

Control for Programmable Superconducting Quantum Systems

Rol, M.A.

DOI

[10.4233/uuid:0a2ba212-f6bf-4c64-8f3d-b707f1e44953](https://doi.org/10.4233/uuid:0a2ba212-f6bf-4c64-8f3d-b707f1e44953)

Publication date

2020

Document Version

Final published version

Citation (APA)

Rol, M. A. (2020). *Control for Programmable Superconducting Quantum Systems*. [Dissertation (TU Delft), Delft University of Technology]. <https://doi.org/10.4233/uuid:0a2ba212-f6bf-4c64-8f3d-b707f1e44953>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

```
219 return gates
220
221
222 def CNOT_like_PTM(idx):
223     """
224     Returns the pauli transfer matrix for gates of the cnot like
225     (q0)  --C1---S1--  --C1---S1-----
226           |           -> |
227     (q1)  --C1---S1--  --C1---S1^Y90--
228     """
229     assert(idx < 5184)
230     idx_0 = (idx // 216) % 24
231     idx_1 = (idx // 216) % 3
232     idx_2 = (idx // 576) % 3
233     idx_3 = (idx // 1728) % 3
234
235     C1_q0 = np.kron(np.eye(4), C1[idx_0])
236     C1_q1 = np.kron(C1[idx_1], np.eye(4))
237     CZ
238     S1_q0 = np.kron(np.eye(4), S1[idx_2])
239     S1y_q1 = np.kron(np.dot(C1[idx_3], Y90), np.eye(4))
240     return np.linalg.multi_dot(list(reversed(gates[idx_0, idx_1, idx_2, idx_3])))
```

Control for Programmable Superconducting Quantum Systems

def gates(idx):



Michiel Adriaan Rol

Propositions

accompanying the dissertation

CONTROL FOR PROGRAMMABLE SUPERCONDUCTING QUANTUM SYSTEMS

by

Michiel Adriaan RoL

1. Quantum computing is not understood well enough to precisely define the responsibilities of each layer in a full-stack quantum computer (Chapter 2 of this thesis).
2. There will always be a demand for ad-hoc characterization protocols (Chapter 3 of this thesis).
3. Projecting the qubit in a known basis is a faster alternative to initializing in $|0\rangle$ for many experiments (Chapter 4 of this thesis).
4. The key challenge in flux-pulsing-based two-qubit gates is addressing distortions of the pulse shape (Chapters 6 and 7 of this thesis).
5. If abstraction is done right, it becomes hard to understand why the initial problem was challenging.
6. Simulating 10 transmons accurately [1] is more useful than simulating 40 qubits efficiently [2].
7. Breakthroughs in quantum computing can only occur when people actively look beyond the narrow confines of their field.
8. Good science is good engineering.
9. The shortest path to success is the path to a short success.
10. Political correctness actively hinders the ideals of inclusion and tolerance it claims to support.

These propositions are regarded as opposable and defensible, and have been approved as such by the promotors Prof. dr. L. DiCarlo & Prof. dr. ir. L.M.K. Vandersypen.

[1] T. E. O'BRIEN, B. M. TARASINSKI, and L. DICARLO. *Density-matrix simulation of small surface codes under current and projected experimental noise*. npj Quantum Information, **3** (39), 2017. (see pages: 1, 2)

[2] T. HÄNER and D. S. STEIGER. *0.5 petabyte simulation of a 45-qubit quantum circuit*. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '17. Association for Computing Machinery, New York, NY, USA, 2017. (see pages: 1, 2)

Stellingen

behorende bij het proefschrift

CONTROL FOR PROGRAMMABLE SUPERCONDUCTING QUANTUM SYSTEMS

door

Michiel Adriaan ROL

1. De quantum computer wordt niet goed genoeg begrepen om de verantwoordelijkheden van iedere laag in een full-stack quantum computer exact te specificeren (Hoofdstuk 2 van dit proefschrift).
2. Er zal altijd vraag zijn naar ad-hoc karakterisatie protocollen (Hoofdstuk 3 van dit proefschrift).
3. Het projecteren van een qubit in een bekende basis is een sneller alternatief dan het initialiseren in $|0\rangle$ voor verscheidene experimenten (Hoofdstuk 4 van dit proefschrift).
4. De grootste uitdaging in op flux pulsen gebaseerde twee qubit gates is het adresseren van puls vervormingen (Hoofdstuk 6 en 7 van dit proefschrift).
5. Als men succesvol abstraheert wordt het lastig om te begrijpen waarom het oorspronkelijke probleem uitdagend was.
6. Het nauwkeurig simuleren van 10 transmons [1] is nuttiger dan het efficiënt simuleren van 40 qubits [2].
7. Doorbraken in quantum informatica kunnen zich alleen voordoen wanneer men verder kijkt dan de strakke kaders van het eigen vakgebied.
8. Goede wetenschap is goede techniek.
9. De kortste weg naar succes is de weg naar een kort succes.
10. Politieke correctheid vormt een actieve belemmering voor de idealen van inclusiviteit en tolerantie die het claimt na te streven.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotoren Prof. dr. L. DiCarlo & Prof. dr. ir. L.M.K. Vandersypen.

[1] T. E. O'BRIEN, B. M. TARASINSKI, and L. DICARLO. *Density-matrix simulation of small surface codes under current and projected experimental noise*. npj Quantum Information, **3** (39), 2017. (see pages: 1, 2)

[2] T. HÄNER and D. S. STEIGER. *0.5 petabyte simulation of a 45-qubit quantum circuit*. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '17. Association for Computing Machinery, New York, NY, USA, 2017. (see pages: 1, 2)

**CONTROL FOR PROGRAMMABLE SUPERCONDUCTING
QUANTUM SYSTEMS**

CONTROL FOR PROGRAMMABLE SUPERCONDUCTING QUANTUM SYSTEMS

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof. dr. ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op vrijdag 9 oktober 2020 om 12:30 uur

door

Michiel Adriaan Rol

Ingenieur in Technische Natuurkunde,
Technische Universiteit Delft, Nederland,
geboren te Amsterdam, Nederland.

Dit proefschrift is goedgekeurd door de promotoren

Prof. dr. L. DiCarlo

Prof. dr. ir. L. M. K. Vandersypen

Samenstelling promotiecommissie:

Rector Magnificus,

Prof. dr. L. DiCarlo

Prof. dr. ir. L. M. K. Vandersypen

voorzitter

Technische Universiteit Delft, promotor

Technische Universiteit Delft, promotor

Onafhankelijke leden:

Prof. dr. M. H. Devoret

Yale University,

New Haven, Connecticut, Verenigde Staten

Dr. R. J. Blume-Kohout

The University of New Mexico,

Albuquerque, New-Mexico, Verenigde Staten

Prof. dr. F. K. Wilhelm-Mauch

Universität des Saarlandes,

Saarbrücken, Duitsland

Dr. C. G. Almudever

Technische Universiteit Delft,

Delft, Nederland

Prof. dr. B. M. Terhal

Technische Universiteit Delft,

Delft, Nederland

Prof. dr. G. A. Steele,

Technische Universiteit Delft,

Delft, Nederland, reservelid



- Printed by:* Gildeprint, Enschede – www.gildeprint.nl
- Cover:* Artist's impression of a programmable superconducting quantum system. Icons represent the different layers of a full-stack quantum computer. The diagram on the back shows a distorted flux pulse that detunes the transmon as part of the Cryoscope. Using the Cryoscope, it is possible to include corrections in the pulse shape that result in a close to ideal detuning of the Net-Zero pulse shape used for two-qubit gates. Images in the background show parts of the different layers of the full-stack quantum computer. At the bottom, we show an optical micrograph of the Chimaera device of which results are featured in this thesis. The front features a photo of the inside of a dilution refrigerator that was used in some of the experiments. The PCB shown on the back is part of the QWG (QuTech Arbitrary Waveform Generator) that was developed at QuTech. The code snippets shown at the top are from PycQED and relate to randomized benchmarking experiments.
- Image credits:* The full-stack icons were developed by Stan van Nieuwamerongen as part of the Quantum Infinity demonstrator project. The Chimaera device was designed by Marc Beekman and Nadia Haider, and fabricated by Nandini Muthusubramanian and Alessandro Bruno, photo by Nandini Muthusubramanian. The fridge was wired by Brian Tarasinski and Jules van Oven, photo by Jacob de Sterke. The QWG was designed by Duije Duerloo, Wouter Vlothuizen, Jacob de Sterke, Raymond Vermeulen and Raymond Schouten, photo by Jacob de Sterke. All works shown were developed in the group of Leo DiCarlo at QuTech. Cover design by Adriaan Rol.

Copyright © 2020 by M.A. Rol

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, without prior permission from the copyright owner.

Casimir PhD Series, Delft-Leiden 2020-25

ISBN 978-90-8593-451-6

An electronic version of this dissertation is available at

<http://repository.tudelft.nl/>.

CONTENTS

Summary	xiii
Samenvatting	xv
Preface	xvii
1 Introduction	1
1.1 Qubits and quantum computing	1
1.1.1 Superposition, entanglement and measurements	1
1.1.2 Interpretations of quantum mechanics	3
1.1.3 The power of entanglement	4
1.1.4 The fragility of qubits	4
1.1.5 Quantum error correction and fault tolerance	5
1.2 Outline of this thesis	7
2 Building a superconducting quantum system	9
2.1 The full-stack quantum computer	9
2.1.1 Limitations of the full-stack quantum computer	10
2.1.2 Building a full-stack quantum computer	13
2.2 The QuSurf architecture	14
2.2.1 Superconducting transmon qubits	14
2.2.2 Designing for fault tolerance	19
2.2.3 Choice of two-qubit gate	19
2.2.4 Scalable surface code	21
2.2.5 Simulating the error correcting cycle	23
2.2.6 Controlling the device	23
3 Assessing the performance of superconducting quantum processors	27
3.1 Introduction	28
3.2 Characterizing errors	28
3.2.1 The black-box approach	29
3.2.2 The experimentalist's approach	30
3.3 Standard operations	32
3.3.1 Operation errors	32
3.3.2 Quantities of interest	33
3.3.3 Characterizing operations	34

3.4	Other error sources	35
3.4.1	Leakage	36
3.4.2	Non-Markovian errors	41
3.5	Crosstalk	44
3.5.1	Idle crosstalk	44
3.5.2	Microwave-drive crosstalk: pulse spillover	49
3.5.3	AC-flux crosstalk	51
3.5.4	Measurement crosstalk	52
3.6	Conclusion & outlook	54
4	Restless tuneup of high-fidelity qubit gates	57
4.1	Introduction	58
4.2	The concept and benefits of restless tuning.	58
4.3	Experimental results	61
4.3.1	Experimental comparison of restless and restful cost functions	61
4.3.2	Signal and noise in restless tuning	62
4.3.3	Gate optimization with restless tuning	63
4.3.4	Gate optimization robustness	64
4.4	Conclusions	64
4.5	Methods	65
4.5.1	Setup for numerical optimization	65
4.5.2	Signal and noise of the restless cost function	66
4.5.3	Modeling	66
4.5.4	Measurement of T_1 fluctuations	70
4.5.5	Relation to experiment	70
4.5.6	Gate set tomography and randomized benchmarking fidelities	71
4.5.7	Verification of conventional and restless tuneup	72
5	Active resonator reset in the nonlinear dispersive regime of circuit QED	73
5.1	Introduction	74
5.2	Experimental results	75
5.2.1	Device characterization	75
5.2.2	Measurement tune-up and the effect of leftover photons	76
5.2.3	AllXY as a photon detector	76
5.2.4	Tune-up and comparison of two methods for active photon depletion	77
5.2.5	Benchmarking depletion methods with a QEC emulation: a flipping ancilla	78
5.2.6	Optimization of the depletion pulse length	80
5.2.7	Benchmarking depletion methods with a QEC emulation: a non-flipping ancilla	81
5.3	Conclusions	81
5.4	Methods	82
5.4.1	Experimental setup	82

5.4.2	Photon number calibration.	83
5.4.3	Numerical optimization of depletion pulses	84
5.4.4	Constant excited state QEC emulation.	84
5.4.5	Theoretical models.	86
6	Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor	89
6.1	Introduction	90
6.2	Concept	92
6.2.1	Correcting distortions.	93
6.2.2	Verifying distortion corrections	94
6.2.3	Cryoscope sensitivity.	94
6.3	Conclusions	97
6.4	Methods	97
6.4.1	Device and experimental setup.	97
6.4.2	Limitations of the Cryoscope.	97
6.4.3	Cryoscope signal-to-noise ratio	102
6.4.4	Real-time predistortion filters	104
6.4.5	Using Cryoscope to measure arbitrary shapes	106
7	A fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits	109
7.1	Introduction	110
7.2	Concept	111
7.3	Experimental results	112
7.3.1	Repeatability	112
7.3.2	Echo-effect	113
7.3.3	Leakage	113
7.3.4	Performance	115
7.3.5	Performance limitations	116
7.4	Conclusions	116
7.5	Methods	117
7.5.1	Device parameters.	118
7.5.2	Flux pulse parametrization	118
7.5.3	Simulation structure	119
7.5.4	Conditional oscillation experiment	126
7.5.5	Optimal performance.	127
7.5.6	Net-Zero pulses as a Mach-Zehnder interferometer	128
7.5.7	Leakage modification for randomized benchmarking	131
8	Controlling a multi-qubit processor	135
8.1	Basic lab infrastructure – design philosophy	136
8.1.1	Extensibility.	136

8.1.2	Usability	137
8.1.3	Automatability	137
8.1.4	Open source	137
8.2	Basic lab infrastructure – core concepts	138
8.2.1	Instruments and Parameters	139
8.2.2	Measurement Control	140
8.2.3	Data.	142
8.2.4	Analysis	144
8.2.5	Pulse sequencing	145
8.3	Automated characterization and calibration.	146
8.3.1	Graph-based tuneup	147
8.3.2	System emulation	148
8.3.3	System-level analysis.	150
8.4	Summary	152
9	Conclusion & outlook	153
9.1	Conclusions	154
9.2	Reflections on the QuSurf architecture	155
9.3	Towards a kilo-qubit system	157
9.3.1	The application problem	158
9.3.2	The fabrication problem.	160
9.3.3	The calibration problem.	162
A	QuSurf Protocol Definitions for LogiQ Metrics	165
A.1	Gate errors	165
A.1.1	Single-qubit randomized benchmarking	165
A.1.2	Two-qubit randomized benchmarking with leakage modification	166
A.1.3	Two-qubit interleaved randomized benchmarking	167
A.1.4	Single-qubit and two-qubit gate-set tomography	167
A.2	Idle error rates	167
A.2.1	Idle decay	167
A.3	Measurement errors	168
A.3.1	Measurement assignment errors	168
A.3.2	Measurement QNDness	168
A.4	Reset and initialization errors	170
A.4.1	Reset and initialization error	170
A.5	Crosstalk errors: Idle crosstalk	171
A.5.1	Residual-ZZ interaction	171
A.5.2	Residual exchange interaction	172
A.6	Crosstalk errors: Weight-1 operation crosstalk.	172
A.6.1	Microwave cross driving.	172
A.6.2	Simultaneous single-qubit RB	173
A.6.3	Measurement-induced dephasing	173

A.6.4 Measurement detection crosstalk	174
A.7 Other error sources	174
A.7.1 Leakage	174
A.7.2 Non Markovianity	175
Acknowledgements	177
Curriculum Vitæ	183
List of Publications	185
References	187

SUMMARY

The discovery of quantum mechanics in the 20th century forms the basis of many of the technologies that define our lives today. The ability to engineer and manipulate individual quantum systems – even create artificial atoms – promises a similar revolutionary leap in technology. A quantum technology of particular interest is quantum computing, which has the potential to solve problems that are intractable for classical computers, opening up new domains of computation.

Meanwhile, an attractive approach to creating engineered quantum systems is circuit quantum electrodynamics (cQED). Where research initially focused on understanding the physics of cQED devices, focus has shifted to building systems capable of performing useful computations. However, this remains extremely challenging, in part due to the inherently fragile nature of the individual quantum bits, but also due to difficulties in controlling and scaling up these systems.

This thesis focuses on the control aspects of building an extensible full-stack quantum computer based on superconducting transmon qubits. We define the demonstration of quantum fault-tolerance as our target application to give focus to our efforts. The QuSurf architecture for a full-stack quantum computer presented in this thesis is designed with this application in mind. We provide a detailed study of the error sources present in this system and give an overview of the relevant characterization techniques.

In the second part of this thesis, we address several key challenges in the control of a quantum computer.

To realize high-fidelity coherence limited gates, we present a novel tuneup protocol that achieves a tenfold speedup over the state-of-the-art. This is realized by eliminating the need for qubit initialization. We demonstrate this protocol by calibrating single-qubit gates to a coherence limited Clifford fidelity of 99.9% in one minute.

Performing repeated parity checks, as is required for quantum error correction, requires reusing qubits quickly after they have been measured. By introducing a numerically optimized depletion pulse we are able to speed up the depletion of measurement photons in a readout resonator without having to rely on specific symmetry conditions. Using this technique speed up photon depletion by more than six inverse resonator linewidths, reducing the error rate in an emulated ancilla parity check by a factor 75.

Flux-pulsing based two-qubit gates are the fastest two-qubit gates. However, they are also very technically demanding. The key challenge in performing these gates is addressing the distortions that control signals experience as they traverse various electrical components. We have developed Cryoscope (short for cryogenic oscilloscope) to characterize and correct these distortions. Cryoscope is an in-situ technique that uses the qubit to sample control

pulses of arbitrary shape. Even when correcting distortions to within $\sim 0.1\%$ two-qubit gates are history-dependent due to the long timescale upon which some of these distortions act. We have invented Net-Zero, a new type of flux-pulsing based two-qubit gate, to address this problem. It makes use of a symmetry condition of the transmon to have net-zero integral, making the gate resilient to long-timescale distortions. The gate suppresses leakage out of the computational subspace to 0.1% by making use of leakage interference and has a built-in echo effect that enhances the coherence of the gate, achieving a two-qubit gate fidelity of 99.1%.

Custom software is required to perform the physics experiments needed to build and operate a quantum computer. PycQED is an open-source software framework we have developed for this purpose. We discuss the design choices and concepts of PycQED before turning our focus to characterization and calibration. Here we introduced dependency graphs as a useful abstraction and system emulation as an essential development tool for automating the characterization and calibration process.

We conclude the thesis by reflecting on the limitations of our architecture and providing an outlook on the grand challenges of building a useful kilo-qubit sized quantum computer. We define these challenges as The Application Problem, The Fabrication Problem, and The Calibration Problem.

SAMENVATTING

De ontdekking van quantum mechanica in de 20e eeuw staat aan de basis van de technologieën die ons leven bepalen. De mogelijkheid om individuele quantum systemen te ontwerpen en manipuleren – zelfs kunstmatige atomen te creëren – belooft te leiden tot een vergelijkbare revolutionaire sprong in technologie. Een bijzonder interessante quantumtechnologie is quantumcomputing. Quantumcomputing heeft de potentie om problemen op te lossen die onhandelbaar zijn voor klassieke computers en maakt daarmee nieuwe soorten berekeningen mogelijk.

Een veelbelovende benadering om kunstmatige quantumsystemen te maken is circuit quantum electrodynamics (cQED). Waar onderzoek zich oorspronkelijk richtte op het begrijpen van de fysica van cQED systemen, is de aandacht inmiddels verschoven naar het ontwerpen van systemen die nuttige berekeningen kunnen uitvoeren. Dit blijft echter een enorme uitdaging, deels vanwege de inherent kwetsbare aard van de individuele quantumbits, maar ook vanwege problemen bij het aansturen en opschalen van deze systemen.

Dit proefschrift richt zich op de aansturingsvraagstukken van het bouwen van een uitbreidbare full-stack quantumcomputer op basis van supergeleidende transmon-qubits. Om focus aan te brengen in het onderzoek definiëren we het demonstreren van quantum fault-tolerance als de beoogde toepassing van deze quantumcomputer. We presenteren de QuSurf architectuur die hiervoor ontworpen is en geven een gedetailleerde studie van de foutenbronnen die in dit systeem aanwezig zijn alsmede een overzicht van de relevante karakterisatietechnieken.

In het tweede deel van dit proefschrift gaan we in op een aantal belangrijke uitdagingen bij de aansturing van een quantumcomputer.

Om hoge kwaliteit coherentie gelimiteerde gates te realiseren, presenteren we een nieuw kalibratieprotocol dat een tienvoudige versnelling bereikt ten opzichte van de state-of-the-art. Dit wordt gerealiseerd door de noodzaak voor qubit-initialisatie te elimineren. We demonstrenen dit protocol door single-qubit-gates te kalibreren tot een coherentie beperkte Clifford fidelity van 99,9% in één minuut.

Voor het uitvoeren van herhaalde pariteitscontroles, zoals vereist is voor quantumfoutcorrectie, moeten qubits snel opnieuw bruikbaar zijn nadat ze zijn gemeten. Door een numeriek geoptimaliseerde depletion-puls te introduceren, is het mogelijk om meetphotonen die in de uitleesresonator achterblijven na een meting versneld te verwijderen. Deze puls is niet afhankelijk van specifieke symmetrieomstandigheden. Met behulp van deze techniek wordt de photon-depletion versneld met meer dan zes inverse lijnbreedtes van de resonator, waardoor het foutpercentage bij een geëmuleerde ancilla-pariteitscontrole met een factor 75 wordt verminderd.

Op flux-pulsen gebaseerde twee-qubit-gates zijn de snelste twee-qubit-gates. Op technisch gebied zijn deze gates echter ook zeer uitdagend. De belangrijkste uitdaging bij het uitvoeren van deze gates is het adresseren van de vervormingen die aansturing-signalen ervaren wanneer ze verschillende elektrische componenten doorkruisen. We hebben de Cryoscope (afkorting van cryogene oscilloscoop) ontwikkeld om deze vervormingen te karakteriseren en te corrigeren. De Cryoscope is een in-situ techniek die de qubit gebruikt om controlepulsen van willekeurige vormen te kunnen bemonsteren. Zelfs wanneer vervormingen worden gecorrigeerd tot op $\sim 0.1\%$, zijn twee qubit-gates geschiedenisafhankelijk. Dit komt door de lange tijdschaal waarop sommige van deze vervormingen werken. Om dit probleem aan te pakken hebben we de Net-Zero twee-qubit-gate uitgevonden, een nieuw type op flux-pulsen gebaseerde twee-qubit-gate. Deze puls maakt gebruik van een symmetrieconditie van de transmon om een netto-nul integraal te hebben, waardoor de gate ongevoelig is voor vervormingen op lange tijdschalen. De gate onderdrukt lekkage uit de computational-subspace tot 0.1% door gebruik te maken van lek-interferentie en heeft een ingebouwd echo-effect dat de coherentie van de gate verbetert, waardoor een twee-qubit-gate-kwaliteit van 99.1% wordt bereikt.

Op maat gemaakte software is vereist om de natuurkundige experimenten uit te voeren die nodig zijn om een quantumcomputer te bouwen en aan te sturen. PycQED is een open-source softwareframework dat we voor dit doel hebben ontwikkeld. We bespreken de ontwerpkeuzes en concepten van PycQED voordat we onze aandacht richten op karakterisering en kalibratie. Hier hebben we afhankelijkheidsgraven geïntroduceerd als een nuttige abstractie en systeemulatie als een essentiële ontwikkelingstool voor het automatiseren van het karakteriserings- en kalibratieproces.

We sluiten het proefschrift af door te reflecteren op de beperkingen van onze architectuur en vooruit te blikken op de grote uitdagingen van het bouwen van een bruikbare quantumcomputer van kilo-qubit-formaat. We definiëren het toepassingsprobleem, het fabricageprobleem en het kalibratieprobleem als de grote uitdagingen in het opschalen van quantumcomputers.

