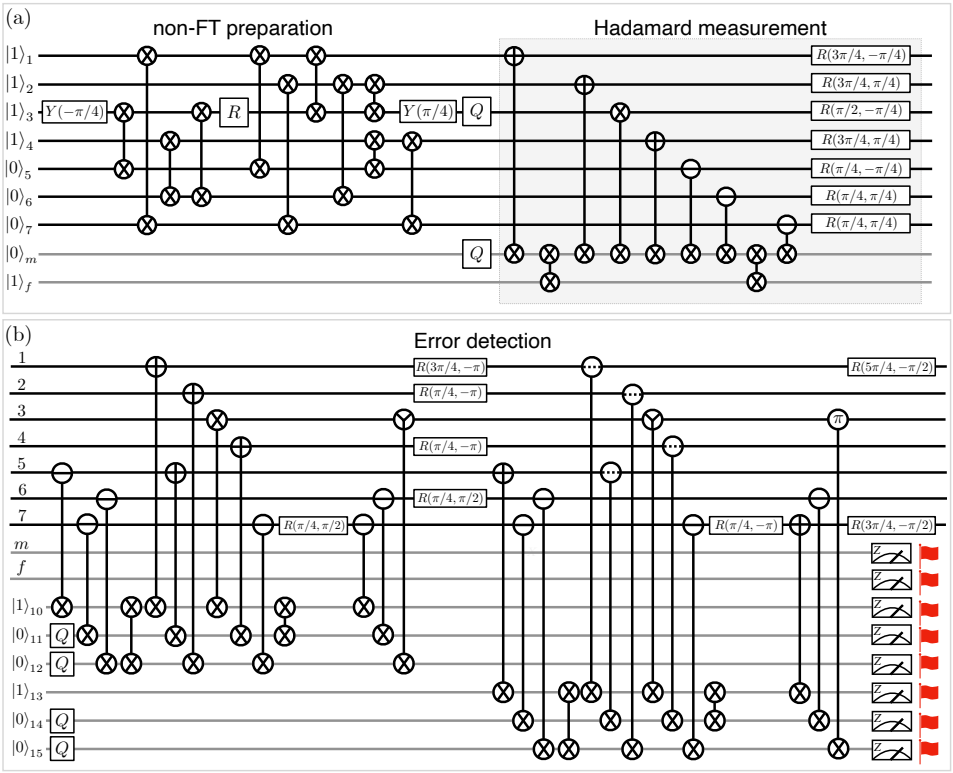


FIGURE 4.10: The three blocks for FT magic state preparation from Fig. 4.8(a) are compiled into a single circuit containing only the two-qubit MS gates $\mathrm{MS}_{\varphi_1,\varphi_2}(-\pi/2)$ given in Fig. 4.1(b) and local rotations in the XY-plane. (a) Circuit to perform the first two preparation steps in conjunction. (b) Circuit to perform the error detection block.

circuits contribute a finite value to the infidelity since

$$\langle H|X|H\rangle = \frac{1}{\sqrt{2}}, \quad \langle H|Y|H\rangle = 0 \quad \text{and} \quad \langle H|Z|H\rangle = \frac{1}{\sqrt{2}} \tag{4.67}$$

as opposed to Pauli states of X- and Z-type where only the opposite-type logical operator, Z_L or X_L respectively, and Y_L can induce logical flips. The logical magic state infidelity clearly decreases with every preparation step over approximately one order of magnitude from 8.4(1)% to 0.9(1)% in simulations (10.3(8)% to 0.6^{+14}_{-5}% in the experiment). The probability to accept the prepared state, i.e. to measure all flags as +1, decreases from 72% after the second preparation step to 27% for the full FT magic state preparation circuit. It is much lower than the acceptance rate for the $|0\rangle_L$ state due to the much larger circuit. It is notable that for this figure of merit, the respective experimentally observed values of approximately 57% and 14% deviate significantly from the simulation. In the table below, we list some quantities of the flag circuits used for logical state preparation alongside with the relative discrepancy $\Delta\varepsilon$ of acceptance rates in the simulations and in the experiments. We observe that, clearly, this discrepancy grows with the number of entangling gates of the circuit.

encoding	no. entangling gates	no. flags	simulation	experiment	$\Delta\varepsilon$
$\|0\rangle_{L,\mathrm{ft}}$	8	1	85%	79%	7%
$\|H\rangle_{L,\mathrm{nf}}$ & M_H	20	2	72%	57%	21%
$\|H\rangle_{L,\mathrm{ft}}$	48	8	27%	14%	48%

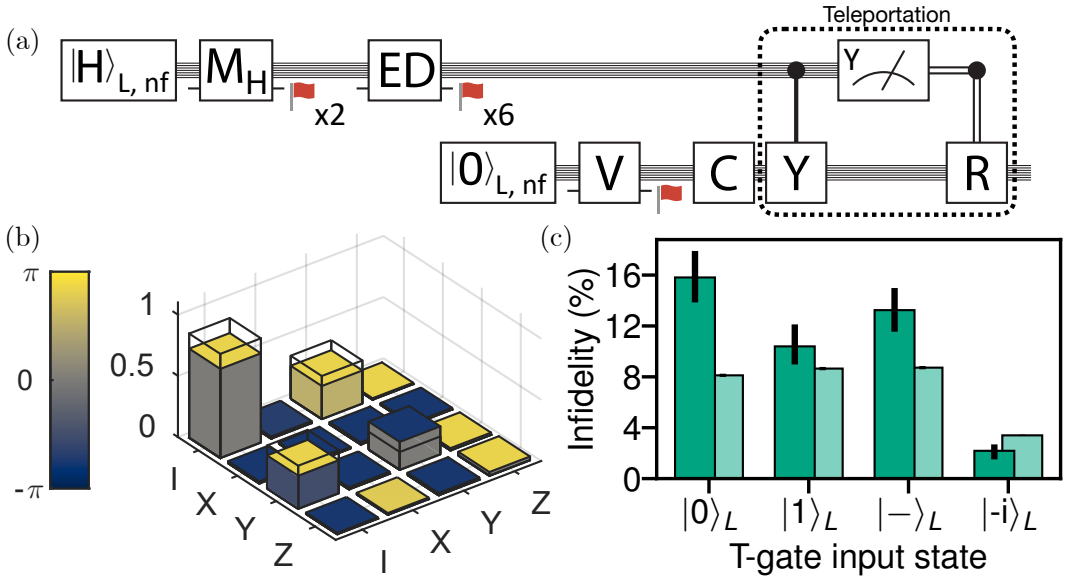


FIGURE 4.11: Fault-tolerant T-gate realization. (a) Logical circuit blocks that realize FT magic state preparation and FT Pauli state preparation on two logical qubits followed by a gate teleportation circuit that applies the logical T-gate to the second logical qubit. Conditional application of $R \equiv R_{Y_L}(\pi/2)$ is performed in software. Apart from magic state preparation, the logical circuit solely consists of logical Clifford operations that are transversal in the Steane code. (b) Logical process matrix of the experimental T-gate. The wireframes correspond to ideal values according to Eq. (4.70). (c) The logical T-gate is applied to various logical Pauli states and the lowest infidelity (simulations and experimental results are depicted lighter and darker, respectively) is achieved for the $|-\mathrm{i}\rangle_L$ state since it is an eigenstate of T_L. The simulations use 10^5 MC shots. Error bars indicate the 68% confidence interval in the figure and in the main text.

When FT preparation of the logical magic state has succeeded, we use gate teleportation as seen in the rightmost part of the circuit in Fig. 4.11(a) to realize the logical T-gate fault-tolerantly on an input logical Pauli state $|\psi\rangle_L$. All flag qubits are in the $|0\rangle$ state in case of acceptance of the logical magic state. Thus, no re-cooling of the ions is required to keep using the previously measured auxiliary qubits as the second qubit register. A target Pauli state $|\psi\rangle_L$ can hence be prepared on them subsequently. The controlled-Y gate then injects the magic state as the circuit takes the state

$$\begin{aligned}
|H\rangle_L|\psi\rangle_L &= (\cos(\pi/8)|0\rangle_L + \sin(\pi/8)|1\rangle_L)(\alpha|0\rangle_L + \beta|1\rangle_L) \\
&\xrightarrow{\mathrm{CY}} \cos(\pi/8)|0\rangle_L(\alpha|0\rangle_L + \beta|1\rangle_L) + \mathrm{i}\sin(\pi/8)|1\rangle_L(\alpha|1\rangle_L - \beta|0\rangle_L) \\
&= \frac{1}{\sqrt{2}}\Big(|+\mathrm{i}\rangle \mathrm{e}^{-\mathrm{i}\frac{\pi}{8}Y_L}|\psi\rangle_L + |-\mathrm{i}\rangle \mathrm{e}^{\mathrm{i}\frac{\pi}{8}Y_L}|\psi\rangle_L\Big) \\
&= \frac{1}{\sqrt{2}}\Big(|+\mathrm{i}\rangle T_L|\psi\rangle_L + |-\mathrm{i}\rangle \mathrm{e}^{\mathrm{i}\frac{\pi}{4}Y_L} T_L|\psi\rangle_L\Big)
\end{aligned} \tag{4.68}$$

and then collapses the second logical qubit to the state $T_L|\psi\rangle_L$ in case the measurement of the first register in the logical Y-basis yields $+1$. For a -1 measurement, the additional logical rotation $R_{Y_L}(\pi/2) = \mathrm{e}^{-\mathrm{i}\frac{\pi}{4}Y_L}$ has to be applied to the second register. The logical process matrix χ of the T-gate is defined in the Pauli basis (see Eq. (3.6)) via

$$T\rho T^\dagger = \sum_{i=0}^{3}\sum_{j=0}^{3} \chi_{ij}\sigma_i\rho\sigma_j \tag{4.69}$$

and reads

$$\chi = \frac{1}{2}\begin{pmatrix} 1+1/\sqrt{2} & 0 & \mathrm{i}/\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 \\ -\mathrm{i}/\sqrt{2} & 0 & 1-1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{4.70}$$

The four states $|0\rangle_L$, $|1\rangle_L$, $|-\rangle_L$ and $|-\mathrm{i}\rangle_L$, which form a tomographically complete set [Hra+04], are used to reconstruct χ as shown in Fig. 4.11(b). The logical fidelities for these states are given by F_H, $F_{-H} = \frac{1}{2}\left(1 - \frac{\langle X_L\rangle + \langle Z_L\rangle}{\sqrt{2}}\right)$, $F_{ZH} = \frac{1}{2}\left(1 - \frac{\langle X_L\rangle - \langle Z_L\rangle}{\sqrt{2}}\right)$ and $F_{-\mathrm{i}}$ since the effect of the T-gate on the respective input states is

$$|H\rangle_L = T_L|0\rangle_L, \quad |-H\rangle_L = T_L|1\rangle_L, \quad Z_L|H\rangle_L = T_L|-\rangle_L, \quad |-\mathrm{i}\rangle_L = T_L|-\mathrm{i}\rangle_L \tag{4.71}$$

up to global phase factors. Because $|-\mathrm{i}\rangle_L$ is an eigenstate to T_L, the procedure yields the lowest infidelity for the state $T_L|-\mathrm{i}\rangle_L$: Intuitively, a state that aligns to a large extent with the logical Y-axis is less prone to noise on the rotation angle for a rotation about the logical Y-axis than states whose logical Bloch vectors are at a larger angle with the logical Y-axis. The infidelities for all four states are given in Fig. 4.11(c), which together yield a mean infidelity of 10(1)% in the experiment. The simulated values qualitatively agree with the observed experimental values.

4.1.3 Extensions of logical state preparation

Following the demonstrated experimental success described above, we conducted further theoretical studies, published in [Heu+23a], on how FT state preparation procedures for logical Pauli and magic states could be improved for future hardware architectures. In the following, we first show numerical evidence that the break-even point of advantageous FT circuit operation over physical qubits is within reach for near-term quantum processors. Next, we provide a resource comparison of these non-deterministic state preparation schemes and their deterministic counterparts and discuss regimes of usefulness for either type when the goal is to prepare multiple logical qubits. Lastly, we further analyze the role of crosstalk on entangling gates during state preparation and find that by appropriate qubit arrangement the Pauli state preparation circuit can be made FT toward X-type crosstalk.

4.1.3.1 Reaching break-even with non-deterministic schemes

We deem the above-described FT logical state preparation circuits non-deterministic since, in the presence of noise, it cannot be determined a priori whether or not all flags will be clear and a state will be accepted in a single given circuit run. Using these circuits, we extend the noise model with parameters given in Tab. 4.1 to also include dephasing noise on idling locations, according to Eq. (4.9). Also phase-averaged crosstalk on single-qubit and MS gates parametrized as in Eqs. (4.23) and (4.42) respectively is included. These are as well considerable sources of noise and the parameters of this extended noise model are listed in Tab. 4.4.

In this section, we provide an overall idea of scaling behavior of logical failure rates dependent on the physical fault rates by uniformly scaling all physical fault rates with a single parameter λ like

$$(p_1, p_2, p_i, p_m, \dots) \to \lambda \cdot (p_1, p_2, p_i, p_m, \dots). \tag{4.72}$$

Operation	Physical fault rate	Numerical value
Single-qubit gates	p_1	5×10^{-3}
Two-qubit gates	p_2	2.5×10^{-2}
Initializations	p_i	4.5×10^{-3}
Measurements	p_m	4.5×10^{-3}
Idling during single-qubit gates	$p_{\text{idle},1}$	7.5×10^{-5} ($t = 15\,\mu\text{s}$)
Idling during MS gates	$p_{\text{idle},2}$	1.0×10^{-5} ($t = 200\,\mu\text{s}$)
Idling during measurements	$p_{\text{idle},m}$	1.5×10^{-5} ($t = 300\,\mu\text{s}$)
Single-qubit gate crosstalk, $\theta = \pi$	$p_{\text{c}_1}(\pi)$	2.5×10^{-4} ($\varepsilon = 1\%$)
Single-qubit gate crosstalk, $\theta = \pi/2$	$p_{\text{c}_1}(\pi/2)$	6.2×10^{-5} ($\varepsilon = 1\%$)
Single-qubit gate crosstalk, $\theta = \pi/4$	$p_{\text{c}_1}(\pi/4)$	1.5×10^{-5} ($\varepsilon = 1\%$)
MS gate crosstalk	p_{c_2}	6.2×10^{-5} ($\varepsilon = 1\%$)

TABLE 4.4: Experimentally-informed fault rates for the extended noise model used in our numerical simulations. For idling noise, the fault rates are determined by the duration t of the respective operation (see Eq. (4.9)) and for crosstalk, the fault rates are determined by the crosstalk ratio ε and the rotation angle θ of the respective gate (see Eqs. (4.20) and (4.42)) [Heu+23a].

This allows us to evaluate how much improvement of physical operations is required to outperform non-FT implementations and physical qubits with our FT circuits for logical state preparation. As laid out in Sec. 2.4.5, we compare logical failure rates of FT state preparations to their physical, unencoded counterparts [Got16b], i.e. the failure rate of initialization to the physical $|0\rangle$ state (followed by a physical Y-rotation of angle $\pi/4$) for the FT Pauli (magic) state preparation circuit. The said comparison is done for the specific hardware setup with the same physical fault rates for both logical and physical qubits. We point out that one may also compare logical performance to the best possible physical qubit operation fidelity achieved in *any* trapped-ion based quantum computer. Another criterion to determine break-even is to compare logical failure rates to the dominant physical fault rate, which in our case is the MS gate fault rate p_2, as done previously, for instance, in [Tro+18; GMB19].

Uniform scaling of all physical fault rates with the scaling parameter λ ranging over the interval $\lambda \in [10^{-4}, 10^{1}]$ is shown in Fig. 4.12 with $\lambda = 1$ representing the rates given in Tab. 4.4. The simulated numerical values are practically obtained by using direct MC (see Sec. 3.4) at the largest physical fault rates. We sample for decreasing values of λ until either the target relative uncertainty is reached or a previously specified maximum number of shots has been run. If direct MC does not obtain a sufficiently accurate data point anymore, the simulation at the given value of λ is repeated using subset sampling (see Sec. 3.5). The maximum weight subset $w_{\max}$ is chosen such that the cutoff error δ at the given value of λ makes up at most half of the target relative uncertainty. Subset sampling is then performed uniformly over all relevant subsets until the sampling error is also at most half of the target relative uncertainty or until the maximum number of samples has been run. The fault-free subset is never actually sampled since its subset failure rate and associated sampling uncertainty are both identical with zero. Fault path counting (see Sec. 3.3) is performed to determine the subset failure rate of all weight-1 subsets exactly, except for fault categories that respect fault tolerance, for instance Pauli faults on MS gates for an FT circuit, where we also set subset failure rates and sampling uncertainties to zero.

For preparation procedures of both the Pauli state $|0\rangle_L$ in Fig. 4.12(a) and the magic state $|H\rangle_L$ in Fig. 4.12(b), we observe that comparison to the physical preparation

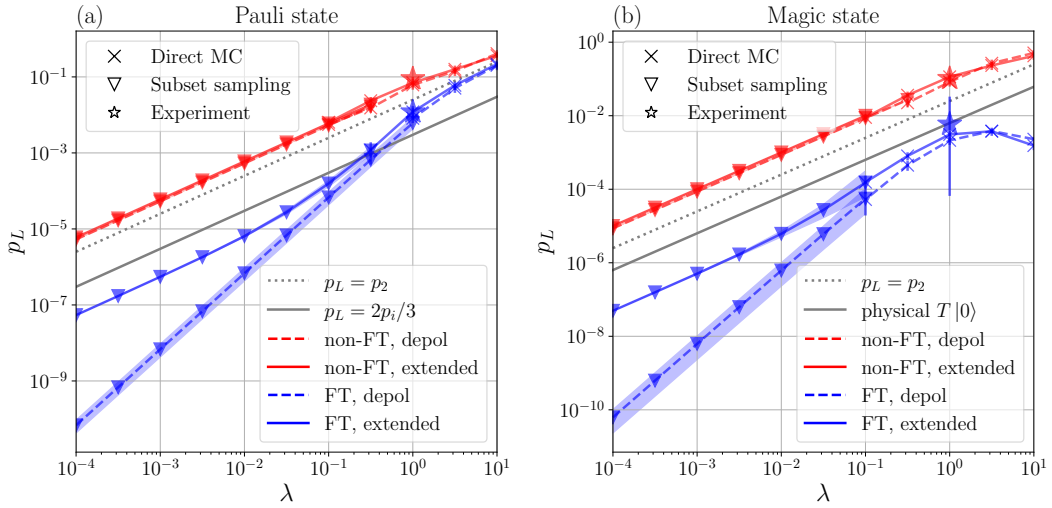


FIGURE 4.12: Logical failure rates of state preparation circuits are obtained via numerical simulations where all physical fault rates are scaled uniformly by the parameter λ. The point $\lambda = 1$ (star markers) corresponds to the fault rates given in Tab. 4.1 for the depolarizing noise model (depol) and Tab. 4.4 for the extended noise model. Failure rates of physical operations are shown in gray (lines without markers). The non-FT failure rates are larger than the FT failure rates for all shown values of λ. We use direct MC (cross markers) and subset sampling (triangle markers) with subsets up to $w_{\max} = 3$. (a) For FT Pauli state preparation, both noise models predict lower failure rates than the failure rate $2p_i/3$ of physical qubit initialization for moderate improvements $\lambda \lesssim 0.3$. Only for low λ, the quadratic scaling observed for depolarizing noise transits to a linear scaling caused by crosstalk in the extended noise model, which in general does not respect fault tolerance. We run between 10^2 and 10^4 samples for estimators obtained by both simulation methods. (b) For FT magic state preparation, the simulations suggest that break-even is reached already: For all simulated values of λ the logical failure rate is smaller than the corresponding rate for the physical magic state initialization (and the physical MS gate fault rate). For $\lambda > 1$, logical failure rates of the FT preparations decrease again, which should be related to the fact that most runs are discarded in this regime. We sample at least 10^2 times for each MC data point and subset failure rates of the non-FT circuits and the FT circuit with extended noise. For the FT circuit with depolarizing noise we use at least 10^3 shots for each subset failure rate. We sample at most 10^4 times for the non-FT circuits and up to 10^5 times for the FT circuits. The leftmost MC data point of the FT depolarizing line is obtained from 2×10^5 shots.

failure rate is the stricter criterion than comparing to the MS gate fault rate p_2 for our physical fault rates. While the initialization of a physical qubit fails with probability $2p_i/3$ given depolarizing noise of strength p_i, the failure rate of the physical magic state preparation is determined by both p_i and p_1. It is also clear that for all curves FT implementations always reach lower logical failure rates than the corresponding non-FT circuits over the full range of λ. Using the depolarizing noise model, both FT circuits show quadratic scaling behavior as $\lambda \to 0$. Since crosstalk breaks fault tolerance in general, we observe that for the extended noise model the linear term dominates the scaling as $\lambda \to 0$, just as for non-FT circuits, for both states. However, even in the presence of crosstalk, the FT circuits are capable of achieving lower logical failure rates than physical operations. The advantage at low λ, where the lines are parallel, is approximately one order of magnitude for both states. In a relatively large interval $10^{-1} \lesssim \lambda \lesssim 10^1$, the depolarizing and extended noise approximately agree in their prediction of logical failure rates. Especially the break-even point where $p_L^{\mathrm{FT}} = 2p_i/3$ for FT Pauli state preparation is found at $\lambda \approx 0.3$ for both noise models.

For FT magic state preparation, both noise models suggest that physical magic state preparation is outperformed for any value of λ. From this we conclude that not only can the break-even point be achieved with moderate improvements of the experiment but also that in the current regime of physical fault rates it is sufficient to rely on predictions made by an experimentally-informed depolarizing noise model. The extended noise model is only needed in a regime where crosstalk dominates noise on operations, especially when the MS gate fault rate becomes much smaller.

Additionally, we performed simulations of coherent overrotations (given by Eq. (4.10)) and crosstalk (given by Eq. (4.44)) to estimate the reliability of the Pauli twirling approximation that we employ for our incoherent noise models. Coherent overrotations on MS gates in the FT Pauli state preparation increase the infidelity from 0.76(5)% (0.82(6)%) in simulations with depolarizing (incoherent overrotation) noise to 1.16(7)% for the experimental parameters at $\lambda = 1$. The infidelity further increases with coherent crosstalk to 1.41(7)%. All values are in accordance with the experimentally measured value of 1.2^{+5}_{-4}%.

Logical failure rates for FT magic state preparation monotonically increase for increasing λ until $\lambda \approx 1$. For values of $\lambda \gtrsim 1$, i.e. physical fault rates that are larger than in the current experimental setup anyway, we observe that logical failure rates start to unexpectedly decrease again. We attribute this to a low acceptance rate (or, equivalently, a large flag rate) in the presence of strong noise so that, although only few states are accepted due to noise, the fidelity of the accepted states apparently becomes larger. In the next section, we shed more light on acceptance rates and compare the above-described non-deterministic schemes to their deterministic counterparts.

4.1.3.2 Comparison to deterministic schemes

In the absence of noise, our non-deterministic FT state preparation schemes always prepare the desired logical states correctly at the first trial and states never need to be discarded. For increasingly strong noise, acceptance rates decrease from unity and a larger fraction of states will be discarded, which might render the protocol experimentally unfeasible, for instance due to cycle time constraints. Eventually, with too low acceptance rates, it might become advantageous to use a deterministic scheme, which always returns the desired state at the price of larger infidelities due to the typically larger qubit and gate overhead than in the corresponding non-deterministic FT scheme.

While triggering a flag leads us to discarding the state in the non-deterministic schemes [Got16a; CC19], we may also make use of the flag information, combined with further operations, to transform all uncorrectable errors that could have been induced during state preparation into correctable errors. As long as all uncorrectable errors can be distinguished up to stabilizer-equivalence by additional operator measurements, the combined measurement information from all auxiliary qubits can be used to correct those errors that are caused by single faults in the circuit so that the FT property is respected.

In this section we elaborate a new scheme for deterministic FT Pauli state preparation[5], first published in [Heu+23a], and provide a compilation of the deterministic FT magic state preparation scheme by [CC19] into MS gates. We then compare logical

[5]This is the scheme already used in Sec. 3.5.2.2.

Data qubit error E	K_1^Z	K_2^Z, K_3^Z	R_2	R_2E	K_{123}^Z	R_1	R_1E
X_1X_3	$+$	$+-$	X_1X_3	I	$-$	X_7	$X_5K_2^X$
X_4X_5	$+$	$-+$	X_4X_5	I	$-$	X_7	$X_6K_1^X$
X_6X_7	$+$	$-+$	X_4X_5	K_1^X	$-$	X_7	X_6
X_6	$-$	$+-$	X_1X_3	$X_4K_1^XK_2^X$	$+$	I	X_6
X_5	$-$	$-+$	X_4X_5	X_4	$+$	I	X_5
X_3	$+$	$--$	X_3	I	$+$	I	X_3

TABLE 4.5: Recovery operations based on the flag error set are used instead of Tab. 2.3 if and only if the flag is triggered. All of these errors can be transformed into correctable errors or corrected directly, by measuring either the single stabilizer $K_{123}^Z = Z_1Z_2Z_4Z_7$ or the two stabilizers $K_2^Z = Z_1Z_3Z_5Z_7$ and $K_3^Z = Z_2Z_3Z_6Z_7$ or all three stabilizer generators, including $K_1^Z = Z_4Z_5Z_6Z_7$. The recovery R_1 is applied when only K_{123}^Z is measured. R_2 is the recovery operation when K_2^Z and K_3^Z are measured. The operator E can be applied directly when all three stabilizers are measured.

failure rates as well as repetition and MS gate overhead of both schemes.

Pauli state

To prepare $|0\rangle_L$ fault-tolerantly in a deterministic way, we reconsider the circuit given in Fig. 4.5. By exhaustive numerical search, we find that, given a triggered flag, the only errors E that can result from any single fault in the circuit are the ones listed in Tab. 4.5. With a clear flag, it is already assured that no uncorrectable error can be on the state. These weight-1 and weight-2 errors, which make up the so-called flag error set (FES), can be corrected after measuring additional Z-stabilizers. Note that weight-1 errors need not be corrected since they do not cause logical failure in the Steane code. Remarkably, we find that it is enough to only measure $K_{123}^Z = K_1^ZK_2^ZK_3^Z = Z_1Z_2Z_4Z_7$ using an additional auxiliary qubit as depicted with logical building blocks in Fig. 4.13(a). The weight-2 errors X_1X_3 and X_4X_5 (the error X_6X_7 is stabilizer-equivalent to X_4X_5 by application of K_1^X) have odd overlap with both K_{123}^Z and the complementary operator $Z_3Z_5Z_6$, which is measured by the flag. Thus, they will flip both auxiliary qubits while any harmless weight-1 error can flip at most one of the auxiliary qubits. The recovery operation X_7 can be conditioned on both auxiliary qubits being in the $|1\rangle$ state to transform X_1X_3 and X_4X_5 into the correctable errors $X_1X_3X_7 = X_5K_2^X$ and $X_4X_5X_7 = X_6K_1^X$ respectively[6]. Note that no additional qubit or gate overhead is required as for a general FT stabilizer measurement but a single auxiliary qubit is enough to measure K_{123}^Z: No single X-fault can propagate back to more than one data qubit and there are no uncorrectable Z-errors for preparation of $|0\rangle_L$ in the Steane code. If a single fault triggers the flag and then an additional single fault happens during the stabilizer measurement, the overall fault configuration would be of order 2 so FT is not broken (see Sec. 2.4.1). Of course, it is also possible to condition other recovery operations on the outcomes of two or three Z-stabilizer measurements, as Tab. 4.5 shows.

In Fig. 4.14, we compare FT state preparation fidelities and repetition or MS gate overhead for non-deterministic and deterministic schemes with the extended noise model. For the $|0\rangle_L$ state, all four variants show a quadratic scaling with the uniform

[6]This feedback operation can also be performed by a Toffoli gate without measurements as we lay out in Sec. 4.2.2. A single auxiliary qubit is sufficient if one is able to first measure $Z_3Z_5Z_6$ and condition a reset of the auxiliary qubit and second measurement of $Z_1Z_2Z_4Z_7$ on the first measurement result being -1, as in Sec. 3.5.2.2.

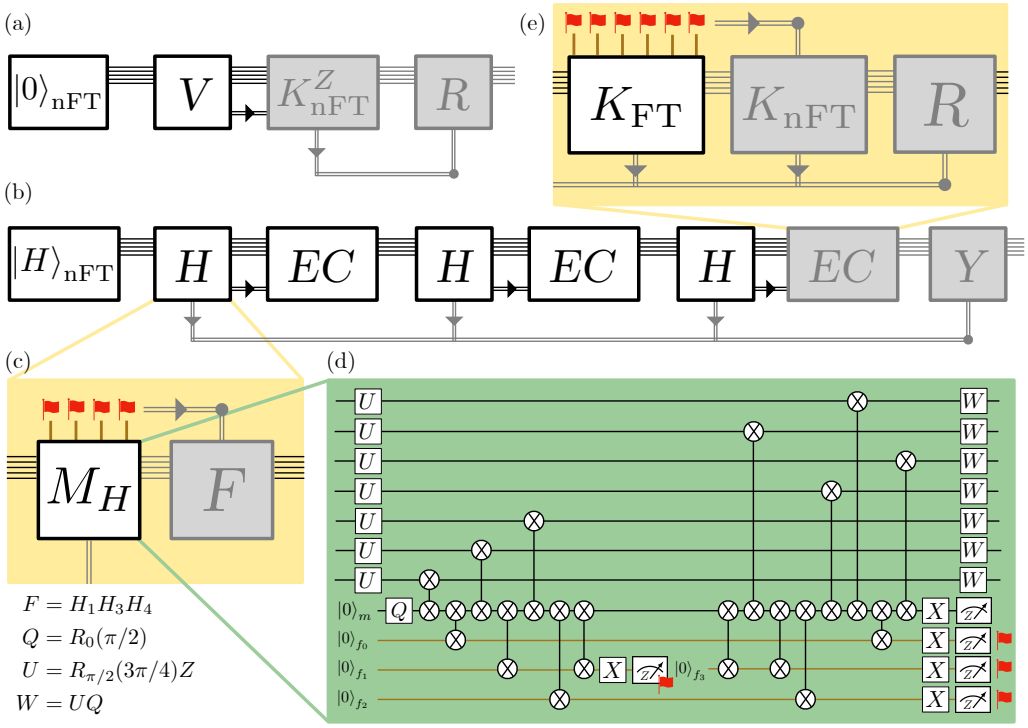


FIGURE 4.13: Logical building blocks act on registers of data qubits and auxiliary qubits to deterministically and fault-tolerantly prepare Pauli and magic states. Gray blocks are only run conditioned on mid-circuit measurements. (a) To prepare $|0\rangle_L$ fault-tolerantly, an additional stabilizer measurement (K^Z_{nFT}) is run and a recovery operation R is applied according to Tab. 4.5 when the verification flag (V) is triggered after non-FT state preparation (see Fig. 4.5). Measurements with single auxiliary qubits are sufficient to preserve fault tolerance. (b) Sequence for deterministic FT magic state preparation due to [CC19]. Non-FT magic state preparation is followed by three rounds of Hadamard measurements and FT EC. The last EC block can be omitted if all measurements in the third Hadamard block yield $+1$. A logical Y-operator is applied when the Hadamard measurements indicate -1 in the second and the third round. (c) FT measurement of the logical Hadamard operator requires four flag qubits that are used to correct all dangerous errors (feedback block F) that can happen within the block. (d) If the flag pattern $f_0, f_2, f_3 \in \{-+-, --+, ---\}$, the operation $F = H_1H_3H_4$ must be applied immediately after M_H. An example of how this guarantees error distinguishability is given in Fig. 4.15. (e) If and only if the FT parallel syndrome readout (block K_{FT}, see Fig. 4.16) flags, we proceed by measuring the syndrome with single auxiliary qubits (K_{nFT}, see Fig. 4.17). The recovery R is chosen from the Hadamard error set (Tab. 4.6) when any flag of M_H is triggered. Otherwise, R is determined by the flag error set $\{X_3X_7, X_4X_6, Z_3Z_7, Z_4Z_6\}$ if a matching syndrome, $-++$ or $++-$ for X- or Z-stabilizers respectively, is measured, otherwise the standard Steane code look up table 2.3 is applied.