PREFACE

When writing this thesis, I made a rookie mistake. Instead of simply combining my papers and writing a minimal introduction, theory, and conclusion chapter, I decided to write several chapters containing new content. In these chapters, I have focused on subjects that are relevant when attempting to build a quantum computer such as systems architecture, error sources, and the tools required to build such a system. I have tried to give as much insight into these subjects as possible, focusing not just on the conclusions but also on the design considerations that went into them. And although these chapters do not fit the template of a traditional experimental physics paper, I believe these subjects to be highly relevant to any quantum engineer.

I am very glad that I ignored the advice of my friends and wrote these extra chapters. I hope you will enjoy reading them as much as I did writing them.

Adriaan Rol
Delft, June 2020

INTRODUCTION

Nature isn't classical, dammit, and if you want to make a simulation of nature, you'd better make it quantum mechanical, and by golly it's a wonderful problem, because it doesn't look so easy.

— Richard Feynman (1982) [1]

1.1 Qubits and quantum computing

The idea of using systems that behave according to the laws of quantum mechanics to perform computations is often credited to Richard Feynman [1]. Quantum mechanics is a theory that was developed in the early 20th century that is best known for the strangeness of some of its core concepts and the associated interpretations. Central in quantum mechanics are the concepts of (1) *superposition*, (2) *entanglement*, and (3) *measurements*. In this section, we will explain these three concepts using the language of quantum information before addressing the question of whether these phenomena can, at least in principle, be used to perform useful computations. For a thorough introduction to quantum information, we refer the reader to Nielsen and Chuang [2].

1.1.1 Superposition, entanglement and measurements

Where a classical bit would be described by its state, which can be either 0 or 1, a quantum mechanical bit or *qubit* can be in a combination of the $|0\rangle$ and $|1\rangle$ states known as a superposition. Such a state is described by a state vector,

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle, \quad (1.1)$$

where α and β are complex valued coefficients with norm $|\alpha|^2 + |\beta|^2 \leq 1$. Any pure single qubit state ($\| |\psi\rangle \|^2 = 1$) can be visualized as a vector on the surface of a Bloch sphere (Figure 1.1). In this representation, $|\psi\rangle = \cos(\theta/2) |0\rangle + e^{i\phi} \sin(\theta/2) |1\rangle$ in standard polar coordinates and the north and south poles correspond to the $|0\rangle$ and $|1\rangle$ state, respectively. Operations or gates on single qubits are rotations of the state vector around a fixed axis. For example, the equivalent to the classical NOT operation, the X gate, is a rotation of 180° around the x axis that transforms $|0\rangle$ in $|1\rangle$ and vice versa. A single-qubit gate that has no classical analogue is the Hadamard (H) gate. The Hadamard gate corresponds to a rotation

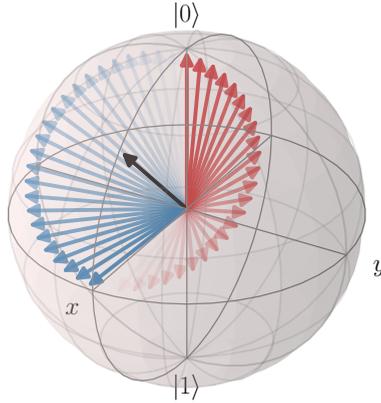


Figure 1.1.: Bloch sphere visualization of the effect of two consecutively applied Hadamard gates to the $|0\rangle$ state. The blue state vectors show the evolution during the first Hadamard gate, transferring $|0\rangle$ to $|+\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$. The red state vectors show the evolution during the second Hadamard gate, transferring $|+\rangle$ back to $|0\rangle$. The black arrow corresponds to the axis of rotation of H .

of 180° around the $\frac{1}{\sqrt{2}}(\hat{z} + \hat{x})$ axis and creates an equal superposition from both $|0\rangle$ and $|1\rangle$. The effect of H is visualized in Figure 1.1.

Entanglement arises naturally when we consider operations on multiple qubits. A shorthand notation is typically used when describing a multi-qubit system, $|\psi\phi\rangle = |\psi\rangle_C \otimes |\phi\rangle_T$, where the subscripts denoting which qubits the states correspond to are omitted. The simplest two-qubit gate is the Controlled-NOT or CNOT gate. This gate applies an X gate on a target qubit conditional on the state of a control qubit, so that it takes the state $|00\rangle \rightarrow |00\rangle$ and $|10\rangle \rightarrow |11\rangle$. If we prepare a control qubit in an equal superposition of $|0\rangle$ and $|1\rangle$: $\frac{1}{\sqrt{2}}(|0\rangle_C + |1\rangle_C)$ and a target qubit in $|0\rangle_T$, and we apply a CNOT gate, a special state known as a Bell state is created:

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle). \quad (1.2)$$

A Bell state is an *entangled* state. A state is entangled if it cannot be expressed as a product of the individual states of its components.

Whenever a qubit is measured, the measurement returns either $+1$ or -1 corresponding to the *eigenvalues* of the measurement operator. The probabilities of these outcomes correspond to the projection of the state $|\psi\rangle$ on to the *eigenstates* of the measurement operator. The most commonly used measurement operator is the Z operator which has two eigenstates, $|0\rangle$ with eigenvalue $+1$ and $|1\rangle$ with eigenvalue -1 . A measurement using the Z operator is also called a measurement in the computational basis or a measurement along the z axis. When performing a measurement in the computational basis, the probability of finding the system in $|0\rangle$ is given by $P(M_Z = +1) = |\langle 0|\psi\rangle|^2 = |\alpha|^2$. This is known as

the Born rule [3]. Crucially, the state after the measurement is either $|0\rangle$ or $|1\rangle$ depending on the declared outcome. The information in the complex probability amplitude is erased when projecting onto the measured state.

1.1.2 Interpretations of quantum mechanics

To appreciate how counterintuitive it is that, colloquially speaking, looking at a system changes the state of that same system, it is useful to try and interpret what is going on. To show how absurd quantum theory is Einstein, Podolsky and Rosen devised a thought experiment called the EPR paradox [4]. In this thought experiment a Bell state is created consisting of two particles that are physically separated from each other. The *Copenhagen Interpretation* of quantum mechanics asserts that when one of the two particles is measured, the other particle instantly collapses into the corresponding state. Taking the Bell state from Equation (1.2), if the first qubit is measured in the Z basis and found to be in $|0\rangle$ the second qubit is instantly projected into $|0\rangle$ as well. If, on the other hand, the first measurement was performed in the X basis¹ and the qubit is found to be in $|+\rangle$, the second qubit is instantly projected into $|+\rangle$ as well. A consequence is that if the physical separation is large, e.g., one particle is at the moon while the other is on earth, the information about the chosen measurement basis and the outcome on earth must travel faster than the speed of light to ensure that a measurement on the moon shortly after the measurement on earth returns the “correct” correlated outcome. Such “spooky action at a distance”, Einstein and collaborators argued, is absurd. The most natural interpretation of quantum mechanics they argued would be that it is incomplete [4]. At the time it was argued that the system was already in a definite state, described by *hidden variables*, before the measurement. It was only our incomplete theoretical understanding that forces us to work with a probabilistic theory in which measurements are special.

Although it is now understood that Einstein’s spooky action at a distance cannot be used to send information faster than the speed of light² and does not contradict relativity, the paradox did expose the fundamental non-classical nature of quantum theory. In 1964, John Bell [5] derived an inequality on the correlations between measurements of distant particles that any theory of nature that obeys locality and realism must obey but that is violated by quantum mechanics. Since then there have been numerous experiments [6–9] that have measured violations of the Bell inequalities, as a consequence local hidden variable theories have largely been abandoned by the scientific community. In addition to the earlier mentioned Copenhagen interpretation of quantum mechanics there are many other interpretations of quantum mechanics ranging from the *relative state interpretation* [10] which asserts that the apparent collapse of the wavefunction is a consequence of the observer getting entangled with the system being observed, to the position, best summarized by “Shut up and calculate!”, that interpretations are pointless as the math already describes what is going on [11].

¹The X basis has eigenstates $|+\rangle = \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and $|-\rangle = \frac{|0\rangle-|1\rangle}{\sqrt{2}}$.

²In order to send information one would need to be able to deterministically force a measurement outcome on the other side which is not possible. It is only possible to correlate bases and measurement outcomes after the fact which still requires a classical communication channel.

1.1.3 The power of entanglement

The amount of (classical) information required to represent n qubits is vast. Because of entanglement there is information not only in the state of individual (isolated) qubits but also in non-classical correlations between them. A useful analogy to grasp the nature of this extra information is that of a quantum book [12]. We can imagine a 100 page classical book and compare this to a 100 page quantum book. If we have read the first 10 pages of the classical book, we have obtained about 10% of the information contained in the book. If on the other hand, we have read out the state of all the qubits on the first 10 pages of the quantum book, we have only gleaned a tiny fraction of the information contained in the quantum book as most information is contained in the correlations between the characters on different pages.

The quantum book analogy is nice because it not only emphasizes the exponential nature of quantum information but simultaneously highlights an important limitation. By reading out the first 10 pages of the quantum book we have simultaneously destroyed a large fraction of the information contained in the book. If all pages are read out, the quantum book is reduced to a classical book containing only classical correlations. Although the quantum book can in principle contain more information than the classical book only a tiny amount is accessible. To be specific, at most 1 bit of classical information is accessible per qubit. This limitation is known as *Holevo's bound* [13]. Working around the limitations imposed by Holevo's bound is one of the main challenges when designing quantum algorithms. In fact, to many people it was not obvious that this could be done at all until Shor discovered his now famous factoring algorithm [14]. In addition to Shor's algorithm which has applications in cryptography, there are applications in chemistry [15], machine learning [16] and material science [17].

1.1.4 The fragility of qubits

Part of the reason quantum mechanics is considered strange is that the predicted phenomena are not observed in our daily (classical) lives. This transition from quantum to classical is described by decoherence. A perfectly isolated quantum system, as described in the preceding sections, is described by unitary evolution according to the Schrödinger equation and is completely coherent. A measurement, effectively an interaction with the environment, forces the system in a classical state and completely decoheres the system. Realistic quantum systems are never perfectly isolated from the environment. Through interactions with the system the environment can, in effect, monitor some of the system observables resulting in decoherence [18].

A key challenge when attempting to control quantum systems is to find the right balance between a good isolation from the environment while still allowing the desired interactions required for control and measurement. To perform any quantum algorithm with high fidelity, the coherence time must be significantly longer than the time it takes to perform the operations. There are several hardware platforms in which basic quantum algorithms have been performed, such as nuclear magnetic resonance systems [19–21], trapped ions [22, 23] and optical systems [24, 25]. Notable solid-state platforms include superconducting transmon

qubits [26–28], NV[−] centers in diamond [29, 30], and more recently quantum-dot-based spin qubits [31, 32].

Although these demonstrations show that it is possible to perform small quantum algorithms, they do not show that it is possible to perform a useful quantum computation. Even if we ignore the practical problems involved in scaling up these systems and assume that all error rates remain constant³, there is a more fundamental problem. To take full advantage of the increased qubit count of larger devices one needs to perform more operations, increasing the circuit depth. The circuit depth, and as a consequence the power of quantum computation, is directly limited by the error rate of the system [33–35]. Although a recent demonstration of quantum supremacy [36] has shown that it is possible to use this limited coherence to perform a task that cannot be performed on a classical computer, and recent advances in hybrid classical quantum algorithms [37–39] suggest that there may be useful applications for noisy-intermediate-scale quantum computers [33], this is nowhere near as impressive as the exponential speedups promised by universal quantum computation.

1.1.5 Quantum error correction and fault tolerance

To reach the full potential of quantum computing, quantum bits must be protected from errors. Where classical error correction relies on copying information and measuring the state to know what correction to perform, this is not possible in quantum information. Unlike classical bits, qubits cannot be copied due to the no-cloning theorem [40, 41]. Additionally, measurements are projective, thereby destroying the encoded information. In a quantum error correcting code the information of a single logical qubit is encoded in an entangled state consisting of multiple physical qubits. Instead of measuring the state of the encoded qubit directly, specific measurements are performed that are designed to only extract information on errors and no information on the logical qubit state. By repeatedly performing these measurements, it is possible to continuously monitor the occurrence of errors. By using a model that describes the probabilities of specific errors occurring, it is possible to decode the error syndrome and reconstruct the original state.

The simplest example of a quantum error correcting code is the linear bit-flip (or phase-flip) code [42–44]. In this code, a logical qubit is encoded in an entangled state consisting of multiple physical data qubits (see Figure 1.2). In the three-data-qubit bit-flip code, a logical qubit is encoded as: $|\psi\rangle_L = \alpha|0\rangle_L + \beta|1\rangle_L = \alpha|000\rangle + \beta|111\rangle$. To determine if an error has occurred, a Z parity check is performed on adjacent qubits that measures if two qubits are the same in the Z basis⁴. The simplest way to perform a parity check is to use an ancilla qubit. By performing a CNOT between each data qubit and the ancilla, the ancilla will be flipped for each data qubit in $|1\rangle$. An ancilla measurement will return $+1$ if both qubits are the same in the Z basis, and -1 if one is different. From the outcomes of these measurements it is possible to determine the required corrections and recover the original state.

As an example, let us assume a bit-flip (X) error happens on the last qubit so that the resulting state is $\alpha|001\rangle + \beta|110\rangle$ and the ancilla measurements will return $+1$ and -1 .

³Which is a bad assumption, see also Chapter 3.

⁴By instead measuring the X parity, this code corrects for phase-flip (Z) errors.

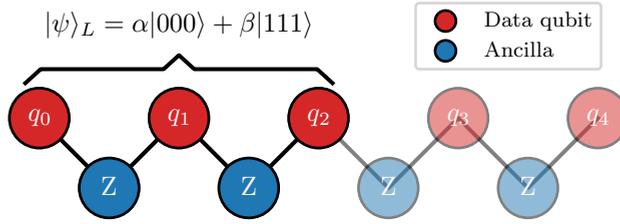


Figure 1.2.: Connectivity required for the bit-flip code. A logical qubit is encoded in an entangled state of multiple data qubits. By performing Z parity measurements on adjacent qubits it is possible to diagnose and correct bit-flip (X) errors. The Z parity of data qubits can be mapped to an ancilla qubit by performing CNOT operations between the data qubits and the ancilla qubit. The parity can be measured by measuring the state of the ancilla without gaining any information on the coefficients α and β that define $|\psi\rangle_L$. Solid circles denote the qubits required for a distance $d = 3$ bit-flip code, by adding the transparent qubits the distance of the code is increased to $d = 5$ allowing the code to correct multiple errors.

There are two possible states that are consistent with the observed error syndrome, $|\psi_1\rangle = \alpha|001\rangle + \beta|110\rangle$ and $|\psi_2\rangle = \alpha|110\rangle + \beta|001\rangle$ corresponding to one and two errors having occurred respectively. Because a single error is more likely than two errors⁵, the third bit will be flipped, recovering the original state. If however, two errors had occurred, the code would not have been able to detect it and the “correction” would have completed a chain of operations that performs a logical bit-flip, resulting in a logical error. The *distance* d of a code is defined as the shortest chain of operations that performs a logical operation. To be robust against multiple errors occurring per measurement cycle, one would need to increase the distance of the code. Stabilizer codes, of which the bit-flip code is one, are robust to $k = \lfloor \frac{d-1}{2} \rfloor$ errors per cycle⁵ [45]. A code is said to operate below the *fault-tolerance threshold* if increasing the distance of the code results in a reduced error rate on the encoded logical qubit. To achieve fault tolerance, the individual error rates need to be small enough that the errors introduced by adding more qubits are smaller than the protection gained by increasing the distance.

In general, the errors that a qubit experiences are more complex than simple bit-flip errors. A qubit can, for example, be slightly detuned resulting in a phase error of a few degrees. A remarkable property of stabilizer codes is that every time a round of measurements is performed, the system is projected into a state where an error either has or has not occurred, effectively turning a *continuum* of errors into a set of *discrete* errors. Because an error channel on a single qubit can be expressed as a combination of the identity I , the bit flip X , the phase flip Z and the combination of bit- and phase-flip XZ , an error correcting code is capable of *universal* error correction if it can detect and correct both bit- and phase-flip errors [2]. A popular example of a code capable of universal error correction is the surface code [46–48].

⁵Under the error model of independent and identical error rates per qubit.

For more information on quantum error correction we refer to Chapter 10 of Nielsen and Chuang [2] and to Terhal [45]. We recommend Fowler *et al.* [48] for a review on the surface code in particular.

1.2 Outline of this thesis

Where this chapter focused on introducing quantum computing and motivating why one would be interested in creating a quantum computer, the rest of this thesis deals with building and controlling a quantum computing system based on superconducting transmon qubits.

In Chapter 2 we discuss our approach to building a quantum computer. We start by giving the full-stack view of quantum computing and discussing the advantages and disadvantages of this way of looking at the system. In Section 2.2 we introduce the **QuSurf architecture**, our architecture for a quantum computer and motivate the design choices based on the target application of realizing quantum fault tolerance.

Assessing the performance of superconducting quantum processors discusses different error sources affecting system performance and how to characterize these. We discuss how to characterize standard operations as well as other error sources such as leakage, non-Markovianity, and several forms of crosstalk.

In **Restless tuneup of high-fidelity qubit gates**, we present a tuneup protocol for single-qubit gates with a tenfold speedup over traditional methods that achieves 0.999 average Clifford fidelity in one minute. Restless tuneup eliminates the overhead introduced by qubit initialization by performing a closed-loop optimization of a cost function based on correlations between measurement outcomes.

In Chapter 5 we present two schemes for actively depleting photons that remain after a measurement from a readout resonator, reducing the wait time before a qubit can be used after a measurement.

Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor introduces Cryoscope, a technique for characterizing and correcting linear-dynamical distortion in a control line, a prerequisite to performing high-fidelity two-qubit gates based on flux-pulses.

Chapter 7 introduces the Net-Zero gate, a fast, high-fidelity two-qubit gate that relies on leakage interference to suppress leakage out of the computational subspace. By making use of a symmetry condition of the transmon, the pulse can have a net-zero integral, making the gate resilient to long-timescale distortions thereby making the gate repeatable and solving an important challenge for superconducting quantum systems.

Chapter 8 discusses the software required to control a multi-qubit processor. The chapter starts by describing the design philosophy and the core concepts of PycQED, an open-source software package we developed for controlling quantum computing experiments. We then discuss automated characterization and calibration and what is required to scale this to larger systems.

In the **conclusion & outlook**, we reflect on the chosen approach to building a prototype quantum computer and discuss what we would do differently were we to design a system

today. In Section 9.3 we provide an outlook on what we believe to be the main challenges when scaling to a useful kilo-qubit sized system. We frame these challenges in terms of three grand problems: [the application problem](#), [the fabrication problem](#), and [the calibration problem](#).

So far, we have motivated quantum computing by focusing on the fundamentally different nature of quantum mechanics and the potential applications of quantum computing. However, there is a big gap between designing quantum algorithms and building a system capable of executing such algorithms. In this chapter, we discuss our approach to building a full-stack quantum computer based on superconducting transmon qubits. In the first section, we discuss the advantages and disadvantages of the full-stack view of quantum computing and use it to motivate our approach to building a quantum computer. In the second part of this chapter, we propose an architecture for a quantum computer based on superconducting transmon qubits and motivate the design choices based on the target application of realizing quantum fault tolerance.

2.1 The full-stack quantum computer

Taking inspiration from classical computing, a quantum computer is envisioned to consist of several layers, each with their own function, challenges and experts [49, 50] (Figure 2.1). This layering is intended to achieve a separation of concerns so that experts in each domain can work on their own problems without worrying about details of other layers. Different proposals of full-stack quantum computing differ on the layers that are identified and where the boundaries between the layers are, but all share a similar structure. At the highest level of the stack lie the applications, the quantum algorithms being developed and implemented by quantum developers. These algorithms are expressed in a language such as Q# [51] and converted into primitive instructions [52, 53] by a compiler that takes the hardware constraints into account [54, 55]. These instructions are interpreted by electronics which create the control signals used to control the quantum device located inside a dilution refrigerator. The stack shown in Figure 2.1 is adapted from [49] to reflect all major components in our architecture. A unique feature of this stack is that it explicitly defines the instruction set as a layer that forms the connection between the software and hardware.

From this picture of a full-stack quantum computer it follows that, in order to build such a machine one requires specialized quantum engineers at all levels of the stack. One needs clean-room engineers who can design and fabricate the quantum devices and improve the yield, coherence times and parameter targeting. One needs cryogenic engineers to operate and develop the fridges that support ever larger devices. Electrical engineers are needed for the control electronics that achieve state of the art specifications and support the operations required by the instruction set. Computer engineers define that instruction set and develop

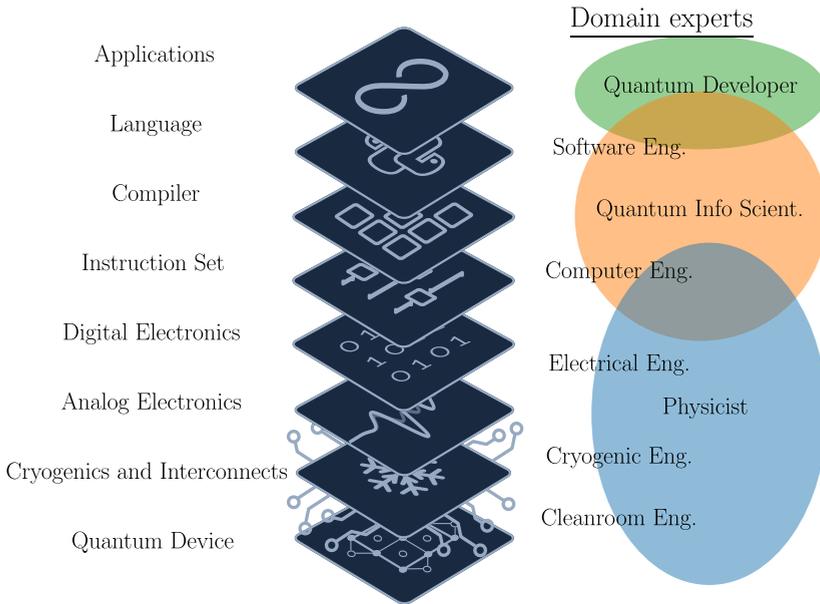


Figure 2.1.: A full-stack quantum computer consists of different layers each with different responsibilities. The icons in the middle represent these layers and are adapted from the Quantum Infinity project. Different engineering disciplines (right) are required to build a full-stack quantum computer while different scientific disciplines are active on different layers of the stack. Due to the inter-disciplinary nature of this field there is significant overlap between the different disciplines.

the compiler infrastructure required to support a high level language suitable to the quantum developers to express their algorithms.

2.1.1 Limitations of the full-stack quantum computer

Although this neatly ordered vision of a full-stack quantum computer is not incorrect, it can be misleading. It suggests that we understand how to build a quantum computer well enough to be able to precisely define the responsibilities of each layer. While physicists are used to abstraction in theoretical models, it is relatively easy to modify these models until they provide a useful description of a system. Making these changes when physically building a system is significantly more difficult. This is particularly problematic when the price in performance of choosing the wrong abstraction can be the difference between being able and not being able to perform a certain task. Here we give three examples of the consequences of inconveniently chosen abstractions.

Non-computational states

Even though physical realizations of qubits generally contain more than two states, it is often a good idea to ignore the *non-computational* states in order to keep things simple, both from a theoretical and an experimental point of view. However, when performing quantum error correction, leakage to non-computational states¹ can no longer be ignored [56–62]. A simple model for leakage replaces qubits with qutrits (three-level systems) but otherwise uses the same language of quantum information.

The problem is that this seemingly innocent change to our model propagates through the stack. The quantum assembly language used to express quantum circuits assumes two-level systems and as such does not allow the control pulses required for active de-leaking strategies (see Section 3.4.1). Being forced to define these operations at a lower (hardware) level makes the abstraction more of a hindrance than a help.

Similarly, by imposing that measurements must return either 0 or 1, valuable information is thrown out. In transmon systems, in particular, the second excited state can typically be measured directly [63, 64] which can be quite valuable for quantum error correction [42, 60] and characterization protocols. If one is willing to relax the constraints even further, one could return continuous outcome values for each individual measurement shot or use machine learning techniques to assign a probability for a state transition occurring during the measurement [65].

The native gate set

Another commonly made abstraction is that of the gate set. In classical computing, a small set of gates (e.g. AND, OR, NOT) can be used to compute any arbitrary classical function. A similar set of primitive gates exists for quantum computing. A set of gates is said to be *universal* if any unitary may be approximated to arbitrary accuracy by a quantum circuit containing only those gates. A universal set of quantum gates can be created from several single-qubit gates and a single two-qubit gate². This gate set is then supported on a physical system through the *native gate set* consisting of several single-qubit rotations around axes in the equatorial plane and a two-qubit gate such as the Conditional-phase (CZ) gate.

An implicit assumption is that this small discrete set of operations contains all the operations that are to be supported on a particular system. Although this is a reasonable model when dealing with a (hypothetical) fault-tolerant quantum computer, it is very limiting for near-term quantum computers. There are many algorithms that require significant overhead to decompose into a limited gate set while the quantum hardware itself often supports a contin-

¹For a detailed discussion on leakage see Section 3.4.1.

²The *standard set* of universal gates consists of the Hadamard gate $H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, the $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$ and $S = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/2} \end{pmatrix}$ gates that perform rotations around the z -axis of 45 deg and 90 deg respectively, and the two-qubit CNOT gate [2].

uum of operations. In a recent work [66], we have modified our system stack to use a *dynamic gate set* to support parametrized operations in the form of the single-qubit $R_{xy}(\varphi, \theta)$ gate

$$R_{xy}(\varphi, \theta) = \begin{pmatrix} \cos(\theta/2) & -ie^{-i\varphi} \sin(\theta/2) \\ -ie^{i\varphi} \sin(\theta/2) & \cos(\theta/2) \end{pmatrix}, \quad (2.1)$$

where φ and θ are continuous variables. In this enhanced stack, the gate set was determined at compile time, taking into account the limitations of hardware and ensuring the right waveforms were uploaded to the control electronics before executing the program. This allowed us to perform an algorithm that models a disorder-induced metal-insulator transition and observe the expected dynamics. Although this demonstration is trivial from a physics perspective³, it has significant impact on the system architecture, affecting the language, the compilation and the way the electronics is operated. This approach could also be used to support more complex operations such as parametrized two-qubit interactions [67, 68].

Variational algorithms

A more explicit example of the limitation of a fixed gate set can be found in variational quantum algorithms. These algorithms are inspired by classical variational approaches for finding low-energy states of a quantum system such as a molecule. The rough idea is to define a trial wavefunction (ansatz) as a function of some parameters and find the minimum expectation value of the energy with respect to these parameters. This minimized ansatz is then an approximation to the lowest energy eigenstate. Recently, it has been realized that emulating this technique using a parametrized quantum circuit has several advantages [37–39]. Specifically, one can directly produce a state using n qubits which would take an exponential number (in n) of complex numbers classically. Because the state is directly created, one can extract properties from the state directly, either for optimization purposes [69, 70] or because these are inherently interesting [71–74].

Variational algorithms are a particularly interesting example as these completely violate the current layers of abstraction, requiring knowledge of the hardware at the level of algorithms. One could perform these algorithms using a dynamic gate set that is determined at compile time but this comes with severe limitations. For one, one would need to recompile the quantum circuit for every iteration of the optimizer, introducing a significant overhead. By viewing parametrized gates not just as a freedom to express a fixed circuit more efficiently but as a means of parameterizing a circuit, it is possible to delay specifying the parameters until the very last compilation step. This can be achieved by supporting *symbolic* variables as opposed to *floating point* variables as a means of expressing parametrized operations [75]. Depending on how well parametrized operations are integrated into the compiler, it is possible to find an optimal circuit representation of the desired unitary in terms of the gate parametrization.

³Changing the phase φ of the gate corresponds to changing the phase of the microwave pulse while changing the rotation angle θ corresponds to the amplitude.

2.1.2 Building a full-stack quantum computer

The reason to highlight these limitations is not because one should always think of qutrits instead of qubits or that all gates should be parametrized. Not only are there more examples we have not discussed⁴, the requirements will also keep changing as our understanding of quantum computing matures. Because qubits are so fragile and hard to control, the development of quantum computing cannot afford the limitations that come with choosing the wrong abstractions. At the same time, a lot of the developments in quantum algorithms would not have been possible without the abstractions of quantum information science. To balance the bottom-up nature of taking advantage of all relevant details of the hardware with the top-down nature of making a system capable of supporting abstract algorithms, I advocate an approach to building quantum computers consisting of three parts: vision, focus, and enough slack.

A grander *vision* is needed to provide direction and inspiration for a project on this scale. Such a vision needs to be specific enough to be useful yet general enough to be able to adapt to an evolving understanding of quantum computing. In that sense, the image of the full-stack quantum computer is a good vision. It is generic enough to be able to hold up to a radical insight such as the emergence of variational algorithms while at the same time providing a clear direction to the overall project as being a circuit-based quantum computer.

Because the limitations of a chosen paradigm do not become evident until using it, it is important to build prototypes. The consequences of design choices become evident when integrating different layers into a fully functional quantum computer. To keep *focus*, clear and specific intermediate goals must be defined as milestones. The success of these prototypes should be judged by two criteria: did it achieve the intended goal and is the approach extensible? The first criterion is to limit the scope and to prevent priorities from shifting unconsciously, while the second is to ensure that the right lessons are learned from these projects. It is important to remember that these prototypes are tools in service of the grander vision, achieving the milestones should never compromise the ability to sustain development.

Most of the problems described in Section 2.1.1 can be traced back to a lack of interest in the opinion of experts with different specializations. The resulting miscommunication often runs a lot deeper than disagreeing on the resulting specifications. Some engineers tend to think the world *is* like the model, and if it is not, it is the hardware engineer's (physicist's) job to ensure it is. Some physicists, on the other hand, tend to be so aware of the limitations of a model that they do not realize the consequences of changing *their* model, resulting in ever-changing specifications. To avoid this, it is important to understand not only what points are being made but also what arguments are given to support them and why. Understanding different viewpoints often allows identifying seemingly irrelevant details that can have far-reaching consequences and force one to think about the problem differently.

However, if there is too much focus it is not possible to overcome these differences. By providing some *slack* it is possible for people to try and understand problems on a deeper level and reflect on the chosen approach. This facilitates that intermediate steps remain aligned with the grander vision.

⁴E.g., non-markoviantiy (Section 3.4.2), calibrations (Chapter 8), etc.

There is an obvious tension between a focus on milestones on one hand, and providing room for reflection and curiosity on the other. It is not clear a priori where the ideal balance lies, and it will likely change as our understanding of quantum computing matures. It is worth noting, however, that the need for curiosity and reflection is explicitly driven by the limitations of our understanding of quantum computing systems.

2.2 The QuSurf architecture

The work presented in this thesis deals with challenges in the control of quantum computers based on superconducting transmon qubits [76, 77]. With single- and two-qubit gate fidelities at the fault-tolerance threshold [78, 79], multi-qubit systems publicly available in the cloud [80–82] and the recent demonstration of quantum supremacy [36] superconducting qubits are firmly among the most mature platforms for quantum computing. In this section we start by introducing superconducting transmon qubits before discussing the goals of the LogiQ project and the Quantum Infinity project. We use these projects to motivate some key design considerations that resulted in the QuSurf architecture. The QuSurf architecture takes its name from the identically named consortium in the IARPA LogiQ project consisting of TU Delft, TNO, ETH Zürich and Zurich Instruments [83].

2.2.1 Superconducting transmon qubits

One of the simplest superconducting circuits is that of a dissipationless LC resonator [Figure 2.2(a)]. The Hamiltonian of this circuit is identical to that of a particle in a one-dimensional quadratic potential, known as the Quantum Harmonic Oscillator (QHO)

$$\hat{H}_{\text{QHO}} = 4E_C \hat{n}^2 + \frac{1}{2} E_L \hat{\phi}^2, \quad (2.2)$$

where $E_C = e^2/(2C)$ is the charging energy required to add an electron with charge e to the island with capacitance C and $E_L = (\Phi_0/2\pi)^2/L$ is the inductive energy with $\Phi_0 = h/2e$ the superconducting magnetic flux quantum⁵. The operators $\hat{n} \equiv \hat{Q}/2e$ and $\hat{\phi} \equiv 2\pi\hat{\Phi}/\Phi_0$, corresponding to the reduced charge and reduced flux, form a canonical conjugate pair. The eigenstates of the Hamiltonian are an infinite series of equally spaced levels E_k , with $E_{k+1} - E_k = \hbar\omega_r = \sqrt{8E_C E_L} = \hbar/\sqrt{LC}$.

Defining a qubit in the number of excitations of the QHO would be a natural choice but it is difficult to individually address any transition because of the equidistant level spacing. To make the QHO a suitable qubit one needs to add a non-linear element to the system so that the transition frequencies are sufficiently different. The Josephson junction [84, 85] is a superconductor-insulator-superconductor junction that forms a dissipationless non-linear element. By replacing the inductor in the QHO with a Josephson junction (JJ) we obtain the modified Hamiltonian,

$$\hat{H}_{\text{Transmon}} = 4E_C \hat{n}^2 - E_J \cos \hat{\phi}, \quad (2.3)$$

⁵Note that there are two conventions for the Charging energy that are commonly used: the energy to add a single electron $E_C = e^2/(2C)$ or the energy required to add a Cooper pair $E_{C'} = (2e)^2/(2C)$. Using the Cooper pair convention $E_{C'}$ results in the following Hamiltonian: $\hat{H}_{\text{QHO}'} = E_C \hat{n}^2 + \frac{1}{2} E_L$.

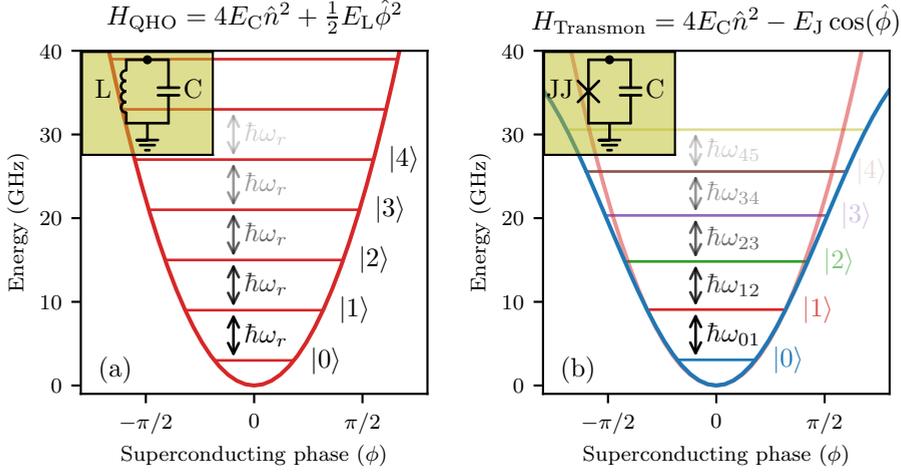


Figure 2.2.: (a) Energy potential of the Quantum Harmonic Oscillator (QHO) and (b) the Transmon. The QHO has a parabolic potential, resulting in equally spaced energy levels with spacing $\hbar\omega_r$. By replacing the inductor with a Josephson junction a nonlinearity is introduced in the Hamiltonian.

where $E_J = I_c \Phi_0 / 2\pi$ is the Josephson energy and I_c the critical current of the junction. In the limit $E_J \gg E_C$, also known as the transmon limit, the energy potential is essentially a weakly anharmonic oscillator [Figure 2.2(b)]. In this limit the *computational subspace* can be defined using the two lowest levels, $|0\rangle$ and $|1\rangle$ with transition frequency $\omega_q = \omega_{01} \approx (\sqrt{8E_J E_C} - E_C) / \hbar$. The anharmonicity is $\alpha = \hbar\omega_{12} - \hbar\omega_{01} \approx -E_C$. A key advantage of the transmon [76] over the Cooper-pair box ($E_J < E_C$) [86, 87] is that charge noise is exponentially suppressed in the transmon. Relaxation times of planar transmon qubits are typically on the order of $T_1 \sim 20 \mu\text{s}$ with similar dephasing times, though we have observed relaxation times in excess of $100 \mu\text{s}$ in test devices (unpublished) and an average $T_1 \sim 45 \mu\text{s}$ for a fully connected 7-qubit device (Chapter 3). Coherence times $T_1 \sim 500 \mu\text{s}$ have been reported [88] and, although no details on the device and experiment have been made public, it is a very promising result.

Flux tunable transmons

By making a small modification to the transmon circuit ω_q can be made flux tunable. A common approach is to replace the single Josephson junction with a loop consisting of two Josephson junctions, a superconducting quantum interference device (SQUID) [89]. For identical Josephson junctions this results in the Hamiltonian,

$$\hat{H}_{\text{Transmon}} = 4E_C \hat{n}^2 - \underbrace{2E_J |\cos(\pi \hat{\Phi}_{\text{ext}} / \Phi_0)|}_{E'_J(\hat{\Phi}_{\text{ext}})} \cos \hat{\phi}, \quad (2.4)$$

where $\hat{\Phi}_{\text{ext}}$ is the external flux through the SQUID-loop. Equation (2.4) corresponds directly to Equation (2.3) with the key difference that the effective E'_J , and as a result ω_q , is flux tunable.

The price of this tunability is that the qubit becomes sensitive to flux noise. This sensitivity to flux can be reduced by making the junctions asymmetric at the cost of the tunable frequency range [90, 91]. By introducing flux tunability it is possible to statically tune qubits to specific operating conditions such as required for frequency multiplexing of microwave controls [92] and gates designed for specific resonance conditions [93], or detune from areas of reduced coherence [36, 94]. Conditional operations can be performed by dynamically controlling the flux to pulse close to interactions [26, 95] or by modulating the flux to parametrically activate an interaction [96].

Single-qubit gates

Single-qubit gates are performed by applying microwave pulses at ω_q . The angle of rotation θ can be controlled by changing the amplitude of the pulse while the phase φ of the pulse corresponds to the phase of the rotation axis in the equatorial plane of the Bloch sphere [Equation (2.1)]. A common pulse shaping technique to reduce leakage to the second excited state is to use a Gaussian envelope for the in-phase and a derivate of Gaussian shape for the out-of-phase quadrature of the microwave pulses [97, 98]. The anharmonicity of the transmon sets a lower bound on the duration of single-qubit operations and is typically chosen to allow single-qubit gates within 20 ns that achieve gate fidelities in excess of 99.9% [78, 99] (Chapter 4). Rotations around the z axis can be performed by updating the phase of the microwave pulses [21, 100] or, in the case of flux-tunable transmons, by temporarily detuning the qubit using a short flux pulse.

Dispersive readout

Transmon qubits are typically not read out directly, instead the qubit is coupled to a superconducting resonator (QHO). The state of the qubit can then be determined by probing the (qubit-dependent) state of the resonator using a microwave pulse. The qubit⁶ resonator interaction is described by the Jaynes-Cummings Hamiltonian [64, 76, 101, 102],

$$\hat{H}_{\text{JC}} = \hbar\omega_r \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) - \frac{\hbar\omega_q}{2} \hat{\sigma}_z + \hbar g \left(\hat{\sigma}_+ \hat{a} + \hat{\sigma}_- \hat{a}^\dagger \right), \quad (2.5)$$

where ω_r and ω_q are the resonator and qubit frequencies, \hat{a}^\dagger and \hat{a} the creation and annihilation operators of the resonator, and $\hat{\sigma}_+$ and $\hat{\sigma}_-$ denote adding and removing an excitation from the qubit, and g is the coupling strength.

In the *dispersive limit*, where detuning between the resonator and the qubit $\Delta = \omega_q - \omega_r$ is significantly larger than the coupling strength g and the resonator linewidth κ , i.e., $|\Delta| \gg g, \kappa$ there is no direct energy exchange between the resonator and the qubit. Instead the qubit

⁶The qubit approximation of a transmon corresponds to ignoring all states with more than 1 excitation.

and resonator transitions "push" each other away. In this limit the dispersive approximation can be used after which the Hamiltonian takes the form,

$$\hat{H}_{\text{disp}} = (\hbar\omega_r + \chi\hat{\sigma}_z) \left(\hat{a}^\dagger \hat{a} + \frac{1}{2} \right) - \frac{\hbar\tilde{\omega}_q}{2} \hat{\sigma}_z, \quad (2.6)$$

where $\chi = g^2/\Delta$ is the qubit-state dependent *dispersive shift* of the resonator frequency. The transmon, however, is not a two-level system. Taking the second excited state into account, the dispersive shift becomes $\chi = \chi_{01} + \frac{\chi_{12}}{2} = -\frac{g^2}{\Delta} \left(\frac{1}{1+\Delta/\alpha} \right)$, with $\chi_{ij} \equiv \frac{g_{ij}^2}{\hbar\omega_{ij} - \hbar\omega_r}$, g_{ij} is the coupling of the $i \rightarrow j$ transition to the resonator. The qubit frequency also changes slightly $\hbar\tilde{\omega}_q = \hbar\omega_q + g^2/\Delta$ due to zero-point fluctuations of photons in the resonator. The dispersive approximation breaks down when the number of photons in the resonator approaches the critical photon number $n_{\text{crit}} \equiv \Delta^2/(4g^2)$.

Coupling transmons

Physical interactions between qubits are required to perform entangling gates. Interactions between transmon qubits are commonly created by coupling to a common *coupling bus* resonator [26] or by coupling them directly to each other [103]. More complex designs make the coupling strength tunable by introducing extra control lines [104–106].

The Hamiltonian of two transmons, q_0 and q_1 , capacitively coupled to a resonator (circuit shown in Figure 2.3) consists of three terms corresponding to the individual subsystems and two terms that describe the interaction between each transmon and the resonator,

$$\begin{aligned} \hat{H} = & \overbrace{\sum_i \hbar\omega_i^{(0)} |i\rangle_0 \langle i|_0}^{q_0} + \overbrace{\sum_j \hbar\omega_j^{(1)} |j\rangle_1 \langle j|_1}^{q_1} + \overbrace{\hbar\omega_r \hat{a}^\dagger \hat{a}}^{\text{res.}} \\ & + \sum_i g_{i,i+1}^{(0)} \left(a_r^\dagger |i\rangle_0 \langle i+1|_0 + a_r |i+1\rangle_0 \langle i|_0 \right) \\ & + \sum_j g_{j,j+1}^{(1)} \left(a_r^\dagger |j\rangle_1 \langle j+1|_1 + a_r |j+1\rangle_1 \langle j|_1 \right), \end{aligned} \quad (2.7)$$

where the superscript $k \in \{0, 1\}$ denotes the subsystems the terms belong to. $g_{i,i+1}^{(k)}$ are the level dependent coupling strengths of the transmons to the resonator. Taking the dispersive

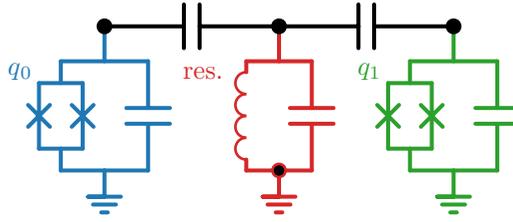


Figure 2.3.: Circuit diagram of two flux-tunable transmons q_0 (blue) and q_1 (green) coupled through a coupling resonator (red).

approximation, $g_{i,i+1}^{(k)} \ll \Delta_{i,i+1}^{(k)} = \hbar\omega_{i,i+1}^{(k)} - \hbar\omega_r$ and truncating at $i \leq 2$, the Hamiltonian can be rewritten as,

$$\begin{aligned}
 \hat{H} = & \overbrace{\sum_i \hbar\omega_i^{(0)} |i\rangle_0 \langle i|_0 + \chi_{i,i+1}^{(0)} |i+1\rangle_0 \langle i+1|_0}^{q_0} \\
 & + \overbrace{\sum_j \hbar\omega_j^{(1)} |j\rangle_1 \langle j|_1 + \chi_{j,j+1}^{(1)} |j+1\rangle_1 \langle j+1|_1}^{q_1} \\
 & + \hat{a}^\dagger \hat{a} \left(\hbar\omega_r + \underbrace{\sum_{k \in \{0,1\}} \left[-\chi_{01}^{(k)} |0\rangle_k \langle 0|_k + \left(\chi_{01}^{(k)} - \chi_{12}^{(k)} \right) |1\rangle_k \langle 1|_k \right]}_{\text{Coupling term}} \right) \\
 & + \sum_{i,j} \frac{g_{i,i+1}^{(0)} g_{j,j+1}^{(1)} \left(\Delta_{i,i+1}^{(0)} + \Delta_{j,j+1}^{(1)} \right)}{2\Delta_{i,i+1}^{(0)} \Delta_{j,j+1}^{(1)}} \left(|i\rangle_0 \langle i+1|_0 \otimes |j+1\rangle_1 \langle j|_1 \right. \\
 & \quad \left. + |i+1\rangle_0 \langle i|_0 \otimes |j\rangle_1 \langle j+1|_1 \right). \tag{2.8}
 \end{aligned}$$

Similar to Equation (2.6) it is possible to identify the transmon dependence in the resonator term as well as the change in the individual transmon terms. Of particular interest is the direct transmon-transmon exchange coupling that emerges. When there are no photons in the resonator, the resonator term drops out and the system Hamiltonian corresponds to that of two directly coupled transmons,

$$\begin{aligned}
 H = & \overbrace{\sum_i \hbar\tilde{\omega}_i^{(0)} |i\rangle_0 \langle i|_0}^{q_0} + \overbrace{\sum_j \hbar\tilde{\omega}_j^{(1)} |j\rangle_1 \langle j|_1}^{q_1} \\
 & + \underbrace{\sum_{i,j} J_{ij} \left(|i\rangle_0 \langle i+1|_0 \otimes |j+1\rangle_1 \langle j|_1 + |i+1\rangle_0 \langle i|_0 \otimes |j\rangle_1 \langle j+1|_1 \right)}_{\text{Coupling term}}, \tag{2.9}
 \end{aligned}$$

with effective coupling,

$$J_{ij} = \frac{g_{i,i+1}^{(0)} g_{j,j+1}^{(1)} \left(\Delta_{i,i+1}^{(0)} + \Delta_{j,j+1}^{(1)} \right)}{2\Delta_{i,i+1}^{(0)} \Delta_{j,j+1}^{(1)}}. \quad (2.10)$$

Besides transmons there exist many more superconducting circuits that can be used for quantum information processing, for a review of superconducting quantum circuits we refer to Gu *et al.* [107] and Blais *et al.* [108] for a review of superconducting qubits in particular we refer to Kjaergaard *et al.* [27] and Krantz *et al.* [28].

2.2.2 Designing for fault tolerance

The quantum computing platform discussed in this thesis is designed with two larger projects in mind. In the LogiQ project the goal is to demonstrate an error corrected logical qubit with a lifetime longer than that of its constituent physical qubits. The goal of the Quantum Infinity project is to build a full-stack quantum computer with a programmable interface accessible through the cloud. Strict performance thresholds have to be met to successfully protect a logical qubit while mature interfaces and programmability are required for a full-stack quantum computer. Because of the complementary nature of these projects it makes sense to develop a single architecture for both applications. Due to the strict performance requirements for quantum error correction, the architectural choices are motivated primarily by the LogiQ project.

One of the most promising approaches to demonstrating quantum fault tolerance is based on the surface code [46–48] due to the feasible grid-like nearest-neighbor connectivity requirement and a high error threshold of $\sim 1\%$ per operation [109, 110]. The smallest surface code capable of universal error correction is the distance $d = 3$ Surface-17 consisting of 9 data qubits and 8 ancilla qubits and is shown in Figure 2.4. For a demonstration of fault tolerance a distance $d = 5$ code is required resulting in the 49 qubit Surface-49 also shown in Figure 2.4. See Section 1.1.5 for a discussion on the relation between code distance, error correction, and fault tolerance.

Although the surface code defines the connectivity of the device it does not define device parameters such as qubit frequencies, resonator frequencies and coupling strengths that are required to design, analyze and fabricate a device.