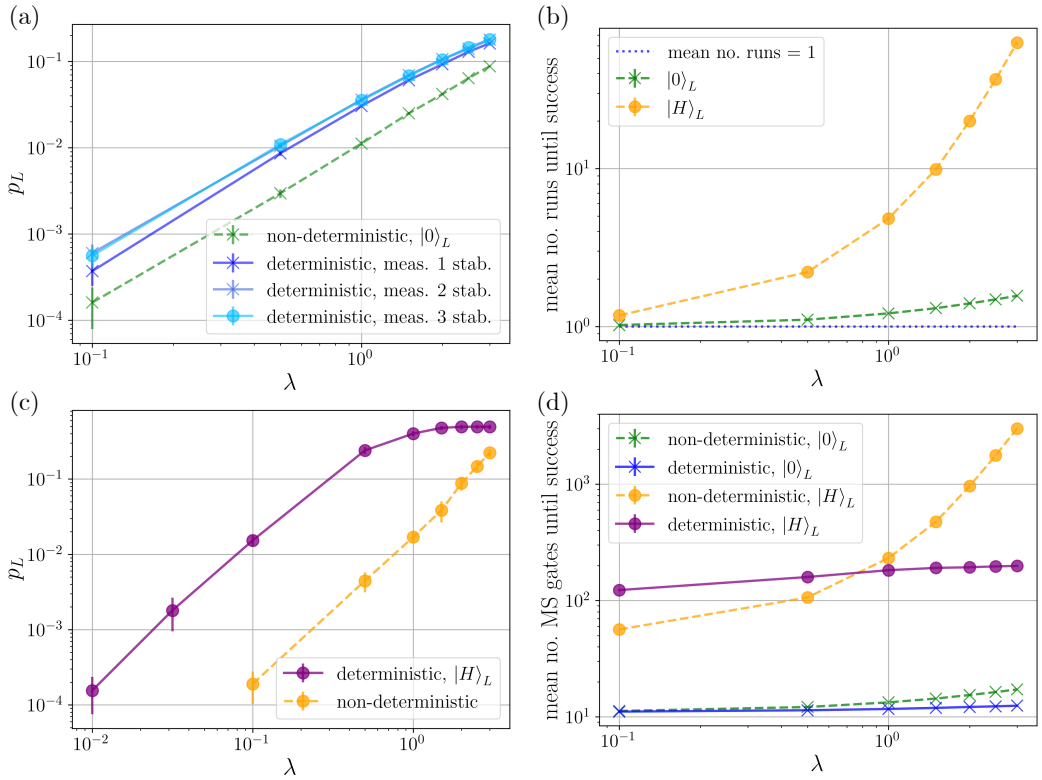


FIGURE 4.14: Trade-off between fidelity and repetition overhead for the non-deterministic and deterministic protocols to fault-tolerantly prepare $|0\rangle_L$ and $|H\rangle_L$ with the extended noise model. (a) Uniform scaling of logical failure rates of FT $|0\rangle_L$ preparation variants, where different numbers of stabilizers are measured, with scaling factor λ obtained via 10^5 MC samples. The offset between the two schemes is approximately a factor of 3. (b) Average number of repetitions for both non-deterministic schemes until all flags are clear so that the state is accepted. Deterministic schemes succeed after one trial by construction. (c) Logical failure rates for FT $|H\rangle_L$ preparation are approximately two orders of magnitude apart in the quadratic scaling regime $\lambda \lesssim 0.5$. For the deterministic scheme we use $\{10^3, 10^4, 10^5\}$ samples for the data points at $\{\lambda > 0.1,\ 0.01 < \lambda \leq 0.1,\ \lambda = 0.01\}$. For the non-deterministic scheme we use $\{10^3, 10^4, 10^5\}$ samples for the data points at $\{\lambda > 1,\ 0.1 < \lambda \leq 1,\ \lambda = 0.1\}$. (d) MS gate overhead to run circuits until the state is accepted. For non-deterministic schemes, this means that all flags are clear. Deterministic schemes realize different circuit sequences depending on in-sequence measurement outcomes. For the non-deterministic FT magic state preparation the increase is over two orders of magnitude. It requires fewer MS gates on average than the deterministic FT magic state preparation protocol for $\lambda \lesssim 0.8$.

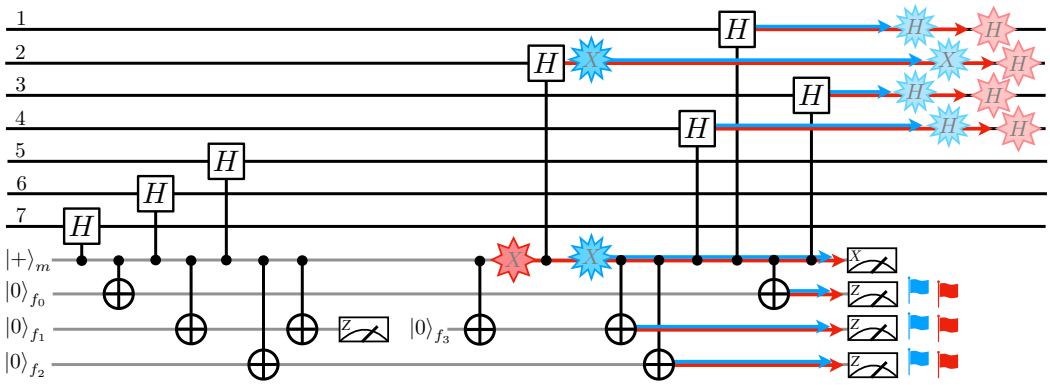


FIGURE 4.15: Flag-FT circuit for measuring the logical Hadamard operator according to [CC19]. The faults X_m (red, 8-cornered star) and X_2X_m (blue, 12-cornered star) are shown with their respective propagated errors $H_1H_2H_3H_4$ and $X_2H_1H_3H_4$.

scaling parameter λ in Fig. 4.14(a). The infidelity of the non-deterministic protocol is approximately a factor of 3 lower than for the deterministic variants. A lower infidelity is expected since the scheme uses fewer noisy gates. The average number of repetitions needed to accept a $|0\rangle_L$ state, shown in Fig. 4.14(b), only moderately grows over the interval $0.1 \leq \lambda \leq 3$ from 1.020(1) to 1.563(6). This repetition overhead translates to the average MS gate overhead given in Fig. 4.14(d), which grows from approximately 11 to 17.2 MS gates. For the deterministic protocol the increase of average MS gate overhead is smaller, from approximately 11 to 12.5 MS gates since, in case of a triggered flagged, the protocol only requires 4 more MS gates for the stabilizer measurement instead of repeating the whole preparation circuit consisting of 11 MS gates as for the non-deterministic protocol. Given that an experiment permits the moderately larger number of MS gates to be performed on average, we conclude that using the non-deterministic state preparation protocol is advantageous since the gain from suppressing the logical failure rate outweighs the necessary repetition overhead in the examined range of λ.

Magic state

The deterministic procedure to fault-tolerantly prepare $|H\rangle_L$ has originally been given in [CC19] and is drawn with logical building blocks in Fig. 4.13(b). We provide a compilation to MS gates for usage in the considered ion trap architecture and analyze how it fares compared to the non-deterministic scheme with regards to logical failure rate and repetition or MS gate overhead.

Analogously to the non-deterministic scheme, the deterministic procedure begins with preparing $|H\rangle_L$ non-fault-tolerantly. Next, the logical Hadamard operator is measured fault-tolerantly (M_H) using a measurement auxiliary qubit and four flag qubits f_0, f_1, f_2 and f_3 with an in-sequence feedback block F that is shown in Fig. 4.13(c). The detailed MS gate compilation of the M_H block is given in Fig. 4.13(d) and is obtained from the CNOT version in Fig. 4.15. Subsequently a full round of FT EC is performed. This means that, first, we run the error detection block (K_{FT}, see Fig. 4.16) and in case of a non-trivial measurement outcome we also run an additional syndrome readout block (K_{nFT}, see Fig. 4.17) to infer an appropriate recovery operation from the syndrome. These two steps, M_H (eventually followed by F) and

EC, must be repeated three times where in the last repetition the EC block can be omitted in case the last Hadamard measurement block yields no non-trivial measurement outcomes. The last step is conditional application of Y_L in case the three logical Hadamard measurement outcomes are either $+--$ or $---$.

The CNOT circuit of the logical Hadamard measurement in Fig. 4.15 illustrates with an example that the operation $F = H_1H_3H_4$ must be applied immediately after M_H in order to make sure that all possible errors are distinguishable if the flag pattern, i.e. the flag qubit measurement outcome of f_0, f_2, f_3, is $-+-, --+$ or $---$. To see this, note that a Hadamard error on the data qubits can follow from, e.g., a single X-fault on the measurement qubit. A Hadamard error is a coherent superposition of Pauli errors since, for example,

$$H_1H_3 = \frac{1}{2}(X_1X_3 + X_1Z_3 + Z_1X_3 + Z_1Z_3). \tag{4.73}$$

The fault X_2X_m after the fourth controlled-H gate in Fig. 4.15 for instance leads to the error $E_1 = X_2H_1H_3H_4$ and causes the same flag pattern $-+--$ as the fault X_m just

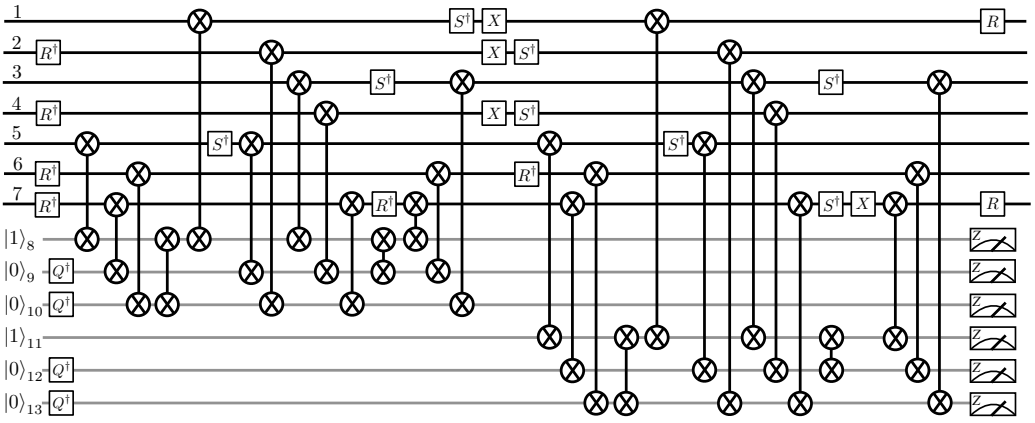


FIGURE 4.16: Parallel stabilizer readout K_{FT} by interleaved measurement of all six stabilizers. Auxiliary qubits simultaneously act as measurement and flag qubits, as used for the deterministic FT magic state preparation protocol (see Fig. 4.13). The circuit is analogous to Fig. 4.10(b) but only X-type MS gates are used.

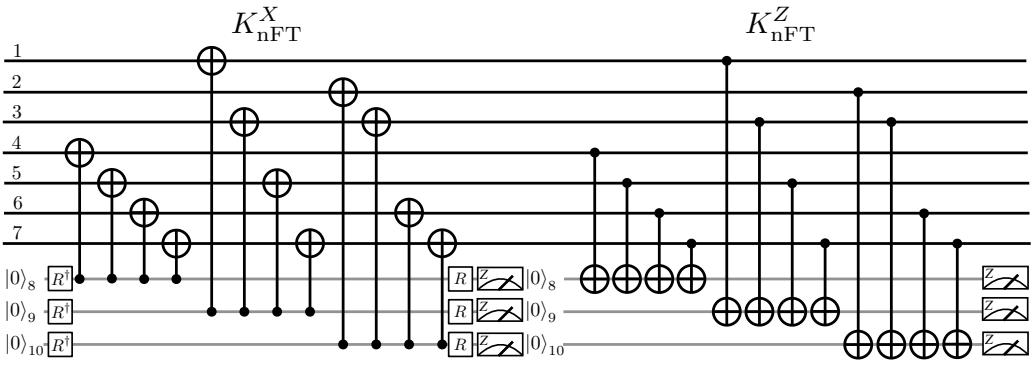


FIGURE 4.17: Sequential stabilizer readout K_{nFT}. The circuit is used for non-FT stabilizer readout of the six-bit syndrome as part of the deterministic FT magic state preparation protocol shown in Fig. 4.13. The first half reads out the X-stabilizers (K^X_{nFT}) and the second half reads out the Z-stabilizers (K^Z_{nFT}). Each CNOT gate is directly compiled into the sequence of MS gates and local rotations as given in Eq. (4.6). The circuit is analogous to the one in Fig. 3.21.

before this gate, which leads to the error $E_2 = H_1H_2H_3H_4$. The error E_2 is stabilizer-equivalent to $E_2' = H_5H_6H_7$ here because $H^{\otimes 7}$ stabilizes $|H\rangle_L$. Now E_1 and E_2' can be collapsed to $X_1X_2X_3X_4$ and $X_5X_6X_7$ by the EC block. They differ by a logical operator and cannot be distinguished by the flag qubits and cause the same syndrome $+++\,-++$. Applying F however transforms the errors to $FE_1 = X_2$ and $FE_2 = H_2$, which can both be corrected. The two errors may also be collapsed to $X_1X_2X_3Z_4$ and $Z_1Z_2Z_3Z_4$ respectively and, again, yield the same syndrome $-++\,+++$ since they only differ by a logical Y-operator but applying F prevents this unintended behavior.

The concrete procedure to determine the recovery operation as illustrated as part of the flowchart diagram in Fig. 4.18 works as follows: If any flag f of M_H is triggered and the syndrome is not trivial, a recovery operation according to the Hadamard look up table 4.6 is applied. Note that the full six-bit syndrome is necessary to correctly identify these errors. If the Hadamard flags are clear or the syndrome is not in the Hadamard look up table, we check if the syndrome is consistent with an error from the flag error set $\mathrm{FES} = \{X_3X_7, X_4X_6, Z_3Z_7, Z_4Z_6\}$ if also any flag from K_{FT} has been triggered. Recall that the flag error set is formed by all errors that can possibly result from any (dangerous) fault in the K_{FT} block given any triggered flag. Weight-2 X- and Z-errors can be corrected independently in this case, informed by the respective Z- and X-syndromes measured by K_{nFT}. If the syndromes measured by K_{FT} and K_{nFT} agree, the recovery operation according to the standard look up table 2.3 can be applied. A detailed derivation of the scheme is provided in the Appendix of [CC19].

The difference in terms of the number of required entangling gates between the deterministic and the non-deterministic FT magic state preparation scheme is much more pronounced than for the Pauli state. Due to the much longer protocol, we also expect larger logical failure rates at a given scaling parameter λ. These are shown in Fig. 4.14(c). The deterministic scheme displays the expected quadratic scaling behavior only for $\lambda \lesssim 0.5$ and differs by approximately two orders of magnitude from the logical failure rate of the non-deterministic protocol in this regime. The repetition overhead of the non-deterministic protocol, which we show in Fig. 4.14(b), grows more rapidly with λ compared to Pauli state preparation. At $\lambda = 1$ ($\lambda = 3$, which would be worse than current rates) on average, approximately 5 (63) repetitions are necessary until the magic state is accepted. As plotted in Fig. 4.14(d), the average number of required MS gates for deterministic magic state preparation grows from approximately 113 MS gates at $\lambda = 0.1$ to approximately 198 MS gates at $\lambda = 3$ because flags are triggered and, as a consequence, the full EC sequence is realized more often, whereas the non-deterministic protocol only uses approximately 56 MS gates on average at $\lambda = 0.1$ but at $\lambda = 3$ approximately 3010 MS gates are required on average. For the experimental fault rates at $\lambda = 1$, the deterministic protocol needs approximately 182 MS gates on average until the state is accepted, which is only slightly less than the approximately 232 MS gates that would be run on average for the non-deterministic protocol.

The gain in logical fidelity that might come along with using a non-deterministic protocol must be traded-off against the potentially large repetition or MS gate overhead. This trade-off must be considered for any experimental realization of the two FT state preparation protocols that we presented and the particular quantum algorithm one aims to implement. The repetition overhead should be kept sufficiently small or one must be able to tolerate the potentially large repetition overhead in order to profit from the lower infidelities of the non-deterministic protocols. On the other hand, deterministic protocols offer the advantage that their run time has a strict upper bound,

which can be useful when multiple logical qubits are to be prepared and the higher infidelities can be tolerated.

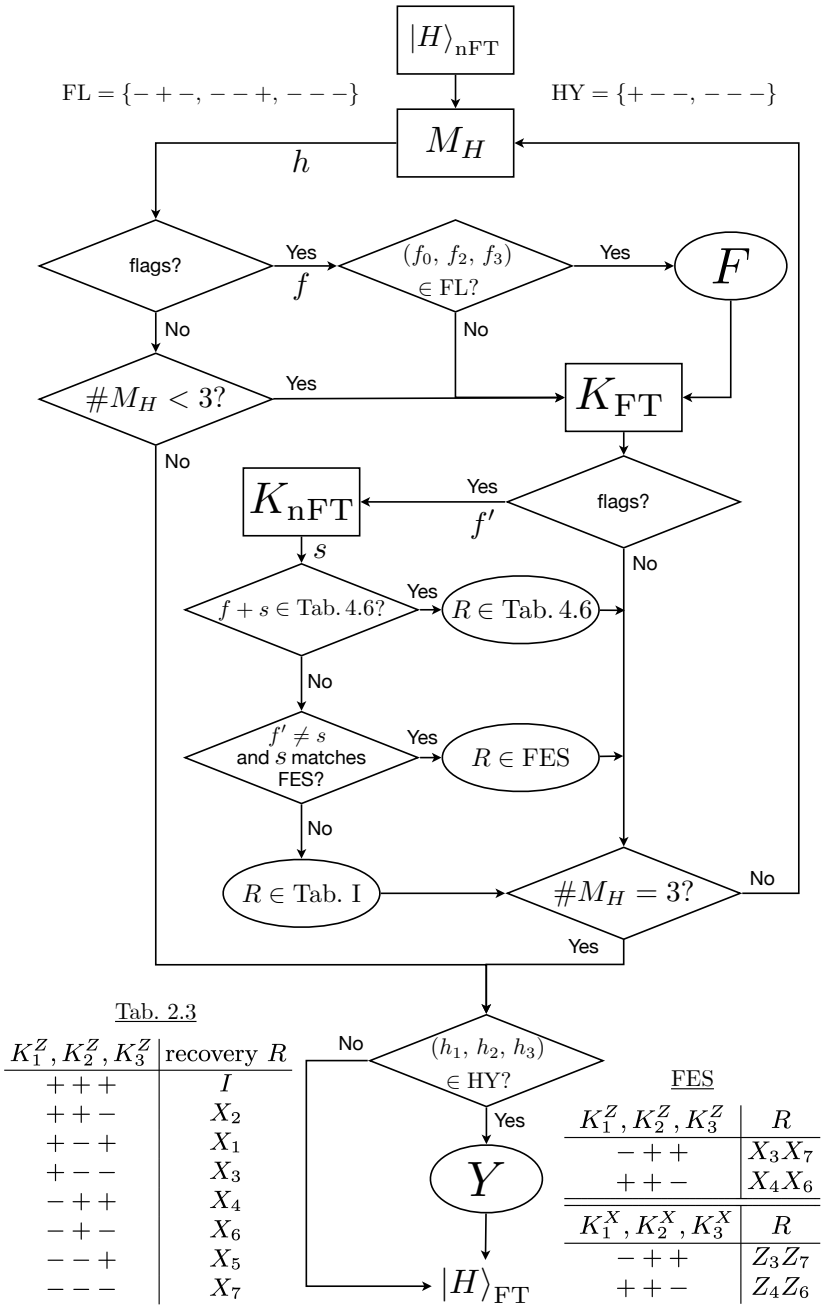


FIGURE 4.18: Flowchart for the deterministic FT magic state preparation procedure. In-sequence measurements determine the circuits of the noisy protocol chosen dynamically during run time (rectangles). The measurement outcome of an individual M_H circuit is labeled h. Measurements of flag circuits are shown as outputs f and f'. Denoted by s is the syndrome output by K_{nFT}. Additional corrections need to be applied (ovals) depending on the intermediate measurement results: R is drawn from the Steane look up table 2.3, the Hadamard error set (Tab. 4.6) or the flag error set (FES) and F and Y are given in Fig. 4.13. The sets of measurement results that cause application of F and Y are labeled FL and HY respectively.

f_0, f_1, f_2, f_3	K_1^X, K_2^X, K_3^X	K_1^Z, K_2^Z, K_3^Z	R
$--++$	$+++$	$++-$	X_2
$---+$	$+++$	$++-$	X_2
$-+-+$	$+++$	$++-$	X_2
$-+--$	$+++$	$++-$	X_2
$--++$	$++-$	$+++$	Z_2
$---+$	$++-$	$+++$	Z_2
$-+-+$	$++-$	$+++$	Z_2
$-+--$	$++-$	$+++$	Z_2
$-+++$	$+++$	$+--$	X_3
$-+++$	$+--$	$+++$	Z_3
$-+++$	$+++$	$---$	X_7
$-+++$	$---$	$+++$	Z_7
$-+++$	$+++$	$++-$	X_1X_3
$-+++$	$+--$	$+-+$	X_1Z_3
$-+++$	$+-+$	$+--$	X_3Z_1
$-+++$	$++-$	$+++$	Z_1Z_3
$--++$	$--+$	$-++$	X_4Z_5
$--++$	$-++$	$--+$	X_5Z_4
$-+++$	$+++$	$+-+$	X_6X_7
$--++$	$+++$	$+-+$	X_6X_7
$-+++$	$---$	$-+-$	X_6Z_7
$--++$	$---$	$-+-$	X_6Z_7
$-+++$	$-+-$	$---$	X_7Z_6
$--++$	$-+-$	$---$	X_7Z_6
$-+++$	$+-+$	$+++$	Z_6Z_7
$--++$	$+-+$	$+++$	Z_6Z_7
$-+++$	$+-+$	$++-$	$X_1X_3Z_1$
$-+++$	$++-$	$+-+$	$X_1Z_1Z_3$
$-+++$	$+++$	$-+-$	$X_1X_3X_4$
$-+--+$	$+++$	$-+-$	$X_1X_3X_4$
$-+++$	$+--$	$--+$	$X_1X_4Z_3$
$-+--+$	$+--$	$--+$	$X_1X_4Z_3$
$-+++$	$+-+$	$---$	$X_3X_4Z_1$
$-+--+$	$+-+$	$---$	$X_3X_4Z_1$
$-+++$	$++-$	$-++$	$X_4Z_1Z_3$
$-+--+$	$++-$	$-++$	$X_4Z_1Z_3$
$-+++$	$-++$	$++-$	$X_1X_3Z_4$
$-+--+$	$-++$	$++-$	$X_1X_3Z_4$
$-+++$	$---$	$+-+$	$X_1Z_3Z_4$
$-+--+$	$---$	$+-+$	$X_1Z_3Z_4$
$-+++$	$--+$	$+--$	$X_3Z_1Z_4$
$-+--+$	$--+$	$+--$	$X_3Z_1Z_4$
$-+++$	$-+-$	$+++$	$Z_1Z_3Z_4$
$-+--+$	$-+-$	$+++$	$Z_1Z_3Z_4$
$---+$	$+++$	$-++$	$X_5X_6X_7$
$-+--+$	$+++$	$-++$	$X_5X_6X_7$
$--++$	$+++$	$-++$	$X_5X_6X_7$
$---+$	$-+-$	$++-$	$X_5X_7Z_6$
$-+--+$	$-+-$	$++-$	$X_5X_7Z_6$
$--++$	$-+-$	$++-$	$X_5X_7Z_6$
$---+$	$--+$	$+-+$	$X_6X_7Z_5$
$-+--+$	$--+$	$+-+$	$X_6X_7Z_5$
$--++$	$--+$	$+-+$	$X_6X_7Z_5$
$---+$	$+--$	$---$	$X_7Z_5Z_6$
$-+--+$	$+--$	$---$	$X_7Z_5Z_6$
$--++$	$+--$	$---$	$X_7Z_5Z_6$
$---+$	$---$	$+--$	$X_5X_6Z_7$
$-+--+$	$---$	$+--$	$X_5X_6Z_7$
$--++$	$---$	$+--$	$X_5X_6Z_7$
$---+$	$+-+$	$--+$	$X_5Z_6Z_7$
$-+--+$	$+-+$	$--+$	$X_5Z_6Z_7$
$--++$	$+-+$	$--+$	$X_5Z_6Z_7$
$---+$	$++-$	$-+-$	$X_6Z_5Z_7$
$-+--+$	$++-$	$-+-$	$X_6Z_5Z_7$
$--++$	$++-$	$-+-$	$X_6Z_5Z_7$
$---+$	$-++$	$+++$	$Z_5Z_6Z_7$
$-+--+$	$-++$	$+++$	$Z_5Z_6Z_7$
$--++$	$-++$	$+++$	$Z_5Z_6Z_7$
$-+++$	$-+-$	$+-+$	$X_6X_7Z_6$
$-+++$	$+-+$	$-+-$	$X_6Z_6Z_7$
$--++$	$+--$	$++-$	$X_5X_7Z_5Z_6$
$--++$	$++-$	$+--$	$X_5X_6Z_5Z_7$
$-+++$	$-++$	$-+-$	$X_1X_3X_4Z_4$
$-+++$	$---$	$--+$	$X_1X_4Z_3Z_4$
$-+++$	$--+$	$---$	$X_3X_4Z_1Z_4$
$-+++$	$-+-$	$-++$	$X_4Z_1Z_3Z_4$

TABLE 4.6: Look up table for flag-FT measurement of the logical Hadamard operator in the deterministic scheme given in Fig. 4.13(b). The full six-bit syndrome needs to be considered in order to choose the appropriate recovery operation R in contrast to the situation where X- and Z-type recoveries are applied independently in standard EC on the Steane code.

4.1.3.3 Multiple logical qubits

When multiple logical qubits are prepared for use in a quantum algorithm on the logical level, e.g. magic state distillation, scheduling aspects may become important. With L deterministic state preparation protocols that are run in parallel, all L logical states will be prepared after constant time. When using a non-deterministic state preparation protocol for all L logical qubits, the overall preparation time is limited by the one logical qubit, which takes the most repetitions until the state is accepted. Given an average number of repetitions $a(\lambda)$, which depends on the noise strength parametrized by λ, it is advantageous, on average, to use a non-deterministic protocol with run time t_n as long as

$$a(\lambda)t_n < t_d \tag{4.74}$$

where t_d is the run time of the corresponding deterministic protocol. In case that state preparation cannot be performed in parallel or, more generally, when successful state preparations do not coincide in time, logical qubits that are already accepted can wait until all states are accepted. Before feeding the states into a subsequent logical building block one may perform an additional round of QEC on the logical qubits.

Assuming that we are capable of preparing L logical qubits but only $k \leq L$ are required for a quantum algorithm of interest, we may ask whether to use the non-deterministic or the deterministic FT state preparation protocol. For a non-deterministic protocol with flag rate f and average number of trials $a = \frac{1}{1-f}$ we can expect that, on average, $k = L(1 - f^a)$ out of L logical qubits are accepted or, equivalently, that we need to run the protocol

$$a = \frac{\log(1 - k/L)}{\log(f)} \tag{4.75}$$

times on average until k out of L logical qubits are accepted. Suppose that all logical qubit preparation blocks are independent[7]. Then we can use a cumulative binomial distribution with success probability $1 - f^a$ to express the probability $P_{\geq k}$ that *at least* k out of L logical qubits are accepted after a runs as

$$P_{\geq k} = \sum_{j=k}^{L} \binom{L}{j} (1 - f^a)^j (f^a)^{L-j} = 1 - \sum_{j=0}^{k-1} \binom{L}{j} (1 - f^a)^j (f^a)^{L-j}. \tag{4.76}$$

For a fixed probability of success $P_{\geq k}$, say 95%, we can numerically evaluate Eq. (4.76) to extract the number of circuit runs a that is needed in order to prepare k out of L logical qubits, which is done for FT Pauli and magic states in Fig. 4.19. Here we highlight regions of advantage for fixed numbers k and L for either protocol variant. As long as at most $a^* = \lfloor t_d/t_n \rfloor$ preparation trials are needed, we consider using the respective non-deterministic protocol advantageous over using the deterministic scheme. If the non-deterministic scheme takes more than a^* attempts, there is no saving in the number of entangling gates over the deterministic variant. We refer to the number of entangling gates as a proxy of circuit run time since MS gates take the longest time in the experimental system, so that, for Pauli (magic) states we have $t_d/t_n = 15/11 \approx 1.4$ ($t_d(\lambda = 1)/t_n = 232/48 \approx 4.8$ where we use the average number of MS gates at $\lambda = 1$). The comparison for Pauli state preparation with a flag rate of $f = 0.17$ is shown in Fig. 4.19(a). For instance, it is sensible to use the

[7] This assumption may be violated in case one encodes $k > 1$ logical qubits in a single $[[n, k, d]]$ QEC code, e.g. a high-rate qLDPC code.

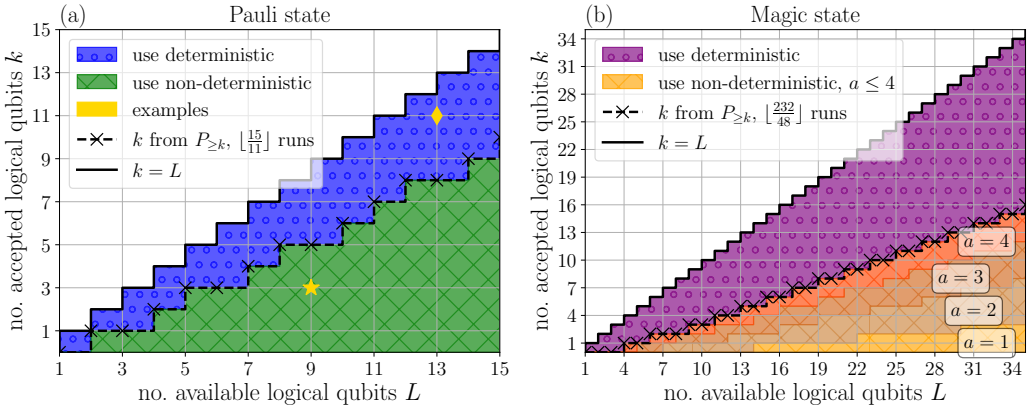


FIGURE 4.19: Non-deterministic FT logical state preparation schemes (crosses) can be advantageous in terms of run time over their deterministic variants (small circles). We deem this the case if the expected number of entangling gates required to accept k out of L logical qubits with probability $P_{\geq k} \geq 95\%$ according to Eq. (4.76) is lower than for the deterministic schemes. These always prepare $k = L$ logical qubits (solid black lines). Boundaries between regions of advantage are marked as dashed lines with crosses. (a) For the Pauli state, running the non-deterministic scheme twice already needs more MS gates than the deterministic version. Examples are marked in yellow (see main text). (b) For the magic state, despite the large flag rate of $f = 0.8$, the large gate overhead of the deterministic scheme permits us to run up to 4 rounds of the non-deterministic variant. Regions of k after 1 through 4 trails are marked in shades of orange.

non-deterministic protocol if at least 3 out of 9 logical qubits should be accepted with a probability of at least 95% but the deterministic protocol should be used if 11 out of 13 accepted states are required. Figure 4.19(b) shows the regions of advantage for the expected number k of accepted logical magic states for up to four runs. The deterministic schemes always prepare $k = L$ logical qubits. Note, however, that a comparison of failure rates may also be indicative for choosing between a deterministic and a non-deterministic scheme.