2.2.3 Choice of two-qubit gate

The choice of two-qubit gate in particular has far reaching consequences for the design of the system. There exist three classes of two-qubit gates for superconducting qubits, flux-pulsing based, microwave driven and flux-pulse modulated, all which have achieved gate fidelities of 99 %.

Two-qubit gates can be performed by tuning certain transitions close to resonance [26, 78, 95, 104, 111, 112]. The duration of these *flux-pulsing based* two-qubit gates is directly limited by the effective coupling strength between the transmons, making these the fastest two-qubit

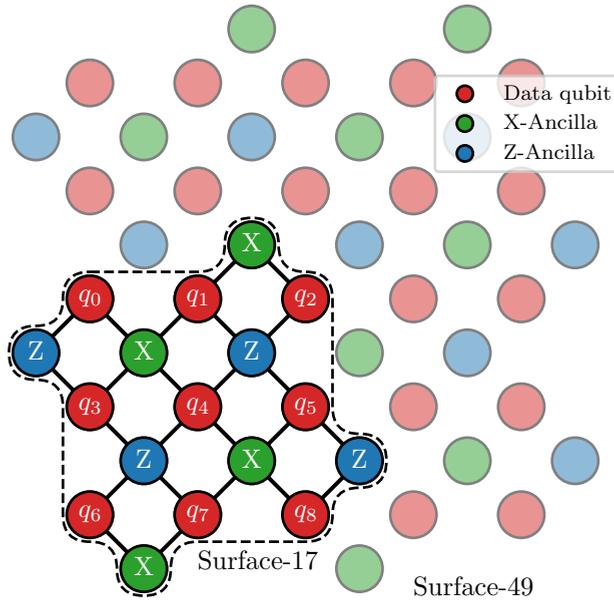


Figure 2.4.: Topology of the distance $d = 3$ Surface-17 (solid) and distance $d = 5$ Surface-49 (transparent) surface code devices. Green and blue circles denote ancilla qubits used to perform parity checks. Red circles denote data qubits.

gates at typical durations of 30 – 60 ns. Although these gates are the fastest, this does not automatically make them the best two-qubit gates. Because pulsing into the required resonance conditions involves detuning from the sweetspot, the performance is typically slightly lower than one would expect based on the gate duration. More importantly, flux-pulsing based two-qubit gates require precise pulse shaping to achieve their performance limits. High-fidelity flux-pulsing based two-qubit gates are very technically challenging because control pulses suffer distortions as they traverse various electrical components on their way to the device. Additionally, when operating in multi-qubit systems, care has to be taken to avoid collisions with unwanted transitions while detuning from the sweetspot.

Alternatively, one can use fixed frequency qubits to perform two-qubit gates using only microwave drives [93, 113–117]. These gates are typically performed by driving one qubit at the frequency of another and are significantly slower ($\sim 150 - 500$ ns) than flux-pulsed gates. The biggest advantage of an all-microwave control scheme is that flux control can be omitted. This reduces the complexity in fabrication, eliminates a potential relaxation channel and simplifies the control electronics. One of the biggest challenges in microwave-based gates is to avoid interactions with undesired transitions. Frequency crowding [118] is particularly problematic when fixed-frequency transmons are coupled to multiple neighbors [119] as required for the surface code.

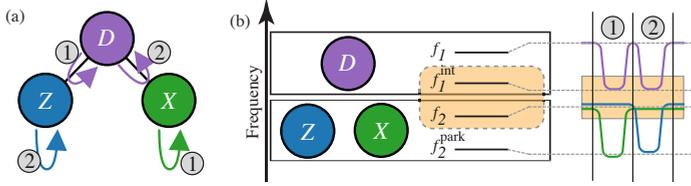


Figure 2.5.: (a) A high frequency qubit, D , is coupled to two low-frequency qubits, Z and X . (b) To perform a CZ gate between D and Z , D detunes from f_1 to the interaction frequency f_1^{int} while X moves to parking frequency f_2^{park} to avoid an unwanted interaction (1). Similarly, to perform a CZ gate between D and X , Z detunes while X is kept fixed (2).

A third approach is to modulate the frequency of the qubit or a coupling element to induce a coupling between the qubits [96, 106, 120]. These *parametrically driven* gates can be seen as a hybrid between the flux-pulsing based and the all-microwave gates. Some of the frequency crowding issues of the all-microwave gates are avoided as the coupling terms are only activated when the drive is activated. The gates are less susceptible to distortions than flux-pulsing based gates because the drives operate at frequencies of ~ 100 MHz, making them significantly easier to realize. However, parametrically driven gates are significantly slower (~ 180 ns) than flux-pulsing based gates while sacrificing the main advantages of an all-microwave architecture.

Due to their speed and easy to understand frequency collisions, flux-pulsing based gates are the choice for our architecture. Addressing frequency collisions is discussed in the next section while addressing distortions is a major subject in Chapters 6 and 7.

2.2.4 Scalable surface code

To perform X or Z stabilizer measurements the parity of multiple data qubits can be mapped to an ancilla qubit using CZ gates. The flux pulsing based CZ gate [26, 95] makes use of an avoided crossing between the $|11\rangle$ and the $|02\rangle$ state of two coupled transmons⁷. By slowly detuning the high-frequency qubit close to the interaction zone and back, the qubit experiences a frequency shift ζ conditional on the low-frequency qubit being in the excited state that can be used to perform a CZ gate.

An additional complexity is added when multiple transmons are coupled together. When a high-frequency qubit, D , is coupled to two qubits, Z and X , operating at a lower frequency (Figure 2.5), D cannot be tuned to the interaction with either Z or X without also interacting with the other. In order to perform a CZ gate between D and Z , X needs to be detuned to avoid an undesired interaction. While X is detuned it cannot be involved in a CZ gate with another high frequency qubit. Avoiding these types of frequency collisions becomes more difficult when more qubits are coupled together and more frequencies are involved.

⁷Avoiding transitions to the non-computational $|02\rangle$ state is an area of active research and is discussed in detail in Chapter 7.

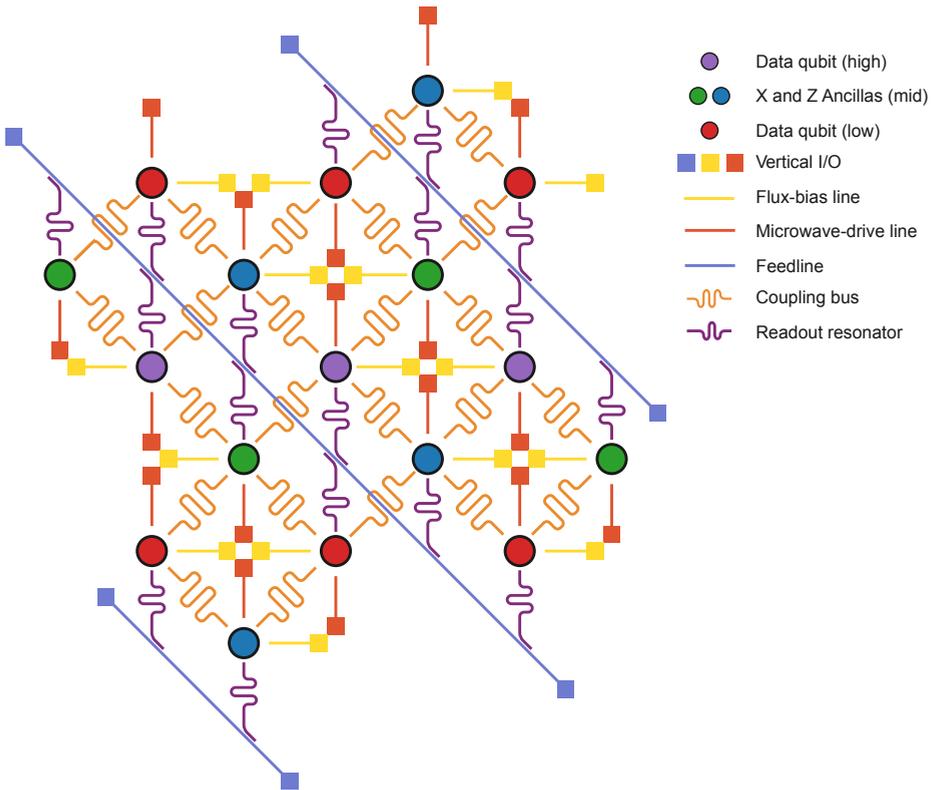


Figure 2.6.: Surface-17 device topology, figure adapted from [121]. Data qubits are placed at either a high (purple) or low (red) frequency while ancilla qubits (green and blue) are placed at the middle frequency. Qubits are coupled to nearest neighbors using coupling buses (orange) and have dedicated microwave-drive lines (red) and flux-bias lines (yellow). Multiple readout resonators (purple) can be read out simultaneously using a common feedline (blue).

Versluis *et al.* [121] have proposed a scalable scheme for performing the error-correction cycle of the surface code. This scheme consists of an 8-qubit unit cell containing only three qubit frequency groups. A unique detuning pattern for each qubit in the unit cell, known as *the flux dance*, allows for a pipelined execution of all parity checks while avoiding all undesired interactions. This 8-qubit unit cell can be instantiated in the design of the Surface-17 device. Figure 2.6 shows the connectivity of the Surface-17 design. Ancilla qubits are all placed at the same frequency while data qubits are placed at a frequency either above or below.

Several innovations in device design and fabrication are required to realize this design. Traditionally all control signals are provided from the sides of the chip. Because the area of a chip scales quadratically while the chip perimeter scales linearly with the lateral chip dimension, 3D integration is required to provide signals at multiple locations on the chip. There are several approaches to realizing 3D integration compatible with high coherence transmons such as through-chip VIAs [121, 122], flip-chip devices [123] and pogo pins [124].

The constraints on the design can be further relaxed by allowing for on-chip transmission-line crossovers known as air bridges [43] and using bus resonators for coupling to provide extra space between the on-chip elements. A high-connectivity transmon known as the Starmon [121] (shown in Figure 2.8) was designed to meet the connectivity requirements for this architecture.

2.2.5 Simulating the error correcting cycle

An important performance metric for error correcting codes is the fault-tolerance threshold. If the error rate is below the fault-tolerance threshold, increasing the distance of the code will result in a reduced logical error rate. For the surface code, the fault-tolerance threshold is typically estimated to be at $\sim 1\%$ [48, 109, 110, 125]. However, the theoretical nature of the error models used in these studies makes it hard to relate these thresholds to experimental parameters such as T_1 . By making the circuit explicit it became possible to perform a simulation with physically motivated error models for each operation. To this end, the full density matrix simulator *quantumsim* [126] was created and used to simulate the experiment [79].

One of the key learnings was that coherence limited gates at a $T_1 = 30 \mu\text{s}$ are good enough to operate slightly below the fault-tolerance threshold. In Chapter 4 we demonstrate a new tuneup protocol for high-fidelity (99.9%) coherence limited single-qubit gates. The challenges involved in performing fast ($< 40 \text{ ns}$) high-fidelity (99.1%) and repeatable two-qubit gates are discussed in Chapters 6 and 7. Because the measurement is the slowest of all the operations used, reducing the measurement time directly improves the circuit performance. In Chapter 5 we reduce the duration of measurements by actively depleting remnant photons in the resonator that prevent operations on the ancilla. To further speed up the simultaneous readout of multiple qubits, high-bandwidth parametric amplifiers [127–129] and Purcell filters [130] are required. For an exhaustive discussion of readout in the QuSurf architecture we refer to Bultink [131].

Although this study [79] included physically motivated models for individual operations, effects such as leakage and crosstalk were not included. Gaining a better understanding of the relevant error sources to improve the models and understand how these effect the error correcting cycle is a subject of ongoing research [62, 132, 133]. In Chapter 3 we give an overview of known error sources, how these can be characterized and their expected effect on quantum error correction.

2.2.6 Controlling the device

The most straightforward approach to controlling a 7–17 qubit device would be to brute force a solution based around the commercial electronics that were used to control previous-generation (2–5 qubits) devices [43, 92]. Although such an approach would probably allow reaching the immediate goal of protecting a qubit against errors, such an approach is not extensible to larger systems (e.g., 49 qubits), a key motivation for pursuing these goals in the first place. The QuSurf control system has been designed to reduce the complexity and monetary cost of larger systems and is shown in Figure 2.7.

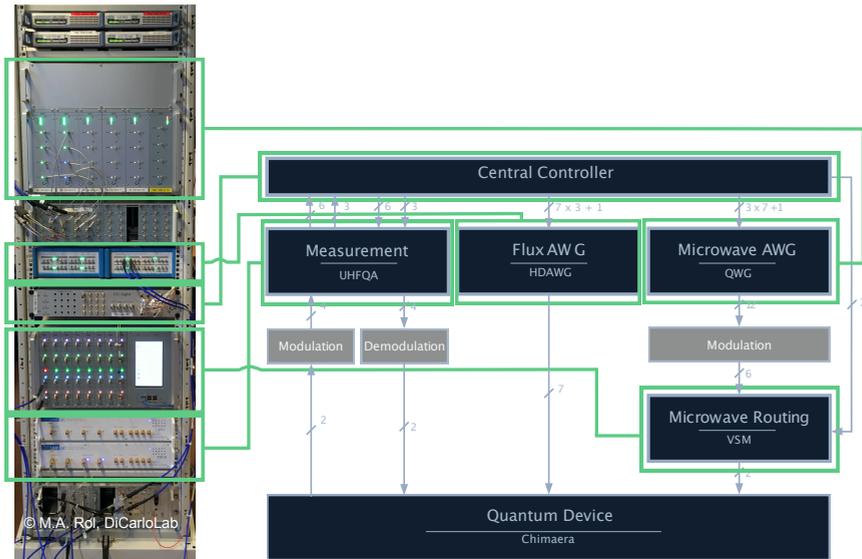


Figure 2.7.: The QuSurf control system consists of a central controller that is used to control dedicated AWGs for microwave and flux control as well as UHFQAs for the readout. A vector switch matrix can be used for routing microwave pulses to different qubits.

The QuSurf control system is a centralized architecture based on *codewords*. A *central controller* (CC) is solely responsible for determining when what operations are performed and communicates with dedicated electronics that execute these operations. The communication with the slave devices is based on codewords that correspond to predefined operations. The pulses for single-qubit control are generated using the QuTech Waveform Generator (QWG). Because the architecture only requires three frequencies for a device of arbitrary size it is possible to make economical use of QWGs and microwave sources. In Asaad *et al.* [92] we have demonstrated the ability to independently control multiple same-frequency qubits using a shared arbitrary waveform generator (AWG). A vector switch matrix (VSM) enables selective broadcasting of primitive control pulses that are tailored for the individual qubits⁸. Flux control, used for two-qubit gates and to detune individual qubits, is performed using a Zurich Instruments HDAWG. The HDAWG contains a special module to correct linear-dynamical distortions in the flux control lines in real time. Correcting distortions in real time is crucial for a fully programmable quantum computer where operations are determined in real time as is required for control flow and feedback [52, 53]. Readout is performed using a Zurich Instruments UHFQA (Quantum Analyzer) that both generates the readout pulses and performs the signal analysis in real-time.

The CC is a programmable device that allows specifying algorithms using a Quantum Instruction Set Architecture (QISA) [52]. The first generations of the CC, the CC-Light and the QCC, support executable Quantum Assembly (eQASM), a QISA that supports explicit

⁸The VSM was the first patent filed by QuTech.

timing control and feedback as well as reconfiguration of the supported operations at compile time [53]. For the QuSurf architecture, we make a distinction between comprehensive feedback, which involves changing the control flow of a program, and fast feedback, which involves masking a deterministic microwave pulse using the VSM, both are supported using the CC. Although eQASM is relatively readable, it is often more convenient to express algorithms using a high-level language. To this end, experiments can be expressed using the python API of OpenQL [54, 134], a compiler framework developed by the Quantum Computer Engineering group in Delft. For a more detailed discussion on QISA, eQASM and the microarchitecture implementation we refer to Fu [134].

A particularly exciting consequence of integrating all different parts of the QuSurf architecture is that one ends up with a working prototype of a full-stack quantum computer (shown in Figure 2.8). In the Quantum Infinity project, we have provided a web interface in which a user could directly execute their own quantum algorithms on our platform. The Quantum Infinity platform has been used for student projects, live demonstrations during the yearly review meetings of the Intel-QuTech partnership, and has resulted in several papers [66, 135].

More interesting than the achievements of this project are the lessons one can learn from building a prototype full-stack quantum computer, both on inter-disciplinary collaboration as well as on a technical level. To give an explicit example, the QuSurf control system is well suited for the intended applications of running quantum error correcting codes and exposing a programmable interface to users. Due to the high connectivity and the copy-pastable device design, it is also very suitable for executing most algorithms. However, the constraints imposed by being codeword based make the system not particularly well suited for calibrations or NISQ applications that require custom rotations or variational parameters. Although these tasks have been performed on the system [66, 67] they required working around the architecture rather than being aided by it. Addressing these limitations is actively being pursued in the latest iteration of the CC and the compiler framework by our collaborators.

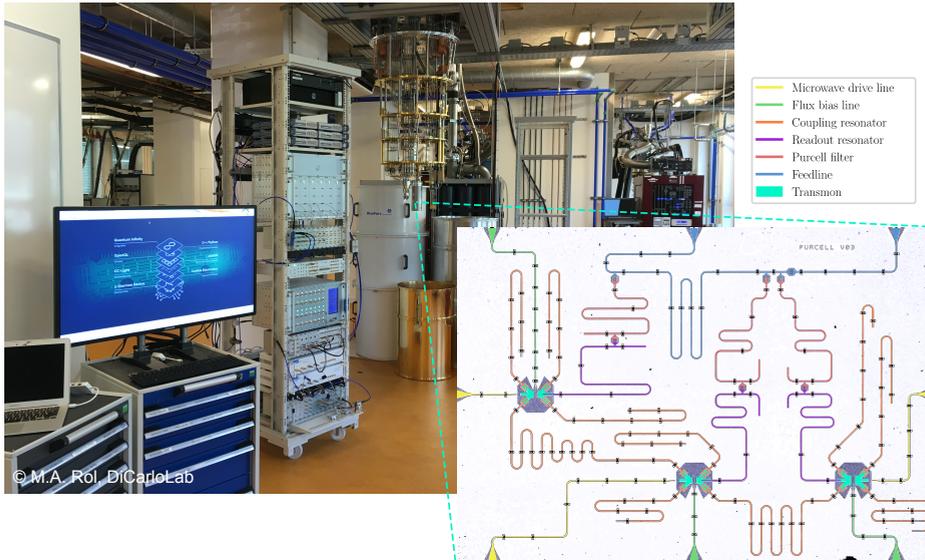


Figure 2.8.: A full-stack quantum computer in DiCarlo lab. From left to right, a monitor showing the web interface of the Quantum Infinity, a rack containing the control electronics of the QuSurf architecture, an opened dilution refrigerator that contains the quantum chip when cooled down. The inset shows a false-colored image of the 3-qubit Purcell device (image adapted from [61]) that was used in the experiments in Chapters 6 and 7.

ASSESSING THE PERFORMANCE OF SUPERCONDUCTING QUANTUM PROCESSORS

Demonstrating Quantum Fault Tolerance requires error rates on the order of $\sim 10^{-2} - 10^{-3}$ for all physical operations. In order to achieve these error rates and to predict device performance, a detailed understanding of the relevant physical error sources is required. There are two major approaches to characterizing errors, a general black-box approach and an ad-hoc approach tailored to system-specific effects. In this chapter, we provide an overview of relevant known error sources in a transmon processor and their effect on physical and logical qubit operations. For each of these error sources we address the question of how to measure and quantify their effect. Extra attention will be given to measuring specific types of non Markovianity and leakage. We will end by contrasting the ad-hoc experimentalists approach to the traditional Quantum Characterization, Verification, and Validation (QCVV) approach of system agnostic characterization routines as well as discussing ways where these can complement each other.

This chapter is based on the feasibility study: "Crosstalk and Other Error Sources in the QuSurf Architecture" [136] that was conducted as part of the IARPA LogiQ program and on a talk titled "Metrics for fault tolerance, an experimentalists approach" that was delivered at the first Assessing the Performance of Quantum Computers (APQC) conference in Estes Park Colorado (2019) [137].

3.1 Introduction

When building a quantum computer it is important to have a detailed understanding of the physical error sources that affect the system and the effect these have on performance. Quantum Characterization, Verification, and Validation (QCVV) is the procedure for estimating the quality of physical quantum systems for use as information processors [138]. Characterization is about determining the effect of operations and noise on a quantum system. Verification and validation are similar terms that have a specific meaning in the context of engineering. Verification is about determining if a device meets the specifications or requirements. Validation on the other hand is about determining if a device can fulfill its intended purpose. It is possible that a device passes the verification stage but does not pass the validation. This can be because of effects that were not taken into account when setting the requirements. As an example, all qubits in a device can have a $T_1 > 50 \mu\text{s}$ which is larger than the specified $T_1 > 30 \mu\text{s}$ but still fail at validation because the device suffers from an effect such as crosstalk that was not taken into account.

Specifying the flux dance [121] and simulating its performance [79] were essential in the design process as these specified the application and set the initial requirements. The error models used in [79] were in large part based on the experience in modeling the experiments described in Chapters 4 and 5. Although these experiments validated the models used in simulation, it was clear that the resulting requirements were initial estimates as relevant effects, such as leakage (Section 3.4.1) and crosstalk (Section 3.5), were not included. These effects were not included because they were difficult to model accurately and efficiently or were unknown at the time. This difficulty in setting hard requirements emphasizes the scientific nature of this field and proves a recurring point of tension in the communication between scientists and engineers.

This chapter gives an overview of our current understanding of what the relevant error sources are, how to measure and quantify their effect, and how these affect the performance of the intended application of quantum error correction. We start by discussing two approaches to QCVV, the system-agnostic black-box approach and the more ad-hoc approach we call the experimentalist's approach. Both approaches are used in the next section when addressing these three questions for the dominant error sources. Data of several devices are provided to illustrate the effects and the characterization methods. We end on a brief discussion on the black-box and experimentalist's approaches as well as an outlook for the use of QCVV for generic and large-scale quantum computing applications. Details on the protocols discussed in this chapter can be found in Appendix A.

3.2 Characterizing errors

There are two distinct approaches to characterizing quantum devices. The first approach, which we refer to as the *black-box approach*, attempts to develop characterization protocols that are as general as possible and rely on as few assumptions as possible. The second approach is an ad-hoc approach we call the *experimentalist's approach*. In this approach, spe-

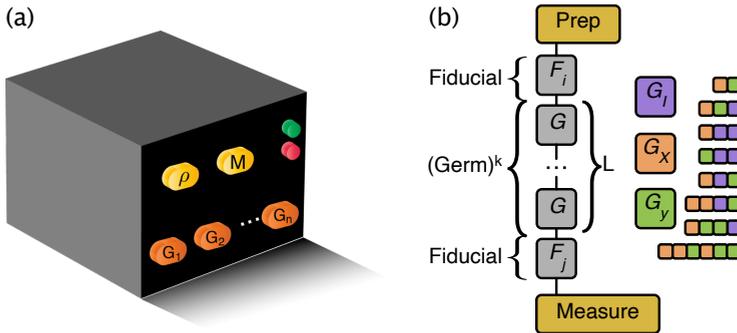


Figure 3.1.: (a) A quantum system is described as a black box with several inputs, visualized as buttons, corresponding to the operations of initialization, measurement and gates. When the measurement button is pressed an output, indicated by the lights, shows the outcome of the operation. (b) Schematic representation of a GST sequence used to characterize operations. Figure adapted from [139].

cific protocols are created to characterize specific error sources. These protocols typically rely on system specific assumptions or approximations and are not intended to be generalized.

3.2.1 The black-box approach

In the black-box approach to quantum characterization the system is described completely by its inputs and outputs [Figure 3.1 (a)] [139]. The inputs correspond to the different operations that can be performed such as initialization, measurement and gates. Whenever the measurement button(s) is (are) pressed an output indicates the outcome of the measurement. The goal is to determine an accurate description of all these operations while relying on a minimal set of assumptions. Typical assumptions are that the system consists of qubits (two-level systems) and that operations are Markovian¹.

Using these assumptions it is possible to construct protocols like randomized benchmarking (RB) [141, 142] and gate set tomography (GST) [139, 143]. RB, which is often considered the gold standard in qubit characterization [144], measures a single error probability ϵ_{RB} that closely approximates the average process infidelity ϵ of a gate. GST, on the other hand, attempts a more complete characterization by determining the completely-positive-trace-preserving (CPTP) maps that describe the individual operations while treating initialization and measurement as vectors. Advantages of these protocols are that the results have rigorous theoretical backing and are expressed in system independent metrics expressed in the language of quantum information.

However, there are some limitations to these protocols. Some of these are practical in nature, for example, both RB and GST require experimental sequences consisting of millions of pulses and require large amounts of data acquisition for characterizing just one or two qubits. In particular, GST also requires excessive data processing which can be on the order

¹We follow the definition of Markovianity of [140], we discuss this in more detail in Section 3.4.

of multiple hours on a desktop computer. While these practical considerations make these protocols experimentally challenging to perform for small systems, it makes them unsuitable for characterizing crosstalk effects on larger devices. A more fundamental problem is that the assumptions on which these protocols are based (Markovianity and two-level systems) are violated in most real-world applications. In transmon based systems, for instance, T_1 relaxation is a major error mechanism. It is commonly known that T_1 is not a constant [94, 99, 145], violating the assumption of Markovianity. Another common error source is leakage to the second excited state. This violates the assumption of two-level systems². Besides these explicit assumptions, it is often implicitly assumed that there is no post-measurement state, something that is of vital importance when reusing ancilla qubits in quantum error correcting codes.

To address these limitations there are many extensions that modify RB such as interleaved randomized benchmarking (IRB) [146] and character benchmarking [147, 148] to measure gate-specific error rates, simultaneous randomized benchmarking (SRB) [149] to measure crosstalk, specialized leakage modifications [150, 151] and many more. Similar modifications exist for GST in order to characterize drift [152, 153] or idle crosstalk [153]. Understanding the subtle differences between the assumptions made and implementing these protocols for a physical system requires specialized knowledge. A good open-source library of characterization protocols is pyGSTi [153, 154] which is focused mainly on GST and its modifications and is maintained by Sandia National Laboratories. A commercial alternative is provided by Quantum Benchmark, a Canadian startup company.

3.2.2 The experimentalist's approach

Although the general approach to QCVV of black-box methods is one of its biggest advantages it can also be a limitation. Sometimes there are implicit assumptions that make these protocols unsuitable for their intended purpose and other times one is interested in a system-specific physical quantity.

An example of hidden assumptions is the leakage protocol of Wood and Gambetta [151] which claims to work without the ability to directly measure leakage. The population out of the computational subspace can then be estimated using the property that probabilities must sum to 1. However, in most systems this property is already used to construct the measurement in the first place³. Although the required modification is relatively straightforward (see A.7.1) it is system specific and requires an understanding of how the protocol can be changed without affecting the outcome.

In order to give a meaningful interpretation to characterization results it is often desirable to express these results in terms of a parameter that has physical significance and units such as T_1 (s) or an effective coupling strength J (Hz). Although these parameters can

²Strictly speaking leakage manifests itself as a non-Markovian error when modeling the system as consisting of two-level systems. However, this error is transformed into a Markovian error by expanding the model to include leaked states.

³In a typical experiment the declared state corresponds to $|0\rangle$ if the measured signal is below some threshold $|0\rangle$, with probability p , and otherwise it corresponds to $|1\rangle$, with probability $1 - p$.

in principle be extracted from GST results, the experiments tend to be very costly in terms of acquisition time and analysis to get very accurate results. This is particularly problematic when these parameters are required to calibrate the system. By making system-specific assumptions it is possible to devise simple experiments tailored to characterize specific effects.

The *experimentalist's approach*⁴ consists of three parts. First, a model is created that describes how the system, including the effect of interest, behaves. This model often contains assumptions such as some other effect, like crosstalk, being negligible. These assumptions are supported by the literature or verified in other independent experiments. Next, some experiment is designed that is as simple as possible and sensitive to the effect of interest. Finally, the experiment is compared to the model, if the data cannot be explained by the model the model is rejected. It is possible for the model to be rejected for two reasons, either the model cannot qualitatively describe the observed behavior or the parameters that it estimates are considered unphysical. If the model is rejected, one has to go back to the first step and create a new model.

A key aspect in the experimentalist's approach is designing experiments. Over the years a lot of standard techniques have been devised to independently measure specific effects and correct for known biases. It is beyond the scope of this work to give an exhaustive overview of these techniques, many of which date back to the days of NMR [21], but it is useful to give an example and highlight some of them as these are very similar as those used in GST and RB.

The simplest example of such a protocol is the characterization of the idle operation I . Errors during idle are modeled as the result of relaxation and dephasing. The relaxation, expressed in terms of T_1 , can be measured by first preparing the qubit in $|1\rangle$, performing the idle operation and then performing a measurement. By repeating the idle operation multiple times, the effect can be amplified. It is then possible to extract T_1 by fitting the probability of measuring $|1\rangle$ to a decaying exponential of the form $Ae^{-t/T_1} + B$. At the same time the protocol is robust to state preparation and measurement errors (SPAM) as these only affect the coefficients A and B and not the decay constant. This protocol is insensitive to the dephasing time because of the design of the experiment. The techniques employed here, amplification by repetition (of the idle operation) and isolation of the effect by design are also used in black box approaches such as RB, where gate errors can be determined in a SPAM resilient way by fitting a decaying exponential, and GST, where germs are repeated to amplify specific errors.

It is of course also possible that the model (in this example exponential T_1 relaxation) is wrong. In flux qubits, for instance, non-exponential relaxation has been observed [155]. Because of the simplicity of these experiments it is straightforward to perform sanity checks and independent experiments to verify assumptions and explore other models. Note that for this particular example, the non-exponential decay would only show up as a model violation in a GST analysis due to its non-Markovian nature.

⁴We emphasize that the experimentalist's approach is not something new. Rather, it is a label we apply to contrast what is common practice in experimental physics to the more theory driven black-box approach.

3.3 Standard operations

The quality of gates depends on the coherence of the qubits and the accuracy and precision of the pulses realizing these gates. The effect of a gate can be expressed as a CPTP map Λ that acts on a density matrix ρ . If the system is in a mixed state $\{p_i, |\psi_i\rangle\}$ the state of the system can be described by the density matrix $\rho = \sum_i p_i |\psi_i\rangle \langle \psi_i|$, where p_i denotes the probability that the system is in state $|\psi_i\rangle$. A particularly convenient representation of Λ is as a Pauli transfer matrix (PTM)

$$(R_\Lambda)_{ij} = \frac{1}{d} \text{Tr} \{ P_i \Lambda(P_j) \}, \quad (3.1)$$

where $\{P_i, P_j\} \in \mathcal{P}^{\otimes n}$ are elements of the n -qubit Pauli set with $\mathcal{P} = \{I, X, Y, Z\}$. Using the superoperator formalism [156] a density matrix ρ is expressed as a vector in the Pauli basis $|\rho\rangle\rangle$ with components $\frac{1}{\sqrt{d}} \text{Tr}\{P_k \rho\}$. Applying the map amounts to a simple multiplication

$$|\Lambda(\rho)\rangle\rangle = R_\Lambda |\rho\rangle\rangle, \quad (3.2)$$

$$R_{\Lambda_2 \cdot \Lambda_1} = R_{\Lambda_2} R_{\Lambda_1}. \quad (3.3)$$

Measurements can be described using a POVM⁵ [2]. A POVM is a set of positive operators $\{E_m\}_{m \in I}$ where I is the set of possible outcomes such that $\sum_{m \in I} E_m = \mathbb{I}$. The expected value p of a POVM E_m for a state ρ is $p(m) = \langle\langle E_m | \rho \rangle\rangle$. The inner product is defined as $\langle\langle A | B \rangle\rangle = \text{Tr}\{A^\dagger B\}$.

As an example, let us consider the measurement of a qubit in the computational basis as discussed in Section 1.1.1. This measurement is described by the POVM $\{E_{+1}, E_{-1}\} = \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$. The probability of observing $m = +1$ is given by

$$p(+1) = |\langle 0 | \psi \rangle|^2 = \langle 0 | \psi \rangle \langle \psi | 0 \rangle = \langle 0 | \rho | 0 \rangle, \quad (3.4)$$

which is equivalent to

$$p(+1) = \langle\langle E_{+1} | \rho \rangle\rangle = \langle\langle 0 | \rho \rangle\rangle \quad (3.5)$$

in the superoperator formalism. By virtue of the sum of the POVM, $p(-1) = 1 - \langle\langle 0 | \rho \rangle\rangle$.

3.3.1 Operation errors

Operation errors are often separated in coherent errors and incoherent errors. Coherent errors such as an over rotation or leakage are typically related to control errors and can be addressed by changing the control pulses by better calibrations and optimal control. Incoherent errors are caused by decoherence effects and can typically⁶ not be addressed by changing the control parameters. For transmon-based systems the dominant decoherence effects are relaxation and dephasing. The single qubit PTM for relaxation is given by

⁵Positive Operator-Valued Measure.

⁶Exceptions are when e.g., the dephasing rate depends on a control parameter such as in the CZ gate discussed in Chapter 7.

$$R_{T_1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \sqrt{1-p} & 0 & 0 \\ 0 & 0 & \sqrt{1-p} & 0 \\ p & 0 & 0 & 1-p \end{pmatrix} \quad (3.6)$$

with $p = 1 - e^{-t/T_1}$ the probability of relaxation. The single qubit PTM of pure dephasing is given by

$$R_{T_\phi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-p & 0 & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.7)$$

with $p = 1 - e^{-t/T_\phi}$ the probability of dephasing⁷.

3.3.2 Quantities of interest

PTMs however, contain a lot of information that can make them hard to interpret or quantitatively compare to PTMs of different qubits or systems. There are several metrics that are often used to quantify gate performance. The average process fidelity

$$F_{\text{avg}}(\Lambda_{\text{targ}}, \Lambda) = \frac{\text{Tr} \left(R_{\Lambda_{\text{targ}}}^{-1} R_{\Lambda} \right) + d}{d(d+1)}, \quad (3.8)$$

which expresses the distance of a gate to the target operation. The average process infidelity corresponds to the average error $\epsilon = 1 - F_{\text{avg}}$ and is often used to predict circuit performance. In the presence of coherent errors the worst-case error can be much larger than ϵ . The worst-case error is described by the diamond norm [157],

$$\epsilon_{\diamond}(\Lambda_{\text{targ}}, \Lambda) = \|\Lambda_{\text{targ}} - \Lambda\|_{\diamond}. \quad (3.9)$$

A useful metric to quantify the coherence of errors is the *unitarity* [158–160],

$$U(\Lambda) = \frac{1}{d^2 - 1} \text{Tr} \left(R_{\Lambda'} R_{\Lambda'}^{\dagger} \right), \quad (3.10)$$

where $R_{\Lambda'}$ is the PTM of Λ with the identity component subtracted⁸ and d is the dimension of the Hilbert space. The unitarity provides an upper bound to the gate fidelity:

$$\left(\frac{dF_{\text{avg}} - 1}{d - 1} \right)^2 \leq U. \quad (3.11)$$

This bound is important for two reasons. When the bound is saturated, the gates are coherence limited, indicating that the only way to improve performance is by increasing the coherence. Second, when the bound is saturated $\epsilon_{\diamond} = \epsilon$ circuit performance can be predicted by

⁷Transmons experience both dephasing and relaxation effects simultaneously. The experimentally measured quantity T_2 is related to T_1 and T_ϕ through $\frac{1}{T_2} = \frac{1}{T_\phi} + \frac{1}{2T_1}$.

⁸For a single qubit PTM $R_{\Lambda'}$ corresponds to the lower 3×3 block of R_{Λ} without the first row and column.

Metric	Value	Method	Device	Reference
Single-qubit gate error ϵ_g	$3.7 \cdot 10^{-4}$	Single qubit RB A.1.1	Chimaera S7	–
Two-qubit gate error ϵ_{CZ}	$9 \cdot 10^{-3}$ $1.3 \cdot 10^{-2}$	Interleaved RB A.1.3	Purcell 3Q Chimaera S7	[112] –
Initialization/reset ϵ_{in}	$1.9 \cdot 10^{-3}$	Readout histograms A.4.1	Chimaera S7	–
Readout/detection $\epsilon_{avg.ass.}$	$1.1 \cdot 10^{-2}$	Readout histograms A.3.1	Chimaera S7	–
T_1 (mean)	$44.8 \mu s$	T_1 relaxation experiment	Chimaera S7	–
T_2 (mean)	$87.1 \mu s$	Echo experiment	Chimaera S7	–

Table 3.1.: Performance of standard operations. The characterization protocols used are referred to in the method column. The device on which the performance was measured is listed in the device column and if these results have been published, a citation is given in the reference column.

multiplying the survival probabilities $(1 - \epsilon)$ of the individual gates in the circuit. Alternatively, one can predict the coherence limited performance based on direct measurements of T_1 and T_2 (see Chapter 4).

3.3.3 Characterizing operations

The most straightforward way of determining these quantities is by performing process tomography [2] to determine the PTM. However, it is impossible to accurately reconstruct a PTM using this method because of SPAM errors. GST [139, 143] was developed to provide an accurate and robust characterization of all operations in a gateset including state preparation and measurement. However, GST can be quite demanding especially considering that one is often only interested in a single error parameter (ϵ).

An alternative approach is to use randomized benchmarking [141, 142] to determine ϵ . In Clifford based randomized benchmarking (see [A.1.1](#)), a series of random gates is sampled from the Clifford group. A final gate is added that inverts the action of all preceding gates. By changing the number of Cliffords and averaging over many randomizations, an exponentially decaying curve can be measured. The decay constant of this curve is related to ϵ while the offset and amplitude are related to SPAM errors. If the error rates of different Cliffords are known to differ significantly, such as when characterizing two-qubit gates, one can use interleaved randomized benchmarking [146] (see [A.1.3](#)). Purity benchmarking [158–160] can be used to measure the unitarity of operations. Measurements are typically characterized by preparing $|0\rangle$ and $|1\rangle$ and collecting measurement statistics in the form of histograms (see [A.3.1](#)).

Although the performance shown in Table 3.1 would allow for quantum error correction many relevant effects are not captured by the characterization methods used. Characterization of the two-qubit CZ gate is particularly problematic as performance is typically limited by leakage and repeatability, a non-Markovian error (see also Chapter 7). Single-qubit gates

suffer from this problem to a lesser extent as leakage is generally low due to the use of DRAG [97, 98] pulses and performance is typically coherence limited (see Chapter 4).

Describing measurements statistics in terms of a POVM is a very powerful and general framework. However, it does not give any information on the post-measurement state. As QEC relies on intermediate measurements of ancilla qubits, it is important to characterize not only the measurement statistics but also the correlations to the post-measurement state. An extension to GST which models measurement as a *quantum instrument* [161] is in development [162], which would allow a full characterization of measurements including correlations to the post most-measurement state. Alternatively one can determine the correlations between measurements directly using repeated measurements (A.3.2).

In the next sections, we give an overview of error sources that are not characterized using the standard RB and GST protocols as well of the different kinds of crosstalk observed in our system. For each of these error sources we discuss how to measure these and how these affect the performance of the surface code. Whenever possible, we convert the error rate to a single scalar value ϵ . When an error is coherent, it is converted to an effective error $\epsilon = \theta^2$ per qubit where θ is the angle per clock cycle in radians⁹. For idle operations, the cycle time is taken as 40 ns, the duration of the longest gate. When we refer to cycles or *rounds* of error correction, we follow [121] and take the act of completing a parity check on all ancilla qubits as a single cycle or round. We take the time it takes to complete a single round as the cycle time.

3.4 Other error sources

There are several error sources that violate the typically made assumptions of two-level Markovian systems (Figure 3.2). Because transmons are not two-level systems, leakage to the second-excited state can occur during single-qubit gates, two-qubit gates and measurement [Figure 3.2 (a-c)]. Following [140], we discern two different kinds of non-Markovian errors. A quantum processor is *stable* if the outcome distribution of any quantum circuit can be described by a fixed distribution over the outputs. Fluctuations in error rates are an example of an error violating stability [Figure 3.2(d)]. Parameter drift is a special case of this type of error in which it is possible to relate the observed error directly to a (control) parameter¹⁰. Distinguishing parameter drift from generic fluctuations in error rates is important as more error mitigation techniques exist for the former (see Section 3.4.2). A processor is *Markovian* if each circuit layer, consisting of all operations that are performed at a specific time step, is stable and can be described by a CPTP map. The effect of post-measurement photons and

⁹This conversion is motivated by the property of many circuits to turn coherent errors into depolarizing errors where the average gate fidelity is a reasonable metric to quantify performance. Following [163], we note that for single-qubit gates $\epsilon_{\text{avg.}}(\exp(-i\theta\hat{n} \cdot \vec{\sigma}/2), \mathbb{1}) = 2 \sin^2(\theta/2)/3 \approx \theta^2/6$. The conversion used is slightly more conservative and chosen to be consistent with the reporting requirements in the IARPA LogiQ program.

¹⁰Note that it is not possible to distinguish parameter drift from other types of fluctuating error rates using only the observed measurement outcomes.

history dependence of flux operations [Figure 3.2 (e)] are two examples of errors that violate Markovianity but not stability¹¹.

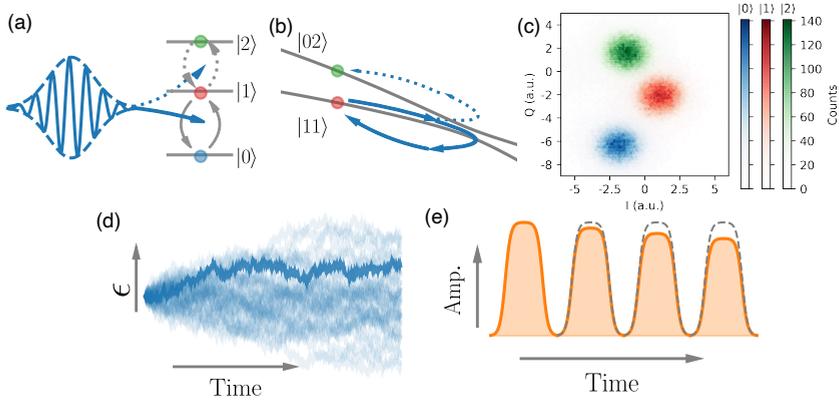


Figure 3.2.: Schematic representation of different error sources violating the two-level (a-c) or Markovian (d,e) assumptions of typical QCVV protocols. (a) Microwave control pulses can drive transitions to the 2nd excited state. (b) A flux pulsing based CZ gate makes use of an avoided crossing with the non-computational $|02\rangle$ state. (c) Performing a measurement can induce transitions to a non-computational state. (d) Low-frequency noise on control parameters or coherence can cause fluctuations in error rates. (e) Operations can be history dependent due to e.g., pulse-distortions acting on timescales longer than the pulse duration.

3.4.1 Leakage

An important error source for transmon qubits is leakage out of the computational subspace spanned by $|0\rangle$ and $|1\rangle$. When leakage occurs, one of the higher energy states is occupied. Typically, this is the second excited state $|2\rangle$. Leakage out of the computational subspace is expressed in terms of the leakage rate L_1 [151]. The leakage rate quantifies the average population lost from a quantum system of interest to states outside the computational subspace. To assess the effect of leakage on an error correcting code, a related quantity, the seepage L_2 [151], is required. The seepage rate quantifies the return of population from outside to inside the computational subspace.

Leakage is a particularly problematic error because once a qubit is leaked, it can remain so for multiple error correcting cycles, leading to multiple errors caused by a single leakage event. An error similar to leakage, qubit loss, signifies losing a qubit entirely. While this can occur in platforms such as ion traps, this does not occur for transmons.

¹¹A system that is Markovian is by definition also stable. However, if a system is non Markovian, this does not imply it is unstable.

Microwave-gate leakage

Leakage can occur during microwave operations if the $|1\rangle \leftrightarrow |2\rangle$ transition is driven [Figure 3.2 (a)] and is characterized using the leakage modification for RB (A.7.1). Leakage in microwave gates can be suppressed by using DRAG pulses [97], as is standard practice. For single-qubit gates, leakage rates are typically $L_1 \sim 10^{-5}$ [79, 92, 164]. Seepage rates for single-qubit gates are dominated by relaxation from $|2\rangle$ to $|1\rangle$ and have typical values $L_2 \sim 2 \times 10^{-3}$. The relaxation time of the second-excited state to the first-excited state ($T_1^{2 \rightarrow 1}$) is typically smaller than T_1 [63, 165].

As a leaked qubit will stay leaked for on average ~ 20 rounds (bounded by $T_1^{2 \rightarrow 1}$), the probability of a qubit being in a leaked state for a given cycle is $\sim 2 \times 10^{-4}$, which is 50 times lower than the dominant error rate in the system (T_1 decay). For this reason, leakage caused by microwave gates is deemed insignificant. Should the need arise, it is possible to reduce the magnitude of the leakage by adding extra control knobs to standard DRAG pulses [100, 166] or by slowing down these pulses.

Flux-gate leakage

More significant leakage can occur during flux pulses when tuning in and out of the $|11\rangle - |02\rangle$ avoided crossing [Figure 3.2 (b)], as is done when performing a controlled phase gate (CZ) [26, 95]. Population exchange can be suppressed using specialized pulse-shaping techniques [68, 112, 167]. Similar to microwave-gate leakage, flux-gate leakage is characterized using the modified RB protocol (A.7.1). However, because leakage can be a dominant error source for flux-pulsing based two-qubit gates, it is important to measure this during calibrations. As a full randomized benchmarking experiment is too slow, the conditional oscillation experiment described in Section 7.5.4 is used to measure leakage during calibrations. Leakage rates for two-qubit gates in our system are $\sim 1 \cdot 10^{-3}$ [61, 112]. Previously reported simulations [132] indicate that the memory break even point can be reached at $\sim 5 \cdot 10^{-3}$ as is shown in Figure 3.3.

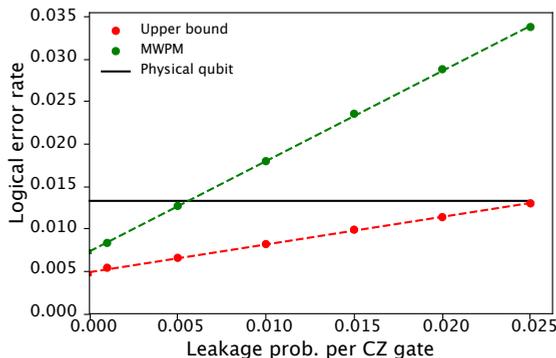


Figure 3.3.: The effect of CZ leakage on the logical error rate. Figure from [132].

Metric	Value	Method	Device	Reference
Single-qubit gate leakage L_1	$1.3 \cdot 10^{-4}$	Single qubit RB A.1.1, A.7.1	Purcell 3Q	[61]
Two-qubit gate leakage L_1	$1.1 \cdot 10^{-3}$	Interleaved RB A.1.3, A.7.1	Purcell 3Q	[61, 112]
Measurement induced leakage L_1	$\sim 1.5\text{--}2.5 \cdot 10^{-3}$	Estimated	Purcell 3Q	[61]

Table 3.2.: Leakage of standard operations. The characterization protocols used are referred to in the method column. The device on which the performance was measured is listed in the device column and if these results have been published, a citation is given in the reference column.

Leakage during CZ gates is limited by dephasing and remaining distortions (see Section 7.3.5). Besides addressing these problems, typical strategies to reduce leakage are slowing down the CZ pulses or increasing the qubit-qubit coupling. Slowing down the pulses directly reduces leakage by improving the adiabaticity of the pulses but comes at the price of an increased error rate due to decoherence. Alternatively, the coupling strength can be increased. This has the same effect as slowing down the pulses but instead increases the residual coupling, leading to an increase in the crosstalk error rate (see Section 3.5.1).