4.1.3.4 Crosstalk-resistant state preparation

As a last part of this study, we again turn toward the role of crosstalk in the non-deterministic FT Pauli state preparation circuit from Fig. 4.5. We remind the reader that crosstalk on entangling gates, in general, breaks fault tolerance. It induces a correlated weight-2 fault that does not respect the FT circuit design principle, since neighboring qubits, which are not acted upon by the ideal entangling gate, are involved. As a consequence, an otherwise FT circuit will, in general, display a logical failure rate $p_L \sim p$ as $p \to 0$ in the presence of such crosstalk as we saw in Fig. 4.12 for the extended noise model. However, it is essential to preserve the quadratic scaling behavior of the logical failure rate in order to ensure that the break-even point of advantage of the logical qubit over the physical qubit can be reached in practice. If the effect of crosstalk is too strong, its linear scaling may prevent us from ever outperforming the physical qubit no matter how small other fault rates are.

Let us now show that it is actually possible to make the Pauli preparation circuit also resistant against X-type crosstalk, given by Eq. (4.45). As a starting point, we notice that the single-qubit rotations that are introduced by decomposition of the CNOT

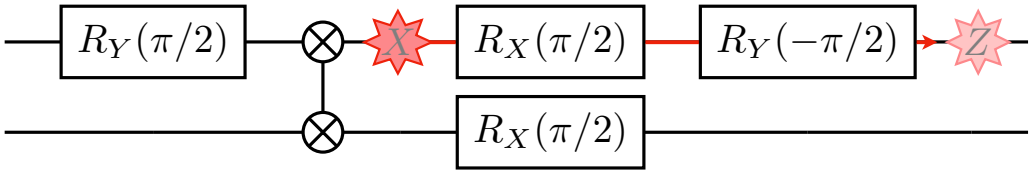


FIGURE 4.20: Decomposition of a CNOT gate into an MS gate $\mathrm{MS}_{0,0}(-\pi/2)$ and local rotations according to Eq. (4.6). The upper qubit acts as the control qubit and the lower qubit is the target qubit. An X-fault on the first qubit directly after the MS gate, which occurs as part of crosstalk faults (see Fig. 4.3), is transformed into a Z-fault by the Y-rotation. X-faults on the target qubit or on neighboring qubits could only be rotated if other Y-rotations are performed on them as part of subsequent CNOTs.

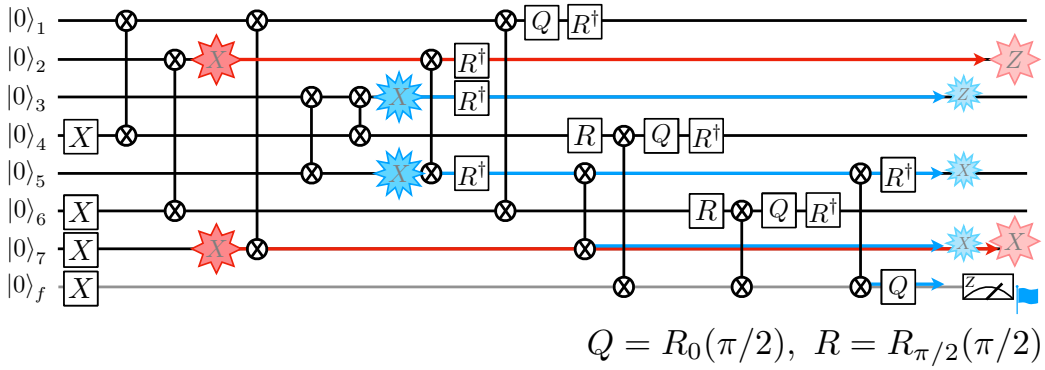


FIGURE 4.21: Crosstalk-resistant circuit for FT Pauli state preparation. All X-type crosstalk faults are either transformed into correctable errors by the circuit (such as X_2X_7 after the second MS gate, red 8-cornered stars) or trigger the flag (such as X_3X_5 after the fifth MS gate, blue 12-cornered stars).

gate into MS gates according to Eq. (4.6), rotate X-faults on a control qubit to Z-faults, as illustrated in Fig. 4.20. Since the Steane code can simultaneously correct both a single X- and a single Z-error, some of the crosstalk faults may propagate to correctable errors. As an example, the X_2X_7 fault after the second MS gate in Fig. 4.21 propagates to Z_2X_7, which can be corrected and does not lead to failure. By an exhaustive numerical check of all possibilities to label the physical qubits, we find that the circuit in Fig. 4.21, where stabilizers now have support on qubits (1,4,6,7), (2,5,6,7) and (3,4,5,7), is capable of neutralizing all crosstalk faults and is thus deemed "crosstalk-resistant" (CTR). Of course, not all errors resulting from crosstalk faults have to be correctable. It is sufficient that they, if uncorrectable, trigger the flag. Consider the fault X_3X_5 after the fifth MS gate as an example. While this fault does not trigger the flag in the non-CTR circuit in Fig. 4.5, it also propagates to the flag in our CTR circuit in Fig. 4.21.

In a comparative code study with realistic noise modeling for ion traps, a graph-based method to choose optimal qubit mappings has been explored to achieve robustness against crosstalk [Deb+20]. For the magic state preparation circuit, we were unable to find a CTR circuit. It is unclear if such circuits can be found in general or if, in longer circuits with many entangling gates, there are just too many crosstalk locations and not enough possible qubit mappings that are meaningfully distinct. For example, for the transversal measurement of the Hadamard operator, rearranging the data qubits cannot be expected to alter the effect of crosstalk faults.

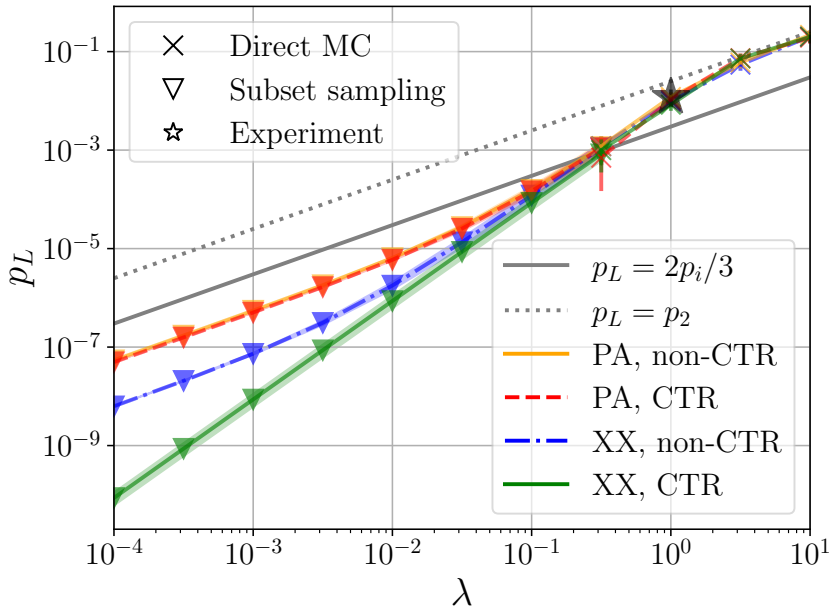


FIGURE 4.22: Simulated logical failure rates of FT Pauli state preparation variants with a uniform scaling factor λ for all physical fault rates. The crosstalk-resistant (CTR) and non-CTR circuits are investigated with both X-type crosstalk (XX) and phase-averaged (PA) crosstalk. Direct MC (crosses) and subset sampling (triangles) are employed in their preferential domain of physical fault rates. The experimentally measured value is at $\lambda = 1$ (star). All curves coincide in this regime. In the low-λ limit, crosstalk causes linear scaling behavior except for the CTR circuit with X-type crosstalk, which keeps scaling quadratically. Failure rates of physical operations are shown for comparison as in Fig. 4.12. The MC data points and subset failure rates are obtained using 10^2 to 10^4 samples and all subsets up to $w_{\max} = 3$ are taken into account for subset sampling.

We numerically demonstrate the CTR property of the Pauli encoding circuit from Fig. 4.21 with a scaling analysis of logical failure rates analogous to the previous non-deterministic state preparation study. It is apparent from Fig. 4.22, that the quadratic scaling behavior is maintained for all values of the scaling parameter λ under the extended noise model with X-type crosstalk and the experimental parameters from Tab. 4.4. Using the same noise model with the non-CTR circuit from Fig. 4.5, the logical failure rate displays linear scaling in the low-λ limit where crosstalk dominates other sources of noise. However, when phase-averaged (PA) crosstalk is used (see Eq. (4.42)) there is almost no difference between the simulated logical failure rates of the two circuits and both scale linearly as $\lambda \to 0$. In this case, since XX, XY, YX and YY faults are applied, we did not find a way to make sure that no crosstalk fault propagates to an uncorrectable error without triggering the flag.

Recall that crosstalk in the considered ion trap quantum processor is not only of X-type [Heu+23a]. As such, we conclude that minimizing crosstalk strength in physical operations is imperative in order to uphold the advantageous quadratic scaling of FT circuit design for as low physical fault rates as possible. We stress that, as previously, in a large interval of approximately one order of magnitude around the experimental values at $\lambda = 1$ all curves in Fig. 4.22 predict the same logical failure rates so crosstalk on entangling gates is not a major source of failure in this particular setup at current noise levels for the shown non-FT and FT state preparation applications. Crosstalk could be avoided or suppressed by using physical gate operations that are inherently

insensitive to crosstalk. This could be achieved by composite pulses or active suppression schemes that make use of additional laser fields to destructively interfere with the unwanted leakage light at neighbor ion positions [Fan+22; TV23].

4.2 Measurement-free fault-tolerant QEC

Experimental implementations of fault-tolerant QEC have been achieved in both ion traps [RA+21] and superconducting transmon architectures [Kri+22; Zha+22; Goo23b] recently. As already mentioned in Chapter 1, for most quantum computing platforms, it poses a large obstacle to perform mid-circuit measurements and condition in-sequence quantum feedback operations on the classical measurement information. This ability is, however, required by state-of-the-art FT QEC schemes to put such devices into practical use for performing quantum algorithms at the logical level where they are protected against noise, as we saw in the previous sections. In-sequence measurements and feed-forward are especially hard in ion traps and neutral-atom architectures where readout of physical qubits is slow [Xu+21; Pog+21] and laser cooling must be applied during or after each detection, which is also much slower (on the order of milliseconds at least [Sin+22; Lis+23; Pin+21; RA+21]) than actual gate operations (that can be performed in a time on the order of microseconds in ion traps [Bru+19] or even less with neutral atoms [Win+23]). There exist, however, techniques to avoid the need of slow recooling, such as using multiple species of ions/atoms [Bli+02; Bar+03; Sin+22] and shuttling ions/atoms into dedicated readout zones so that heating of the whole setup is prevented [KMW02; Chi+05; Hil+22; Mos+23; Blu+22]. Although mid-circuit measurements [Gra+23; Nor+23; Lis+23] and real-time feedback [Hui+23; Sin+23] were recently demonstrated in neutral-atom platforms, large scale experiments with in-sequence logic might require additional techniques. For instance, cavity-enhanced fluorescence imaging may be faster and lead to less heating because fewer photons need to be scattered [Boc+10; Dei+22].

In this section, we present a novel measurement-free fault-tolerant quantum error correction scheme [HLM23], inspired by Steane-type EC, that can be used for any distance-3 CSS code. This makes the scheme especially suitable for state-of-the-art and near-term devices. It is fully FT toward circuit-level depolarizing noise on all circuit locations and thus enables practical FT QEC in hardware architectures that do not support mid-circuit measurements. We illustrate our scheme using the Steane code but note that it could equally well be applied to, for instance, a distance-3 surface code[8].

Although measurements are not fundamentally needed for fault tolerance, some dissipative element is always required for QEC in order to remove entropy from the system [Sch+11; Bar+11; Sch+13a]. This could for instance be a supply of fresh auxiliary qubits from a sufficiently large reservoir or the ability to perform qubit reset. Since measurements are persistently challenging to perform in practice, a vast amount of work has been dedicated to finding measurement-free circuit designs for, but not only for, FT QEC. Non-FT measurement-free schemes include a surface code implementation [Erc+18] but also autonomous error correction protocols that use engineered dissipation [HMM22] or methods using bosonic QEC codes [TCV20; Ger+21] can be classified as such schemes. Measurement-free FT gate implementations for the T-gate and the Toffoli gate have been suggested [Boy+10]. The authors of [PSBT10]

[8] A discussion of applying the scheme to the distance-3 surface code, where not only all weight-1 but also some weight-2 errors can be corrected [TS14], is provided in [HLM23].

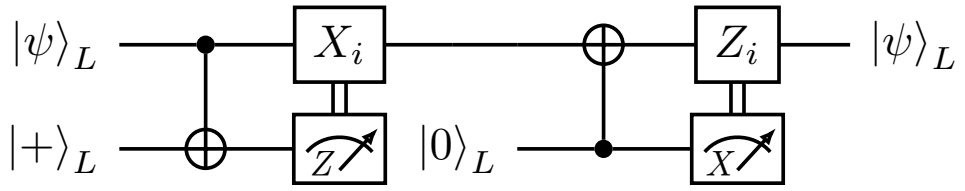


FIGURE 4.23: Illustration of Steane-type EC on the logical level. In two rounds, one for X- and one for Z-corrections, logical auxiliary qubits are coupled to a logical data qubit $|\psi\rangle_L$ via transversal CNOT gates and measured transversally in the opposite Pauli bases. Pauli corrections X_i and Z_i on individual physical qubits are applied according to the measured syndromes, which are obtained by classical post-processing of the measurement bitstrings.

devise an FT EC scheme for the 9-qubit Bacon-Shor code, which is extendable to general CSS codes, and give a competitive threshold by means of concatenation. It has later been established that measurement-free EC does not need to be substantially inferior to conventional, measurement-based QEC in principle [CJ17a; CJ17b]. With some restrictive additional assumptions about the form of the noise, measurement-free schemes that tolerate a subset of all possible faults have been formulated [CJS16; Per+23].

In the following, we first review Steane-type EC in Sec. 4.2.1. Our scheme requires logical auxiliary qubits, so we provide a deterministic measurement-free FT encoding circuit for the Steane code in Sec. 4.2.2. Next, we lay out the actual QEC scheme in Sec. 4.2.3 and discuss practical implementations as well as possible advantageous modes of operation over conventional syndrome-measurement-based QEC in Sec. 4.2.4. Lastly, we discuss scaling up the scheme to larger-distance codes in Sec. 4.2.5.

4.2.1 Steane-type error correction

Steane-type error correction [Ste97] is one of the earliest proposals of fault-tolerant QEC, alongside with Shor-type EC discussed in Sec. 2.4.3. The scheme makes use of two logical qubits that are encoded in the same QEC code, as illustrated in Fig. 4.23, in order to correct one type of Pauli errors each in two subsequent blocks.

Let the first logical qubit be in an arbitrary logical state $|\psi\rangle_L$ formed by n physical qubits. The second logical qubit is prepared in the logical $|+\rangle_L$ state and connected to the target of a subsequent transversal CNOT gate. It is used to correct X-errors on $|\psi\rangle_L$, to which the control of the transversal CNOT is connected. This way, there will be no backaction on $|\psi\rangle_L$ in the fault-free case since $\mathrm{C_1NOT_2}\,|\psi\rangle \otimes |+\rangle = |\psi\rangle \otimes |+\rangle$. However, X-errors can propagate through the logical, transversal CNOT from $|\psi\rangle_L$ to $|+\rangle_L$. Then, all physical qubits of $|+\rangle_L$ are measured in the Z-basis. The first half of Fig. 4.23 illustrates this. The Z-measurements of the second logical qubit reveal no information about $|\psi\rangle_L$ but just collapse $|+\rangle_L$ to any of the constituent states of either $|0\rangle_L$ or $|1\rangle_L$[9]. Correctable X-errors can be uniquely identified by decoding the classical measurement bitstring. Subsequently, the X-errors are corrected by classically-controlled feedback operations. The procedure can be performed analogously to correct Z-errors with all X- and Z-type states and operations interchanged

[9]Note that if the logical auxiliary qubit were instead prepared in $|0\rangle_L$, this procedure would measure a logical operator $Z_L = Z^{\otimes n}$ on the first logical qubit and thus collapse a general superposition of the form $|\psi\rangle_L = \alpha|0\rangle_L + \beta|1\rangle_L$.

and reversing the control and target of the logical CNOT gate as the second half of Fig. 4.23 shows.

As for Shor-type EC, the correct preparation of logical auxiliary qubits must be verified [AGP06; CDT09; PR13; CR18a]. The order, in which X- and Z-type errors are corrected, can potentially play a role for the quantitative performance of the scheme if the predominant noise is asymmetric. For example, Z-faults that may originate from dephasing of the auxiliary state $|+\rangle_L$ will propagate back to the logical data qubit during an X-correction block. Due to the transversal CNOT gates, it is clear that Steane-type EC can only be applied to CSS codes, opposing Shor-type (and Knill-type) EC that work for any stabilizer code[10].

To perform a full cycle of Steane-type EC on the Steane code, only 14 physical CNOT gates are needed for the transversal CNOTs and an additional 22 physical CNOT gates are required for flag-FT preparation of the logical auxiliary qubits. This is less than the 52 physical CNOT gates necessary for the flag-FT error-correction procedure described in Sec. 4.1.3.2. For Steane-type EC, the two parallel qubit measurements per cycle can be deferred to the end of the circuit if enough physical qubits are practically available.

4.2.2 Deterministic encoding circuit without measurements

Logical auxiliary qubits that are required for Steane-type EC can be prepared fault-tolerantly using any technique most suited for a particular hardware platform under consideration.

The most standard way to initialize a logical qubit is to prepare all n physical qubits in the state $|0\rangle^{\otimes n}$, then measure all stabilizers to obtain a random syndrome and apply Pauli corrections based on the syndrome in order to bring the logical qubit quantum state into the $+1$ eigenspace of all stabilizers, i.e. the code space (cf. Sec. 3.5.2.2). This method is only suitable if measurements can be performed fast and with low infidelities. In superconducting transmon qubits this method can be employed since entangling gates can typically not be executed between arbitrary pairs of physical qubits. In contrast, we previously used the more resource-friendly flag-FT initialization circuit by [Got16a] (see Fig. 4.4(a)) owing to all-to-all qubit connectivity offered by the considered ion trap setup.

As we discussed in Sec. 4.1.3.2, the flag-FT initialization circuit for the Steane code can be made deterministic by adding one extra auxiliary qubit and four additional CNOT gates. Recall that, by the flag look up table 4.5, only a single recovery operation X_7 conditioned on both auxiliary qubits being in the $|1\rangle$ state is necessary in order to render the FT state preparation circuit deterministic. This feedback operation does not require measurement and classical processing of the measurement information. Due to its simplicity, the recovery can be applied coherently via a single Toffoli gate as depicted in Fig. 4.24. The two auxiliary qubits are the control qubits and a single data qubit is targeted in order to apply the conditional X_7 recovery. The circuit in Fig. 4.24 uses this circumstance to enable FT preparation of $|0\rangle_L$ in the Steane code without measurements[11]. Note that any high-weight error that could result from the

[10] While two transversal CNOT gates are necessary in Steane-type EC to correct X- and Z-errors separately, Knill-EC requires only a single logical CNOT gate but two logical auxiliary qubits are needed [AGP06].

[11] A circuit to initialize a logical qubit in the distance-3 surface code fault-tolerantly and without measurements has been given by [GHK23].

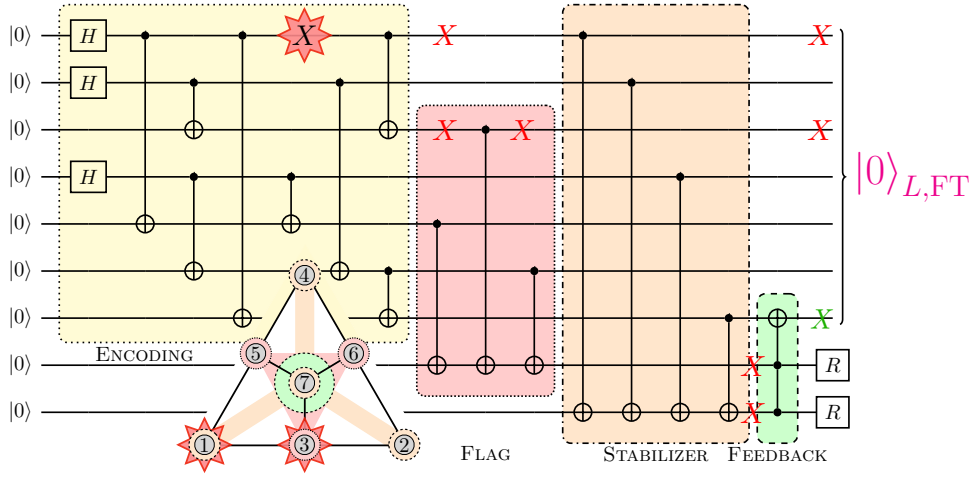


FIGURE 4.24: Circuit to fault-tolerantly prepare $|0\rangle_L$ in the Steane code (inset) without measurements. The encoding and flag blocks have been given by [Got16a]; stabilizer and feedback blocks are our extension. Expectation values of $Z_3Z_5Z_6$ ("flag", red dotted circles on code graph) and $Z_1Z_2Z_4Z_7$ ("stabilizer", orange dashed circles on code graph) are coherently mapped to two auxiliary qubits. The single X_1 fault (red star) that was caught by the flag in Fig. 4.4(a) now flips both auxiliary qubits and thus triggers the coherent X_7 feedback operation (green box) facilitated by a Toffoli gate (large green circle on code graph). The resulting operator $X_1X_3X_7$ is stabilizer-equivalent to X_5, a correctable error. The two auxiliary qubits can be reset (R) or discarded in the end.

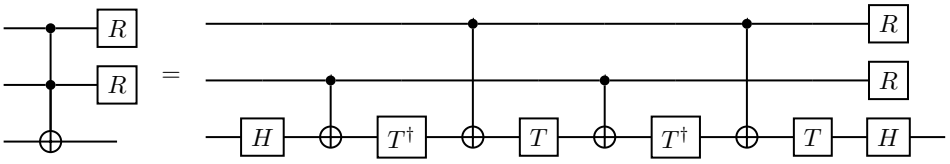


FIGURE 4.25: Simplified decomposition of the Toffoli gate followed by reset of the two control qubits. Four CNOT gates as well as Hadamard gates H, reset R and the standard Z-type T- and $T^\dagger$-gates are required.

Toffoli gate can at most cause a weight-1 error on the logical qubit and the auxiliary qubits are discarded anyway. The auxiliary qubit state before the Toffoli gate can only be different from $|0\rangle \otimes |0\rangle$ if a fault has happened already. In this case, an additional fault during the Toffoli gate that could potentially lead to a wrong correction and thus cause logical failure would be an overall fault configuration of order 2 so that fault tolerance is respected.

Physically, the Toffoli gate can be executed as a native multi-qubit gate or decomposed into CNOT gates and single-qubit gates. The former does not break fault tolerance since no fault from the 3-qubit Pauli group (see Eq. (3.13)) can lead to logical failure. For the latter, we can modify the well-known decomposition into 6 CNOT gates from [NC10] because the final state of the auxiliary qubits is irrelevant for this application. In fact, only 4 CNOT gates are required if the Toffoli is followed by reset of the control qubits as shown in Fig. 4.25. Let us remark that the reduced Toffoli gate decompositions by [Bar+95; Mas16] are not feasible for use in our scheme. Those gate decompositions perform the Toffoli operation up to a relative phase between certain computational basis states, which would lead to erroneous phase flips when the gate is applied, even without noise, to a superposition state such as $|0\rangle_L$.

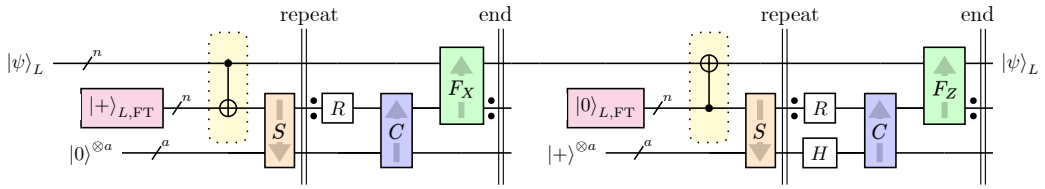


FIGURE 4.26: Measurement-free fault-tolerant quantum error correction cycle built from logical blocks. After FT initialization of logical auxiliary qubits, a logical CNOT gate propagates errors that are mapped (S) to a third register of a physical qubits. The syndrome is copied (C) once for each syndrome-correction pair after resetting the physical qubits (R). The repetition refers to all $2^a - 1$ possible non-trivial syndromes, for which the corresponding feedback operation F_σ of Pauli type $\sigma \in \{X, Z\}$ is applied.

4.2.3 Quantum error correction unit

With fault-tolerantly prepared logical auxiliary qubits available, we are now in a position to describe our measurement-free (MF) scheme on the logical level. The need for measurements in Steane-type EC can be circumvented with sufficient qubit and gate overhead, which is small enough to allow for implementation of the scheme in state-of-the-art or near-term devices, as we demonstrate in the following. The scheme can be applied to any distance-3 CSS code since it only requires that a transversal CNOT gate is available.

4.2.3.1 Logical level

Completely analogous to Steane-type EC, the MFEC scheme is subdivided into two half-cycles; one for X-corrections and one for Z-corrections. Each block begins with coherently mapping errors from the logical data qubit $|\psi\rangle_L$ to the logical auxiliary qubit with the help of a transversal, logical CNOT gate, which is depicted with logical blocks in Fig. 4.26. Since the target (control) of the logical CNOT is connected to the logical auxiliary qubit in the $|+\rangle_L$ ($|0\rangle_L$) state when correcting $X(Z)$-errors, there is no backaction on the logical data qubit state $|\psi\rangle_L$ in the fault-free case, i.e. the $Z_L(X_L)$ operator is not measured. Then, instead of performing bitwise measurements as in Steane-type EC, the $Z(X)$-syndrome of the logical auxiliary qubit is coherently mapped (block S in Fig. 4.26) to an additional register of a physical auxiliary qubits, each in the $|0\rangle$ ($|+\rangle$) state, and afterwards the logical auxiliary qubit is reset (R). Now, the syndrome state can be coherently copied (C) back to the second qubit register from where a feedback operation F applies recovery operations appropriate for the syndrome. The syndrome state must be coherently copied from the third register as many times as needed in order to hard-code all possible feedback operations, i.e. a suitable look up table. Note that, for the Z-correction block, a layer of a physical Hadamard gates H is required to transform the syndrome information into the computational basis.

4.2.3.2 Application to the Steane code

The Steane code serves as an illustrative example for the application of our measurement-free FT QEC scheme, which we show on the physical qubit level in Fig. 4.27. The logical CNOT gate is realized by $n = 7$ physical CNOT gates. Since the X- and Z-type corrections in the look up table 2.3 for the Steane code are each based on a three-bit syndrome, $a = \frac{n-k}{2} = 3$ physical auxiliary qubits are required to store

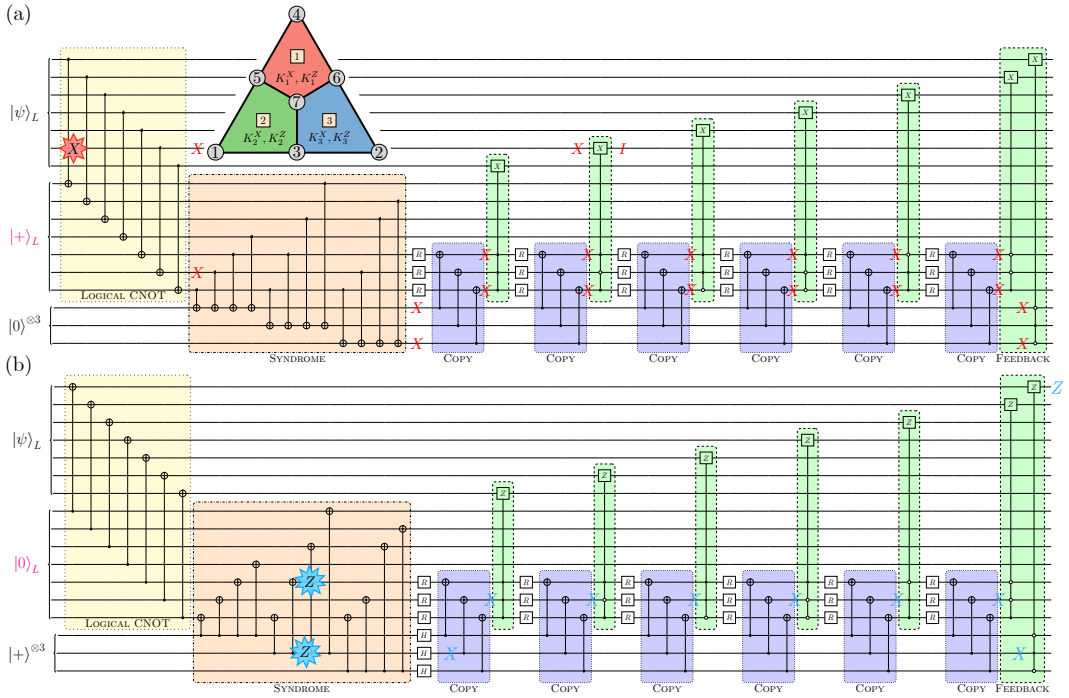


FIGURE 4.27: Circuit on the level of physical operations for MF FT QEC in the Steane code. Logical auxiliary qubit initialization can be done using any suitable protocol. No distinctly FT routine is required for the syndrome mapping step (orange) since potentially dangerous faults are erased by the resets (R) anyway. White (black) circles on the control qubits of the feedback gates condition application of the target operation on the $|0\rangle$ ($|1\rangle$) state. For the last feedback operation in each half-cycle, copying from the third to the second register can be omitted. (a) X-correction block. As an example, the error X_6 propagates to the logical auxiliary qubit and then flips the syndrome qubits to the state $|101\rangle$. The state is copied to all feedback gates but it only matches the control structure of the second C_3NOT gate so that it applies the recovery operation X_6 and corrects the error. The other feedback gates have no effect on the logical data qubit. (b) Z-correction block. The single fault $Z_{12}Z_{16}$ (blue stars), for instance, only propagates to a correctable weight-1 error Z_1 through the last C_3Z gate.

the syndrome, which is mapped from the logical auxiliary qubit[12]. Also, 3 physical CNOT gates are consequently used to transversally implement each copy step. Each feedback operation is performed by a C_3NOT or C_3Z gate and their control patterns encode the look up table 2.3 of the Steane code. The feedback operations can be considered as quasi-transversal in the sense that any syndrome is uniquely connected to a single physical data qubit in the first register and there is no connection between the individual syndromes due to the intermediate resets.