Measurement-induced leakage

Measurement-induced state leakage has been observed and reported in some experiments by the UCSB/Google group [42, 64]. It is mainly linked to strong measurement powers above the critical photon number, where the dispersive model breaks down [101]. We have studied measurement-induced leakage by rapidly interleaving measurement-and-depletion operations with coherent operations, thereby emulating an ancilla engaged in multi-round parity checks (Chapter 5), as well as performing error correction experiments [61]. In these experiments, we quantify performance by extracting the mean number of rounds to a measurement outcome that deviates from the ideal result, termed rounds-to-event (RTE). Measurement-induced leakage would take the qubit out of the computational subspace prohibiting coherent operations and thereby reducing the obtained RTE. The modeling of the experiments does not include measurement induced leakage. The good agreement observed between the model and experiment demonstrates the non-demolition character of the measurement and a negligible leakage rate. These experiments have since been performed on multiple devices and have shown similar results. Alternatively, one can determine the measurement induced leakage directly using the method described in [64].

Mitigating the impact of leakage

At currently observed leakage rates, leakage will remain one of the dominant contributions to the Surface-17 logical error rate ϵ_L . We have two main strategies to mitigate the effect of leakage on ϵ_L . The first is to optimize the QEC decoder to account for leakage, thus minimiz-

ing its effect on the decoder efficiency. The second is to implement de-leaking strategies on data and ancilla qubits.

The indirect detection of leakage events in a QEC cycle was recently demonstrated [61]. In Varbanov *et al.* [62] the effect of leakage, including the distinguish-ability of the 2nd excited state, is studied in detail.

De-leaking strategies involve forcing leaked qubits to re-enter the computational subspace, cutting the leakage event short. These strategies may either be blind (i.e., indiscriminantly performed to all qubits) or targeted to qubits identified as leaked via the techniques discussed in [61, 62]. Any de-leaking strategy needs to ensure that it does not create more errors than it corrects. In general, if a leaked qubit is expected to stay leaked for another $\tau_{\text{cycle}}/T_1^{2 \rightarrow 1} (\approx 20)$ rounds, a de-leakage strategy with a p_d probability of de-leaking a leaked qubit, a p_e probability of creating an error within the computational subspace, and a p_l probability of leaking an un-leaked qubit will only reduce the overall error rate when

$$p_d \tau_{\text{cycle}}/T_1^{2 \rightarrow 1} \gg p_l \tau_{\text{cycle}}/T_1^{2 \rightarrow 1} + p_e. \quad (3.12)$$

This is a particular issue for blind de-leaking strategies, as p_d per qubit scales with the probability of a qubit being leaked, while p_e and p_l will not (indeed, p_l increases with the probability of a qubit being un-leaked). Targeted strategies avoid this as the probability of a targeted qubit being leaked can be kept rather high. However, blind strategies have the advantage that they maximize p_d over the entire surface. As our qubits are decoherence limited, any strategy that necessitates additional time between QEC cycles quickly increases p_e , so maximizing the usage of this time is ideal. Blind strategies have an additional advantage of simplicity; the strategy requires minimal setup and configuration, which may be a drain on computational resources.

De-leaking of ancilla qubits may be performed via active reset techniques [63, 168]. These strategies may be performed blindly on ancilla qubits, as they have negligible risk of leaking un-leaked ancilla qubits, and the error in our knowledge of the resulting ancilla state is no worse than that post-measurement. The largest cost of the active reset is the additional time between successive QEC cycles, during which data qubits are prone to T_1 and T_2 noise (which contributes to p_e in Eq. 3.12). Optimizing the frequency of resets per QEC cycle is an interesting optimization direction for future research. Blindly resetting data qubits is not a viable de-leaking strategy, as the reset destroys the quantum information stored on the logical qubit. Targeted resetting of leaked data qubits may provide a decrease in leakage if they can be identified sufficiently well. This could then be performed during the reset of ancilla qubits, and would come at zero additional time cost.

A more appealing alternative to resetting data qubits is a de-leakage scheme that does not affect the state of the qubit within the computational subspace. This would allow for blind data qubit de-leakage schemes. Alternatively, this could be combined with a targeted scheme to further drive down the chance of introducing any additional errors (leakage or none). The most well-known such scheme is that of [57], where data and ancilla qubits are swapped by a CZ gate such that the ancilla qubit is never excited into the $|2\rangle$ state. More recent versions of this scheme [59] are not feasible on the Surface-17 layout, as they require additional qubits.

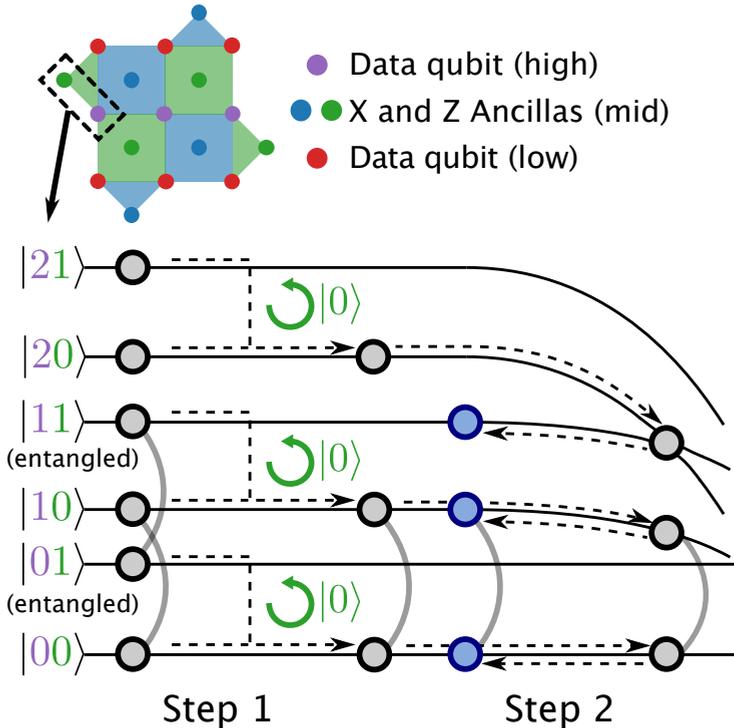


Figure 3.4.: De-leaking of data qubits by driving excitations onto nearby ancilla qubits without affecting data within the computational subspace. (Top) Surface-17 layout showing an example ancilla-data qubit pair (any ancilla neighboring a high-frequency data qubit is sufficient). (Bottom) Diagram showing the de-leaking scheme, lines correspond to different energy levels labeled by the number of excitations in the data (purple) and ancilla (green) qubit. Circles and dashed arrows denote the effect of the different steps on a system starting out in a particular state. (Bottom-left) In the first step of the process, the ancilla qubit is reset to $|0\rangle$ via active reset [63, 168]. Note that a leaked ancilla will also be returned to $|0\rangle$ by these schemes (not shown). This process does not affect any entanglement (grey line) between the $|0\rangle$ and $|1\rangle$ states of the data qubit. (Bottom-right) Once the ancilla qubit is in the $|0\rangle$ state, the data qubit is driven to the $|11\rangle - |02\rangle$ crossover point and attempts to swap an excitation onto the ancilla qubit. If the data qubit was leaked, this returns the system to within the computational subspace, and cycles of QEC may immediately continue. If the data qubit was not leaked, it is unaffected by this process up to a single-qubit phase that can be corrected.

An alternative de-leakage scheme is one where an excitation from a high-frequency data qubit in the $|2\rangle$ state is swapped onto a neighboring ancilla qubit, without affecting the state of the qubit within the computational basis. Such a process is described in Figure 3.4. First, the ancilla qubit is reset to the $|0\rangle$ state. Then, the neighboring high-frequency data qubit¹² is

¹²This strategy would need alteration for low-frequency data qubits as the level structure is different, but we do not expect them to be affected significantly by leakage as they do not suffer from the dominant flux-gate leakage.

fluxed to the $|02\rangle - |11\rangle$ transition point for a time t_{swap} chosen to swap the two states. If the data qubit was leaked, this will return the system to $|11\rangle$, which is part of the computational subspace $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. If the data qubit was not leaked, the $|11\rangle$ state is unoccupied prior to the gate so this should not affect unleaked data qubits (save for a correctable single-qubit phase). In this case, the ancilla qubit remains in $|0\rangle$, and may be used for the next QEC cycle without additional measurement or reset. If the ancilla qubit is in $|1\rangle$ post de-leaking, the data qubit will be out of the code space, and the next stabilizer measurement will be random, so careful measurement and reset is unnecessary anyway. This implies that the time for this procedure is dominated by the initial ancilla reset. Care must be taken to minimize any error in resetting the ancilla qubit, as this scheme will drive residual excitations onto the data qubit. Future simulations are required to determine the optimal rate at which these de-leakage schemes would operate, what (if any) mitigation power they will have, and whether a significant gain is obtained from targeted de-leakage of data qubits versus blind de-leakage.

3.4.2 Non-Markovian errors

System stability

Although quantum processors are typically assumed to be stable this is always an approximation, as a consequence the error rates of a quantum processor fluctuate over time [Figure 3.2(d)]. The simplest characterization protocol for system stability is to repeatedly measure a quantity of interest and perform basic frequency domain analysis as required (see e.g., Section 4.5.4). A more systematic analysis of parameter drift is proposed in [152]. Some of these fluctuations can be traced back to fluctuations in physical parameters. In this case the system can be stabilized by reducing the sensitivity to the parameter in question or by stabilizing the physical parameter. One can also recalibrate the system to adapt to changes in parameters. For this approach to be successful, the time it takes to recalibrate the system must be short compared to the scale of the drift.

An example of a fluctuation related to a physical parameter were the single-qubit phase errors when creating a Bell state using a CZ gate. When first investigating CZ gates we faced challenges in the stability of our gates. Because the duration of the gate (240 ns) and the required detuning (1.15 GHz) were quite large, the phase acquired by the fluxed qubit (QR) was very sensitive to any change in the amplitude of the control pulse. The stability of a Bell state created using this CZ gate was measured during a period of 13 hours both with and without intermediate recalibration of the pulse. Because the changes in acquired phase correlated with the day night cycle the suspicion arose that that this effect was related to the temperature in the lab. To test this hypothesis, the same experiment was repeated with a heat gun pointing at the AWG. Both the acquired phase and the temperature go up as the heat gun is activated after ~ 30 mins (Figure 3.6) and go back down after it is turned off. By correlating the change in temperature to the acquired phase [Figure 3.6(c)] it was possible to identify temperature fluctuations as the cause of this error.

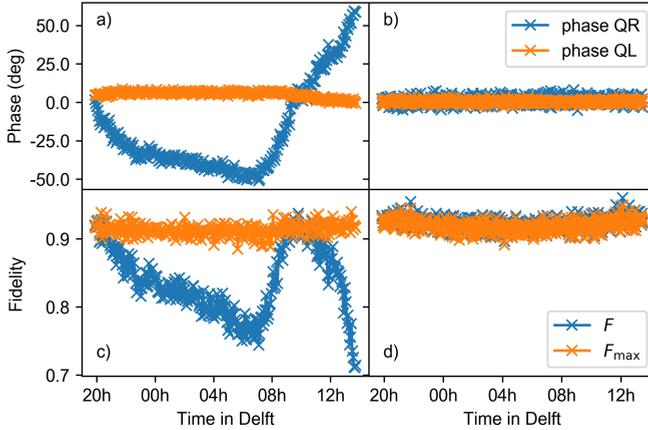


Figure 3.5.: Stability of Bell state generation measured during a period of 13 hours. Single-qubit phase error on the flux-pulsed qubit (QR) and static qubit (QL) with respect to the target Bell state (a,b). Fidelity to target Bell state (F) and Fidelity corrected for single-qubit phase errors (F_{\max}). Panels (a,c) are without recalibration and (b,d) are with recalibration.

After identifying the cause of these fluctuations, several steps were taken to stabilize the system. The sensitivity to the temperature was reduced by modifying the amplifiers of the AWG to be first order insensitive to changes in temperature around the operating temperature. Additionally, the qubit-qubit coupling was increased to enable shorter gates which are less sensitive to changes in amplitude. After these improvements system stability was no longer a critical issue. If required it is possible to further improve the system stability by e.g., thermally stabilizing the room temperature control electronics.

Often, it is more practical to simply recalibrate parameters that are known to drift over time. This can be done by either interleaving calibration stages [99, 169] between experiments or by including an in-situ calibration that runs during the error correcting cycle [170]. The origins of other fluctuations in error rates, such as fluctuations in coherence [94, 99, 145, 171–173], are harder to control. As a consequence, it is not straightforward to reduce these fluctuations. The effect of the remaining fluctuations on QEC can be partially mitigated by using a decoder that adaptively changes its error model on experimentally observed error rates [174].

History dependent operations

Ideally, every application of an operation behaves identical independent of the history of the system. Non-Markovian errors are particularly problematic as standard characterization techniques rely on amplifying errors by repeating operations and averaging in order to gather statistics. In this section, we focus on a particular type of non-Markovian errors: (usually coherent) gate errors that depend deterministically on operations that preceded the gate in question. Examples of such errors include timing calibration errors, where pulses overlap

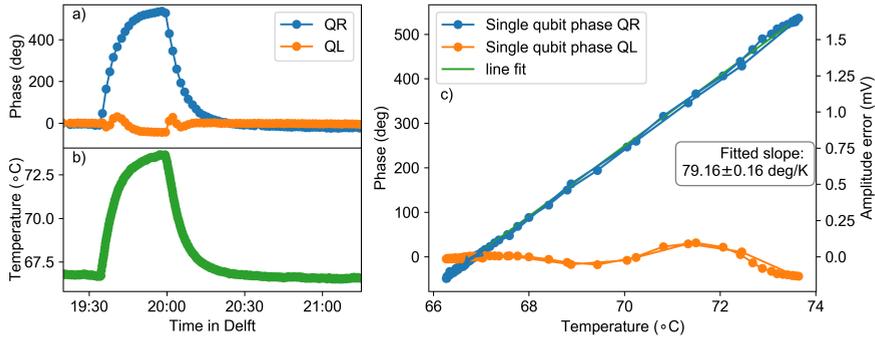


Figure 3.6.: Temperature dependence of single-qubit phases during Bell state generation. Acquired phases of the fluxed qubit (QR) and static qubit (QL) as a function of time. Temperature as a function of time (b) and phase correlated against temperature (c).

with preceding pulses, gate errors due to residual measurement photons (Chapter 5) and the history dependence of flux pulses (Chapter 7).

Overlapping operations are the simplest kind of history dependence and can be removed by ensuring that operations start after preceding operations are finished. In the case of microwave operations calibrating timings is standard procedure and these errors should not occur. Should a small overlap remain between operations, then this will directly affect the error rate as measured using randomized benchmarking. Gate set tomography will detect this type of error as a significant model violation. There are currently no signs that this error significantly affects error rates.

Residual photons are slightly more complex to understand but are effectively the same problem; the next operation can only start when the preceding (measurement) operation is finished. In Chapter 5 we investigated in detail when a measurement operation ends, which is not at the end of the stimulating pulse, and how to reduce this time using active photon depletion. This trades the non-Markovian overlap error for increased decoherence, which is Markovian. In [79], we optimize the competition between these two effects, and find that the optimal trade-off incurs mostly decoherence errors, which we do not discuss here.

Flux pulses are more problematic as the electrical pulses used to generate flux pulses typically suffer from distortions that can have timescales significantly longer than the coherence time of the transmon. As such, defining the end of the operation after any residual effects are gone, as is done in the case of microwave and measurement operations, is not a feasible approach. Instead, these residual effects have to be corrected for by modifying individual pulse shapes in order to realize identical operations [78, 91]. Because of the relatively long timescales of these distortions, errors induced by consecutive pulses can accumulate, making it both more important to correct for these effects as well as harder to characterize.

Metric	Value	Method	Device	Reference
idle crosstalk ϵ_{ZZ}	$5 \cdot 10^{-5}$	Residual ZZ A.5.1	Purcell 3Q	[61]
Microwave crosstalk ϵ_{mwX}	$1 \cdot 10^{-3}$	SRB A.6.2	Purcell 3Q	[61]
AC-flux crosstalk	–	–	–	–
Measurement ind. dephasing ϵ_{ϕ_M}	$< 4 \cdot 10^{-3}$	A.6.3	Purcell 3Q	[61]
Measurement det. crosstalk ϵ_{mdX}	$1.3 \cdot 10^{-3}$	A.6.4	Purcell 3Q	[61]

Table 3.3.: Crosstalk characterization in the 3 qubit Purcell device [61]. The characterization protocols used are referred to in the method column. The device on which the performance was measured is listed in the device column and if these results have been published, a citation is given in the reference column.

A custom method is designed to measure the history dependence of flux pulses and discussed in Section 7.3.1. The single-qubit phase θ can be translated into an effective error on the subsequent gate using:

$$\epsilon_{\text{hist}} = \theta^2. \quad (3.13)$$

For a target CZ gate infidelity of $\epsilon_{CZ} = 0.01$, a phase error of ~ 5 deg can be tolerated. By correcting for distortions using the Zurich Instruments HDAWG real-time predistortion filters in combination with a net-zero pulse shape, the measured phase error can be reduced to $\theta \lesssim 1$ deg, independent of T_{sep} . The sensitivity to distortions can be further reduced by using a double sweetspot transmon made with asymmetric junctions [91]. By designing the transmon such that the lower sweetspot is slightly below the interaction zone of the CZ gate, the phase acquired is less sensitive to errors and noise in the applied flux.

3.5 Crosstalk

Crosstalk is an imprecise but widely used term from classical electronics that refers to any signal or circuit unintentionally affecting another signal or circuit. In experimental quantum computing the term is used to describe a range of physical phenomena. Following Sarovar *et al.* [140] we define crosstalk errors as errors due to undesired dynamics that violate either *locality* or *independence*. Locality effectively means that no correlations between qubits are created unless they are explicitly involved in an entangling gate while independence corresponds to the notion that qubits are only affected by operations acting on those specific qubits. For the exact definitions of locality and independence we refer to [140]. Here we discuss two types of crosstalk, idle crosstalk resulting from always on interactions and operation crosstalk. An overview of measured error rates due to different types of crosstalk is given in Table 3.3.

3.5.1 Idle crosstalk

There are two forms of idle crosstalk in the QuSurf architecture. These are residual ZZ interactions between nearest-neighbor qubits and residual exchange interactions ($XX + YY$)

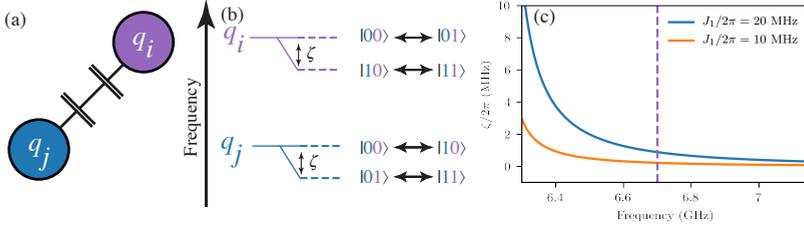


Figure 3.7.: (a) Two neighboring qubits q_i and q_j are coupled through a coupling resonator. (b) At their idle frequencies, both qubit frequencies are slightly dependent on the state of the other qubit with strength ζ . (c) Residual coupling ζ for two qubits q_i and q_j as a function of the frequency of q_i for two different coupling strengths J_1 . Dashed line indicates the idle frequency (6.7 GHz) of q_i , q_j is kept at 6.0 GHz.

between next-nearest-neighbor same-frequency qubits. Both of these interactions violate locality [140]. In addition to these forms of idle crosstalk it is possible for other crosstalk terms to be present due to spurious couplings.

Residual ZZ interaction

The coupling between nearest-neighbor transmons q_i and q_j by a bus resonator is used to implement CZ gates by flux detuning the higher-frequency qubit so that the states $|11\rangle$ and $|20\rangle$ are near resonance. At the idle frequencies, the qubits are ideally far detuned from this interaction zone, but there remains a residual idle Hamiltonian of the form $H = \zeta_{ij} |11\rangle \langle 11|$. Since the interaction is always on, the single-qubit terms can be taken as a renormalization of the qubit frequency [Fig. 3.7(a,b)]. Using perturbation theory the frequency shift $\zeta_{ij} = \hbar\omega_{11} - \hbar\omega_{01} - \hbar\omega_{10}$ can be approximated as

$$\zeta_{ij} = -J_2^2 \left(\frac{1}{\hbar\omega_{20} - \hbar\omega_{11}} + \frac{1}{\hbar\omega_{02} - \hbar\omega_{11}} \right), \quad (3.14)$$

where the subscripts in ω_{kl} are used to label the states based on the number of excitations in qubits i and j and $J_2 \approx \sqrt{2}J_1$ is the effective coupling between qubits i and j .

During an idle time τ_I , this results in a coherent correlated ZZ error with angle $\theta = \frac{\zeta_{ij}\tau_I}{4}$. For coherent errors, the error accrued over a time equal to the two-qubit gate duration ($\tau_I = \tau_{CZ} = 40$ ns) is reported as

$$\epsilon_{ZZ} = \theta^2 = \left(\frac{\zeta_{ij}\tau_I}{4} \right)^2. \quad (3.15)$$

From Equation (3.14) we can determine the residual ZZ coupling for different idle frequencies and coupling strengths [Figure 3.7(c)]. At $\zeta/2\pi = 890$ kHz and $\zeta/2\pi = 330$ kHz for the residual coupling between mid-high and low-high frequency qubits this error source is not negligible. Importantly, this error source depends strongly on the idle frequency of the qubits, even a small error in qubit frequency targeting can result in a relatively large residual

	$J_1/2\pi$	$\omega_i/2\pi$	$\omega_j/2\pi$	$\zeta/2\pi$	ϵ_{ZZ}
Design	20 MHz	6.0 GHz	6.7 GHz	890 kHz	$3.1 \cdot 10^{-3}$
Design	20 MHz	6.0 GHz	4.9 GHz	330 kHz	$4.4 \cdot 10^{-4}$
Mitigation ω_j	20 MHz	6.0 GHz	7.0 GHz	410 kHz	$6.5 \cdot 10^{-4}$
Mitigation ω_j	20 MHz	6.0 GHz	4.7 GHz	230 kHz	$2.2 \cdot 10^{-4}$
Mitigation J_1	10 MHz	6.0 GHz	6.7 GHz	220 kHz	$2.0 \cdot 10^{-4}$
Mitigation J_1	10 MHz	6.0 GHz	4.9 GHz	80 kHz	$2.7 \cdot 10^{-5}$
Mitigation J_1, ω_j	10 MHz	6.0 GHz	7.0 GHz	100 kHz	$4.1 \cdot 10^{-5}$
Mitigation J_1, ω_j	10 MHz	6.0 GHz	4.7 GHz	59 kHz	$1.4 \cdot 10^{-5}$

Table 3.4.: Residual ZZ coupling of nearest neighbor qubits for designed couplings and idle frequencies as well as for three mitigation strategies.

coupling. Besides using echo pulses to mitigate the impact of this residual coupling, there are two mitigation strategies that reduce the magnitude of this effect.

The first mitigation strategy is to increase the separation of idle frequencies. By changing the idle frequencies to $\omega_L/2\pi = 4.7$ GHz, $\omega_M/2\pi = 6.0$ GHz, $\omega_H/2\pi = 7.0$ GHz the couplings reduce to 410 kHz and 230 kHz respectively (Table 3.4). The downside of this approach is a larger detuning from the flux-insensitive point (the sweetspot) during two-qubit gates, which leads to increased dephasing from flux noise and an increased sensitivity to pulse distortions.