As an example, propagation of the correctable error X_6 is shown in Fig. 4.27(a). It propagates to the logical auxiliary qubit and then causes the physical auxiliary qubits to flip from $|000\rangle$ to $|101\rangle$. After the syndrome mapping, 3 physical qubits of the second register are each reset to the state $|0\rangle$ and the syndrome state is repeatedly copied to these 3 qubits by 3 physical CNOT gates. Only for the second feedback gate C_3NOT the 101-control structure matches the syndrome state so that consequently the operation X_6 is applied and the error X_6 is corrected. All other C_3NOT gates do not apply their target operation since their control structure is not matched by the

[12]In [HLM23], a variant of the scheme that decodes the logical auxiliary qubit and thus saves the additional a physical qubits is provided.

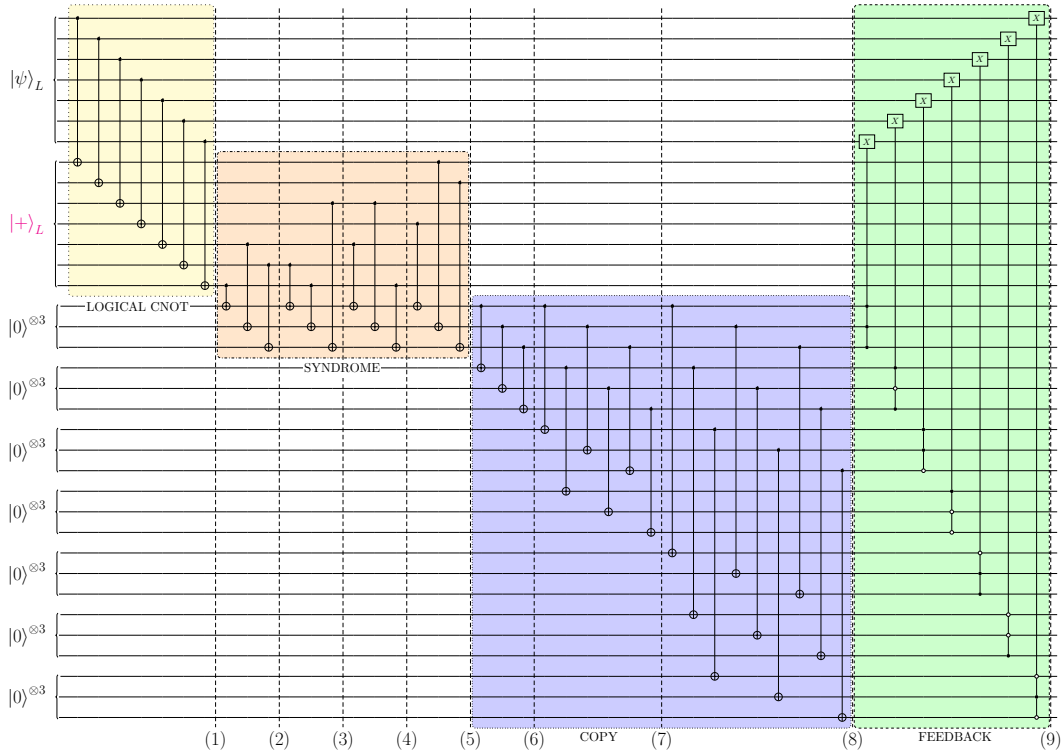


FIGURE 4.28: The repetition overhead of the MFEC scheme can be transformed into a qubit overhead. Correction of X-errors as in Fig. 4.27(a) can be scheduled in nine time steps if gates can be executed in parallel. Step 1 applies the logical CNOT. The syndrome mapping step is performed from time step 2 to 5. From time step 6 to 8, the syndrome is copied six times. The seven feedback operations can all be performed in parallel in the last time step. The Z-correction block can be parallelized analogously.

propagated error. A single fault in the second or third register can cause at most a correctable error of weight 1 on $|\psi\rangle_L$ because at most a single feedback operation can be erroneously applied since all are conditioned on different non-trivial syndromes. An example is given in Fig. 4.27(b). The input state $|\psi\rangle_L$ is assumed to be ideal in this case because the scheme is only FT toward a single fault. If instead $|\psi\rangle_L$ carries a weight-1 error before applying the MF QEC block already, the then-assumed fault-free scheme correctly maps the error to the according syndrome and applies the appropriate correction. A fault configuration of at least order 2 must happen to cause failure of the scheme.

Note that the ability to perform physical qubit resets and apply the copy and feedback steps repeatedly is not strictly required. It is also possible, as pointed out before and sketched for the X-correction block in Fig. 4.28, to supply a sufficient number of physical auxiliary qubits in the beginning and copy the syndrome state to $2^a - 1$ blocks directly. Then, each feedback operation can be conditioned on a different set of 3 control qubits. This way, the repetition overhead can be transformed into a qubit overhead, which might be more feasible for some experimental platforms. For instance, the 35 qubits used in Fig. 4.28 can be loaded into a neutral-atom tweezer array [Blu+22], while, in an ion trap, it could be more feasible to work with 17 qubits and perform intermediate resets.

4.2.3.3 Resources

Let us now discuss resource requirements for our scheme applied to any CSS code with parameters $[[n, k = 1, d = 3]]$. We require that $2n$ physical qubits are available and a qubits are used to store the syndrome of each Pauli type. The procedure for FT logical qubit initialization might demand additional physical qubits. In case the X- and Z-syndromes have the same length, as is the case for color codes and surface codes, the number of syndrome qubits is $a = \frac{n-k}{2}$. In total, $\mathcal{O}(n)$ physical qubits are needed for the implementation of the scheme as in Fig. 4.27. When we transform the repetition overhead into a qubit overhead, if, for instance, the cycle time would become too long to implement the repetitions or if no reset operation is available, the total number of qubits would increase to $2n + a(2^a - 1) = \mathcal{O}(2^{\frac{n}{2}})$, as in Fig. 4.28. We still operate two logical qubits but for each of the $2^a - 1$ feedback operations we employ a block of a physical qubits.

Suppose for simplicity that the X-correction block and the Z-correction block are symmetric and thus need the same number of CNOT gates. The logical CNOT gate in each block is realized by n physical CNOT gates. The syndrome mapping step consists of $a \times s$ physical CNOT gates, assuming that all stabilizers have the same Pauli weight s. Subsequently, a CNOTs are used to copy the syndrome coherently $2^a - 2$ times. To decompose an a-qubit-controlled feedback operation, it is known that $u = 2^{a+1} - 3$ two-qubit gates are sufficient [Bar+95] and we require $2^a - 1$ feedback operations. In total, at most $2 \times (n + a(s + 2^a - 2) + u(2^a - 1))$ CNOT gates are required to implement the QEC cycle.

For the Steane code, operating on $n = 7$ physical qubits with $a = 3$ stabilizers of weight $s = 4$, these requirements amount to a total of 256 physical CNOT gates. We show in Sec. 4.2.4.1 that the multi-qubit-controlled gate decomposition can be simplified due to the subsequent reset of the control qubits so that u is reduced from 13 to 8 and thus 186 CNOT gates are sufficient for the Steane code QEC cycle. For the MF and FT preparation of two logical auxiliary qubits, 38 additional CNOT gates are required using the circuits in Figs. 4.24 and 4.25.

As Fig. 4.28 illustrates for the Steane code, the logical CNOT gate can be run in a single time step, if one is able to execute physical CNOT gates in parallel. Then the stabilizer mapping needs s time steps and copying the syndrome from one block to a total of $2^a - 1$ blocks can be done in a time steps. The feedback operations can be performed in a single time step in principle. Furthermore, the respective syndrome mapping and copying steps of the X- and Z-correction blocks may also be carried out in parallel. Reducing the circuit depth of the QEC cycle is desirable because coherence times in near-term devices are limited.

4.2.4 Implementations

In what follows we discuss several practical implementations of our MFEC scheme. We use various sets of basis gates that are native to different quantum computing platforms. As one option, we decompose all multi-qubit gates of the scheme into two-qubit gates. The coherent quantum feedback operations can also be implemented using native multi-qubit gates if they are practically available with physical fault rates that are lower than the expected overall error of their decompositions, which we discuss as another option. Multi-qubit gates may also be used to decrease the gate count of the syndrome mapping step. Notably, multi-qubit gates in these positions need not invalidate the fault tolerance property of the scheme. However, it is indispensable that

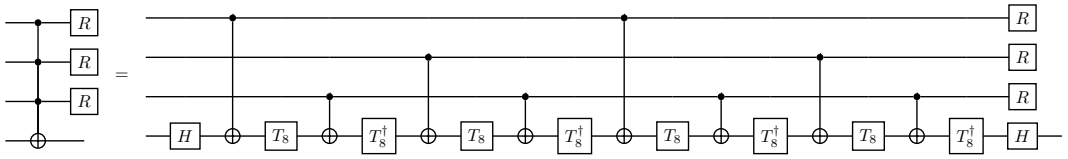


FIGURE 4.29: Simplified decomposition of the C_3NOT gate followed by reset of the three control qubits. Eight CNOT gates as well as Hadamard gates H, reset R and the rotation $T_8 = \exp(-\mathrm{i}\,\pi/16\,Z)$ are required.

transversal CNOT gates are used for fault propagation and for coherently copying the syndrome state; they must not be replaced by multi-qubit gates. Single faults of high Pauli weight in these locations, according to Eq. (3.13), could lead to uncorrectable errors and destroy fault tolerance. Then, we identify parameter regimes of advantageous operation of the MFEC scheme compared to measurement-based schemes in general and in particular compared to the state-of-the-art flag EC protocol that we already employed in Sec. 4.1.3.2.

4.2.4.1 Decomposition into two-qubit gates

The MFEC unit for the Steane code from Fig. 4.27 consists of mostly two-qubit gates already. Only the 3-qubit-controlled feedback gates need to be decomposed, in a way analogous to the Toffoli gate decomposition in Fig. 4.25. As stated in Sec. 4.2.3.3, 13 CNOT gates are sufficient to exactly decompose the C_3NOT gate according to [Bar+95]. We note that, since the state of the control qubits after the gate is reset and thus irrelevant, a decomposition with fewer CNOTs is possible. Consider the circuit equivalence in Fig. 4.29. Using a sequence of eight Z-rotations, each of angle $\pm\pi/8$, we can intersperse only 8 CNOT gates in a way that all rotations cancel exactly if the state of the 3 control qubits is different from $|111\rangle$. When all control qubits are activated however, the target operations of the CNOT gates align all Z-rotations into the same direction so that, in effect, a full π-rotation about the Z-axis is performed by the circuit[13]. The outer two Hadamard gates then transform the Z-flip to an X-flip so the overall effect of the circuit is the C_3NOT operation followed by reset of the control qubits.

We now provide an overall idea of the performance of the MFEC scheme decomposed into CNOT gates by simulating logical failure rates with a depolarizing noise model with a single parameter p on all operations, i.e. single-qubit gates, two-qubit gates and initializations. Logical failure rates are compared to the flag EC scheme of [Rei20] used in Sec. 4.1.3.2 where we also assume depolarizing noise on measurements. In accordance with our findings from Sec. 4.1.3, we assume that this model captures the essence of the noisy circuits for current and anticipated capabilities of experimental hardware. We also model idling noise since the MFEC unit contains many more gate operations than the flag EC unit and we discuss the role of idling noise in more detail in Sec. 4.2.4.3. Here, we choose a depolarizing channel with a noise strength of $p/100$ on respective idling locations during all operations (this includes measurements), which is a common assumption when simulating QEC blocks [CB18; CC19] and reflects the orders of magnitude of physical fault rates given in Tab. 4.4.

Recall the flag EC scheme, employed as a subroutine in Sec. 4.1.3.2, which works as follows. First, all stabilizers are measured using six auxiliary qubits in an interleaved

[13]This would implement the C_3Z gate.

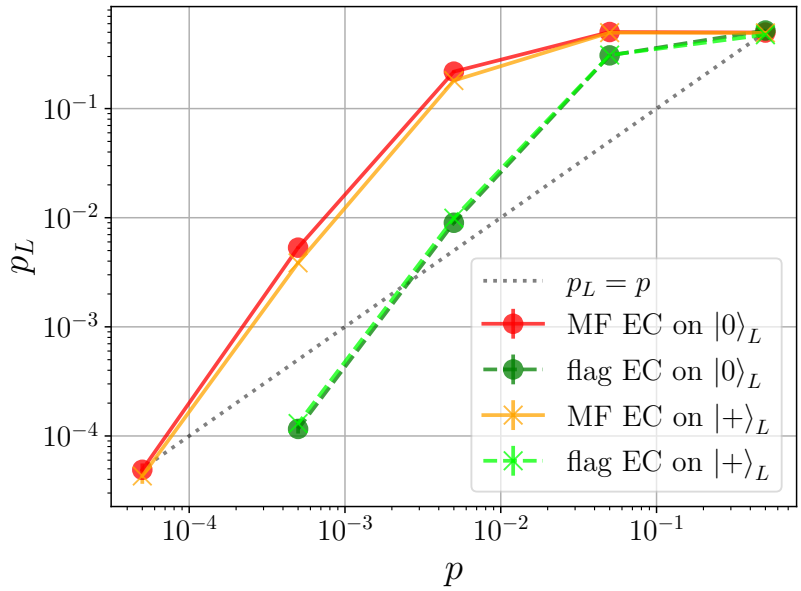


FIGURE 4.30: Comparison of logical failure rates for MFEC and flag EC for two logical input states $|0\rangle_L$ and $|+\rangle_L$. A single-parameter depolarizing noise model with strength p on single-qubit gates, two-qubit gates, initializations and measurements, as well as on idling locations with strength $p/100$, is used. Here, the MFEC scheme is decomposed entirely into two-qubit gates.

way as the last circuit block in Fig. 4.8(a) shows. No uncorrectable error can be present on the data qubits if all measurements yield +1. If any flag is triggered, we must distinguish if faults have propagated from within the circuit to the data qubits or faults have been present on the input state already. A second step of sequential stabilizer measurements with single auxiliary qubits (Fig. 4.17[14]) clarifies the origin of the non-trivial syndrome. If the measurement outcomes from the two blocks are different, the first one is interpreted as a flag event and the appropriate two-qubit correction can be applied according to the flag error set if the subsequently measured syndrome matches any possible two-qubit error. Otherwise, or if the two measured syndromes agree, the standard look up table 2.3 of the Steane code is applied.

Logical failure rates for MFEC and flag EC are shown in Fig. 4.30. We clearly observe that, as expected for FT schemes, indeed the logical failure rates $p_L \sim p^2$ scale quadratically as $p \to 0$ for both logical input states $|\psi\rangle_L \in \{|0_L\rangle, |+\rangle_L\}$. The logical failure rates differ by approximately one order of magnitude and the values of p where the EC schemes outperform the physical fault rate $p_L = p$ lie at approximately $p = 6 \times 10^{-5}$ for the MFEC scheme and at approximately $p = 3 \times 10^{-3}$ for the flag scheme. This gap may be narrowed by using advantageous compilations of the MFEC scheme in practice, which we discuss next.

4.2.4.2 Native multi-qubit gates in neutral-atom platforms and ion traps

Different types of multi-qubit gates are natively available in both neutral-atom platforms and ion trap quantum processors. For the former, mechanisms that are based on

[14] In the actual simulations, six auxiliary qubits are used instead of three and all measurements are performed at the end of the block. Instead of the single-qubit Y-type rotations, we used Hadamard gates in the simulations for this section.

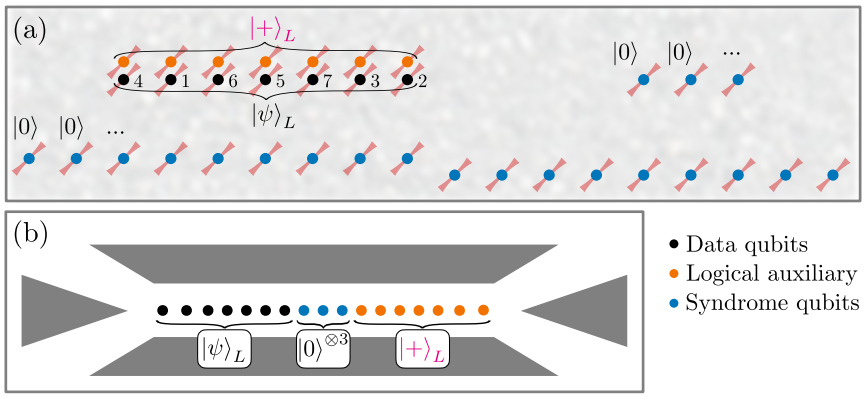


FIGURE 4.31: Sketched initial configurations for implementations of the MFEC scheme for the Steane code in (a) a neutral-atom tweezer array and (b) a static, linear ion trap. A complete shuttling schedule for neutral-atom platforms is given in [HLM23].

Impl.	p_1 [%]	p_2[%]	p_3 [%]	p_4 [%]	p_5 [%]	p_i[%]	p_m [%]	$p_{\text{idle},x}$
CNOT	0.007	0.3	-	-	-	0.0002	0.3	$p_x/100$
NATF	0.007	0.3	1	1	1	0.0002	0.3	$p_x/100$
MIMS	0.007	0.3	1	1	1	0.0002	0.3	$p_x/100$

TABLE 4.7: Physical fault rates used in the multi-parameter depolarizing noise model to numerically simulate three different MFEC implementations correspond to $\lambda = 1$ in Fig. 4.36. Fault rates at idling positions during an operation of type $x \in \{1, 2, 3, 4, 5, i, m\}$ are assumed to be 1% of the respective operation fault rate. Single-qubit operation and two-qubit gate fault rates p_1, p_2, p_i and p_m are state-of-the-art values [RA+21]. Multi-qubit gate fault rates p_3, p_4 and p_5 are based on theory proposals [Ber+17; Ras+20; Esp+21; PDP22].

the interactions between atoms excited to Rydberg states, i.e. the Rydberg blockade, are typically used to perform these gates. With global laser pulses, C_2Z gates have been realized in experiments [Lev+19; Eve+23] and there exist theoretical proposals to implement C_nNOT gates [ISM11; PDP22; Eve+23; KM20], $CNOT_n$ gates [Mül+09; KM20] and C_nNOT_m gates [You+21] up to local rotations. For the latter, apart from the widely-used MS gates [Sch+13a], the iToffoli gate can be implemented directly with a varying number of control qubits [Mon+09; Esp+21]. Possible configurations of atoms/ions in actual devices are sketched in Fig. 4.31.

A native C_3NOT gate facilitates implementation of the feedback operations in Fig. 4.27. We can expect an improvement of MFEC performance if the C_3NOT gate can be executed fast enough and with a large fidelity. Note that also the number of idling locations is reduced if the decomposition into CNOTs can be avoided. We estimate that using the C_3NOT gate is advantageous if its fault rate $p_4 < 8p_2 + 10p_1$ compared to a decomposition according to Fig. 4.29 with respective fault rates. The number of CNOT gates and single-qubit gates on the target qubit in Fig. 4.29 provide a lower bound to estimate the expected error rate of the decomposition since, also, faults from the control qubits may propagate through the CNOTs to the target qubit. This rough estimation is fulfilled by the physical fault rates listed in Tab. 4.7, which are based on state-of-the-art values [RA+21; Pin+21] and theory proposals [Ber+17; Ras+20; Esp+21; PDP22]. A native Toffoli gate is also advantageous to use for preparation of the logical auxiliary qubits since, with the values in Tab. 4.7, $p_3 < 4p_2 + 6p_1$ (see Fig. 4.25) by the same reasoning.

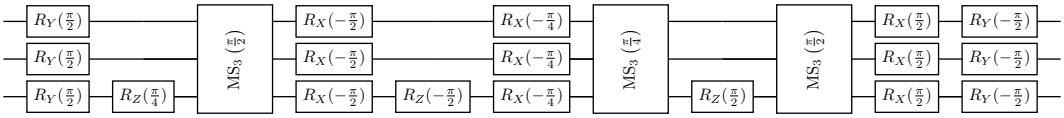


FIGURE 4.32: Exact decomposition of the Toffoli gate into 3-qubit X-type MS gates (see Eq. (4.1)) and local rotations according to [Mar$^+$16]. The upper two wires are the control qubits and the lower wire represents the target qubit. The last layer of local rotations on the control qubits could be omitted when used as the feedback operation in Fig. 4.24 due to the subsequent reset.

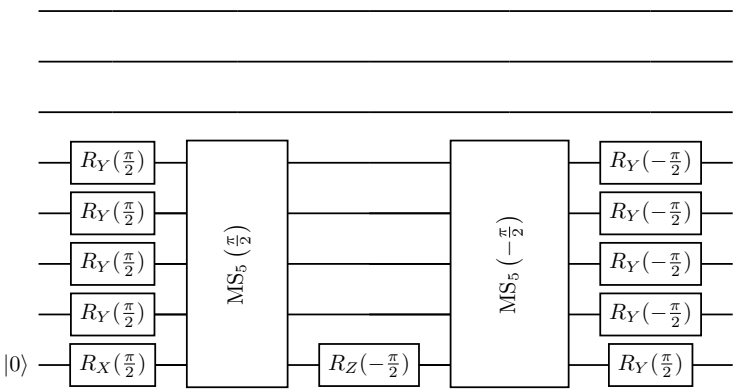


FIGURE 4.33: The expectation value of the stabilizer K_1^Z can be mapped to an auxiliary qubit by two 5-qubit X-type MS gates with $\theta = \pm\pi/2$ and single-qubit rotations as shown in [Mül$^+$11; Ber$^+$17]. If the Y-rotations on the data qubits are omitted, the circuit maps the eigenvalue of the corresponding X-type stabilizer K_1^X onto the auxiliary qubit.

With q-ion X-type MS gates, described by the unitary in Eq. (4.1) with $\vec{\varphi} = \vec{0}$, we can facilitate both the syndrome mapping and feedback steps of Fig. 4.27 [Bar$^+$11; Mül$^+$11]. We use the previously known compilation of the Toffoli gate into 3-qubit MS gates [Mar$^+$16], reproduced in Fig. 4.32, to implement the feedback operation of the state preparation circuit in Fig. 4.24. Regarding the syndrome mapping step, we can employ two 5-qubit MS gates to map the expectation value of each weight-4 stabilizer onto a single auxiliary qubit [Mül$^+$11; Ber$^+$17] so that six multi-qubit gates are needed in total to map a three-bit syndrome (see Fig. 4.33). We stress again that fault tolerance is respected since the 5-qubit depolarizing channel can only cause a high-weight error on the logical auxiliary qubit, which is reset anyway, and a wrong syndrome, which can at most add a weight-1 error to the logical data qubit. Note that this means that using a single 10-qubit gate that acts on all physical qubits of the logical auxiliary qubit and the three syndrome qubits would comply with fault tolerance in principle. However, we do not expect the syndrome mapping to be less noisy when six 5-qubit MS gates are employed instead of twelve CNOT gates since $6p_5 > 12p_2$ with the parameters in Tab. 4.7.

Lastly, we implement the 4-qubit feedback operation as the sequence of 4-qubit MS gates and local rotations, shown in Fig. 4.34, that we found numerically with a variational circuit ansatz in pennylane [Ber$^+$22]. In this framework, we used the 64 rotation angles that parametrize the four layers of 4-qubit MS gates and arbitrary rotations of all three Pauli types as input parameters $\vec{x}$ to a variational circuit. The cost function was chosen as the weighted average of the Pauli-X and -Z expectation values of the target qubit of a C_3NOT gate for application of the parametrized circuit unitary $U(\vec{x})$ to all 16 computational basis states and additional 16 states with the control qubits

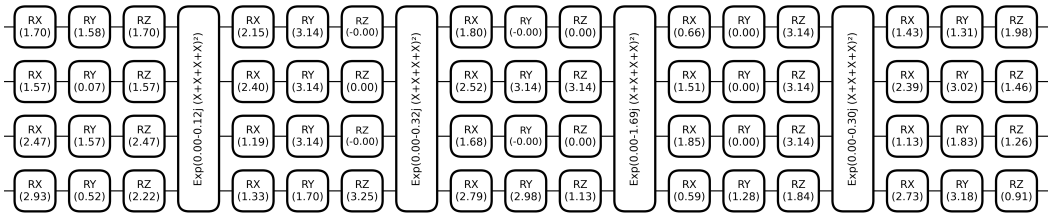


FIGURE 4.34: Decomposition of the C_3NOT gate into 4-qubit X-type MS gates and local rotations followed by reset of the control qubits. The upper three wires are the control qubits and the lower wire represents the target qubit. Rotation angles are given with higher precision in Tab. 4.8. The last layer of local rotations on the control qubits could be omitted when used for the feedback operations in Fig. 4.27(a) due to the subsequent reset.

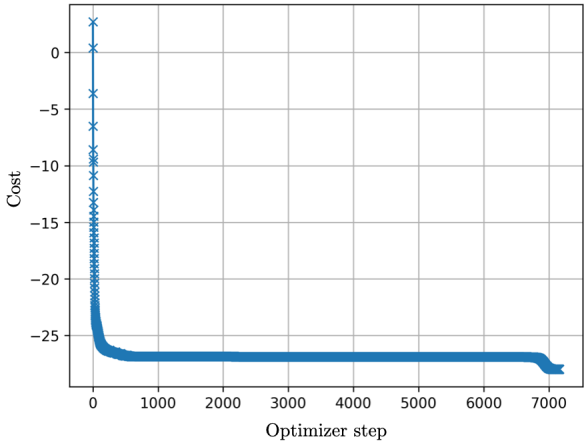


FIGURE 4.35: The cost function in Eq. (4.77) converges to the target value of -28 for the variational circuit in Fig. 4.34.

in the computational basis and the target qubit in the $|\pm\rangle$-basis. The weights of the states $U(\vec{x})|1110\rangle$, $U(\vec{x})|1111\rangle$, $U(\vec{x})|111+\rangle$ and $U(\vec{x})|111-\rangle$ are chosen as $-7/2$ and all other states are weighted by $-1/2$ so that the global minimum of the cost function

$$\begin{aligned} C(\vec{x}) = -\frac{1}{2}\Bigg(\sum_{abc\neq 111} & \langle abc0|U^\dagger(\vec{x})Z_4U(\vec{x})|abc0\rangle - \langle abc1|U^\dagger(\vec{x})Z_4U(\vec{x})|abc1\rangle \\ & + \langle abc+|U^\dagger(\vec{x})X_4U(\vec{x})|abc+\rangle - \langle abc-|U^\dagger(\vec{x})X_4U(\vec{x})|abc-\rangle \\ & - 7\,\langle 1110|U^\dagger(\vec{x})Z_4U(\vec{x})|1110\rangle + 7\,\langle 1111|U^\dagger(\vec{x})Z_4U(\vec{x})|1111\rangle \\ & + 7\,\langle 111+|U^\dagger(\vec{x})X_4U(\vec{x})|111+\rangle - 7\,\langle 111-|U^\dagger(\vec{x})X_4U(\vec{x})|111-\rangle \Bigg) \end{aligned} \tag{4.77}$$

is $C = -28$; a, b and c can take the values 0 and 1. The choice of the weights is motivated by putting an emphasis on the desired effect of $U(\vec{x})$ on the target qubit in case all control qubits are in the $|1\rangle$ state. Their exact values are the result of numerically experimenting until the global minimum of the cost function is actually found in practice for a combination of variational parameters $\vec{x}$. We use the `AdagradOptimizer` in pennylane until the minimal value of the cost function is found with an absolute accuracy of 10^{-4} as shown in Fig. 4.35. The converged angles for all 64 gates of the

time step	rotation	angle θ	time step	rotation	angle θ
1	$R_X^{(1)}$	1.7016306974989441	40	$R_X^{(1)}$	0.6597193711414078
2	$R_Y^{(1)}$	1.5795679517126637	41	$R_Y^{(1)}$	0.000127747992253436
3	$R_Z^{(1)}$	1.7016364361060135	42	$R_Z^{(1)}$	3.141326508787639
4	$R_X^{(2)}$	1.5706691398072636	43	$R_X^{(2)}$	1.5086374816792918
5	$R_Y^{(2)}$	0.07488703273604898	44	$R_Y^{(2)}$	0.0003195950440647524
6	$R_Z^{(2)}$	1.5708162845196991	45	$R_Z^{(2)}$	3.141618529495557
7	$R_X^{(3)}$	2.4733498110482617	46	$R_X^{(3)}$	1.851040827063517
8	$R_Y^{(3)}$	1.570978087668805	47	$R_Y^{(3)}$	$2.5330048710337304 \cdot 10^{-5}$
9	$R_Z^{(3)}$	2.473338798284629	48	$R_Z^{(3)}$	3.141620662584241
10	$R_X^{(4)}$	2.925198421469473	49	$R_X^{(4)}$	0.5851416542930774
11	$R_Y^{(4)}$	0.5184554247895697	50	$R_Y^{(4)}$	1.2817745160285958
12	$R_Z^{(4)}$	2.2155872069814624	51	$R_Z^{(4)}$	1.8434910130310886
13	MS	0.47618449796926465	52	MS	1.1852630143343514
14	$R_X^{(1)}$	2.1467351234080563	53	$R_X^{(1)}$	1.4334708713034108
15	$R_Y^{(1)}$	3.141888995985973	54	$R_Y^{(1)}$	1.3149000591854494
16	$R_Z^{(1)}$	−0.0007547728309335194	55	$R_Z^{(1)}$	1.9799754514097745
17	$R_X^{(2)}$	2.400499912253685	56	$R_X^{(2)}$	2.385180572177377
18	$R_Y^{(2)}$	3.1415882232243812	57	$R_Y^{(2)}$	3.0169184347659774
19	$R_Z^{(2)}$	$6.7559904951046 \cdot 10^{-5}$	58	$R_Z^{(2)}$	1.4620953315141385
20	$R_X^{(3)}$	1.185074722683019	59	$R_X^{(3)}$	1.1258670832503659
21	$R_Y^{(3)}$	3.141691576333686	60	$R_Y^{(3)}$	1.8289082653712643
22	$R_Z^{(3)}$	$-3.562993822558777 \cdot 10^{-5}$	61	$R_Z^{(3)}$	1.257061399856186
23	$R_X^{(4)}$	1.3267674839439592	62	$R_X^{(4)}$	2.726979125875662
24	$R_Y^{(4)}$	1.697674066992338	63	$R_Y^{(4)}$	3.182404050384048
25	$R_Z^{(4)}$	3.2542541431385135	64	$R_Z^{(4)}$	0.9113006716946862
26	MS	1.2961753988001792			
27	$R_X^{(1)}$	1.80332241033805			
28	$R_Y^{(1)}$	$-9.14307205449397 \cdot 10^{-6}$			
29	$R_Z^{(1)}$	0.00013217255162876435			
30	$R_X^{(2)}$	2.5206093101178855			
31	$R_Y^{(2)}$	3.141617687717394			
32	$R_Z^{(2)}$	3.141423898780103			
33	$R_X^{(3)}$	1.6829531013144419			
34	$R_Y^{(3)}$	$-1.2744701675664226 \cdot 10^{-5}$			
35	$R_Z^{(3)}$	$1.1892895790349708 \cdot 10^{-5}$			
36	$R_X^{(4)}$	2.7949214025522897			
37	$R_Y^{(4)}$	2.9814542086299864			
38	$R_Z^{(4)}$	1.134224555725302			
39	MS	6.779626443860337			

TABLE 4.8: Angle parameters for the C_3NOT gate decomposition in Fig. 4.34.

variational circuit are listed in Tab. 4.8. All other CNOT gates in the MFEC circuit are compiled into 2-qubit MS gates and local rotations according to Eq. (4.6).

For the numerical simulations of the multi-qubit-gate implementations we use a multi-parameter depolarizing noise model where all physical fault rates

$$\vec{p} = (p_1, p_2, p_3, p_4, p_5, p_i), \tag{4.78}$$

i.e. single- and multi-qubit gate fault rates as well as the physical qubit initialization fault rate, are scaled uniformly with a scaling parameter λ (as in Sec. 4.1.3). The point at $\lambda = 1$ corresponds to the fault rates provided in Tab. 4.7. On the respective idling locations, we use the dephasing channel from Eq. (3.12) with assumed noise strengths of $\vec{p}/100$ for simplicity. This order of magnitude is estimated rather pessimistically from the gate times and coherence times achieved in neutral-atom platforms. A two-qubit gate, for example, that takes time $t_2 \approx 250\,\text{ns}$ on a qubit with coherence time $T_2^* \approx 4\,\text{ms}$ ($T_2 \approx 1\,\text{s}$ when spin echo techniques are applied) and has a fault rate $p_2 = 5 \times 10^{-3}$ corresponds to a ratio $p_2/p_{\text{idle},2} \approx 160$ (4×10^4) according to Eq. (4.9) [Gra$^+$22; Blu$^+$22; Eve$^+$23]. The estimation also holds for ion traps, where, for instance, the 5-qubit MS gate with fault rate $p_5 = 5 \times 10^{-2}$ and duration $t_5 = 60\,\mu\text{s}$ corresponds to a ratio $p_5/p_{\text{idle},5} \approx 166$ when the coherence time is $T_2 = 100\,\text{ms}$ [Ber$^+$17]. With longer coherence times, such as $T_2 = 2\,\text{s}$ in the system of

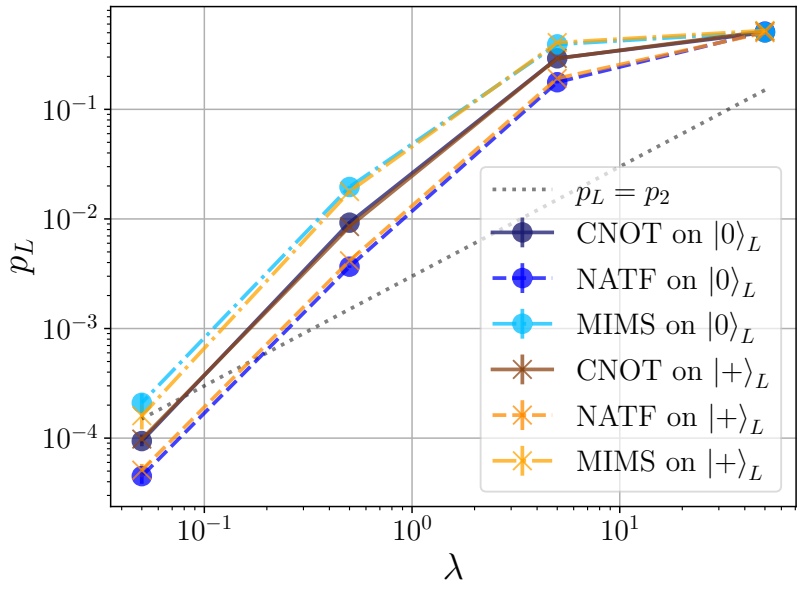


FIGURE 4.36: Logical failure rates for the MFEC scheme compiled into CNOT gates, native multi-qubit-controlled feedback gates (NATF) and multi-ion Mølmer-Sørensen gates (MIMS). We use a multi-parameter depolarizing noise model with dephasing on idling locations with fault rates given in Tab. 4.7 and uniformly vary all physical fault rates with a scaling parameter λ. Logical failure rates on both input states $|0\rangle_L$ and $|+\rangle_L$ can vary up to one order of magnitude depending on the native gates that are used.

[Pin+21; Mos+23], a two-qubit gate fault rate of $p_2 = 2\times10^{-3}$ performed in $t_2 = 25\,\mu s$ corresponds to a ratio $p_2/p_{\text{idle},2} \approx 320$.

Logical failure rates of the MFEC scheme with native multi-qubit-controlled feedback operations or multi-ion MS gates are compared to the decomposition into CNOT gates in Fig. 4.36 for both logical input states $|\psi\rangle_L \in \{|0_L\rangle, |+\rangle_L\}$. All three implementations are capable to reach logical failure rates $p_L < p_2$ lower than the two-qubit gate fault rates when λ is decreased by approximately one order of magnitude. In the regime of low physical fault rates, where the six lines in Fig. 4.36 are (almost) parallel, the implementation with multi-ion MS gates performs approximately a factor of 2 worse than the CNOT decomposition of the scheme, which is qualitatively expected. The implementation with native multi-qubit-controlled gates, however, performs approximately a factor of 5 better than the CNOT decomposition at low λ. So we conclude that, indeed, a large part of the gap between MFEC and flag EC as observed in Fig. 4.30 may be bridged by employing fast high-fidelity multi-qubit-controlled gates. One might be able to compensate the deteriorating effect of the MFEC overhead further by considering hardware-specific noise characteristics such as biased noise or bias-preserving gates [Con+22]. To this end, our numerical simulations of depolarizing noise might be overly pessimistic.

4.2.4.3 Measurement-free advantage

In this section we assess under which conditions MFEC can potentially have an advantage over syndrome-measurement-based QEC, such as the flag EC protocol used above. Denoting the logical failure rates of the two protocol types as p_L^{MF} and p_L^{SM} respectively, we consider MFEC advantageous if $p_L^{\text{MF}}/p_L^{\text{SM}} \leq 1$. We first provide an

analytical estimation of a parameter region of practical advantage and then compare this estimation to numerical simulations with two different noise models. The first one is a simple, few-parameter model, where we trade-off fault rates of idling qubits against fault rates of operations. The second one is a more detailed noise model, which describes circuits that can be implemented in a neutral-atom architecture using native multi-atom Rydberg gates.

Suppose that all operations, for the QEC schemes expressed in terms of CNOT gates, suffer from depolarizing noise of strength p. Furthermore, taking into account the dominant sources of idling noise for each scheme, consider an idling fault rate $p_{\text{idle,op}}$ for idling during operations for the MF protocol and also a distinct idling fault rate $p_{\text{idle},m}$ for idling during measurements in the conventional protocol. We assume that the time t to execute a gate operation or qubit initialization/reset is much smaller than the time it takes to perform a measurement. Idling rates p_{idle} are proportional to the idling time t according to Eq. (4.9) as long as $t/T_2 \ll 1$.

If idling were the only source of failure, i.e. assume for the moment that $p = 0$, and the cycle times τ_{MF} and τ_{SM} for either scheme are much smaller than the coherence time T_2, then the ratio of logical failure rates is approximately given by

$$\frac{p_L^{\text{MF}}}{p_L^{\text{SM}}} = \left(\frac{\tau_{\text{MF}}}{\tau_{\text{SM}}}\right)^2 \tag{4.79}$$

since single faults that occur with leading-order probability $p_{\text{idle}} \sim \tau$ cannot cause failure of either FT protocol. The two total cycle times can be roughly estimated as

$$\tau_{\text{MF}} \simeq t_{\text{ops}} \cdot \#\text{operations} \quad \text{and} \quad \tau_{\text{SM}} \simeq t_{\text{meas}} \cdot \#\text{measurements} \tag{4.80}$$

in case all operations are executed sequentially and the measurement-based protocol's cycle time is mostly determined by the time t_{meas} to perform a measurement, i.e. all operations are much faster than measurements.

In another scenario where operational faults are the only non-vanishing source of noise $p \neq 0$ and there were no idling noise due to very long coherence times $T_2 \to \infty$, the leading-order behavior of logical failure rates is

$$p_L^{\text{MF}} \sim (cp)^2 \quad \text{and} \quad p_L^{\text{SM}} \sim (c'p)^2 \tag{4.81}$$

for $p \ll 1$. The two constants c^2 and c'^2 are determined by the number of bad locations (see Sec. 2.4.1). The ratio of these constants can be used to infer the (dis-)advantage of MFEC over the conventional protocol since

$$\frac{p_L^{\text{MF}}}{p_L^{\text{SM}}} = \left(\frac{c}{c'}\right)^2 \tag{4.82}$$

follows from Eq. (4.81).

Let us now take both operation faults and the respective dominant source of idling noise into account at the same time. The low-p behavior of the logical failure rates is then given by

$$p_L^{\text{MF}} \sim c^2p^2 + \tilde{c}^2 p_{\text{idle,op}}^2 + bpp_{\text{idle,op}} \tag{4.83}$$

$$p_L^{\text{SM}} \sim c'^2p^2 + \tilde{c}'^2 p_{\text{idle},m}^2 + b'pp_{\text{idle},m} \tag{4.84}$$

for both $p, p_{\text{idle}} \ll 1$ since at least two faults in total must occur in order to cause logical failure. The constants $\tilde{c}, \tilde{c}', b$ and b' are determined, just as c and c', by the

number of bad locations for all respective fault combinations of order 2, i.e. two idling faults during operations, two idling faults during measurements, one operation fault and one simultaneous idling fault during an operation and one operation fault and one idling fault during a measurement respectively. The expressions in Eqs. (4.83) and (4.84) can be upper-bounded by assuming that the constants $\tilde{c}, \tilde{c}', b$ and b' represent the *total* number of respective circuit locations, instead of just the *bad* locations, for either protocol. Then, b and b' can be expressed as the products $c\tilde{c}$ and $c'\tilde{c}'$ respectively[15]. With this assumption we can estimate the ratio of logical failure rates as

$$\frac{p_L^{\mathrm{MF}}}{p_L^{\mathrm{SM}}} \sim \frac{(cp)^2 + (\tilde{c}p_{\mathrm{idle,op}})^2 + c\tilde{c}pp_{\mathrm{idle,op}}}{(c'p)^2 + (\tilde{c}'p_{\mathrm{idle},m})^2 + c'\tilde{c}'pp_{\mathrm{idle},m}}. \tag{4.85}$$

This fraction can be expanded around $p = 0$ and $p_{\mathrm{idle,op}} = 0$ where we include all second order terms to obtain

$$\begin{aligned}\frac{p_L^{\mathrm{MF}}}{p_L^{\mathrm{SM}}} = \left(\frac{\tilde{c}^2}{\tilde{c}'^2}\frac{p_{\mathrm{idle,op}}^2}{p_{\mathrm{idle},m}^2} + \mathcal{O}(p_{\mathrm{idle,op}}^3)\right) + p\left(\frac{c\tilde{c}p_{\mathrm{idle,op}}}{\tilde{c}'^2 p_{\mathrm{idle},m}^2} + \mathcal{O}(p_{\mathrm{idle,op}}^2)\right) \\ + p^2\left(\frac{c^2}{\tilde{c}'^2 p_{\mathrm{idle},m}^2} + \mathcal{O}(p_{\mathrm{idle,op}})\right) + \mathcal{O}(p^3).\end{aligned} \tag{4.86}$$

From Eq. (4.86), we expect the advantage of MFEC whenever

$$\frac{p_{\mathrm{idle},m}}{p_{\mathrm{idle,op}}} \geq \sqrt{\frac{\tilde{c}^2}{\tilde{c}'^2} + \frac{c\tilde{c}}{\tilde{c}'^2}\frac{p}{p_{\mathrm{idle,op}}} + \frac{c^2}{\tilde{c}'^2}\frac{p^2}{p_{\mathrm{idle,op}}^2}} \tag{4.87}$$

the ratio of idling rates, which are assumed approximately proportional to the respective idling times t_{meas} and t_{ops}, is larger than a bound set by the ratios $\tilde{c}^2/\tilde{c}'^2$, $c\tilde{c}/\tilde{c}'^2$ and $c^2/\tilde{c}'^2$ of numbers of circuit locations: c ($\tilde{c}$) is the number of operations (idling locations) in the MFEC circuit and $\tilde{c}'$ is the number of idling locations during measurements in the conventional QEC protocol.

In Fig. 4.37, we evaluate the advantage of MFEC over flag EC according to our analytical estimation in Eq. (4.87). For the MF scheme we count

$$c = \text{no. operations (MF)} = 456 \quad \text{and} \tag{4.88}$$
$$\tilde{c} = \text{no. idling during operations (MF)} = 5790 \tag{4.89}$$

locations and for the flag scheme we have

$$c' = \text{no. operations (FL)} = 88 \quad \text{and} \tag{4.90}$$
$$\tilde{c}' = \text{no. idling during measurement (FL)} = 14 \tag{4.91}$$

locations because, again, we assume that all gate operations are run sequentially with exactly one gate per time step and qubit initializations and measurements can be performed in parallel. Given these constants, a set of noise parameters that satisfies Eq. (4.87) with a large margin is

$$\begin{aligned} p &= 5 \times 10^{-4} \\ p_{\mathrm{idle,op}} = 10^{-4} &\overset{T_2 = 10^{-2}\,\mathrm{s}}{\Longleftrightarrow} t_{\mathrm{ops}} = 2 \times 10^{-6}\,\mathrm{s} \\ p_{\mathrm{idle},m} = 6.5 \times 10^{-2} &\Longleftrightarrow t_{\mathrm{meas}} = 1.4 \times 10^{-3}\,\mathrm{s}. \end{aligned} \tag{4.92}$$

[15]Note that one could, in principle, also determine the value of these constants by fault path counting or estimate the fraction of bad locations to total locations via MC simulation.

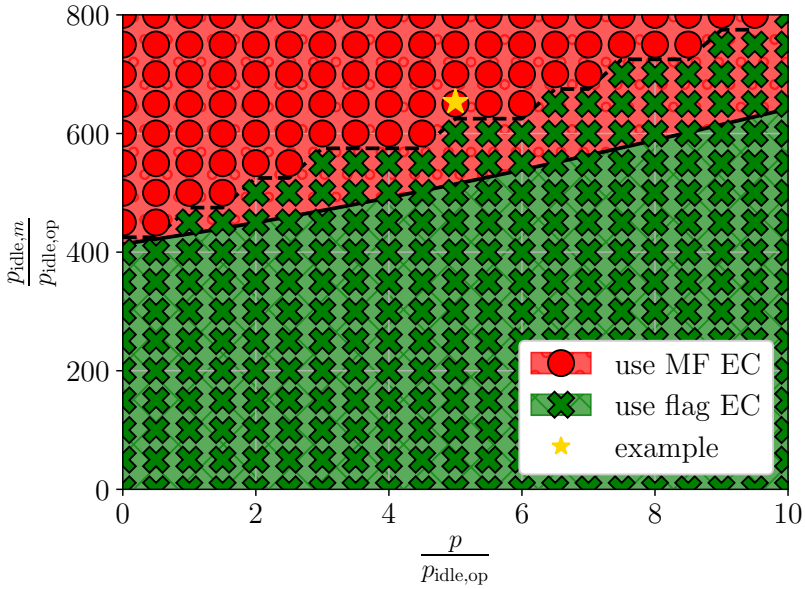


FIGURE 4.37: Regimes of advantageous use for MFEC and flag EC in terms of the parameter ratios $p/p_{\text{idle,op}}$ and $p_{\text{idle},m}/p_{\text{idle,op}}$ are estimated based on numerical simulation (markers) and Eq. (4.87) (filled area) with $\tilde{c}^2/\tilde{c}'^2 = 171041$, $c\tilde{c}/\tilde{c}'^2 = 13471$ and $c^2/\tilde{c}'^2 = 1061$, which follow from Eqs. (4.88)-(4.91). In the simulations, we use absolute values of $p = p_{\text{idle},m} = p_{\text{idle,op}} = 10^{-4}$ at $p/p_{\text{idle,op}} = 1$ and $p_{\text{idle},m}/p_{\text{idle,op}} = 1$. We apply depolarizing noise to all operations, i.e. single-qubit gates, two-qubit gates, initializations and measurements, as well as idling locations. Lines mark the estimated boundaries between regions of advantage. The star marker corresponds to the values in Eq. (4.92).

The logical failure rate for MFEC compiled into CNOT gates with these parameters is estimated from MC simulation to be $p_L^{\text{MF}} = 3.44(25)\%$, which is in fact smaller than $p_L^{\text{FL}} = 3.96(26)\%$, the corresponding value for flag EC. Figure 4.37 also displays numerical data that we acquire by sampling logical failure rates of both schemes using MC simulation until the 95% confidence interval allows us to distinguish which of the two schemes yields the lower failure rate. At $p = 0$, the numerical and analytical estimations agree. Measurements should be at least 400 times slower than operations for MFEC to be advantageous over flag EC. For finite p, the prediction starts to deviate from the numerical results.

To wrap up our analysis of MFEC advantage, we provide numerically simulated estimations of logical failure rates for a neutral-atom platform using the same multi-parameter noise model as in Sec. 4.2.4.2. Our scheme can help to facilitate the practical implementation of FT QEC in these architectures: Performance of measurements without atom-loss and with real-time feedback remains challenging and measurements are generally slow compared to gate operations. A collection of realistic physical fault rates [Blu$^+$22; Eve$^+$23] that we employ to simulate both the MFEC and flag EC protocol in the multi-parameter noise model is given in Tab. 4.9. Multi-qubit gates can be executed within about 100 to 500 ns [Eve$^+$23; Lev$^+$19; Pag$^+$22; JP22] and measurements take 3.5 to 25 ms [Gra$^+$23; Nor$^+$23; Lis$^+$23; Hui$^+$23], which corresponds to ratios $t_{\text{meas}}/t_{\text{ops}}$ from 7×10^3 to 2.5×10^5. The ratio $p/p_{\text{idle,op}}$ that we considered in Fig. 4.37 takes, for instance for two-qubit gates that make up the largest portion of operations in the MFEC scheme, the value $p_2/p_{\text{idle},2} \approx 160$ for the two-qubit gate parameters given in Tab. 4.9 with $T_2^* = 4$ ms [Blu$^+$22]. Thus, we cannot expect

Operation fault rate	Value	Operation time	Idling rate $T_2^* = 4\,\mathrm{ms}$	Idling rate $T_2 = 1\,\mathrm{s}$
p_1	3×10^{-4}	$1\,\mu\mathrm{s}$	1.2×10^{-4}	5×10^{-7}
p_2	5×10^{-3}	$250\,\mathrm{ns}$	3.1×10^{-5}	1.25×10^{-7}
p_3	2×10^{-2}	$500\,\mathrm{ns}$	6.2×10^{-5}	2.5×10^{-7}
p_4	3×10^{-2}	$500\,\mathrm{ns}$	6.2×10^{-5}	2.5×10^{-7}
p_i	2×10^{-3}	$1\,\mu\mathrm{s}$	1.2×10^{-4}	5×10^{-7}
p_m	2×10^{-3}	$500\,\mu\mathrm{s}$, $1\,\mathrm{ms}$	$5.9 \times 10^{-2}, 11.8 \times 10^{-2}$	$2.5 \times 10^{-4}, 5 \times 10^{-4}$

TABLE 4.9: Realistic fault rates of physical operations in a neutral-atom platform and corresponding idling times translated to idling fault rates for dephasing noise using two different coherence times T_2^* and T_2. The values are based on [Blu+22; Eve+23] and used for simulation of MFEC and flag EC implemented with neutral atoms.

Protocol (varied parameters)	$T_2^* = 4\,\mathrm{ms}$	$T_2 = 1\,\mathrm{s}$
MFEC (*realistic*)	5.9(3)%	5.6(1)%
MFEC (*improved*)	1.7(1)%	1.7(1)%
Flag EC ($t_{\mathrm{meas}} = 1\,\mathrm{ms}$)	11.8(4)%	0.66(4)%
Flag EC ($t_{\mathrm{meas}} = 500\,\mu\mathrm{s}$)	4.9(2)%	0.65(4)%

TABLE 4.10: Simulated logical failure rates for MFEC and flag EC in a neutral-atom platform with coherence times T_2^* and T_2. MFEC failure rates are listed for both *realistic* operation times and fault rates and anticipated future parameters that are *improved* by a factor of 2. We assume representative values for the measurement time in flag EC of $500\,\mu\mathrm{s}$ and $1\,\mathrm{ms}$. Strikingly, MFEC outperforms flag EC with improved operations and if coherence times are limited to T_2^*. If T_2 can be achieved, flag EC yields the lowest failure rate. It may, however, still be useful that the MFEC unit can be executed in a shorter time than the flag EC unit in this case.

our previous analytical estimations to hold in this parameter region and we therefore provide numerical simulation results.

The simulated logical failure rates for MFEC and flag EC are given in Tab. 4.10. We consider both the realistic values of Tab. 4.9 and a near-future scenario where all rates are improved by a factor of 2 for MFEC. The flag EC protocol is modeled with optimistic measurement times of $t_{\mathrm{meas}} = 1\,\mathrm{ms}$ and an even further improved value of $t_{\mathrm{meas}} = 500\,\mu\mathrm{s}$. The rather short coherence time of $T_2^* = 4\,\mathrm{ms}$ may be extended by utilizing spin echo techniques to more than $T_2 = 1\,\mathrm{s}$ [Blu+22] and we model both cases in our simulations.

We observe that MFEC fails with a rate of $p_L^{\mathrm{MF}} = 5.9(3)\%$ when the coherence time is T_2^*. In this case and with a measurement time of 1 ms, the logical failure rate of flag EC is as high as $p_L^{\mathrm{FL}} = 11.8(4)\%$ but outperforms MFEC with a value of $p_L^{\mathrm{FL}} = 4.9(2)\%$ when measurements only take $500\,\mu\mathrm{s}$. MFEC again beats flag EC when all physical fault rates are improved by a factor of 2 with a logical failure rate of $p_L^{\mathrm{MF}} = 1.7(1)\%$. The physical fault rate improvement translates to about a factor of 4 for the logical failure rate as expected for an FT protocol. Increasing the coherence time to T_2 causes no significant effect on MFEC failure rates. The dominant source of noise for MFEC in this regime are the physical operations and not the overall duration of the protocol. This is different for the flag scheme, which benefits greatly from increasing the coherence time to T_2 because the measurement times, while still long compared to other operations, then do not take so long that idling during measurements takes up a large fraction of the coherence time. The logical failure rate for flag EC in case that

measurements take 1 ms becomes as low as $p_L^{\mathrm{FL}} = 0.66(4)\%$ and does not significantly improve anymore if the measurement time is further reduced.

We summarize our results by pointing out again that the suggested measurement-free fault-tolerant QEC scheme opens a competitive pathway, especially for neutral-atom platforms, toward beneficial FT QEC and can even outperform state-of-the-art EC schemes in small codes before fast in-sequence measurements combined with in-sequence logic become widely available.

4.2.5 Scaling up

Applying the MFEC scheme to codes of larger distance d is not only of theoretical interest but can also help to suppress logical failure rates further than low-distance codes given that physical fault rates are sufficiently low. It is crucial to ensure that multiple faults, namely fault configurations of order up to $t = \lfloor \frac{d-1}{2} \rfloor$ can never lead to logical failures. This poses additional requirements on the FT construction of the auxiliary system in the scheme.

Consider, as an example that illustrates this necessity, an MFEC unit on the logical level, as in Fig. 4.26, but for a $d = 5$ code where $t = 2$ faults must never cause logical failure. Obviously, a single fault within, e.g., the syndrome mapping step S can cause application of an erroneous feedback operation that could be of Pauli weight 2 since the $d = 5$ code has a look up table that corrects weight-2 errors. An additional fault on the input state, so that the overall fault configuration is of order 2 in accordance with the requirements of fault tolerance of level 2, may however render the overall output error to be of weight 3, which in general will be uncorrectable and lead to logical failure in a $d = 5$ code.

We conjecture that our method can be straightforwardly translated to larger distances by employing t logical CNOT gates for any half-cycle of X-/Z-corrections such that faults are mapped to t logical auxiliary qubits. The syndrome can be mapped from each logical auxiliary qubit to a dedicated set of physical qubits. Applying multi-qubit-controlled operations such that the control qubits are in the same state for all t syndromes may perform the feedback steps. Only one physical data qubit can be targeted by any multi-qubit gate to comply with fault tolerance at level t, i.e. higher-weight feedback operations must be split to several multi-qubit-controlled gates that are conditioned on the same syndrome state.

However, this approach requires further investigation and a rigorous fault tolerance check that must also include the preparation circuits of the logical auxiliary qubits. Adapting techniques from flag fault tolerance to the MF setting could be another possible approach to efficiently restore fault tolerance for higher-distance codes [CB18; CR20; Rei20]. It is expected that, due to the exponentially growing number of feedback operations, our scheme might not be practical for larger-distance codes anyway. Concatenation may provide an alternative route for scale-up worth exploring.

Chapter 5

Conclusions & outlook

We are witnessing the dawn of fault-tolerant universal quantum computation where first primitives such as fault-tolerant logical state preparation, repetitive fault-tolerant quantum error correction cycles and fault-tolerant universal gate set implementations are demonstrated experimentally – some of these in various quantum computing hardware platforms. Reaching the level of fidelity for physical components needed to outperform physical qubits with logical qubits encoded into low-distance quantum error correcting codes is a frontier that needs to be tackled from two sides: On the one hand, experimental capabilities must be improved by identifying the most appropriate physical hardware platforms and devising suitable techniques to implement high-fidelity quantum operations. On the other hand, the theory groundwork should accommodate quantum computer engineering by developing tools to make low-overhead fault-tolerant quantum computation circuitry practically applicable to and useful in real devices.

By dynamical subset sampling, we are providing a technique to the community that allows for efficient numerical simulation of the low physical fault rates of current and anticipated future experiments involving long sequences of quantum circuits with in-sequence measurements and feed-forward of classical information. The problem of lacking robustness that occurs with weak noise for the current standard simulation method – direct Monte Carlo sampling – is circumvented by our method. As for any kind of numerical simulation of a theoretical model, this allows researchers to reliably estimate the performance of quantum protocols under circuit-level noise before investing time and money to implement them with physical hardware in their laboratories.

We have shown that, in practice, the performance of particular algorithms, namely logical Pauli and magic state preparation, is improved by using fault-tolerant circuit designs that are comprised of more noisy physical qubits and gate operations than their non-fault-tolerant counterparts. From a theory perspective, break-even points of practical advantage of fault-tolerant circuit designs are within reach with current quantum computing hardware or only moderate improvements. We have especially considered a particular ion trap setup and identified the entangling gate infidelity as the limiting factor via numerical simulations. A minimalist generic architecture-agnostic depolarizing noise model that is only informed by the experimental infidelities of physical operations is found to be appropriate to predict logical fidelities and quantum state fidelities for a single logical qubit. Thus, it is sufficient for use in our simulations to approximate the experimental findings when, crucially, the overall fault tolerance property is respected. We point out that the microscopic structure of crosstalk on entangling gates, however, can affect its impact on overall failure rates.

Measurement-free fault-tolerant quantum error correction, for which we provide detailed circuits on the physical qubit level, offers a road toward practical error-corrected quantum computation in near-term devices, where in-sequence measurements and feedback operations based on the classical measurement information are slow or unfeasible to conduct. In the long run, it may also be useful as a low-level scheme with fast gate operations for concatenated protocols. The measurement-free scheme is shown to be competitive with a state-of-the-art flag scheme in a regime where the performance of the physical architecture is limited by coherence time. We have demonstrated, using numerical simulations, that the disadvantage that comes from the additional gate overhead necessary in the measurement-free scheme can be partially compensated by employing a beneficial native gate set. The measurement-free scheme can even reach lower logical failure rates than the flag scheme with realistic parameters for gate fidelities and operation times of a neutral-atom setup that uses native multi-qubit gates.

While all building blocks for a digital error-corrected universal quantum computer are now experimentally available, a combination of, for example, the fault-tolerant universal gate set implementation of [Pos+22] with repetitive quantum error correction as in [RA+21] or using other fault-tolerant error correction techniques, is yet pending. With these experimental realizations in sight, we get closer to performing quantum algorithms on logical qubits. It is an interesting open question how different codes that host multiple logical qubits, such as high-rate qLDPC codes [LZ22; PK22; Din+23], could be more suitable for this endeavor. Fault-tolerant building blocks under circuit-level noise for these codes, as well as for Floquet codes [Fah+23; KZ23], are needed to take advantage of their potentially promising features like high thresholds, low-weight stabilizers or single-shot fault tolerance [Bom15b; CTV17; Cam19; DRS21; Gu+23].

For the considered ion trap experiments, a full characterization of logical building blocks, especially the logical CNOT gate where the experimental results most strikingly deviate from the simulations, has not yet been performed. Devising scalable characterization methods is a general open challenge: In order to combine logical building blocks, which also includes measurement-free fault-tolerant quantum error correction, in a large-scale quantum computation, it would be desirable to coarse-grain the effect of noise by employing effective noise channels on the logical level for each logical operation.

A fruitful direction of practical research could be to design fault-tolerant building blocks in more hardware-adapted ways to be most effective for the particular experimental platform they are intended to being used in and employ more sophisticated and validated theoretical noise models [Leu+97; Con+22]. The effect of coherent noise has only superficially been explored in our considerations and needs further quantitative investigation [HDF19]. The accuracy and honesty of the Pauli twirling approximation, which we used extensively, is a subject of active research [Puz+14; Kat19]. Additionally, we have not exhaustively explored the effects of correlated noise between different quantum circuit components that could stem from, for example, magnetic field fluctuations that lead to global dephasing [Pal+22]. Neither did we study effects of non-Markovian noise in depth [TB05; VA17].

With experiments being capable of operating a growing number of physical qubits, we also get closer to operating higher-distance quantum error correcting codes in actual hardware [Goo23b]. Therefore, as a potential path for future investigations, all schemes presented in this thesis, i.e. magic state preparation, dynamical subset

sampling and measurement-free fault-tolerant quantum error correction, could be extended to codes with distance $d \geq 5$. A generalization of fault-tolerant magic state preparation has been given in [CN20]. While generalizing dynamical subset sampling seems to be mostly a combinatorics challenge, the latter requires more general considerations about the nature of fault propagation and accumulation as was pointed out in [CB18]. Other fault-tolerant gadgets such as code switching or lattice surgery may also be possible to realize without measurements.

A comprehensive study that compares scaling up by directly increasing code distances versus concatenation should be conducted. Not only for our measurement-free scheme, but also for fault-tolerant universal gate sets, more extensive research in this direction is required, for instance starting from the analysis in [CJOL17] for concatenation or [CC19] for magic state distillation. Performing magic state distillation on the logical level with low-distance logical qubits seems possible with magic state distillation schemes that consume small numbers of input magic states. For example, with a native controlled-H operation that could be realized transversally between logical qubits encoded in the Steane code, the 10-to-2 scheme by [MEK12] actually only requires two magic states as input. Fault-tolerant measurement of the logical Hadamard operator in the four-qubit code should also be possible with a flag circuit. The distillation capabilities of more recently discovered codes are still to be explored, as has been done, for instance, for the $[[10, 1, 2]]$ code in [VK22].