An alternate approach is to reduce the coupling strength. By reducing the coupling strength from $J_1/2\pi = 20$ MHz to $J_1/2\pi = 10$ MHz the residual couplings are reduced to 220 kHz and 80 kHz. The downside of this approach is that the speed limit for two-qubit gates is increased from 18 ns to 35 ns. Because of the $\sim 25\%$ overhead required for low-leakage two-qubit gates and the single-qubit phase correction pulse, this extends the two-qubit gate time from 40 ns to 60 ns. For each two-qubit gate, the fluxed qubit therefore spends more time away from the flux-insensitive point. The total error-correction cycle time may not significantly increase as it is dominated by parallel running measurements in pipelined fashion [121].

The residual coupling can be measured from the frequency difference of Ramsey oscillations on $q_i q_j$ in state $|0\rangle$ and $|1\rangle$ (A.5.1). In the Chimaera S7 device we measure couplings up to $\zeta = 2.7$ MHz (Figure 3.8). The high couplings can be explained by frequency targeting issues where specifically the high frequency qubits are significantly lower than designed. The agreement between calculated and measured couplings is decent. Here we have only considered residual ZZ coupling of nearest neighbor qubits based on the designed frequencies and couplings. In reality there are spurious couplings between non-nearest neighbor qubits as well as higher order terms that are not taken into account. Providing an accurate prediction of residual coupling to aid the design and fabrication of devices is ongoing research that is currently pursued by our collaborators.

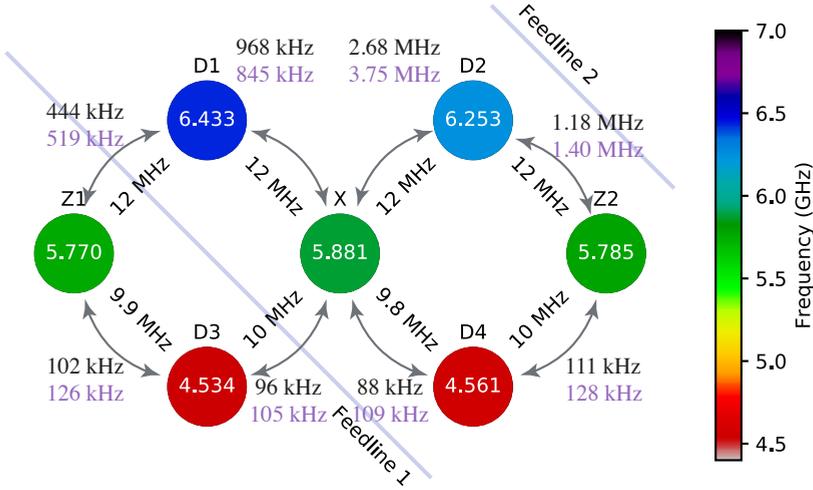


Figure 3.8.: Measured frequencies and couplings of the Chimaera S7 device. Qubit frequencies $\omega/2\pi$ are given in GHz, measured couplings $J_1/2\pi$ are given in MHz. Curved arrows denote residual ZZ couplings $\zeta/2\pi$, purple values are calculated using Equation (3.14), black values are measured (see A.5.1).

Residual exchange interaction

Several next-nearest neighboring qubits in the QuSurf architecture are biased to the same frequency [Figure 3.9(a)]. The residual coupling between these pairs leads to an exchange interaction

$$H = J_R |10\rangle \langle 01| + h.c. \quad (3.16)$$

This exchange can be evidenced by a sinusoidally decaying population in T_1 experiments [92] or from an avoided crossing when tuning one of two qubits [Figure 3.9(b)]. The result is a coherent bit-flip error for both qubits with angle per idle gate $\theta = J_R \tau_I$ and corresponding error

$$\epsilon_{XX} = \theta^2 = (J_R \tau_I)^2. \quad (3.17)$$

The avoided crossings for all four types of next-nearest neighbor interactions [Figure 3.9(b-c)] are obtained by numerically diagonalizing the Hamiltonian. For the interactions HMH and MHM (MLM and LML) we obtain $J_1/2\pi = 571$ kHz (363 kHz). These correspond to $\epsilon_{XX} = 5.2 \cdot 10^{-3}$ ($2.1 \cdot 10^{-3}$).

The same mitigation strategies as for the ZZ errors apply. The results of changing the separation of the idling frequencies, reducing the coupling strength J_1 between nearest neighbor qubits and combining these strategies is summarized in Table 3.5. The same downsides apply as in Sec. 3.5.1.

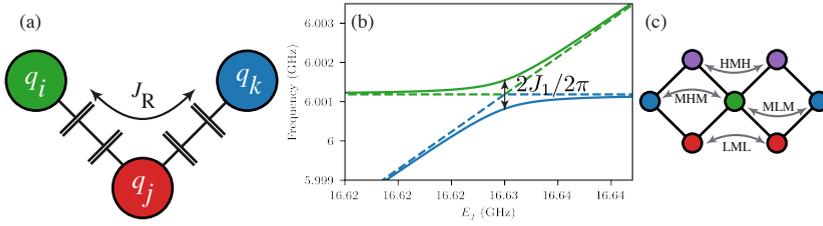


Figure 3.9.: (a) Parasitic exchange interaction between two same-frequency next-nearest neighbors q_i and q_k is mediated by couplings to the common neighbor q_j via bus resonators. (b) Simulated avoided crossing for q_i and q_k mediated by q_j using design parameters for the QuSurf architecture (see Table 3.5). J_R is extracted from the minimum frequency splitting, which corresponds to $2J_1/2\pi$. (c) A Surface-7 patch used to visualize the different contributions to J_R . Labels correspond to the interactions of Table 3.5.

	$J_1/2\pi$	$\omega_L/2\pi$	$\omega_H/2\pi$	$J_R/2\pi$	ϵ_{XX}
Design HMH & MHM	20 MHz	4.9 GHz	6.7 GHz	571 kHz	$5.2 \cdot 10^{-3}$
Design MLM & LML	20 MHz	4.9 GHz	6.7 GHz	363 kHz	$2.1 \cdot 10^{-3}$
Mitigation J_1 HMH & MHM	10 MHz	4.9 GHz	6.7 GHz	143 kHz	$3.2 \cdot 10^{-4}$
Mitigation J_1 MLM & LML	10 MHz	4.9 GHz	6.7 GHz	91 kHz	$1.3 \cdot 10^{-4}$
Mitigation ω HMH & MHM	20 MHz	4.7 GHz	7.0 GHz	445 kHz	$3.2 \cdot 10^{-3}$
Mitigation ω MLM & LML	20 MHz	4.7 GHz	7.0 GHz	307 kHz	$1.5 \cdot 10^{-3}$
Mitigation J_1, ω HMH & MHM	10 MHz	4.7 GHz	7.0 GHz	111 kHz	$2.0 \cdot 10^{-4}$
Mitigation J_1, ω MLM & LML	10 MHz	4.7 GHz	7.0 GHz	77 kHz	$9.3 \cdot 10^{-5}$

Table 3.5.: Residual XX coupling between next-nearest neighbor qubits mediated by a common neighbor qubit for designed parameters as well as for three mitigation strategies. Three letter abbreviations indicate the particular couplings as indicated in Figure 3.9(c). The middle frequency ω_H is kept fixed at 6.0 GHz.

	mw- D_1 (dB)	mw- D_2 (dB)	mw- D_3 (dB)	mw- D_4 (dB)	mw- Z_1 (dB)	mw- Z_2 (dB)	mw- X (dB)
D_1	0.0	> 55.0	> 55.0	> 55.0	> 55.0	> 55.0	19.7
D_2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
D_3	> 52.4	> 52.4	0.0	> 52.4	22.8	> 52.4	25.0
D_4	> 52.2	> 52.2	> 52.2	0.0	> 52.2	20.6	22.0
Z_1	18.7	> 52.1	38.3	> 52.1	0.0	> 52.1	> 52.1
Z_2	> 52.3	25.6	> 52.3	9.4	> 52.3	0.0	> 52.3
X	31.2	25.0	32.2	17.0	35.4	31.0	0.0

Table 3.6.: Microwave tone spill over of a Surface-7 device (not Chimaera) using vertical IO, suppression in dB. Qubit D_2 was not functional.

An alternative mitigation strategy would be to slightly detune (~ 5 MHz) from the resonance condition. Although this breaks the frequency reuse scheme of the QuSurf architecture it virtually eliminates this type of error.

3.5.2 Microwave-drive crosstalk: pulse spillover

In the QuSurf architecture, each qubit has a dedicated microwave drive line. Due to weak coupling to spurious (packaging) modes of these drive lines and of non-targeted qubits, there can be microwave pulse spill-over (Figure 3.10). This spillover leads to a different error depending on the frequency of the drive tone and the subjected qubit: X and Y errors for resonantly driven qubits and Z errors due to a Stark shift for non-same-frequency qubits. Both forms of microwave-drive crosstalk affect the error rate as measured using SRB (A.6.2).

To quantify both forms of crosstalk, we examine the spill-over of each line to each qubit from Rabi experiments. As a metric we measure the microwave isolation for each pair of qubit q_i and microwave drive line D_j , defined as the pulse amplitude A_{ij}^{180} required to perform an $X180$ operation on q_i through D_j , normalized by the required amplitude when using the correct drive line D_i , expressed in power reduction:

$$M_{ij} = 20 \log_{10} \left(\frac{A_{ij}^{180}}{A_{ii}^{180}} \right) [\text{dB}]. \quad (3.18)$$

Resonant microwave cross-driving

Resonant microwave cross-driving describes the effect of spill-over of a microwave signal to a same-frequency qubit. The result is a coherent rotation around an unknown axis in the $X - Y$ plane, with angle

$$\theta = \pi \frac{A_{ij}^{180}}{A_{ii}^{180}}. \quad (3.19)$$

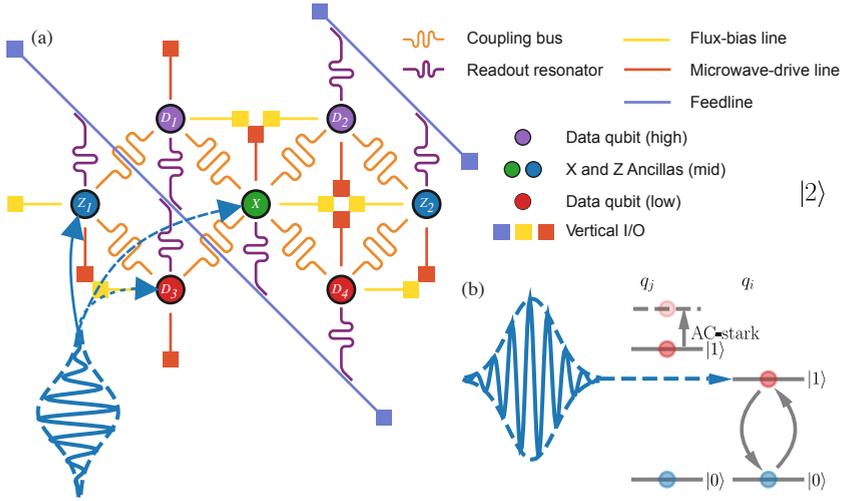


Figure 3.10.: Microwave cross-driving. (a) When applying a microwave pulse to qubit Z1 (solid line) via its dedicated drive line, part of the microwave signal can spill over to other qubits. We distinguish cross-driving of same-frequency qubits (e.g. X, long-dashed line), leading to a rotation around an axis in the $X - Y$ plane and cross-driving of non-same-frequency qubits (e.g. short-dashed line) leading to a Z-type error due to a temporary Stark shift of the qubit frequency. (b) Schematic representation of an AC Stark shift of q_j induced by off-resonant microwave driving of a tone intended for q_i . When a transition couples to an off-resonant microwave signal, the transition is effectively shifted for the duration of this off-resonant drive.

The error rate is then given by

$$\epsilon_{CD_{ij}} = \theta^2 = \left(\pi \frac{A_{ij}^{180}}{A_{ii}^{180}} \right)^2. \quad (3.20)$$

For a Surface-7 device using vertical IO, 35 out of 42 matrix elements were measured (D2 qubit was not measured) and are shown in Table 3.6. Of these, 10 elements concern same-frequency cross-driving. Eight elements were not measurable and given a lower bound of > 52 dB, corresponding to $\epsilon_{CD_{ij}} < 5 \cdot 10^{-5}$. The two measurable elements have $M_{X,Z1} = 35.4$ dB and $M_{X,Z2} = 31.0$ dB correspond to $\epsilon_{CD_{X,Z1}} = 2.8 \cdot 10^{-3}$ and $\epsilon_{CD_{X,Z2}} = 7.8 \cdot 10^{-3}$.

Off-resonance microwave cross-driving

When a transition couples to an electric field such as that of an off-resonant microwave tone, it experiences a change in frequency known as an AC Stark shift [102, 175]:

$$2\pi f_{AC} = 2ng^2/\Delta, \quad (3.21)$$

where f_{AC} is the frequency shift, n the average photon number, g the coupling to the photon field and Δ the frequency difference between the bare frequency and the external field.

We expect this effect to be rather small on account of the detuning between different qubits and the use of dedicated drive lines. Although it is straightforward to measure this effect, this has not been done.

This effect can be measured by performing a Ramsey experiment on qubit q_i while driving qubit q_j using e.g., a randomized benchmarking sequence and comparing the extracted frequency to that when not driving q_j . The frequency difference f_{AC} can be related to a coherent error θ_{AC} and an error rate using:

$$\epsilon_{AC} = \theta^2 = \left(\frac{f_{AC} \cdot \tau_I}{2} \right)^2. \quad (3.22)$$

Should it turn out that the AC Stark shift is a dominant contribution to the error rate, this can be reduced by increasing the frequency separation (Δ) between different frequency qubits or by reducing the coupling to these undesired microwave signals. To reduce the coupling strength, electrostatic simulations can be used to investigate stray couplings. Once any stray couplings have been addressed, the coupling can be further reduced by redesigning the device to have a lower qubit-qubit coupling.

3.5.3 AC-flux crosstalk

Transmons can be detuned by applying a current to a flux-bias line which induces a flux through a SQUID loop. A current applied to a specific flux-bias line should only affect one qubit. However, the on-chip return currents to ground can lead to significant flux crosstalk. The flux crosstalk for DC currents is typically on the order of $\sim 2\%$ but has no consequence for the effective error rate as it can be canceled through calibration. Although one might naively expect the AC-flux crosstalk to be the same as the DC-flux crosstalk, this is not the case. The AC-flux crosstalk is typically significantly lower because of slow (\sim ms timescale) dynamics [176].

AC-flux crosstalk can be measured by performing a Ramsey experiment on a spectator qubit while applying a flux pulse on another qubit. By comparing the difference in extracted frequency when the flux pulse is applied to the frequency extracted when the flux pulse is not applied, the effective flux crosstalk can be extracted. Although this is a straightforward experiment, this experiment has not been performed yet.

The effect of AC-flux crosstalk can be mitigated in several ways. The first solution is to ensure that waveforms have a net-zero integral to prevent any charge buildup [176]. A second solution would be to redesign flux-bias lines to ensure that currents to ground do not pass close by other qubits. Finally, it is possible to characterize the remaining flux crosstalk and create a linear inversion matrix to include a correction for the crosstalk in the pulses generated by the AWG.

3.5.4 Measurement crosstalk

In measurement, we identify two forms of crosstalk, measurement induced dephasing and measurement detection crosstalk.

Measurement-induced dephasing

A measurement pulse applied to a qubit q_j can lead to parasitic measurement of another qubit q_i on the same feedline [130, 177]. When performing measurement, a measurement pulse is applied to one of the feedlines. To allow selective addressing, each readout resonator has a distinct resonance frequency. When targeting one resonator, a parasitic population of another resonator can occur leading to dephasing of the coupled qubit. The effect is quantified by the integrated measurement-induced dephasing [102, 130, 178], with the corresponding phase-flip error per measurement

$$\epsilon_{\phi_M} = \frac{1}{2} \left(1 - e^{-\overline{\Gamma_{m,ij}} \tau_m} \right). \quad (3.23)$$

This dephasing only leads to an operation error if q_i was not intended to be measured. In the QuSurf architecture, it is thus only required to reduce this error if q_i and q_j are of different type (i.e., not both X ancillas, Z ancillas or data qubits). Using the 3-qubit Purcell device [61], the highest observed error was $\epsilon_{\phi_M} = 4 \cdot 10^{-3}$ (see also Table 3.3).

There are several mitigation strategies to reduce measurement-induced dephasing. First, one can consider using pulse shaping techniques to reduce the spectral overlap between the pulse and spurious frequencies. Simulations [130] indicate that using a Gaussian-filtered square-shaped readout pulse can significantly reduce measurement-induced dephasing. Alternatively, one can redesign the frequencies of the readout resonators to be further apart

Measurement detection crosstalk

The measurement detection crosstalk describes the dependence of the measurement outcome of qubit q_i on the state of qubit q_j . We quantify this dependence by first defining the cross-measurement fidelity [130]¹³:

$$F_{ij} = 1 - P(e_i | I_j) - P(g_i | \pi_j), \quad (3.24)$$

where e_i (g_i) denotes the assignment of q_i to the excited (ground) state, π_j (I_j) denotes the preparation with (without) a π pulse on q_j . For the device, we obtain an overall average measurement detection crosstalk error of

$$\epsilon_{\text{mdX}} = \langle 1 - |F_{ij}| \rangle. \quad (3.25)$$

where the average $\langle \cdot \rangle$ is taken over all combinations of assignment (i) and preparation (j) with $i \neq j$.

Similar to [130], we characterize detection crosstalk on a 4-qubit patch of the Chimaera S7 device. In order to characterize the errors produced by detection crosstalk in multi-qubit

¹³Note that this definition differs from the commonly used *average* assignment fidelity of Equation (A.10) by a factor 2.

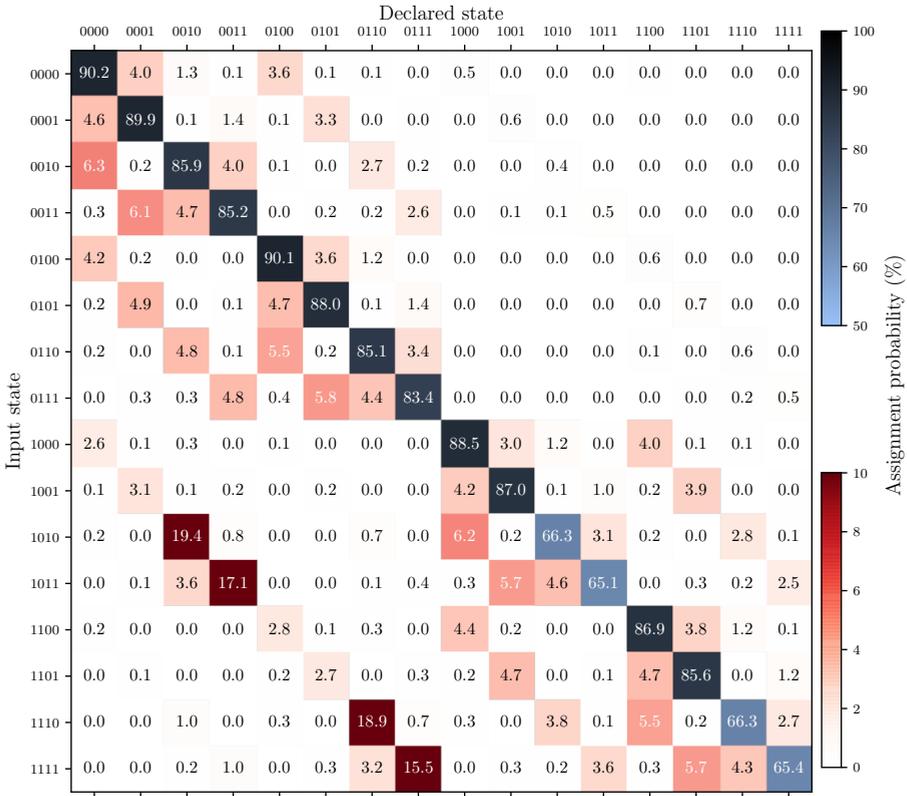


Figure 3.11.: Assignment probability matrix for a four qubit sub-patch of the Chimaera S7 device consisting of qubits X , D_4 , D_2 and Z_2 . Input state and declared state labels are ordered X , D_4 , D_2 , Z_2 .

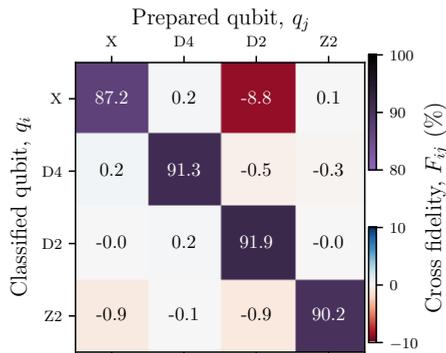


Figure 3.12.: Measured cross-fidelity matrix F_{ij} for a four qubit sub patch of the Chimaera S7 device as defined in Equation (3.24).

quantum algorithms, we prepared qubits X , $D4$, $D2$ and $Z2$ in all $2^4 = 16$ basis states, and measured each state. From these results we construct the 16×16 assignment probability matrix shown in Figure 3.11. Based on the assignment probability matrix, we first investigate how much information about qubit i is contained in the measurement result of qubit j by calculating the F_{ij} . In an ideal measurement, the result of qubit i should be independent of which state we prepare qubit j , and hence $F_{ii} = 1$, and $F_{ij} = 0$ if $i \neq j$ as we prepare qubit i equally in the excited and ground states. We see in Figure 3.12 that all intended qubit readout fidelities (diagonal terms) are greater than 85 %, while for all other terms, with the exception of $F_{X,D2}$, the cross-fidelity is below 1 %. With the exception of the high cross fidelity between X and $D2$, our results are consistent with [130]. The individual qubit readout errors of around 10 % are much larger than the correlated measurement errors. As expected due to the use of Purcell and mode-matched filters, incorrect qubit assignment appears to originate mainly from single-qubit errors such as T_1 decay, improper state distinguishability due to finite SNR, and not so much from spectral overlap between the readout tones. It is possible that the large readout crosstalk is related to the large residual ZZ coupling between qubits X and $D2$.

3.6 Conclusion & outlook

In this chapter we have given an overview of several error mechanisms and characterization techniques relevant for the superconducting quantum systems.

Direct error rates of standard operations, single-qubit, two-qubit gates, and measurement shown in Table 3.1, are typically coherence limited and not preventing a demonstration of quantum fault tolerance. Mature black box characterization protocols (RB A.1.1, IRB A.1.3 and GST A.1.4) exist and are used to characterize single- and two-qubit gates. Measurement is typically characterized using a tailor made protocol (A.3.1 and A.3.2).

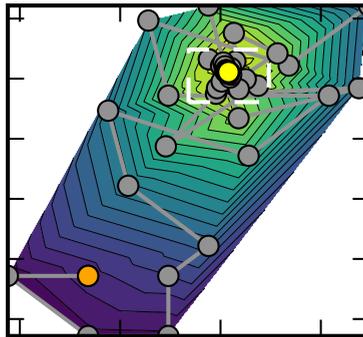
Other error sources such as leakage (Section 3.4.1) and non-Markovian errors appear to be small enough to not significantly impact performance, although understanding how leakage and non-Markovian errors affect the surface code and quantum computation in general is an area of active research. Leakage is typically characterized by making modifications to existing black-box protocols (A.7.1) while characterizing non-Markovian errors requires custom protocols tailored to the specific error source. Parameter drift in particular is an error source that is only investigated when its effects manifest themselves in experiment. Characterization of parameter drift is typically done using ad-hoc analysis although there are attempts at standardization [152]. Parameter drift is particularly relevant as we anticipate that this will become more important as systems become larger, and calibration time increases, and as error rates go down as this causes systems to become more sensitive to small changes in parameters.

Crosstalk errors are particularly problematic, both from a performance as well as a characterization perspective. Performance in the Chimaera S7 device is limited by the strong residual ZZ coupling (Figure 3.8) caused by frequency (mis) targeting of the qubits. By addressing the frequency targeting we expect to be able to resolve this issue although it is not clear how crosstalk will affect devices once these problems are resolved. However, recent

experiments in similar systems [36] indicate that crosstalk in large systems can be small. Characterizing crosstalk on larger devices is challenging because of scaling issues. Finding protocols that do not only quantify the magnitude of crosstalk but also identify what kind of crosstalk and what elements are responsible is an area of active research in the QCVV community.

Although we have initially framed black-box characterization and the experimentalist's approach as opposing viewpoints on QCVV, they are two sides of the same coin. Both approaches attempt to characterize the same systems and rely on similar techniques such as amplifying errors by repeating operations. The biggest advantage of the experimentalist's approach is that it is flexible in dealing with new and unanticipated effects. The black-box approach is more rigorous and has the potential to be more efficient, especially in characterizing crosstalk in large systems. However, system agnostic methods generally do not take advantage of all the knowledge of a particular system.

In an ideal scenario, the experimentalist's approach is used to explore new systems and more rigorous protocols are put in place as understanding solidifies. At the same time, the experimentalist's toolbox is expanded with techniques from the QCVV community. I am particularly optimistic about recent developments in the pyGSTi library [153, 154] in which it is possible to define custom error models (as opposed to a fully generic PTM) and perform model testing on a dataset corresponding to arbitrary experimental sequences. This makes the rigorous GST analysis, at least in principle, accessible in experimental settings. Other challenges relate to a language barrier between theorists and experimentalists, both in relating dimensionless quantities to system specific values with units such as Hz and Volt, but also in understanding what problems are relevant and what techniques are available.



We present a tuneup protocol for qubit gates with tenfold speedup over traditional methods reliant on qubit initialization by energy relaxation. This speedup is achieved by constructing a cost function for Nelder-Mead optimization from real-time correlation of non-demolition measurements interleaving gate operations without pause. Applying the protocol on a transmon qubit achieves 0.999 average Clifford fidelity in one minute, as independently verified using randomized benchmarking and gate set tomography. The adjustable sensitivity of the cost function allows detecting fractional changes in gate error with nearly constant signal-to-noise ratio. The restless concept demonstrated can be readily extended to the tuneup of two-qubit gates and measurement operations.

4.1 Introduction

Reliable quantum computing requires the building blocks of algorithms, quantum gates, to be executed with low error. Strategies aiming at quantum supremacy without error correction [179, 180] require $\sim 10^3$ gates, and thus gate errors $\sim 10^{-3}$. Concurrently, a convincing demonstration of quantum fault tolerance using the circuits Surface-17 and -49 [125, 181] under development by several groups worldwide requires gate errors one order of magnitude below the $\sim 10^{-2}$ threshold of surface code [48, 182].

The quality of qubit gates depends on qubit coherence times and the accuracy and precision of the pulses realizing them. With the exception of a few systems known with metrological precision [183], pulsing requires meticulous calibration by closed-loop tuning, i.e., pulse adjustment based on experimental observations. Numerical optimization algorithms have been implemented to solve a wide range of tuning problems with a cost-effective number of iterations [169, 170, 178, 184–186]. However, relatively little attention has been given to quantitatively exploring the speed and robustness of the algorithms used. This becomes crucial with more complex and precise quantum operations, as the number of parameters and requisite precision of calibration grow.

Though many aspects of tuning qubit gates are implementation independent, some details are specific to physical realizations. Superconducting transmon qubits are a promising hardware for quantum computing, with gate times already exceeding coherence times by three orders of magnitude. Conventional gate tuneup relies on qubit initialization, performed passively by waiting several times the qubit energy-relaxation time T_1 or actively through feedback-based reset [63]. Passive initialization becomes increasingly inefficient as T_1 steadily increases [187, 188], while feedback-based reset is technically involved [189].

Here, we present a gate tuneup method that dispenses with T_1 initialization and achieves tenfold speedup over the state of the art [169] without active reset. Restless tuneup exploits the real-time correlation of quantum-non-demolition (QND) measurements to interleave gate operations without pause, and the evaluation of a cost function for numerical optimization with adjustable sensitivity at all levels of gate fidelity. This cost function is obtained from a simple modification of the gate sequences of conventional randomized benchmarking (CRB) to penalize both gate errors within the qubit subspace and any leakage from it. We quantitatively match the signal-to-noise ratio of this cost function with a model that includes measured T_1 fluctuations. Restless tuneup robustly achieves T_1 -dominated gate fidelity of 0.999, verified using both CRB with T_1 initialization and a first implementation of gate set tomography (GST) [139] in a superconducting qubit. While this performance matches that of conventional tuneup, restless is tenfold faster and converges in one minute.

4.2 The concept and benefits of restless tuning

In many tuneup routines [Figure 4.1(a)], the relevant information from the measurements can be expressed as the fraction ε of non-ideal outcomes (m_n). In conventional gate tuneup, a qubit is repeatedly initialized in the ground state $|0\rangle$, driven by a set of gates ($\{G\}$) whose

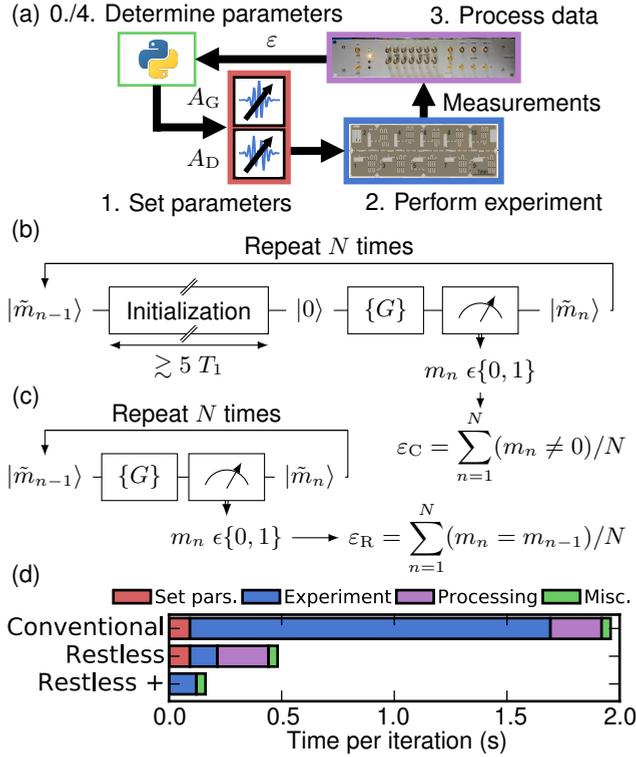


Figure 4.1.: (a) A general qubit gate tuneup loop. In conventional tuneup (b), the qubit is initialized before measuring the effect of $\{G\}$. In restless tuneup (c), the qubit is not initialized, and instead m_{n-1} is used to estimate the initial state ($|\tilde{m}_{n-1}\rangle$). (d) Benchmark of various contributions to the time per iteration in conventional and restless tuneup, without and with technical improvements (see text for details).

net operation is ideally identity, and measured [Figure 4.1(b)]. The conventional cost function is the raw infidelity,

$$\varepsilon_C = \sum_{n=1}^N (m_n \neq 0) / N.$$

The central idea of restless tuning [Figure 4.1(c)] is to remove the time-costly initialization step, by measuring the correlation between subsequent QND measurements and interleaving gate operations without any rest¹. For example, when the net ideal gate operation is a bit flip, we can define the error fraction

$$\varepsilon_R = \sum_{n=2}^N (m_n = m_{n-1}) / N. \quad (4.1)$$

We demonstrate restless tuneup of DRAG pulses [97] on the transmon qubit used in Chapter 5. We choose DRAG pulses (duration $\tau_P = 20$ ns) for their proven ability to reduce gate

¹except 3.25 μ s needed for passive depletion of photons leftover from the 1 μ s measurement [178]

error and leakage [98, 166] with few-parameter analytic pulse shapes. These pulses consist of Gaussian (G) and derivative of Gaussian (D) envelopes of the in- and quadrature-phase components of a microwave drive at the transition frequency f between qubit levels $|0\rangle$ and $|1\rangle$. These components are generated using four channels of an arbitrary waveform generator (AWG), frequency upconversion by sideband modulation of one microwave source, and two I-Q mixers. The G and D components are combined inside a vector switch matrix (VSM) [92] (details in Section 4.5.1). A key advantage of this scheme using four channels is the ability to independently set the G and D amplitudes (A_G and A_D , respectively), without uploading new waveforms to the AWG.

To measure the speedup obtained from the restless method, we must take the complete iteration into account. The traditional iteration of a tuneup routine involves: (1) setting parameters (4 channel amplitudes on a Tektronix 5014 AWG); (2) acquiring $N = 8000$ measurement outcomes; (3) sending the measurement outcomes to the computer and processing them; and (4) miscellaneous overhead that includes determining the parameters for the next iteration, as well as saving and plotting data. In Figure 4.1(d), we visualize these costs for an example optimization experiment. We intentionally penalize the restless method by choosing a large number of gates (~ 550). Even in these conditions, restless sequences reduce the acquisition time from 1.60 to 0.12 s. However, the improvement in total time per iteration (from 1.98 to 0.50 s) is modest due to 0.38 s of overhead.

We take two steps to reduce overhead. The 0.23 s required to send all measurement outcomes to the computer and then calculate the error fraction is reduced to < 1 ms by calculating the fraction in real time, using the same FPGA system that digitizes and processes the raw measurement signals into bit outcomes. The 0.09 s required to set the four channel amplitudes in the AWG is reduced to 1 ms by setting A_G and A_D in the VSM. With these two technical improvements, the remaining overhead is dominated by the miscellaneous contributions (40 ms). This reduces the total time per restless (conventional) iteration to 0.16 s (1.64 s).

A quantity of common interest in gate tuneup is the average Clifford fidelity F_{Cl} , which is typically measured using CRB. In CRB, $\{G\}$ consists of sequences of N_{Cl} random Clifford gates, including a final recovery Clifford gate that makes the ideal net operation identity. Following [190], we compose the 24 single-qubit Clifford gates from the set of π and $\pm\pi/2$ rotations around the x and y axes, which requires an average of 1.875 gates per Clifford. Gate errors make ε_C increase with N_{Cl} as [141, 142]

$$1 - \varepsilon_C = A \cdot (p_{\text{Cl}})^{N_{\text{Cl}}} + B. \quad (4.2)$$

Here, A and B are constants determined by state preparation and measurement error (SPAM), and $1 - p_{\text{Cl}}$ is the average depolarizing probability per gate, making $F_{\text{Cl}} = \frac{1}{2} + \frac{1}{2}p_{\text{Cl}}$. Extracting F_{Cl} from a CRB experiment involves measuring ε_C for different N_{Cl} and fitting Equation (4.2). However, for tuning it is sufficient to optimize ε_C at one choice of N_{Cl} , because $\varepsilon_C(N_{\text{Cl}})$ decreases monotonically with F_{Cl} [169].

In the presence of leakage, CRB sequences and ε_C are not ideally suited for restless tuneup. Typically, there is significant overlap in the readout signals from the first- ($|1\rangle$) and

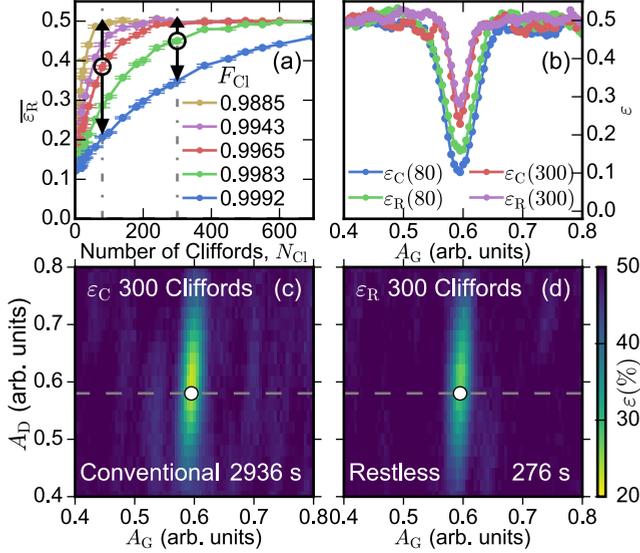


Figure 4.2.: (a) Average error fraction of RRB for different F_{Cl} vs N_{Cl} . (b) ε_C and ε_R as a function of A_G for $N_{Cl} = 80$ and $N_{Cl} = 300$. The curves are denoted by a dashed line in (c-d). (c-d) ε for $N_{Cl} = 300$ as a function of A_G and A_D . White circles indicate minimal ε . Total acquisition time is shown at the bottom right.

second- ($|2\rangle$) excited state of a transmon. A transmon in $|2\rangle$ can produce a string of identical measurement outcomes until it relaxes back to the qubit subspace. If the ideal net operation of $\{G\}$ is identity, the measurement outcomes can be indistinguishable from ideal behavior. Although the leakage on single-qubit gates is typically small ($10^{-5} - 10^{-3}$ per Clifford for the range of A_D considered [92, 166]), a simple modification to the sequence allows penalizing leakage. By choosing the recovery Clifford for restless randomized benchmarking (RRB) sequences so that the ideal net operation of $\{G\}$ is a bit flip, leakage produces an error. This simple modification makes ε_R a better cost function.

4.3 Experimental results

4.3.1 Experimental comparison of restless and restful cost functions

We now examine the suitability of the restless scheme for optimization (Figure 4.2). Plots of the average $\varepsilon_R(N_{Cl})$ [$\overline{\varepsilon}_R(N_{Cl})$] at various F_{Cl} (controlled via A_G) behave similarly to ε_C in CRB. Furthermore, ε_R is minimized at the same A_G as ε_C , with only a shallower dip because of SPAM. The (A_G, A_D) landscapes for both cost functions [Figure 4.2(c-d)] are smooth around the optimum, making them suitable for numerical optimization. The fringes far from the optimum arise from the limited number of seeds (always 200) used to generate the RB sequences. Note that while the landscapes are visually similar, the difference in time required to map them is striking: ~ 50 min for ε_C versus < 5 min for ε_R at $N_{Cl} = 300$.

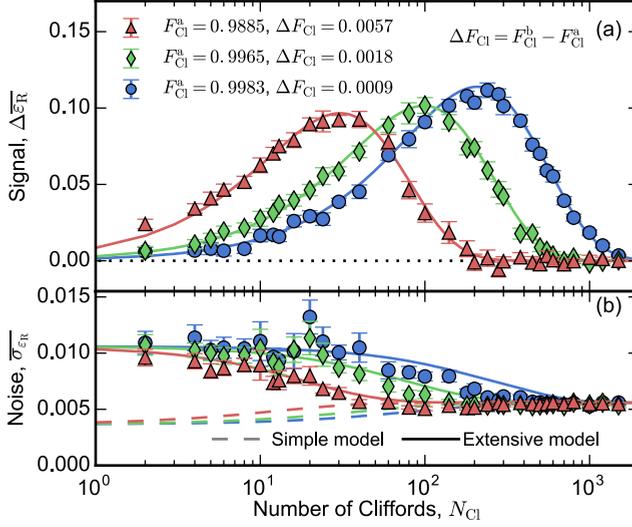


Figure 4.3.: (a) Signal $\Delta \overline{\varepsilon}_R$ for a halving of the gate infidelity, plotted as a function N_{Cl} at $F_{Cl}^a \sim 0.989$ (red), 0.996 (green) and 0.998 (blue). (b) Noise dependence on N_{Cl} at the same fidelity levels. Added curves are obtained from the two models described in the main text.

4.3.2 Signal and noise in restless tuning

The sensitivity of ε_R to the tuning parameters depends on both the gate fidelity and N_{Cl} . This can be seen in the variations between curves in Figure 4.2(a). In order to quantify this sensitivity, we define a signal-to-noise ratio (SNR). For signal we take the average change in the error fraction, $\Delta \overline{\varepsilon}_R = \overline{\varepsilon}_R(F_{Cl}^b) - \overline{\varepsilon}_R(F_{Cl}^a)$, from F_{Cl}^a to $F_{Cl}^b \approx \frac{1}{2} + \frac{1}{2} F_{Cl}^a$ (halving the infidelity). For noise we take $\overline{\sigma}_{\varepsilon_R}$, the average standard deviation of ε_R between F_{Cl}^a and F_{Cl}^b . We find that the maximal SNR remains ~ 15 for an optimal choice of N_{Cl} that increases with F_{Cl}^a (Figure 4.3 and details in Section 4.5.2). This allows tuning in logarithmic time, since reducing error rates $p \rightarrow p/2^M$ requires only M optimization steps.

A simple model describes the measurement outcomes as independent and binomially distributed with error probability ε_R , as per Equation (4.2) with $\varepsilon_C \rightarrow \varepsilon_R$. This model captures all the essential features of the signal. However, it only quantitatively matches the noise at high N_{Cl} . Experiment shows an increase in noise at low N_{Cl} . In this range, ε_R is dominated by SPAM, which is primarily due to T_1 . We surmise that the increase stems from T_1 fluctuations [145] during the acquisition of statistics in these RRB experiments. To test this hypothesis, we develop an extensive model incorporating T_1 fluctuations into the calculation of both signal and noise Section 4.5.2. We find good agreement with experimental results using independently measured values of $\overline{T_1}$ and σ_{T_1} . The good agreement confirms the non-demolition character of the measurement previously reported in [178].

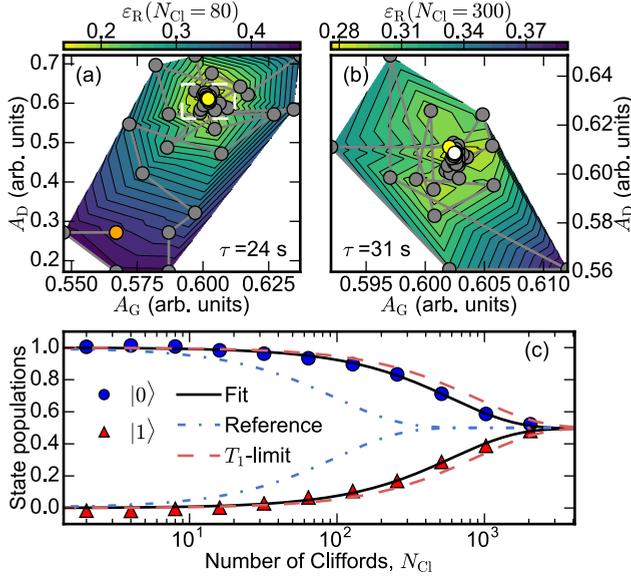


Figure 4.4.: Two-parameter restless tuneup using a two-step optimization, first at $N_{Cl} = 80$ (a) and then at $N_{Cl} = 300$ (b). Contour plots show a linear interpolation of ε_R . The starting point, intermediate result and final result are marked by orange, yellow, and white dots respectively. (c) CRB of tuned pulses ($F_{Cl} = 0.9991$), compared to $F_{Cl}^{(T_1)} = 0.9994$ and $F_{Cl} = 0.995$ for reference.

4.3.3 Gate optimization with restless tuning

Following its validation, we now employ ε_R in a two-step numerical optimization protocol (Figure 4.4). We choose the Nelder-Mead algorithm [191] as it is derivative-free and easy to use, requiring only the specification of a starting point and initial step sizes. The first step using $\varepsilon_R(N_{Cl} = 80)$ ensures convergence even when starting relatively far from the optimum, while the second step using $\varepsilon_R(N_{Cl} = 300)$ fine tunes the result. We test the optimization for four realistic starting deviations from the optimal parameters (A_D^{opt}, A_G^{opt}). A_G is chosen at both approximately 6% above and below A_G^{opt} , selected as a worst-case estimate from a Rabi oscillation experiment. A_D is chosen at both approximately half and double A_D^{opt} . The initial step sizes are $\Delta A_G \approx -0.03A_G^{opt}$, $\Delta A_D \approx -0.25A_D^{opt}$ for the first step, and $\Delta A_G \approx -0.01A_G^{opt}$, $\Delta A_D \approx -0.08A_D^{opt}$ for the second step.

We assess the accuracy of the above optimization and compare to traditional methods. A CRB experiment [Figure 4.4(c)] following two-parameter restless optimization indicates $F_{Cl} = 0.9991$. This value matches the average achieved by both restless and conventional tuneups for the different starting conditions. We also implement GST to independently verify results obtained using CRB. From the process matrices we extract the average GST Clifford fidelity, $F_{Cl}^{GST} = 0.99907 \pm 0.00003$ (0.99909 ± 0.00003) for restless (conventional) tuneup Section 4.5.6, consistent with the value obtained from CRB.

	2-par. (A_G, A_D)		3-par. (A_G, A_D, f)	
	conv.	restl.	conv.	restl.
$\overline{F_{Cl}}$	0.9991	0.9991	0.9990	0.9990
$\sigma_{F_{Cl}}$	$3 \cdot 10^{-5}$	$3 \cdot 10^{-5}$	0.0001	0.0001
$\overline{\tau}$	660 s	59 s	610 s	66 s
σ_τ	110 s	11 s	110 s	13 s
$\overline{N_{it}}$	400	370	370	420
$\sigma_{N_{it}}$	70	70	70	80
$\overline{F_{Cl}^{(T_1)}}$	0.9994		0.9993	
$\overline{T_1}$	21.4 μ s		19.3 μ s	

Table 4.1.: Tuning protocol performance. Mean (overlined) and standard deviations (denoted by σ) of F_{Cl} , time to convergence τ , and number of iterations N_{it} for restless and conventional tuneups with 2 and 3 parameters. Average T_1 measured throughout these runs and corresponding average $F_{Cl}^{(T_1)}$ are also listed.

4.3.4 Gate optimization robustness

The robustness of the optimization protocol is tested by interleaving tuneups with CRB and T_1 measurements over 11 hours (summarized in Table 4.1, and detailed in Section 4.5.7). Both tuneups reliably converge to $F_{Cl} = 0.9991$, close to the T_1 limit [192]:

$$F_{Cl}^{(T_1)} \approx \frac{1}{6} \left(3 + 2e^{-\tau_c/2T_1} + e^{-\tau_c/T_1} \right) = 0.9994, \quad (4.3)$$

with $\tau_c = 1.875 \tau_p$. However, restless tuneup converges in one minute, while conventional tuneup requires eleven.

It remains to test how restless tuneup behaves as additional parameters are introduced. Many realistic scenarios also require tuning the drive frequency f . As a worst case, we take an initial detuning of ± 250 kHz. The initial step size in the first (second) step is 100 kHz (50 kHz). The 3-parameter optimization converges to $F_{Cl} = 0.9990 \pm 0.0001$ for both restless and conventional tuneups. We attribute the slight decrease in F_{Cl} achieved by 3-parameter optimization to the observed reduction in average T_1 .

4.4 Conclusions

In summary, we have developed an accurate and robust tuneup method achieving a tenfold speedup over the state of the art [169]. This speedup is achieved by avoiding qubit initialization by relaxation, and by using real-time correlation of measurement outcomes to build the cost function for numerical optimization. We have applied the restless concept to the tuneup of Clifford gates on a transmon qubit, reaching a T_1 -dominated fidelity of 0.999 in one minute, verified by conventional randomized benchmarking and gate set tomography. We have shown experimentally that the method can detect fractional reductions in gate error with nearly constant signal-to-noise ratio. An interesting next direction is to develop an algorithm that makes

optimal use of this tunable sensitivity while maintaining the demonstrated robustness. The enhanced speed combined with the generic nature of the optimizer would also allow exploring other, more generic non-adiabatic gates without analytic pulse shapes, in a fashion analogous to optimal control theory [193, 194]. Immediate next experiments will extend the restless concept to the tuneup of two-qubit controlled-phase gates [78, 195] exploiting interactions with non-computational states [95], in which leakage errors often dominate ($\sim 10^{-2}$). In this context, we anticipate that the RRB modification and the ε_R cost function will prove essential to reach 0.999 fidelity. Finally, we also envision applying the restless concept to the simultaneous tuneup of single-qubit gates in the many-qubit setting (e.g, a logical qubit).

4.5 Methods

This section presents the hardware configuration used for the numerical tuneup, the characterization and modeling of the signal and noise of restless randomized benchmarking, and the procedure for calculating Clifford gate fidelities from GST process matrices. Finally, it presents the data summarized in Table 1 of the main text.

4.5.1 Setup for numerical optimization

The key hardware components executing the tuneup loop of