Scaling up both the number and distance of logical qubits in actual experiments not only requires theoretical advancement of fault-tolerant circuit design. Development of practical embeddings of physical qubits into the experimental architecture is also necessary. It is currently unclear how to practically design a general hardware layout that is capable of providing the necessary connectivity between physical qubits in order to encode, stabilize and operate logical qubits encoded in a given code and at the same time provides the required connections between encoded blocks. While real all-to-all connectivity as conveniently offered by atomic systems may not generally be needed, the nearest-neighbor connectivity of superconducting transmon systems is a limitation that must be overcome, for instance, by suitable embeddings [Cha$^+$20]. Potential directions of hardware development might lead to ion traps with two-dimensional ion configurations [Hol$^+$20; Jai$^+$20; Jai$^+$23]. One can also arrange various one-dimensional ion crystals and connect them either via shuttling [Hil$^+$22; Akh$^+$23] or photonic links [Wan$^+$20a; Sah$^+$23]. In a recent work, it has been argued theoretically that the noisy interconnection may not be detrimental to the threshold of the featured quantum error correcting code [Ram$^+$23]. A perspective on also scaling up the necessary control electronics for ion trap quantum computers is given in [MAB23]. For superconducting transmon architectures, implementing long-range connectivity appears as a logical next step for actors committed to this platform [Haz$^+$21; Bra$^+$22; Zha$^+$23b; Hon$^+$23]. Scaling up to over one thousand physical qubits is already an actively pursued direction of development, eventually enabling a route toward universality for these systems [IBMapb].

Exciting times with rapidly evolving progress lie ahead of us. Once in-sequence logic becomes widely available, we anticipate more demonstrations of fault-tolerant primitives, for instance in ion traps, such as optimized flag fault-tolerant quantum error correction [Bha$^+$23] or fault-tolerant code switching. With the large numbers of qubits already available in neutral-atom platforms, we expect more complex quantum simulation experiments [KML23] as well as experiments that demonstrate operation of high-rate qLDPC codes, such as proposed in [Xu$^+$23a], and fault-tolerant quantum

error correction with many physical qubits along the lines of Steane-type or Knill-type error correction in the near future. A demonstration of universality on the logical level, using magic state injection as in [Pos+22] or another technique, is still pending using neutral-atoms or other quantum computing platforms. Fault-tolerant advantage for a small quantum algorithm containing both Clifford and non-Clifford gates has been observed most recently in an ion trap quantum computer [Wan+23]. An implementation of Grover's algorithm on the logical level has experimentally been shown to perform better than the algorithm on the physical level in the same superconducting transmon device [PL22].

More fundamentally, we expect the theory toolbox for fault-tolerant quantum computation to be continuously enhanced and systematically characterized, thus offering more possibilities for researchers to put these building blocks together into productive use. Demonstrations of actual computationally relevant quantum algorithms performed directly on error-corrected logical qubits with capabilities unreachable by bare physical qubits shall put an end to the NISQ era in the next decades.

Bibliography

[AB06] S. Anders and H. J. Briegel. "Fast simulation of stabilizer circuits using a graph-state representation". In: *Physical Review A* 73.2 (2006), p. 022334. DOI: 10.1103/PhysRevA.73.022334 (cited on p. 38).

[ABO08] D. Aharonov and M. Ben-Or. "Fault-tolerant quantum computation with constant error rate". In: *SIAM Journal on Computing* 38.4 (2008), p. 1207. DOI: 10.1137/S0097539799359385 (cited on pp. 13, 39).

[Abo+22] M. H. Abobeih, Y. Wang, J. Randall, et al. "Fault-tolerant operation of a logical qubit in a diamond quantum processor". In: *Nature* 606.7916 (2022), p. 884. DOI: 10.1038/s41586-022-04819-6 (cited on pp. 4, 5, 33).

[AG04] S. Aaronson and D. Gottesman. "Improved simulation of stabilizer circuits". In: *Physical Review A* 70.5 (2004), p. 052328. DOI: 10.1103/PhysRevA.70.052328 (cited on p. 38).

[Agh+23] M. Aghaee, A. Akkala, Z. Alam, et al. "InAs-Al hybrid devices passing the topological gap protocol". In: *Physical Review B* 107 (2023), p. 245423. DOI: 10.1103/PhysRevB.107.245423 (cited on p. 22).

[AGP06] P. Aliferis, D. Gottesman, and J. Preskill. "Quantum accuracy threshold for concatenated distance-3 codes". In: *Quantum Information and Computation* 6.2 (2006), p. 97. DOI: 10.26421/QIC6.2-1 (cited on pp. 13, 115).

[Akh+23] M. Akhtar, F. Bonus, F. R. Lebrun-Gallagher, et al. "A high-fidelity quantum matter-link between ion-trap microchip modules". In: *Nature Communications* 14.1 (2023), p. 531. DOI: 10.1038/s41467-022-35285-3 (cited on p. 137).

[AL18] T. Albash and D. A. Lidar. "Adiabatic quantum computation". In: *Reviews of Modern Physics* 90 (2018), p. 015002. DOI: 10.1103/RevModPhys.90.015002 (cited on p. 4).

[And+20] C. K. Andersen, A. Remm, S. Lazar, et al. "Repeated quantum error detection in a surface code". In: *Nature Physics* 16.8 (2020), p. 875. DOI: 10.1038/s41567-020-0920-y (cited on p. 4).

[AP08] P. Aliferis and J. Preskill. "Fault-tolerant quantum computation against biased noise". In: *Physical Review A* 78.5 (2008), p. 052331. DOI: 10.1103/PhysRevA.78.052331 (cited on p. 23).

[Arr+21] J. M. Arrazola, V. Bergholm, K. Brádler, et al. "Quantum circuits with many photons on a programmable nanophotonic chip". In: *Nature* 591.7848 (2021), p. 54. DOI: 10.1038/s41586-021-03202-1 (cited on p. 3).

[Aru+19] F. Arute, K. Arya, R. Babbush, et al. "Quantum supremacy using a programmable superconducting processor". In: *Nature* 574.7779 (2019), p. 505. DOI: 10.1038/s41586-019-1666-5 (cited on p. 3).

[Aws+18] D. D. Awschalom, R. Hanson, J. Wrachtrup, et al. "Quantum technologies with optically interfaced solid-state spins". In: *Nature Photonics* 12.9 (2018), p. 516. DOI: 10.1038/s41566-018-0232-2 (cited on p. 3).

[Bal17] C. J. Ballance. "High-fidelity quantum logic in Ca^+". Springer, 2017. DOI: 10.1007/978-3-319-68216-7 (cited on pp. 76, 79).

[Bar+00] P. A. Barton, C. J. S. Donald, D. M. Lucas, et al. "Measurement of the lifetime of the $3d^2D_{5/2}$ state in $^{40}Ca^+$". In: *Physical Review A* 62 (2000), p. 032503. DOI: 10.1103/PhysRevA.62.032503 (cited on p. 77).

[Bar+03] M. D. Barrett, B. DeMarco, T. Schaetz, et al. "Sympathetic cooling of $^9Be^+$ and $^{24}Mg^+$ for quantum logic". In: *Physical Review A* 68 (2003), p. 042302. DOI: 10.1103/PhysRevA.68.042302 (cited on p. 113).

[Bar+11] J. T. Barreiro, M. Müller, P. Schindler, et al. "An open-system quantum simulator with trapped ions". In: *Nature* 470.7335 (2011), p. 486. DOI: 10.1038/nature09801 (cited on pp. 113, 124).

[Bar+22] K. Barnes, P. Battaglino, B. J. Bloom, et al. "Assembly and coherent control of a register of nuclear spin qubits". In: *Nature Communications* 13.1 (2022), p. 2779. DOI: 10.1038/s41467-022-29977-z (cited on p. 3).

[Bar+95] A. Barenco, C. H. Bennett, R. Cleve, et al. "Elementary gates for quantum computation". In: *Physical Review A* 52.5 (1995), p. 3457. DOI: 10.1103/PhysRevA.52.3457 (cited on pp. 116, 120, 121).

[BCD01] L. D. Brown, T. T. Cai, and A. DasGupta. "Interval estimation for a binomial proportion". In: *Statistical Science* 16.2 (2001), p. 101. DOI: 10.1214/ss/1009213286 (cited on p. 41).

[BCS22] N. Bao, C. Cao, and V. P. Su. "Magic state distillation from entangled states". In: *Physical Review A* 105.2 (2022), p. 022602. DOI: 10.1103/PhysRevA.105.022602 (cited on p. 27).

[BD12] N. Bonesteel and D. DiVincenzo. "Quantum circuits for measuring Levin-Wen operators". In: *Physical Review B* 86.16 (2012), p. 165113. DOI: 10.1103/PhysRevB.86.165113 (cited on p. 22).

[BE21] N. P. Breuckmann and J. N. Eberhardt. "Quantum low-density parity-check codes". In: *PRX Quantum* 2.4 (2021), p. 040101. DOI: 10.1103/PRXQuantum.2.040101 (cited on p. 20).

[Bea+18] S. J. Beale, J. J. Wallman, M. Gutiérrez, et al. "Quantum error correction decoheres noise". In: *Physical Review Letters* 121.19 (2018), p. 190501. DOI: 10.1103/PhysRevLett.121.190501 (cited on p. 14).

[Ben+96] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, et al. "Mixed-state entanglement and quantum error correction". In: *Physical Review A* 54.5 (1996), p. 3824. DOI: 10.1103/PhysRevA.54.3824 (cited on pp. 25, 36).

[Ber13] B. A. Bernevig. "Topological insulators and topological superconductors". Princeton University Press, 2013. DOI: 10.1080/00107514.2014.948923 (cited on p. 19).

[Ber+17] A. Bermudez, X. Xu, R. Nigmatullin, et al. "Assessing the progress of trapped-ion processors towards fault-tolerant quantum computation". In: *Physical Review X* 7.4 (2017), p. 041061. DOI: 10.1103/PhysRevX.7.041061 (cited on pp. 123, 124, 126).

[Ber+19] A. Bermudez, X. Xu, M. Gutiérrez, et al. "Fault-tolerant protection of near-term trapped-ion topological qubits under realistic noise sources". In: *Physical Review A* 100.6 (2019), p. 062307. DOI: 10.1103/PhysRevA.100.062307 (cited on p. 79).

[Ber+22] V. Bergholm, J. Izaac, M. Schuld, et al. "PennyLane: Automatic differentiation of hybrid quantum-classical computations". 2022. arXiv: 1811.04968 [quant-ph] (cited on p. 124).

[BH12] S. Bravyi and J. Haah. "Magic-state distillation with low overhead". In: *Physical Review A* 86.5 (2012), p. 052329. DOI: 10.1103/PhysRevA.86.052329 (cited on pp. 26, 27, 29).

[Bha+22] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, et al. "Noisy intermediate-scale quantum algorithms". In: *Reviews of Modern Physics* 94.1 (2022), p. 015004. DOI: 10.1103/RevModPhys.94.015004 (cited on p. 2).

[Bha+23] D. Bhatnagar, M. Steinberg, D. Elkouss, et al. "Low-depth flag-style syndrome extraction for small quantum error-correction codes". 2023. arXiv: 2305.00784 [quant-ph] (cited on p. 137).

[BHC04] K. R. Brown, A. W. Harrow, and I. L. Chuang. "Arbitrarily accurate composite pulse sequences". In: *Physical Review A* 70.5 (2004), p. 052318. DOI: 10.1103/PhysRevA.70.052318 (cited on p. 79).

[BK05] S. Bravyi and A. Kitaev. "Universal quantum computation with ideal Clifford gates and noisy ancillas". In: *Physical Review A* 71.2 (2005), p. 022316. DOI: 10.1103/PhysRevA.71.022316 (cited on pp. 24–26).

[BK98] S. B. Bravyi and A. Y. Kitaev. "Quantum codes on a lattice with boundary". 1998. arXiv: quant-ph/9811052 [quant-ph] (cited on p. 20).

[Bli+02] B. B. Blinov, L. Deslauriers, P. Lee, et al. "Sympathetic cooling of trapped Cd^+ isotopes". In: *Physical Review A* 65.4 (2002), p. 040304. DOI: 10.1103/PhysRevA.65.040304 (cited on p. 113).

[Blu+22] D. Bluvstein, H. Levine, G. Semeghini, et al. "A quantum processor based on coherent transport of entangled atom arrays". In: *Nature* 604.7906 (2022), p. 451. DOI: 10.1038/s41586-022-04592-6 (cited on pp. 3, 4, 33, 113, 119, 126, 130, 131).

[BMD06] H. Bombín and M. A. Martín-Delgado. "Topological quantum distillation". In: *Physical Review Letters* 97.18 (2006), p. 180501. DOI: 10.1103/PhysRevLett.97.180501 (cited on p. 20).

[BNB19] N. C. Brown, M. Newman, and K. R. Brown. "Handling leakage with subsystem codes". In: *New Journal of Physics* 21.7 (2019), p. 073055. DOI: 10.1088/1367-2630/ab3372 (cited on p. 13).

[Boc+10] J. Bochmann, M. Mücke, C. Guhl, et al. "Lossless state detection of single neutral atoms". In: *Physical Review Letters* 104 (2010), p. 203601. DOI: 10.1103/PhysRevLett.104.203601 (cited on p. 113).

[Bom13] H. Bombín. "An introduction to topological quantum codes". 2013. arXiv: 1311.0277 [quant-ph] (cited on p. 4).

[Bom+13] H. Bombín, R. W. Chhajlany, M. Horodecki, et al. "Self-correcting quantum computers". In: *New Journal of Physics* 15.5 (2013), p. 055023. DOI: 10.1088/1367-2630/15/5/055023 (cited on p. 4).

[Bom15a] H. Bombín. "Gauge color codes: optimal transversal gates and gauge fixing in topological stabilizer codes". In: *New Journal of Physics* 17.8 (2015), p. 083002. DOI: 10.1088/1367-2630/17/8/083002 (cited on p. 31).

[Bom15b] H. Bombín. "Single-shot fault-tolerant quantum error correction". In: *Physical Review X* 5 (2015), p. 031043. DOI: 10.1103/PhysRevX.5.031043 (cited on p. 136).

[Boy+10] P. O. Boykin, T. Mor, V. Roychowdhury, et al. "Algorithms on ensemble quantum computers". In: *Natural Computing* 9.2 (2010), p. 329. DOI: 10.1007/s11047-009-9133-0 (cited on p. 113).

[BP13] P. Brooks and J. Preskill. "Fault-tolerant quantum computation with asymmetric Bacon-Shor codes". In: *Physical Review A* 87.3 (2013), p. 032310. DOI: 10.1103/PhysRevA.87.032310 (cited on p. 13).

[Bra+19] S. Bravyi, D. Browne, P. Calpin, et al. "Simulation of quantum circuits by low-rank stabilizer decompositions". In: *Quantum* 3 (2019), p. 181. DOI: 10.22331/q-2019-09-02-181 (cited on p. 39).

[Bra+22] S. Bravyi, O. Dial, J. M. Gambetta, et al. "The future of quantum computing with superconducting qubits". In: *Journal of Applied Physics* 132.16 (2022). DOI: 10.1063/5.0082975 (cited on pp. 3, 137).

[Bra+23] S. Bravyi, A. W. Cross, J. M. Gambetta, et al. "High-threshold and low-overhead fault-tolerant quantum memory". 2023. arXiv: 2308.07915 [quant-ph] (cited on p. 20).

[Bri+09] H. J. Briegel, D. E. Browne, W. Dür, et al. "Measurement-based quantum computation". In: *Nature Physics* 5.1 (2009), p. 19. DOI: 10.1038/nphys1157 (cited on p. 4).

[Bro13] P. Brooks. "Quantum error correction with biased noise". California Institute of Technology, 2013. URL: https://thesis.library.caltech.edu/7774/ (cited on p. 28).

[Bru+19] C. D. Bruzewicz, J. Chiaverini, R. McConnell, et al. "Trapped-ion quantum computing: Progress and challenges". In: *Applied Physics Reviews* 6.2 (2019), p. 021314. DOI: 10.1063/1.5088164 (cited on pp. 2, 3, 5, 75, 113).

[BS19] H. Bluhm and L. R. Schreiber. "Semiconductor spin qubits—A scalable platform for quantum computing?" In: *2019 IEEE International Symposium on Circuits and Systems*. IEEE. 2019, p. 1. DOI: 10.1109/ISCAS.2019.8702477 (cited on p. 3).

[But+23] F. Butt, S. Heußen, M. Rispler, et al. "Fault-tolerant code switching protocols for near-term quantum processors". 2023. arXiv: 2306.17686 [quant-ph] (cited on pp. ix, 31).

[BV13] S. Bravyi and A. Vargo. "Simulation of rare events in quantum error correction". In: *Physical Review A* 88.6 (2013), p. 062308. DOI: 10.1103/PhysRevA.88.062308 (cited on p. 42).

[BVT21] F. Battistel, B. M. Varbanov, and B. M. Terhal. "Hardware-efficient leakage-reduction scheme for quantum error correction with superconducting transmon qubits". In: *PRX Quantum* 2.3 (2021), p. 030314. DOI: 10.1103/PRXQuantum.2.030314 (cited on p. 37).

[Cai+23] Z. Cai, R. Babbush, S. C. Benjamin, et al. "Quantum error mitigation". 2023. arXiv: 2210.00921 [quant-ph] (cited on p. 4).

[Cam19] E. T. Campbell. "A theory of single-shot error correction for adversarial noise". In: *Quantum Science and Technology* 4.2 (2019), p. 025006. DOI: 10.1088/2058-9565/aafc8f (cited on p. 136).

[CB09] E. T. Campbell and D. E. Browne. "On the structure of protocols for magic state distillation". In: *Theory of Quantum Computation, Communication, and Cryptography*. Springer Berlin Heidelberg, 2009, p. 20. DOI: 10.1007/978-3-642-10698-9_3 (cited on p. 27).

[CB10] E. T. Campbell and D. E. Browne. "Bound states for magic state distillation in fault-tolerant quantum computation". In: *Physical Review Letters* 104.3 (2010), p. 030503. DOI: 10.1103/PhysRevLett.104.030503 (cited on p. 27).

[CB18] C. Chamberland and M. E. Beverland. "Flag fault-tolerant error correction with arbitrary distance codes". In: *Quantum* 2 (2018), p. 53. DOI:

10.22331/q-2018-02-08-53 (cited on pp. 4, 18, 51, 66, 68, 69, 121, 132, 137).

[CC19] C. Chamberland and A. W. Cross. "Fault-tolerant magic state preparation with flag qubits". In: *Quantum* 3 (2019), p. 143. DOI: 10.22331/q-2019-05-20-143 (cited on pp. 28, 29, 51, 93, 100, 102, 104, 106, 121, 137).

[CD+23] A. Carignan-Dugas, D. Dahlen, I. Hincks, et al. "The error reconstruction and compiled calibration of quantum computing cycles". 2023. arXiv: 2303.17714 [quant-ph] (cited on p. 92).

[CDT09] A. W. Cross, D. P. DiVincenzo, and B. M. Terhal. "A comparative code study for quantum fault-tolerance". 2009. arXiv: 0711.1556 [quant-ph] (cited on pp. 2, 115).

[CH17] E. T. Campbell and M. Howard. "Unified framework for magic state distillation and multiqubit gate synthesis with reduced resource cost". In: *Physical Review A* 95.2 (2017), p. 022316. DOI: 10.1103/PhysRevA.95.022316 (cited on p. 27).

[Cha18] C. Chamberland. "New methods in quantum error correction and fault-tolerant quantum computing". 2018. URL: https://www.uwspace.uwaterloo.ca/handle/10012/14049 (cited on pp. 15, 17, 18).

[Cha+20] C. Chamberland, G. Zhu, T. J. Yoder, et al. "Topological and subsystem codes on low-degree graphs with flag qubits". In: *Physical Review X* 10 (2020), p. 011022. DOI: 10.1103/PhysRevX.10.011022 (cited on pp. 20, 31, 137).

[Cha+21] A. Chatterjee, P. Stevenson, S. De Franceschi, et al. "Semiconductor qubits in practice". In: *Nature Reviews Physics* 3.3 (2021), p. 157. DOI: 10.1038/s42254-021-00283-9 (cited on p. 3).

[Che+23] J.-S. Chen, E. Nielsen, M. Ebert, et al. "Benchmarking a trapped-ion quantum computer with 29 algorithmic qubits". 2023. arXiv: 2308.05071 [quant-ph] (cited on p. 2).

[Chi+05] J. Chiaverini, R. B. Blakestad, J. Britton, et al. "Surface-electrode architecture for ion-trap quantum information processing". In: *Quantum Information & Computation* 5.6 (2005), p. 419. DOI: 10.5555/2011670.2011671 (cited on pp. 76, 113).

[CJ17a] B. Cruikshank and K. Jacobs. "High-threshold low-overhead fault-tolerant classical computation and the replacement of measurements with unitary quantum gates". In: *Physical Review Letters* 119.3 (2017), p. 030503. DOI: 10.1103/PhysRevLett.119.030503 (cited on p. 114).

[CJ17b] B. Cruikshank and K. Jacobs. "The role of quantum measurements in physical processes and protocols". In: *Quantum Science and Technology* 2.3 (2017), p. 033001. DOI: 10.1088/2058-9565/aa6d3e (cited on p. 114).

[CJOL17] C. Chamberland, T. Jochym-O'Connor, and R. Laflamme. "Overhead analysis of universal concatenated quantum codes". In: *Physical Review A* 95 (2017), p. 022313. DOI: 10.1103/PhysRevA.95.022313 (cited on pp. 31, 137).

[CJS16] D. Crow, R. Joynt, and M. Saffman. "Improved error thresholds for measurement-free error correction". In: *Physical Review Letters* 117.13 (2016), p. 130503. DOI: 10.1103/PhysRevLett.117.130503 (cited on p. 114).

[CN20] C. Chamberland and K. Noh. "Very low overhead fault-tolerant magic state preparation using redundant ancilla encoding and flag qubits". In:

npj Quantum Information 6.1 (2020), p. 91. DOI: 10.1038/s41534-020-00319-5 (cited on pp. 29, 137).

[Com+17] J. Combes, C. Granade, C. Ferrie, et al. "Logical Randomized Benchmarking". 2017. arXiv: 1702.03688 [quant-ph] (cited on p. 92).

[Con+22] I. Cong, H. Levine, A. Keesling, et al. "Hardware-efficient, fault-tolerant quantum computation with Rydberg atoms". In: *Physical Review X* 12 (2022), p. 021049. DOI: 10.1103/PhysRevX.12.021049 (cited on pp. 127, 136).

[CR18a] C. Chamberland and P. Ronagh. "Deep neural decoders for near term fault-tolerant experiments". In: *Quantum Science and Technology* 3.4 (2018), p. 044002. DOI: 10.1088/2058-9565/aad1f7 (cited on p. 115).

[CR18b] R. Chao and B. W. Reichardt. "Fault-tolerant quantum computation with few qubits". In: *npj Quantum Information* 4.1 (2018), p. 42. DOI: 10.1038/s41534-018-0085-z (cited on pp. 4, 18).

[CR18c] R. Chao and B. W. Reichardt. "Quantum error correction with only two extra qubits". In: *Physical Review Letters* 121.5 (2018), p. 050502. DOI: 10.1103/PhysRevLett.121.050502 (cited on pp. 4, 17, 18).

[CR20] R. Chao and B. W. Reichardt. "Flag fault-tolerant error correction for any stabilizer code". In: *PRX Quantum* 1.1 (2020), p. 010302. DOI: 10.1103/PRXQuantum.1.010302 (cited on pp. 4, 18, 132).

[CS96] A. R. Calderbank and P. W. Shor. "Good quantum error-correcting codes exist". In: *Physical Review A* 54.2 (1996), p. 1098. DOI: 10.1103/PhysRevA.54.1098 (cited on p. 32).

[CTV17] E. T. Campbell, B. M. Terhal, and C. Vuillot. "Roads towards fault-tolerant universal quantum computation". In: *Nature* 549.7671 (2017), p. 172. DOI: 10.1038/nature23460 (cited on pp. 3, 14, 136).

[DA07] D. P. DiVincenzo and P. Aliferis. "Effective fault-tolerant quantum computation with slow measurements". In: *Physical Review Letters* 98.2 (2007), p. 020501. DOI: 10.1103/PhysRevLett.98.020501 (cited on pp. 4, 17).

[Dan+09] C. Dankert, R. Cleve, J. Emerson, et al. "Exact and approximate unitary 2-designs and their application to fidelity estimation". In: *Physical Review A* 80.1 (2009), p. 012304. DOI: 10.1103/PhysRevA.80.012304 (cited on pp. 10, 36).

[Deb+20] D. M. Debroy, M. Li, S. Huang, et al. "Logical performance of 9 qubit compass codes in ion traps with crosstalk errors". In: *Quantum Science and Technology* 5.3 (2020), p. 034002. DOI: 10.1088/2058-9565/ab7e80 (cited on pp. 78, 79, 111).

[Dei+22] E. Deist, Y.-H. Lu, J. Ho, et al. "Mid-circuit cavity measurement in a neutral atom array". In: *Physical Review Letters* 129 (2022), p. 203602. DOI: 10.1103/PhysRevLett.129.203602 (cited on p. 113).

[Den+02] E. Dennis, A. Kitaev, A. Landahl, et al. "Topological quantum memory". In: *Journal of Mathematical Physics* 43.9 (2002), p. 4452. DOI: 10.1063/1.1499754 (cited on pp. 14, 20).

[Dev15] J. L. Devore. "Probability and statistics for engineering and the sciences". Cengage Learning, 2015. ISBN: 978-1305251809 (cited on pp. 41, 45).

[Din+23] I. Dinur, M.-H. Hsieh, T.-C. Lin, et al. "Good quantum LDPC codes with linear time decoders". In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. 2023, p. 905. DOI: 10.1145/3564246.3585101 (cited on p. 136).

[DL98] D. P. DiVincenzo and D. Loss. "Quantum information is physical". In: *Superlattices and Microstructures* 23.3-4 (1998), p. 419. DOI: 10.1006/spmi.1997.0520 (cited on p. 2).

[DMN13] S. J. Devitt, W. J. Munro, and K. Nemoto. "Quantum error correction for beginners". In: *Reports on Progress in Physics* 76.7 (2013), p. 076001. DOI: 10.1088/0034-4885/76/7/076001 (cited on pp. 2, 4, 9, 13, 17, 37, 51).

[DP16] A. De and L. P. Pryadko. "Universal set of dynamically protected gates for bipartite qubit networks: Soft pulse implementation of the [[5,1,3]] quantum error-correcting code". In: *Physical Review A* 93 (2016), p. 042333. DOI: 10.1103/PhysRevA.93.042333 (cited on pp. 27, 28).

[DP17] A. S. Darmawan and D. Poulin. "Tensor-network simulations of the surface code under realistic noise". In: *Physical Review Letters* 119.4 (2017), p. 040502. DOI: 10.1103/PhysRevLett.119.040502 (cited on p. 37).

[Drm+23] P. Drmota, D. Main, D. P. Nadlinger, et al. "Robust quantum memory in a trapped-ion quantum network node". In: *Physical Review Letters* 130 (2023), p. 090803. DOI: 10.1103/PhysRevLett.130.090803 (cited on p. 77).

[DRS21] N. Delfosse, B. W. Reichardt, and K. M. Svore. "Beyond single-shot fault-tolerant quantum error correction". In: *IEEE Transactions on Information Theory* 68.1 (2021), p. 287. DOI: 10.1109/TIT.2021.3120685 (cited on p. 136).

[Dvi+23] T. Dvir, G. Wang, N. van Loo, et al. "Realization of a minimal Kitaev chain in coupled quantum dots". In: *Nature* 614.7948 (2023), p. 445. DOI: 10.1038/s41586-022-05585-1 (cited on p. 22).

[Eas13] B. Eastin. "Distilling one-qubit magic states into Toffoli states". In: *Physical Review A* 87.3 (2013), p. 032321. DOI: 10.1103/PhysRevA.87.032321 (cited on p. 27).

[Ega+21] L. Egan, D. M. Debroy, C. Noel, et al. "Fault-tolerant control of an error-corrected qubit". In: *Nature* 598.7880 (2021), p. 281. DOI: 10.1038/s41586-021-03928-y (cited on pp. 3, 5).

[EK09] B. Eastin and E. Knill. "Restrictions on transversal encoded quantum gate sets". In: *Physical Review Letters* 102.11 (2009), p. 110502. DOI: 10.1103/PhysRevLett.102.110502 (cited on p. 20).

[Eme+07] J. Emerson, M. Silva, O. Moussa, et al. "Symmetrized characterization of noisy quantum processes". In: *Science* 317.5846 (2007), p. 1893. DOI: 10.1126/science.1145699 (cited on p. 36).

[Erc+18] H. E. Ercan, J. Ghosh, D. Crow, et al. "Measurement-free implementations of small-scale surface codes for quantum-dot qubits". In: *Physical Review A* 97.1 (2018), p. 012318. DOI: 10.1103/PhysRevA.97.012318 (cited on p. 113).

[Erh+21] A. Erhard, H. Poulsen Nautrup, M. Meth, et al. "Entangling logical qubits with lattice surgery". In: *Nature* 589.7841 (2021), p. 220. DOI: 10.1038/s41586-020-03079-6 (cited on p. 5).

[Esc+03] J. Eschner, G. Morigi, F. Schmidt-Kaler, et al. "Laser cooling of trapped ions". In: *Journal of the Optical Society of America B* 20.5 (2003), p. 1003. DOI: 10.1364/JOSAB.20.001003 (cited on p. 76).

[Esp+21] J. D. A. Espinoza, K. Groenland, M. Mazzanti, et al. "High-fidelity method for a single-step N-bit Toffoli gate in trapped ions". In: *Physical Review A* 103.5 (2021), p. 052437. DOI: 10.1103/PhysRevA.103.052437 (cited on p. 123).

[Eve+23] S. J. Evered, D. Bluvstein, M. Kalinowski, et al. "High-fidelity parallel entangling gates on a neutral atom quantum computer". 2023. DOI: 10.1038/s41586-023-06481-y (cited on pp. 3, 123, 126, 130, 131).

[Fah+23] A. Fahimniya, H. Dehghani, K. Bharti, et al. "Fault-tolerant hyperbolic Floquet quantum error correcting codes". 2023. arXiv: 2309.10033 [quant-ph] (cited on pp. 31, 136).

[Fan+22] C. Fang, Y. Wang, S. Huang, et al. "Crosstalk suppression in individually addressed two-qubit gates in a trapped-ion quantum computer". In: *Physical Review Letters* 129 (2022), p. 240504. DOI: 10.1103/PhysRevLett.129.240504 (cited on pp. 79, 113).

[Fey82] R. P. Feynman. "Simulating physics with computers". In: *International Journal of Theoretical Physics* 21.6 (1982), p. 467. DOI: 10.1007/BF02650179 (cited on p. 1).

[FF+23] M. Foss-Feig, A. Tikku, T.-C. Lu, et al. "Experimental demonstration of the advantage of adaptive quantum circuits". 2023. arXiv: 2302.03029 [quant-ph] (cited on p. 51).

[Fig+19] C. Figgatt, A. Ostrander, N. M. Linke, et al. "Parallel entangling operations on a universal ion-trap quantum computer". In: *Nature* 572.7769 (2019), p. 368. DOI: 10.1038/s41586-019-1427-5 (cited on p. 79).

[FLW02] M. H. Freedman, M. Larsen, and Z. Wang. "A modular functor which is universal for quantum computation". In: *Communications in Mathematical Physics* 227 (2002), p. 605. DOI: 10.1007/s002200200645 (cited on p. 22).

[Fow+12] A. G. Fowler, M. Mariantoni, J. M. Martinis, et al. "Surface codes: Towards practical large-scale quantum computation". In: *Physical Review A* 86 (2012), p. 032324. DOI: 10.1103/PhysRevA.86.032324 (cited on pp. 23, 26, 29, 30).

[FW20] S. T. Flammia and J. J. Wallman. "Efficient estimation of Pauli channels". In: *ACM Transactions on Quantum Computing* 1.1 (2020), p. 1. DOI: 10.1145/3408039 (cited on p. 92).

[Gae+16] J. P. Gaebler, T. R. Tan, Y. Lin, et al. "High-Fidelity Universal Gate Set for $^{9}Be^{+}$ Ion Qubits". In: *Physical Review Letters* 117 (2016), p. 060505. DOI: 10.1103/PhysRevLett.117.060505 (cited on p. 3).

[GB15] M. Gutiérrez and K. R. Brown. "Comparison of a quantum error-correction threshold for exact and approximate errors". In: *Physical Review A* 91.2 (2015), p. 022335. DOI: 10.1103/PhysRevA.91.022335 (cited on p. 37).

[GC99] D. Gottesman and I. L. Chuang. "Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations". In: *Nature* 402.6760 (1999), p. 390. DOI: 10.1038/46503 (cited on p. 23).

[Ger+21] J. M. Gertler, B. Baker, J. Li, et al. "Protecting a bosonic qubit with autonomous quantum error correction". In: *Nature* 590.7845 (2021), p. 243. DOI: 10.1038/s41586-021-03257-0 (cited on pp. 4, 113).

[GF19] C. Gidney and A. G. Fowler. "Efficient magic state factories with a catalyzed $|CCZ\rangle$ to $2|T\rangle$ transformation". In: *Quantum* 3 (2019), p. 135. DOI: 10.22331/q-2019-04-30-135 (cited on p. 27).

[GHK23] H. Goto, Y. Ho, and T. Kanao. "Measurement-free fault-tolerant logical zero-state encoding of the distance-three nine-qubit surface code in a one-dimensional qubit array". In: *Physical Review Research* 5 (2023), p. 043137. DOI: 10.1103/PhysRevResearch.5.043137 (cited on p. 115).

[Gho95] P. K. Ghosh. "Ion traps (International series of monographs on physics 90)". Clarendon Press, 1995. ISBN: 0-19-853995-9 (cited on p. 76).

[GHZ89] D. M. Greenberger, M. A. Horne, and A. Zeilinger. "Going beyond Bell's theorem". Springer, 1989, p. 69. DOI: 10.1007/978-94-017-0849-4 (cited on p. 16).

[Gid21] C. Gidney. "Stim: A fast stabilizer circuit simulator". In: *Quantum* 5 (2021), p. 497. DOI: 10.22331/q-2021-07-06-497 (cited on p. 38).

[Gid+21] C. Gidney, M. Newman, A. Fowler, et al. "A fault-tolerant honeycomb memory". In: *Quantum* 5 (2021), p. 605. DOI: 10.22331/q-2021-12-20-605 (cited on p. 31).

[GK05] C. Gerry and P. L. Knight. "Introductory quantum optics". Cambridge University Press, 2005. DOI: 10.1017/CBO9780511791239 (cited on pp. 79, 83).

[GMB19] M. Gutiérrez, M. Müller, and A. Bermúdez. "Transversality and lattice surgery: Exploring realistic routes toward coupled logical qubits with trapped-ion quantum processors". In: *Physical Review A* 99.2 (2019), p. 022330. DOI: 10.1103/PhysRevA.99.022330 (cited on pp. 42, 47, 98).

[GNM22] C. Gidney, M. Newman, and M. McEwen. "Benchmarking the planar honeycomb code". In: *Quantum* 6 (2022), p. 813. DOI: 10.22331/q-2022-09-21-813 (cited on p. 31).

[Goo21] Google Quantum AI. "Exponential suppression of bit or phase errors with cyclic error correction". In: *Nature* 595.7867 (2021), p. 383. DOI: 10.1038/s41586-021-03588-y (cited on p. 4).

[Goo23a] Google Quantum AI. "Non-abelian braiding of graph vertices in a superconducting processor". In: *Nature* (2023), p. 1. DOI: 10.1038/s41586-023-05954-4 (cited on pp. 1, 22).

[Goo23b] Google Quantum AI. "Suppressing quantum errors by scaling a surface code logical qubit". In: *Nature* 614.7949 (2023), p. 676. DOI: 10.1038/s41586-022-05434-1 (cited on pp. 4, 5, 113, 136).

[Goo60] L. A. Goodman. "On the exact variance of products". In: *Journal of the American Statistical Association* 55.292 (1960), p. 708 (cited on p. 57).

[Goo62] L. A. Goodman. "The variance of the product of K random variables". In: *Journal of the American Statistical Association* 57.297 (1962), p. 54 (cited on p. 57).

[Got09] D. Gottesman. "An Introduction to quantum error correction and fault-tolerant quantum computation". 2009. arXiv: 0904.2557 [quant-ph] (cited on pp. 2, 3, 13).

[Got14] D. Gottesman. "Fault-tolerant quantum computation with constant overhead". 2014. arXiv: 1310.2984 [quant-ph] (cited on pp. 19, 20).

[Got16a] H. Goto. "Minimizing resource overheads for fault-tolerant preparation of encoded states of the Steane code". In: *Scientific Reports* 6.1 (2016), p. 19578. DOI: 10.1038/srep19578 (cited on pp. 4, 18, 28, 51, 66, 67, 87, 93, 100, 115, 116).

[Got16b] D. Gottesman. "Quantum fault tolerance in small experiments". 2016. arXiv: 1610.03507 [quant-ph] (cited on pp. 19, 98).

[Got97] D. Gottesman. "Stabilizer codes and quantum error correction". California Institute of Technology, 1997. URL: https://thesis.library.caltech.edu/2900/2/THESIS.pdf (cited on p. 9).

[Got98a] D. Gottesman. "The Heisenberg representation of quantum computers". 1998. arXiv: quant-ph/9807006 [quant-ph] (cited on pp. 4, 10, 38).

[Got98b] D. Gottesman. "Theory of fault-tolerant quantum computation". In: *Physical Review A* 57 (1998), p. 127. DOI: 10.1103/PhysRevA.57.127 (cited on p. 3).

[Gra+22] T. M. Graham, Y. Song, J. Scott, et al. "Multi-qubit entanglement and algorithms on a neutral-atom quantum computer". In: *Nature* 604.7906 (2022), p. 457. DOI: 10.1038/s41586-022-04603-6 (cited on pp. 3, 126).

[Gra+23] T. M. Graham, L. Phuttitarn, R. Chinnarasu, et al. "Mid-circuit measurements on a neutral atom quantum processor". 2023. arXiv: 2303.10051 [quant-ph] (cited on pp. 5, 113, 130).

[Gro96] L. K. Grover. "A fast quantum mechanical algorithm for database search". In: *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*. 1996, p. 212. DOI: 10.1145/237814.237866 (cited on p. 2).

[GSF21] H. Goldman, R. Sohal, and E. Fradkin. "Composite particle construction of the Fibonacci fractional quantum Hall state". In: *Physical Review B* 103.23 (2021), p. 235118. DOI: 10.1103/PhysRevB.103.235118 (cited on p. 22).

[Gu+23] S. Gu, E. Tang, L. Caha, et al. "Single-shot decoding of good quantum LDPC codes". 2023. arXiv: 2306.12470 [quant-ph] (cited on p. 136).

[Gup+23] R. S. Gupta, N. Sundaresan, T. Alexander, et al. "Encoding a magic state with beyond break-even fidelity". 2023. arXiv: 2305.13581 [quant-ph] (cited on pp. 24, 87).

[Gut+16] M. Gutiérrez, C. Smith, L. Lulushi, et al. "Errors and pseudothresholds for incoherent and coherent noise". In: *Physical Review A* 94.4 (2016), p. 042338. DOI: 10.1103/PhysRevA.94.042338 (cited on p. 37).

[GZ13] M. R. Geller and Z. Zhou. "Efficient error models for fault-tolerant architectures and the Pauli twirling approximation". In: *Physical Review A* 88.1 (2013), p. 012314. DOI: 10.1103/PhysRevA.88.012314 (cited on p. 36).

[Haa+17] J. Haah, M. B. Hastings, D. Poulin, et al. "Magic state distillation with low space overhead and optimal asymptotic input count". In: *Quantum* 1 (2017), p. 31. DOI: 10.22331/q-2017-10-03-31 (cited on p. 27).

[Haz+21] S. Hazra, A. Bhattacharjee, M. Chand, et al. "Ring-resonator-based coupling architecture for enhanced connectivity in a superconducting multiqubit network". In: *Physical Review Applied* 16 (2021), p. 024018. DOI: 10.1103/PhysRevApplied.16.024018 (cited on p. 137).

[HC17] M. Howard and E. Campbell. "Application of a resource theory for magic states to fault-tolerant quantum computing". In: *Physical Review Letters* 118.9 (2017), p. 090501. DOI: 10.1103/PhysRevLett.118.090501 (cited on p. 27).

[HDF19] E. Huang, A. C. Doherty, and S. Flammia. "Performance of quantum error correction with coherent errors". In: *Physical Review A* 99 (2019), p. 022313. DOI: 10.1103/PhysRevA.99.022313 (cited on p. 136).

[Hel+22] J. Helsen, I. Roth, E. Onorati, et al. "General framework for randomized benchmarking". In: *PRX Quantum* 3.2 (2022), p. 020357. DOI: 10.1103/PRXQuantum.3.020357 (cited on p. 10).

[Hen+20] L. Henriet, L. Beguin, A. Signoles, et al. "Quantum computing with neutral atoms". In: *Quantum* 4 (2020), p. 327. DOI: 10.22331/q-2020-09-21-327 (cited on p. 3).

[Heu+23a] S. Heußen, L. Postler, M. Rispler, et al. "Strategies for a practical advantage of fault-tolerant circuit design in noisy trapped-ion quantum computers". In: *Physical Review A* 107.4 (2023), p. 042422. DOI: 10.1103/PhysRevA.107.042422 (cited on pp. ix, 42, 51, 66, 77–79, 81, 97, 98, 100, 112).

[Heu+23b] S. Heußen, D. Winter, M. Rispler, et al. "Dynamical subset sampling of quantum error correcting protocols". 2023. arXiv: 2309.12774 [quant-ph] (cited on pp. ix, 42, 51).

[HFW20] R. Harper, S. T. Flammia, and J. J. Wallman. "Efficient learning of quantum noise". In: *Nature Physics* 16.12 (2020), p. 1184. DOI: 10.1038/s41567-020-0992-8 (cited on p. 92).

[HH18a] J. Haah and M. B. Hastings. "Codes and protocols for distilling T, controlled-S, and Toffoli gates". In: *Quantum* 2 (2018), p. 71. DOI: 10.22331/q-2018-06-07-71 (cited on p. 27).

[HH18b] M. B. Hastings and J. Haah. "Distillation with sublogarithmic overhead". In: *Physical Review Letters* 120.5 (2018), p. 050504. DOI: 10.1103/PhysRevLett.120.050504 (cited on p. 27).

[HH21a] J. Haah and M. B. Hastings. "Measurement sequences for magic state distillation". In: *Quantum* 5 (2021), p. 383. DOI: 10.22331/q-2021-01-20-383 (cited on p. 28).

[HH21b] M. B. Hastings and J. Haah. "Dynamically generated logical qubits". In: *Quantum* 5 (2021), p. 564. DOI: 10.22331/q-2021-10-19-564 (cited on p. 31).

[Hil+22] J. Hilder, D. Pijn, O. Onishchenko, et al. "Fault-tolerant parity readout on a shuttling-based trapped-ion quantum computer". In: *Physical Review X* 12.1 (2022), p. 011032. DOI: 10.1103/PhysRevX.12.011032 (cited on pp. 4, 68, 113, 137).

[HLM23] S. Heußen, D. F. Locher, and M. Müller. "Measurement-free fault-tolerant quantum error correction in near-term devices". 2023. arXiv: 2307.13296 [quant-ph] (cited on pp. ix, 113, 118, 123).

[HMM22] P. M. Harrington, E. J. Mueller, and K. W. Murch. "Engineered dissipation for quantum information science". In: *Nature Reviews Physics* 4.10 (2022), p. 660. DOI: 10.1038/s42254-022-00494-8 (cited on pp. 4, 113).

[Hol+20] P. C. Holz, S. Auchter, G. Stocker, et al. "2D linear trap array for quantum information processing". In: *Advanced Quantum Technologies* 3.11 (2020), p. 2000031. DOI: 10.1002/qute.202000031 (cited on p. 137).

[Hon+23] Y. Hong, M. Marinelli, A. M. Kaufman, et al. "Long-range-enhanced surface codes". 2023. arXiv: 2309.11719 [quant-ph] (cited on p. 137).

[Hor+12] C. Horsman, A. G. Fowler, S. Devitt, et al. "Surface code quantum computing by lattice surgery". In: *New Journal of Physics* 14.12 (2012), p. 123011. DOI: 10.1088/1367-2630/14/12/123011 (cited on p. 31).

[Hra+04] Z. Hradil, J. Řeháček, J. Fiurášek, et al. "Maximum-likelihood methods in quantum mechanics". In: *Quantum state estimation*. Springer, 2004, p. 59. DOI: 10.1007/978-3-540-44481-7_3 (cited on p. 97).

[HRB08] H. Häffner, C. Roos, and R. Blatt. "Quantum computing with trapped ions". In: *Physics Reports* 469.4 (2008), p. 155. DOI: 10.1016/j.physrep.2008.09.003 (cited on p. 75).

[Hui+23] W. Huie, L. Li, N. Chen, et al. "Repetitive readout and real-time control of nuclear spin qubits in ^{171}Yb atoms". 2023. DOI: 10.1103/PRXQuantum.4.030337 (cited on pp. 5, 113, 130).

[IBMapa] IBM. "On 'quantum supremacy'". www.ibm.com/blogs/research/2019/10/on-quantum-supremacy. (snapshot on 2 Aug 2021) (cited on p. 3).

[IBMapb] IBM. "The IBM quantum development roadmap". www.ibm.com/quantum/roadmap. (snapshot on 4 Sep 2023) (cited on p. 137).

[IP20] J. K. Iverson and J. Preskill. "Coherence in logical quantum channels". In: *New Journal of Physics* 22.7 (2020), p. 073066. DOI: 10.1088/1367-2630/ab8e5c (cited on p. 14).

[Iqb+23] M. Iqbal, N. Tantivasadakarn, R. Verresen, et al. "Creation of non-abelian topological order and anyons on a trapped-ion processor". 2023. arXiv: 2305.03766 [quant-ph] (cited on pp. 1, 22).

[ISM11] L. Isenhower, M. Saffman, and K. Mølmer. "Multibit C_kNOT quantum gates via Rydberg blockade". In: *Quantum Information Processing* 10.6 (2011), p. 755. DOI: 10.1007/s11128-011-0292-4 (cited on p. 123).

[Jai+20] S. Jain, J. Alonso, M. Grau, et al. "Scalable arrays of micro-Penning traps for quantum computing and simulation". In: *Physical Review X* 10.3 (2020), p. 031027. DOI: 10.1103/PhysRevX.10.031027 (cited on p. 137).

[Jai+23] S. Jain, T. Sägesser, P. Hrmo, et al. "Unit cell of a Penning microtrap quantum processor". 2023. arXiv: 2308.07672 [quant-ph] (cited on p. 137).

[Jen+22] A. Jenkins, J. W. Lis, A. Senoo, et al. "Ytterbium nuclear-spin qubits in an optical tweezer array". In: *Physical Review X* 12.2 (2022), p. 021027. DOI: 10.1103/PhysRevX.12.021027 (cited on p. 3).

[JO+12] T. Jochym-O'Connor, Y. Yu, B. Helou, et al. "The robustness of magic state distillation against errors in Clifford gates". 2012. arXiv: 1205.6715 [quant-ph] (cited on p. 28).

[Jon13] C. Jones. "Multilevel distillation of magic states for quantum computing". In: *Physical Review A* 87.4 (2013), p. 042305. DOI: 10.1103/PhysRevA.87.042305 (cited on p. 27).

[JP22] S. Jandura and G. Pupillo. "Time-optimal two- and three-qubit gates for Rydberg atoms". In: *Quantum* 6 (2022), p. 712. DOI: 10.22331/q-2022-05-13-712 (cited on p. 130).

[Kat17] A. Katabarwa. "A dynamical interpretation of the Pauli twirling approximation and quantum error correction". 2017. arXiv: 1701.03708 [quant-ph] (cited on p. 37).

[Kat19] A. Katabarwa. "Near term quantum computation". University of Georgia, 2019. URL: https://esploro.libs.uga.edu/esploro/outputs/9949366062402959 (cited on p. 136).

[KB15] A. Kubica and M. E. Beverland. "Universal transversal gates with color codes: A simplified approach". In: *Physical Review A* 91.3 (2015), p. 032330. DOI: 10.1103/PhysRevA.91.032330 (cited on p. 31).

[KBMD09] H. G. Katzgraber, H. Bombín, and M. Martín-Delgado. "Error threshold for color codes and random three-body Ising models". In: *Physical Review Letters* 103.9 (2009), p. 090501. DOI: 10.1103/PhysRevLett.103.090501 (cited on p. 14).

[KG15] A. Katabarwa and M. R. Geller. "Logical error rate in the Pauli twirling approximation". In: *Scientific Reports* 5.1 (2015), p. 1. DOI: 10.1038/srep14670 (cited on p. 37).

[Kim+23] Y. Kim, A. Eddins, S. Anand, et al. "Evidence for the utility of quantum computing before fault tolerance". In: *Nature* 618.7965 (2023), p. 500. DOI: 10.1038/s41586-023-06096-3 (cited on p. 3).

[Kit01] A. Y. Kitaev. "Unpaired Majorana fermions in quantum wires". In: *Physics-Uspekhi* 44.10S (2001), p. 131. DOI: 10.1070/1063-7869/44/10S/S29 (cited on p. 22).

[Kit03] A. Y. Kitaev. "Fault-tolerant quantum computation by anyons". In: *Annals of Physics* 303.1 (2003), p. 2. DOI: 10.1016/S0003-4916(02)00018-0 (cited on p. 22).

[Kit97] A. Y. Kitaev. "Quantum computations: algorithms and error correction". In: *Russian Mathematical Surveys* 52.6 (1997), p. 1191. DOI: 10.1070/RM1997v052n06ABEH002155 (cited on pp. 8, 19, 22).

[Kja+20] M. Kjaergaard, M. E. Schwartz, J. Braumüller, et al. "Superconducting qubits: Current state of play". In: *Annual Review of Condensed Matter Physics* 11 (2020), p. 369. DOI: 10.1146/annurev-conmatphys-031119-050605 (cited on p. 3).

[KL97] E. Knill and R. Laflamme. "Theory of quantum error-correcting codes". In: *Physical Review A* 55.2 (1997), p. 900. DOI: 10.1103/PhysRevA.55.900 (cited on p. 10).

[KLZ96] E. Knill, R. Laflamme, and W. Zurek. "Threshold accuracy for quantum computation". 1996. arXiv: quant-ph/9610011 [quant-ph] (cited on p. 13).

[KM20] M. Khazali and K. Mølmer. "Fast multiqubit gates by adiabatic evolution in interacting excited-state manifolds of Rydberg atoms and superconducting circuits". In: *Physical Review X* 10 (2020), p. 021054. DOI: 10.1103/PhysRevX.10.021054 (cited on p. 123).

[KML23] M. Kalinowski, N. Maskara, and M. D. Lukin. "Non-abelian Floquet spin liquids in a digital Rydberg simulator". In: *Physical Review X* 13.3 (2023), p. 031008. DOI: 10.1103/PhysRevX.13.031008 (cited on p. 137).

[KMW02] D. Kielpinski, C. Monroe, and D. J. Wineland. "Architecture for a large-scale ion-trap quantum computer". In: *Nature* 417.6890 (2002), p. 709. DOI: 10.1038/nature00784 (cited on pp. 77, 113).

[Kni04] E. Knill. "Fault-tolerant postselected quantum computation: Schemes". 2004. arXiv: quant-ph/0402171 [quant-ph] (cited on pp. 16, 25).

[Kni05] E. Knill. "Quantum computing with realistically noisy devices". In: *Nature* 434.7029 (2005), p. 39. DOI: 10.1038/nature03350 (cited on pp. 11, 16).

[Kok+07] P. Kok, W. J. Munro, K. Nemoto, et al. "Linear optical quantum computing with photonic qubits". In: *Reviews of Modern Physics* 79 (2007), p. 135. DOI: 10.1103/RevModPhys.79.135 (cited on p. 3).

[Kri+22] S. Krinner, N. Lacroix, A. Remm, et al. "Realizing repeated quantum error correction in a distance-three surface code". In: *Nature* 605.7911 (2022), p. 669. DOI: 10.1038/s41586-022-04566-8 (cited on pp. 4, 113).

[Kub18] A. Kubica. "The ABCs of the color code: A study of topological quantum codes as toy models for fault-tolerant quantum computation and quantum phases of matter". California Institute of Technology, 2018. URL: https://thesis.library.caltech.edu/10955/ (cited on p. 20).

[KZ23] R. Kobayashi and G. Zhu. "Fault-tolerant logical gates via constant depth circuits and emergent symmetries on non-orientable topological

stabilizer and Floquet codes". 2023. arXiv: 2310.06917 [quant-ph] (cited on p. 136).

[LA20] L. Lao and C. G. Almudever. "Fault-tolerant quantum error correction on near-term quantum processors using flag and bridge qubits". In: *Physical Review A* 101.3 (2020), p. 032333. DOI: 10.1103/PhysRevA.101.032333 (cited on p. 18).

[Laf+96] R. Laflamme, C. Miquel, J. P. Paz, et al. "Perfect quantum error correction code". 1996. arXiv: quant-ph/9602019 [quant-ph] (cited on p. 25).

[Lan+19] K. A. Landsman, Y. Wu, P. H. Leung, et al. "Two-qubit entangling gates within arbitrarily long chains of trapped ions". In: *Physical Review A* 100 (2019), p. 022332. DOI: 10.1103/PhysRevA.100.022332 (cited on p. 77).

[Lan91] R. Landauer. "Information is physical". In: *Physics Today* 44.5 (1991), p. 23. DOI: 10.1063/1.881299 (cited on p. 1).

[LAR11] A. J. Landahl, J. T. Anderson, and P. R. Rice. "Fault-tolerant quantum computing with color codes". 2011. arXiv: 1108.5738 [quant-ph] (cited on p. 14).

[LB13] D. A. Lidar and T. A. Brun. "Quantum error correction". Cambridge University Press, 2013. DOI: 10.1017/CBO9781139034807 (cited on p. 4).

[LB18] P. H. Leung and K. R. Brown. "Entangling an arbitrary pair of qubits in a long ion crystal". In: *Physical Review A* 98 (2018), p. 032318. DOI: 10.1103/PhysRevA.98.032318 (cited on p. 77).

[Lei+03] D. Leibfried, R. Blatt, C. Monroe, et al. "Quantum dynamics of single trapped ions". In: *Reviews of Modern Physics* 75 (2003), p. 281. DOI: 10.1103/RevModPhys.75.281 (cited on p. 75).

[Lek+17] B. Lekitsch, S. Weidt, A. G. Fowler, et al. "Blueprint for a microwave trapped ion quantum computer". In: *Science Advances* 3.2 (2017), e1601540. DOI: 10.1126/sciadv.1601540 (cited on p. 76).

[Leu+97] D. W. Leung, M. A. Nielsen, I. L. Chuang, et al. "Approximate quantum error correction can lead to better codes". In: *Physical Review A* 56.4 (1997), p. 2567. DOI: 10.1103/PhysRevA.56.2567 (cited on p. 136).

[Lev+19] H. Levine, A. Keesling, G. Semeghini, et al. "Parallel implementation of high-fidelity multiqubit gates with neutral atoms". In: *Physical Review Letters* 123 (2019), p. 170503. DOI: 10.1103/PhysRevLett.123.170503 (cited on pp. 3, 123, 130).

[Li+17] M. Li, M. Gutiérrez, S. E. David, et al. "Fault tolerance with bare ancillary qubits for a [[7, 1, 3]] code". In: *Physical Review A* 96.3 (2017), p. 032341. DOI: 10.1103/PhysRevA.96.032341 (cited on pp. 42, 47, 79).

[Li20] M. Li. "Fault-tolerance on near-term quantum computers and subsystem quantum error correcting codes". Georgia Institute of Technology, 2020. URL: http://hdl.handle.net/1853/62750 (cited on p. 42).

[Lia+23] H.-J. Liao, K. Wang, Z.-S. Zhou, et al. "Simulation of IBM's kicked Ising experiment with Projected Entangled Pair Operator". 2023. arXiv: 2308.03082 [quant-ph] (cited on p. 3).

[Lin+17] N. M. Linke, M. Gutierrez, K. A. Landsman, et al. "Fault-tolerant quantum error detection". In: *Science Advances* 3.10 (2017), e1701074. DOI: 10.1126/sciadv.1701074 (cited on p. 4).

[Lis+23] J. W. Lis, A. Senoo, W. F. McGrew, et al. "Mid-circuit operations using the omg-architecture in neutral atom arrays". 2023. arXiv: 2305.19266 [quant-ph] (cited on pp. 5, 113, 130).

[Lit19] D. Litinski. "Magic state distillation: Not as costly as you think". In: *Quantum* 3 (2019), p. 205. DOI: 10.22331/q-2019-12-02-205 (cited on p. 27).

[Liu+21] Y. Liu, X. Liu, F. Li, et al. "Closing the 'quantum supremacy' gap: achieving real-time simulation of a random quantum circuit using a new sunway supercomputer". In: *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. 2021, p. 1. DOI: 10.1145/3458817.3487399 (cited on p. 3).

[LZ22] A. Leverrier and G. Zémor. "Quantum tanner codes". In: *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science*. 2022, p. 872. DOI: 10.1109/FOCS54457.2022.00117 (cited on p. 136).

[Ma+22] S. Ma, A. P. Burgers, G. Liu, et al. "Universal gate operations on nuclear spin qubits in an optical tweezer array of ^{171}Yb atoms". In: *Physical Review X* 12.2 (2022), p. 021028. DOI: 10.1103/PhysRevX.12.021028 (cited on p. 3).

[MAB23] M. Malinowski, D. Allcock, and C. Ballance. "How to wire a 1000-qubit trapped-ion quantum computer". In: *PRX Quantum* 4 (2023), p. 040313. DOI: 10.1103/PRXQuantum.4.040313 (cited on p. 137).

[Mad+22] L. S. Madsen, F. Laudenbach, M. F. Askarani, et al. "Quantum computational advantage with a programmable photonic processor". In: *Nature* 606.7912 (2022), p. 75. DOI: 10.1038/s41586-022-04725-x (cited on p. 3).

[Mar+16] E. A. Martinez, T. Monz, D. Nigg, et al. "Compiling quantum algorithms for architectures with multi-qubit gates". In: *New Journal of Physics* 18.6 (2016), p. 063029. DOI: 10.1088/1367-2630/18/6/063029 (cited on p. 124).

[Mar+22] J. F. Marques, B. M. Varbanov, M. S. Moreira, et al. "Logical-qubit operations in an error-detecting surface code". In: *Nature Physics* 18.1 (2022), p. 80. DOI: 10.1038/s41567-021-01423-9 (cited on p. 5).

[Mas16] D. Maslov. "Advantages of using relative-phase Toffoli gates with an application to multiple control Toffoli optimization". In: *Physical Review A* 93.2 (2016), p. 022311. DOI: 10.1103/PhysRevA.93.022311 (cited on p. 116).

[Mas17] D. Maslov. "Basic circuit compilation techniques for an ion-trap quantum machine". In: *New Journal of Physics* 19.2 (2017), p. 023035. DOI: 10.1088/1367-2630/aa5e47 (cited on p. 78).

[McA+20] S. McArdle, S. Endo, A. Aspuru-Guzik, et al. "Quantum computational chemistry". In: *Reviews of Modern Physics* 92 (2020), p. 015003. DOI: 10.1103/RevModPhys.92.015003 (cited on p. 2).

[MEK12] A. M. Meier, B. Eastin, and E. Knill. "Magic-state distillation with the four-qubit code". 2012. arXiv: 1204.4221 [quant-ph] (cited on pp. 25–27, 94, 137).

[Mer+14] J. T. Merrill, S. C. Doret, G. Vittorini, et al. "Transformed composite sequences for improved qubit addressing". In: *Physical Review A* 90.4 (2014), p. 040301. DOI: 10.1103/PhysRevA.90.040301 (cited on p. 79).

[MG+22] F. Martínez-García, L. Gerster, D. Vodola, et al. "Analytical and experimental study of center-line miscalibrations in Mølmer-Sørensen gates". In: *Physical Review A* 105.3 (2022), p. 032437. DOI: 10.1103/PhysRevA.105.032437 (cited on p. 83).

[MGE12] E. Magesan, J. M. Gambetta, and J. Emerson. "Characterizing quantum gates via randomized benchmarking". In: *Physical Review A* 85.4 (2012), p. 042311. DOI: 10.1103/PhysRevA.85.042311 (cited on p. 10).

[Mi+22] X. Mi, M. Ippoliti, C. Quintana, et al. "Time-crystalline eigenstate order on a quantum processor". In: *Nature* 601.7894 (2022), p. 531. DOI: 10.1038/s41586-021-04257-w (cited on p. 1).

[Mia+23] K. C. Miao, M. McEwen, J. Atalaya, et al. "Overcoming leakage in quantum error correction". In: *Nature Physics* (2023), p. 1. DOI: doi.org/10.1038/s41567-023-02226-w (cited on p. 37).

[Mon+09] T. Monz, K. Kim, W. Hänsel, et al. "Realization of the quantum Toffoli gate with trapped ions". In: *Physical Review Letters* 102 (2009), p. 040501. DOI: 10.1103/PhysRevLett.102.040501 (cited on p. 123).

[Mon+14] C. Monroe, R. Raussendorf, A. Ruthven, et al. "Large-scale modular quantum-computer architecture with atomic memory and photonic interconnects". In: *Physical Review A* 89.2 (2014), p. 022317. DOI: 10.1103/PhysRevA.89.022317 (cited on p. 77).

[Mon16] A. Montanaro. "Quantum algorithms: An overview". In: *npj Quantum Information* 2.1 (2016), p. 1. DOI: 10.1038/npjqi.2015.23 (cited on p. 2).

[Mos+23] S. Moses, C. Baldwin, M. Allman, et al. "A race track trapped-ion quantum processor". 2023. arXiv: 2305.03828 [quant-ph] (cited on pp. 2, 3, 113, 127).

[MRV23] D. H. Menendez, A. Ray, and M. Vasmer. "Implementing fault-tolerant non-Clifford gates using the [[8,3,2]] color code". 2023. arXiv: 2309.08663 [quant-ph] (cited on p. 87).

[MS99] K. Mølmer and A. Sørensen. "Multiparticle entanglement of hot trapped ions". In: *Physical Review Letters* 82.9 (1999), p. 1835. DOI: 10.1103/PhysRevLett.82.1835 (cited on pp. 76, 82).

[Mül+09] M. Müller, I. Lesanovsky, H. Weimer, et al. "Mesoscopic Rydberg gate based on electromagnetically induced transparency". In: *Physical Review Letters* 102 (2009), p. 170502. DOI: 10.1103/PhysRevLett.102.170502 (cited on p. 123).

[Mül+11] M. Müller, K. Hammerer, Y. L. Zhou, et al. "Simulating open quantum systems: from many-body interactions to stabilizer pumping". In: *New Journal of Physics* 13.8 (2011), p. 085007. DOI: 10.1088/1367-2630/13/8/085007 (cited on p. 124).

[MW01] F. Mintert and C. Wunderlich. "Ion-trap quantum logic using long-wavelength radiation". In: *Physical Review Letters* 87.25 (2001), p. 257904. DOI: 10.1103/PhysRevLett.87.257904 (cited on p. 76).

[Nay+08] C. Nayak, S. H. Simon, A. Stern, et al. "Non-abelian anyons and topological quantum computation". In: *Reviews of Modern Physics* 80.3 (2008), p. 1083. DOI: 10.1103/RevModPhys.80.1083 (cited on p. 22).

[NBN15] X. Ni, O. Buerschaper, and M. Van den Nest. "A non-commuting stabilizer formalism". In: *Journal of Mathematical Physics* 56.5 (2015). DOI: 10.1063/1.4920923 (cited on p. 39).

[NC10] M. A. Nielsen and I. L. Chuang. "Quantum computation and quantum information: 10th anniversary edition". Cambridge University Press, 2010. DOI: 10.1017/CBO9780511976667 (cited on pp. 1, 3, 4, 7, 8, 10, 11, 23, 35, 116).

[ND05] M. A. Nielsen and C. M. Dawson. "Fault-tolerant quantum computation with cluster states". In: *Physical Review A* 71.4 (2005), p. 042323. DOI: 10.1103/PhysRevA.71.042323 (cited on p. 4).

[Nig+14] D. Nigg, M. Müller, E. A. Martinez, et al. "Quantum computations on a topologically encoded qubit". In: *Science* 345.6194 (2014), p. 302. DOI: 10.1126/science.1253742 (cited on pp. 4, 12, 33, 79).

[Nor+23] M. A. Norcia, W. B. Cairncross, K. Barnes, et al. "Mid-circuit qubit measurement and rearrangement in a ^{171}Yb atomic array". 2023. arXiv: 2305.19119 [quant-ph] (cited on pp. 5, 113, 130).

[NP12] J. Napp and J. Preskill. "Optimal Bacon-Shor codes". 2012. arXiv: 1209.0794 [quant-ph] (cited on p. 13).

[OC17] J. O'Gorman and E. T. Campbell. "Quantum computation with realistic magic-state factories". In: *Physical Review A* 95.3 (2017), p. 032338. DOI: 10.1103/PhysRevA.95.032338 (cited on p. 27).

[Ohn+04] T. Ohno, G. Arakawa, I. Ichinose, et al. "Phase structure of the random-plaquette Z_2 gauge model: accuracy threshold for a toric quantum memory". In: *Nuclear Physics B* 697.3 (2004), p. 462. DOI: 10.1016/j.nuclphysb.2004.07.003 (cited on p. 14).

[Oze11] R. Ozeri. "The trapped-ion qubit tool box". In: *Contemporary Physics* 52.6 (2011), p. 531. DOI: 10.1080/00107514.2011.603578 (cited on pp. 75, 76).

[Pae+23] A. Paetznick, C. Knapp, N. Delfosse, et al. "Performance of planar Floquet codes with Majorana-based qubits". In: *PRX Quantum* 4 (2023), p. 010310. DOI: 10.1103/PRXQuantum.4.010310 (cited on pp. 30, 31, 42, 47, 56).

[Pag+22] A. Pagano, S. Weber, D. Jaschke, et al. "Error budgeting for a controlled-phase gate with strontium-88 Rydberg atoms". In: *Physical Review Research* 4 (2022), p. 033019. DOI: 10.1103/PhysRevResearch.4.033019 (cited on p. 130).

[Pal+17] A. Paler, I. Polian, K. Nemoto, et al. "Fault-tolerant, high-level quantum circuits: form, compilation and description". In: *Quantum Science and Technology* 2.2 (2017), p. 025003. DOI: 10.1088/2058-9565/aa66eb (cited on p. 15).

[Pal+22] A. K. Pal, P. Schindler, A. Erhard, et al. "Relaxation times do not capture logical qubit dynamics". In: *Quantum* 6 (2022), p. 632. DOI: 10.22331/q-2022-01-24-632 (cited on pp. 78, 136).

[PCZ22] F. Pan, K. Chen, and P. Zhang. "Solving the sampling problem of the Sycamore quantum circuits". In: *Physical Review Letters* 129.9 (2022), p. 090502. DOI: 10.1103/PhysRevLett.129.090502 (cited on p. 3).

[PDP22] G. Pelegrí, A. J. Daley, and J. D. Pritchard. "High-fidelity multiqubit Rydberg gates via two-photon adiabatic rapid passage". In: *Quantum Science and Technology* 7.4 (2022), p. 045020. DOI: 10.1088/2058-9565/ac823a (cited on p. 123).

[Ped+19] E. Pednault, J. A. Gunnels, G. Nannicini, et al. "Leveraging secondary storage to simulate deep 54-qubit Sycamore circuits". 2019. arXiv: 1910.09534 [quant-ph] (cited on p. 3).

[Per+23] M. A. Perlin, V. N. Premakumar, J. Wang, et al. "Fault-tolerant measurement-free quantum error correction with multi-qubit gates". 2023. arXiv: 2007.09804 [quant-ph] (cited on p. 114).

[Pil+14] C. Piltz, T. Sriarunothai, A. Varón, et al. "A trapped-ion-based quantum byte with 10^{-5} next-neighbour cross-talk". In: *Nature Communications* 5.1 (2014), p. 4679. DOI: 10.1038/ncomms5679 (cited on p. 79).

[Pin+21] J. M. Pino, J. M. Dreiling, C. Figgatt, et al. "Demonstration of the trapped-ion quantum CCD computer architecture". In: *Nature* 592.7853 (2021), p. 209. DOI: 10.1038/s41586-021-03318-4 (cited on pp. 77, 113, 123, 127).

[PK22] P. Panteleev and G. Kalachev. "Asymptotically good quantum and locally testable classical LDPC codes". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. 2022, p. 375. DOI: 10.1145/3519935.3520017 (cited on p. 136).

[PL22] B. Pokharel and D. Lidar. "Better-than-classical Grover search via quantum error detection and suppression". 2022. arXiv: 2211.04543 [quant-ph] (cited on p. 138).

[Pog+21] I. Pogorelov, T. Feldker, C. D. Marciniak, et al. "Compact ion-trap quantum computing demonstrator". In: *PRX Quantum* 2 (2021), p. 020343. DOI: 10.1103/PRXQuantum.2.020343 (cited on pp. 3, 76–78, 86, 113).

[Pos+18] L. Postler, Á. Rivas, P. Schindler, et al. "Experimental quantification of spatial correlations in quantum dynamics". In: *Quantum* 2 (2018), p. 90. DOI: 10.22331/q-2018-09-03-90 (cited on p. 78).

[Pos+22] L. Postler, S. Heußen, I. Pogorelov, et al. "Demonstration of fault-tolerant universal quantum gate operations". In: *Nature* 605.7911 (2022), p. 675. DOI: 10.1038/s41586-022-04721-1 (cited on pp. ix, 5, 33, 75, 79, 86, 89, 136, 138).

[PR13] A. Paetznick and B. W. Reichardt. "Fault-tolerant ancilla preparation and noise threshold lower bounds for the 23-qubit Golay code". 2013. arXiv: 1106.2190 [quant-ph] (cited on p. 115).

[PR+21] P. Parrado-Rodríguez, C. Ryan-Anderson, A. Bermudez, et al. "Crosstalk suppression for fault-tolerant quantum error correction with trapped ions". In: *Quantum* 5 (2021), p. 487. DOI: 10.22331/q-2021-06-29-487 (cited on pp. 37, 79).

[PR21] P. Prabhu and B. W. Reichardt. "Fault-tolerant syndrome extraction and cat state preparation with fewer qubits". In: *16th Conference on the Theory of Quantum Computation, Communication and Cryptography*. 2021, 5:1. DOI: 10.4230/LIPIcs.TQC.2021.5 (cited on pp. 17, 18).

[Pre18] J. Preskill. "Quantum computing in the NISQ era and beyond". In: *Quantum* 2 (2018), p. 79. DOI: 10.22331/q-2018-08-06-79 (cited on p. 2).

[Pre98] J. Preskill. "Reliable quantum computers". In: *Proceedings of the Royal Society A* 454.1969 (1998), p. 385. DOI: 10.1098/rspa.1998.0167 (cited on pp. 2, 14, 16, 23).

[PSBT10] G. A. Paz-Silva, G. K. Brennen, and J. Twamley. "Fault tolerance with noisy and slow measurements and preparation". In: *Physical Review Letters* 105.10 (2010), p. 100501. DOI: 10.1103/PhysRevLett.105.100501 (cited on p. 113).

[Puz+14] D. Puzzuoli, C. Granade, H. Haas, et al. "Tractable simulation of error correction with honest approximations to realistic fault models". In: *Physical Review A* 89.2 (2014), p. 022306. DOI: 10.1103/PhysRevA.89.022306 (cited on pp. 37, 136).

[RA18] C. Ryan-Anderson. "Quantum algorithms, architecture, and error correction". The University of New Mexico, 2018. URL: https://digitalrepository.unm.edu/phyc_etds/203/ (cited on p. 75).

[RA19] C. Ryan-Anderson. "PECOS: Performance estimator of codes on surfaces". `https://github.com/PECOS-packages/PECOS`. 2019 (cited on p. 75).

[RA+21] C. Ryan-Anderson, J. Bohnet, K. Lee, et al. "Realization of real-time fault-tolerant quantum error correction". In: *Physical Review X* 11.4 (2021), p. 041058. DOI: `10.1103/PhysRevX.11.041058` (cited on pp. 4, 5, 33, 51, 68, 113, 123, 136).

[RA+22] C. Ryan-Anderson, N. C. Brown, M. S. Allman, et al. "Implementing fault-tolerant entangling gates on the five-qubit code and the color code". 2022. arXiv: `2208.01863 [quant-ph]` (cited on pp. 5, 33).

[RAK18] A. V. Rynbach, M. Ahsan, and J. Kim. "A quantum computing performance simulator based on circuit failure probability and fault path counting". In: *ACM Journal on Emerging Technologies in Computing Systems* 14.1 (2018), p. 1 (cited on p. 39).

[Ram+23] J. Ramette, J. Sinclair, N. P. Breuckmann, et al. "Fault-tolerant connection of error-corrected qubits with noisy links". 2023. arXiv: `2302.01296 [quant-ph]` (cited on p. 137).

[Ras+20] S. E. Rasmussen, K. Groenland, R. Gerritsma, et al. "Single-step implementation of high-fidelity n-bit Toffoli gates". In: *Physical Review A* 101.2 (2020), p. 022308. DOI: `10.1103/PhysRevA.101.022308` (cited on p. 123).

[Rei05] B. W. Reichardt. "Quantum universality from magic states distillation applied to CSS codes". In: *Quantum Information Processing* 4.3 (2005), p. 251. DOI: `10.1007/s11128-005-7654-8` (cited on pp. 24, 26).

[Rei20] B. W. Reichardt. "Fault-tolerant quantum error correction for Steane's seven-qubit color code with few or no extra qubits". In: *Quantum Science and Technology* 6.1 (2020), p. 015007. DOI: `10.1088/2058-9565/abc6f4` (cited on pp. 4, 18, 93, 94, 121, 132).

[RH07] R. Raussendorf and J. Harrington. "Fault-tolerant quantum computation with high threshold in two dimensions". In: *Physical Review Letters* 98.19 (2007), p. 190504. DOI: `10.1103/PhysRevLett.98.190504` (cited on p. 14).

[RHG06] R. Raussendorf, J. Harrington, and K. Goyal. "A fault-tolerant one-way quantum computer". In: *Annals of physics* 321.9 (2006), p. 2242. DOI: `10.1016/j.aop.2006.01.012` (cited on p. 4).

[RHG07] R. Raussendorf, J. Harrington, and K. Goyal. "Topological fault-tolerance in cluster state quantum computation". In: *New Journal of Physics* 9.6 (2007), p. 199. DOI: `10.1088/1367-2630/9/6/199` (cited on pp. 4, 40).

[Rie+17] L. Riesebos, X. Fu, S. Varsamopoulos, et al. "Pauli frames for quantum computer architectures". In: *DAC '17: Proceedings of the 54th Annual Design Automation Conference 2017*. New York, NY, USA: Association for Computing Machinery, 2017, p. 1. ISBN: 978-1-45034927-7. DOI: `10.1145/3061639.3062300` (cited on p. 11).

[RM15] Á. Rivas and M. Müller. "Quantifying spatial correlations of general quantum dynamics". In: *New Journal of Physics* 17.6 (2015), p. 062001. DOI: `10.1088/1367-2630/17/6/062001` (cited on p. 78).

[Rof19] J. Roffe. "Quantum error correction: an introductory guide". In: *Contemporary Physics* 60.3 (2019), p. 226. DOI: `10.1080/00107514.2019.1667078` (cited on p. 13).

[Roo00] C. F. Roos. "Controlling the quantum state of trapped ions". Universität Innsbruck, 2000. URL: https://www.quantumoptics.at/images/publications/dissertation/roos-diss.pdf (cited on p. 76).

[RT09] G. Rubino and B. Tuffin. "Rare event simulation using Monte Carlo methods". John Wiley & Sons, 2009. DOI: 10.1002/9780470745403 (cited on p. 42).

[Saf16] M. Saffman. "Quantum computing with atomic qubits and Rydberg interactions: progress and challenges". In: *Journal of Physics B: Atomic, Molecular and Optical Physics* 49.20 (2016), p. 202001. DOI: 10.1088/0953-4075/49/20/202001 (cited on p. 3).

[Sah+23] U. Saha, J. D. Siverns, J. Hannegan, et al. "Routing single photons from a trapped ion using a photonic integrated circuit". In: *Physical Review Applied* 19 (2023), p. 034001. DOI: 10.1103/PhysRevApplied.19.034001 (cited on p. 137).

[Sar+20] M. Sarovar, T. Proctor, K. Rudinger, et al. "Detecting crosstalk errors in quantum information processors". In: *Quantum* 4 (2020), p. 321. DOI: 10.22331/q-2020-09-11-321 (cited on p. 79).

[Sat+21] K. J. Satzinger, Y.-J. Liu, A. Smith, et al. "Realizing topologically ordered states on a quantum processor". In: *Science* 374.6572 (2021), p. 1237. DOI: 10.1126/science.abi8378 (cited on pp. 1, 19).

[SC22] N. Shutty and C. Chamberland. "Decoding merged color-surface codes and finding fault-tolerant Clifford circuits using solvers for satisfiability modulo theories". In: *Physical Review Applied* 18.1 (2022), p. 014072. DOI: 10.1103/PhysRevApplied.18.014072 (cited on p. 29).

[Sch+11] P. Schindler, J. T. Barreiro, T. Monz, et al. "Experimental repetitive quantum error correction". In: *Science* 332.6033 (2011), p. 1059. DOI: 10.1126/science.1203329 (cited on p. 113).

[Sch+13a] P. Schindler, M. Müller, D. Nigg, et al. "Quantum simulation of dynamical maps with trapped ions". In: *Nature Physics* 9.6 (2013), p. 361. DOI: 10.1038/nphys2630 (cited on pp. 113, 123).

[Sch+13b] P. Schindler, D. Nigg, T. Monz, et al. "A quantum information processor with trapped ions". In: *New Journal of Physics* 15.12 (2013), p. 123012. DOI: 10.1088/1367-2630/15/12/123012 (cited on p. 76).

[Sch95] B. Schumacher. "Quantum coding". In: *Physical Review A* 51 (1995), p. 2738. DOI: 10.1103/PhysRevA.51.2738 (cited on p. 1).

[Sed+21] J. R. Seddon, B. Regula, H. Pashayan, et al. "Quantifying quantum speedups: Improved classical simulation from tighter magic monotones". In: *PRX Quantum* 2.1 (2021), p. 010345. DOI: 10.1103/PRXQuantum.2.010345 (cited on p. 39).

[SFN15] S. D. Sarma, M. Freedman, and C. Nayak. "Majorana zero modes and topological quantum computation". In: *npj Quantum Information* 1.1 (2015), p. 1. DOI: 10.1038/npjqi.2015.1 (cited on p. 22).

[Shi03] Y. Shi. "Both Toffoli and controlled-NOT need little help to do universal quantum computing". In: *Quantum Information & Computation* 3.1 (2003), p. 84 (cited on p. 8).

[Sho94] P. W. Shor. "Algorithms for quantum computation: Discrete logarithms and factoring". In: *Proceedings 35th annual symposium on foundations of computer science*. IEEE. 1994, p. 124. DOI: https://doi.org/10.1109/SFCS.1994.365700 (cited on pp. 2, 29).

[Sho96] P. W. Shor. "Fault-tolerant quantum computation". In: *Proceedings of 37th Conference on Foundations of Computer Science*. IEEE, 1996, p. 56.

ISBN: 978-0-8186-7594. DOI: 10.1109/SFCS.1996.548464 (cited on pp. 3, 4, 16).

[Sil+08] M. Silva, E. Magesan, D. W. Kribs, et al. "Scalable protocol for identification of correctable codes". In: *Physical Review A* 78.1 (2008), p. 012347. DOI: 10.1103/PhysRevA.78.012347 (cited on p. 36).

[Sim+17] D. S. Simon, G. Jaeger, A. V. Sergienko, et al. "Quantum metrology". Springer, 2017. DOI: 10.1007/978-3-319-46551-7_4 (cited on p. 2).

[Sin+22] K. Singh, S. Anand, A. Pocklington, et al. "Dual-element, two-dimensional atom array with continuous-mode operation". In: *Physical Review X* 12 (2022), p. 011040. DOI: 10.1103/PhysRevX.12.011040 (cited on p. 113).

[Sin+23] K. Singh, C. E. Bradley, S. Anand, et al. "Mid-circuit correction of correlated phase errors using an array of spectator qubits". In: *Science* 380.6651 (2023), p. 1265. DOI: 10.1126/science.ade5337 (cited on pp. 5, 113).

[SM00] A. Sørensen and K. Mølmer. "Entanglement and quantum computation with ions in thermal motion". In: *Physical Review A* 62 (2000), p. 022311. DOI: 10.1103/PhysRevA.62.022311 (cited on pp. 76, 82).

[Soc23] J. Soch. "The book of statistical proofs". https://statproofbook.github.io. 2023 (cited on pp. 41, 45).

[Sol00] R. M. Solovay. "Lie groups and quantum circuits". Presentation at *Mathematics of Quantum computation*, Mathematical Sciences Research Institute, https://www.slmath.org/workshops/189/schedules/12826. 2000 (cited on p. 8).

[SSP14] M. Schuld, I. Sinayskiy, and F. Petruccione. "The quest for a quantum neural network". In: *Quantum Information Processing* 13 (2014), p. 2567. DOI: 10.1007/s11128-014-0809-8 (cited on p. 2).

[SSP15] M. Schuld, I. Sinayskiy, and F. Petruccione. "An introduction to quantum machine learning". In: *Contemporary Physics* 56.2 (2015), p. 172. DOI: 10.1080/00107514.2014.964942 (cited on p. 2).

[Sta+05] S. Stahl, F. Galve, J. Alonso, et al. "A planar Penning trap". In: *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics* 32 (2005), p. 139. DOI: 10.1140/epjd/e2004-00179-x (cited on p. 76).

[Ste03] A. M. Steane. "Overhead and noise threshold of fault-tolerant quantum error correction". In: *Physical Review A* 68.4 (2003), p. 042322. DOI: 10.1103/PhysRevA.68.042322 (cited on p. 40).

[Ste14] A. M. Stephens. "Fault-tolerant thresholds for quantum error correction with the surface code". In: *Physical Review A* 89 (2014), p. 022321. DOI: 10.1103/PhysRevA.89.022321 (cited on p. 20).

[Ste96a] A. Steane. "Multiple-particle interference and quantum error correction". In: *Proceedings of the Royal Society A* 452.1954 (1996), p. 2551. DOI: 10.1098/rspa.1996.0136 (cited on p. 31).

[Ste96b] A. M. Steane. "Error correcting codes in quantum theory". In: *Physical Review Letters* 77.5 (1996), p. 793. DOI: 10.1103/PhysRevLett.77.793 (cited on p. 32).

[Ste97] A. M. Steane. "Active stabilization, quantum computation, and quantum state synthesis". In: *Physical Review Letters* 78.11 (1997), p. 2252. DOI: 10.1103/PhysRevLett.78.2252 (cited on pp. 16, 114).

[Tab+23] C. Taballione, M. C. Anguita, M. de Goede, et al. "20-Mode Universal Quantum Photonic Processor". In: *Quantum* 7 (2023), p. 1071. ISSN: 2521-327X. DOI: 10.22331/q-2023-08-01-1071 (cited on p. 3).

[Tak+17] M. Takita, A. W. Cross, A. D. Córcoles, et al. "Experimental demonstration of fault-tolerant state preparation with superconducting qubits". In: *Physical Review Letters* 119.18 (2017), p. 180501. DOI: 10.1103/PhysRevLett.119.180501 (cited on p. 4).

[TB05] B. M. Terhal and G. Burkard. "Fault-tolerant quantum computation for local non-Markovian noise". In: *Physical Review A* 71 (2005), p. 012336. DOI: 10.1103/PhysRevA.71.012336 (cited on p. 136).

[TCV20] B. M. Terhal, J. Conrad, and C. Vuillot. "Towards scalable bosonic quantum error correction". In: *Quantum Science and Technology* 5.4 (2020), p. 043001. DOI: 10.1088/2058-9565/ab98a5 (cited on pp. 4, 113).

[Ter15] B. M. Terhal. "Quantum error correction for quantum memories". In: *Reviews of Modern Physics* 87.2 (2015), p. 307. DOI: 10.1103/RevModPhys.87.307 (cited on pp. 4, 13).

[Tin+23] J. Tindall, M. Fishman, M. Stoudenmire, et al. "Efficient tensor network simulation of IBM's Eagle kicked Ising experiment". 2023. arXiv: 2306.14887 [quant-ph] (cited on p. 3).

[Tro+18] C. J. Trout, M. Li, M. Gutiérrez, et al. "Simulating the performance of a distance-3 surface code in a linear ion trap". In: *New Journal of Physics* 20.4 (2018), p. 043038. DOI: 10.1088/1367-2630/aab341 (cited on pp. 42, 98).

[TS14] Y. Tomita and K. M. Svore. "Low-distance surface codes under realistic quantum noise". In: *Physical Review A* 90.6 (2014), p. 062320. DOI: 10.1103/PhysRevA.90.062320 (cited on pp. 37, 113).

[TV23] B. T. Torosov and N. V. Vitanov. "Narrowband composite two-qubit gates for crosstalk suppression". In: *Physical Review A* 107 (2023), p. 032618. DOI: 10.1103/PhysRevA.107.032618 (cited on p. 113).

[VA17] I. de Vega and D. Alonso. "Dynamics of non-Markovian open quantum systems". In: *Reviews of Modern Physics* 89 (2017), p. 015001. DOI: 10.1103/RevModPhys.89.015001 (cited on p. 136).

[Vei+14] V. Veitch, S. H. Mousavian, D. Gottesman, et al. "The resource theory of stabilizer quantum computation". In: *New Journal of Physics* 16.1 (2014), p. 013009. DOI: 10.1088/1367-2630/16/1/013009 (cited on p. 27).

[VK22] M. Vasmer and A. Kubica. "Morphing quantum codes". In: *PRX Quantum* 3 (2022), p. 030319. DOI: 10.1103/PRXQuantum.3.030319 (cited on p. 137).

[VKL99] L. Viola, E. Knill, and S. Lloyd. "Dynamical decoupling of open quantum systems". In: *Physical Review Letters* 82.12 (1999), p. 2417. DOI: 10.1103/PhysRevLett.82.2417 (cited on p. 79).

[Vod+22] D. Vodola, M. Rispler, S. Kim, et al. "Fundamental thresholds of realistic quantum error correction circuits from classical spin models". In: *Quantum* 6 (2022), p. 618. DOI: 10.22331/q-2022-01-05-618 (cited on p. 20).

[Vui18] C. Vuillot. "Is error detection helpful on IBM 5Q chips?" In: *Quantum Information and Computation* 18.11-12 (2018), p. 949. DOI: 10.26421/QIC18.11-12-4 (cited on p. 4).

[Vui21] C. Vuillot. "Planar Floquet codes". 2021. arXiv: 2110.05348 [quant-ph] (cited on p. 31).

[Wan+20a] J. Wang, F. Sciarrino, A. Laing, et al. "Integrated photonic quantum technologies". In: *Nature Photonics* 14.5 (2020), p. 273. DOI: 10.1038/s41566-019-0532-1 (cited on p. 137).

[Wan+20b] Y. Wang, Z. Hu, B. C. Sanders, et al. "Qudits and high-dimensional quantum computing". In: *Frontiers in Physics* 8 (2020), p. 589504. DOI: 10.3389/fphy.2020.589504 (cited on p. 1).

[Wan+21] P. Wang, C.-Y. Luan, M. Qiao, et al. "Single ion qubit with estimated coherence time exceeding one hour". In: *Nature Communications* 12.1 (2021). DOI: 10.1038/s41467-020-20330-w (cited on p. 3).

[Wan+23] Y. Wang, S. Simsek, T. M. Gatterman, et al. "Fault-tolerant one-bit addition with the smallest interesting colour code". 2023. arXiv: 2309.09893 [quant-ph] (cited on p. 138).

[WBB22] M. A. Webster, B. J. Brown, and S. D. Bartlett. "The XP stabiliser formalism: A generalisation of the Pauli stabiliser formalism with arbitrary phases". In: *Quantum* 6 (2022), p. 815. DOI: 10.22331/q-2022-09-22-815 (cited on p. 39).

[WH23] D. Winter and S. Heußen. "qsample". https://github.com/dpwinter/qsample. 2023 (cited on pp. ix, 35).

[Whe90] J. A. Wheeler. "Information, physics, quantum: The search for links". 1990, p. 309. ISBN: 9780429500459 (cited on p. 1).

[Wil27] E. B. Wilson. "Probable inference, the law of succession, and statistical inference". In: *Journal of the American Statistical Association* 22.158 (1927), p. 209. DOI: 10.2307/2276774 (cited on p. 41).

[Win+23] K. Wintersperger, F. Dommert, T. Ehmer, et al. "Neutral atom quantum computing hardware: Performance and end-user perspective". In: *EPJ Quantum Technology* 10.1 (2023), p. 1. DOI: 10.1140/epjqt/s40507-023-00190-1 (cited on pp. 3, 5, 113).

[Win+98] D. J. Wineland, C. Monroe, W. M. Itano, et al. "Experimental issues in coherent quantum-state manipulation of trapped atomic ions". In: *Journal of Research of the National Institute of Standards and Technology* 103.3 (1998), p. 259. DOI: 10.6028/jres.103.019 (cited on p. 76).

[Wu+21] Y. Wu, W.-S. Bao, S. Cao, et al. "Strong quantum computational advantage using a superconducting quantum processor". In: *Physical Review Letters* 127.18 (2021), p. 180501. DOI: 10.1103/PhysRevLett.127.180501 (cited on p. 3).

[Xie+22] L. Xie, J. Zhai, Z. Zhang, et al. "Suppressing ZZ crosstalk of quantum computers through pulse and scheduling co-optimization". In: *Proceedings of the 27th ACM International Conference on Architectural Support for Programming Languages and Operating Systems*. 2022, p. 499. DOI: 10.1145/3503222.3507761 (cited on p. 79).

[Xu+21] W. Xu, A. V. Venkatramani, S. H. Cantú, et al. "Fast preparation and detection of a Rydberg qubit using atomic ensembles". In: *Physical Review Letters* 127.5 (2021), p. 050501. DOI: 10.1103/PhysRevLett.127.050501 (cited on p. 113).

[Xu+23a] Q. Xu, J. P. B. Ataides, C. A. Pattison, et al. "Constant-overhead fault-tolerant quantum computation with reconfigurable atom arrays". 2023. arXiv: 2308.08648 [quant-ph] (cited on pp. 20, 137).

[Xu+23b] S. Xu, Z.-Z. Sun, K. Wang, et al. "Digital simulation of projective non-abelian anyons with 68 superconducting qubits". In: *Chinese Physics Letters* (2023). DOI: 10.1088/0256-307X/40/6/060301 (cited on pp. 1, 22).

[YK17] T. J. Yoder and I. H. Kim. "The surface code with a twist". In: *Quantum* 1 (2017), p. 2. DOI: 10.22331/q-2017-04-25-2 (cited on pp. 4, 18).

[You+21] J. T. Young, P. Bienias, R. Belyansky, et al. "Asymmetric blockade and multiqubit gates via dipole-dipole interactions". In: *Physical Review Letters* 127 (2021), p. 120501. DOI: 10.1103/PhysRevLett.127.120501 (cited on p. 123).

[Zen+19] B. Zeng, X. Chen, D.-L. Zhou, et al. "Quantum information meets quantum matter". Springer, 2019. DOI: 10.1007/978-1-4939-9084-9 (cited on p. 19).

[Zha09] Y. Zhao. "Quantum cryptography in real-life applications: assumptions and security". University of Toronto, 2009. URL: https://tspace.library.utoronto.ca/bitstream/1807/19307/1/Zhao_Yi_200911_PhD_Thesis.pdf (cited on p. 2).

[Zha21] B. Zhang. "Improving circuit performance in a trapped-ion quantum computer". Duke University, 2021. URL: https://dukespace.lib.duke.edu/dspace/handle/10161/24374 (cited on p. 78).

[Zha+22] Y. Zhao, Y. Ye, H.-L. Huang, et al. "Realization of an error-correcting surface code with superconducting qubits". In: *Physical Review Letters* 129.3 (2022), p. 030501. DOI: 10.1103/PhysRevLett.129.030501 (cited on pp. 4, 113).

[Zha+23a] H. Zhang, L. Wan, S. Paesani, et al. "Encoding error correction in an integrated photonic chip". In: *PRX Quantum* 4.3 (2023), p. 030340. DOI: 10.1103/PRXQuantum.4.030340 (cited on p. 4).

[Zha+23b] X. Zhang, E. Kim, D. K. Mark, et al. "A superconducting quantum simulator based on a photonic-bandgap metamaterial". In: *Science* 379.6629 (2023), p. 278. DOI: 10.1126/science.ade7651 (cited on p. 137).

[Zho+20] H.-S. Zhong, H. Wang, Y.-H. Deng, et al. "Quantum computational advantage using photons". In: *Science* 370.6523 (2020), p. 1460. DOI: 10.1126/science.abe8770 (cited on p. 3).

[Zho+21] H.-S. Zhong, Y.-H. Deng, J. Qin, et al. "Phase-programmable Gaussian boson sampling using stimulated squeezed light". In: *Physical Review Letters* 127.18 (2021), p. 180502. DOI: 10.1103/PhysRevLett.127.180502 (cited on p. 3).

Declaration of Authorship

I, Sascha Heinrich Heußen, declare that this thesis titled, "Applications of fault-tolerant topological quantum error correction in near-term devices" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date: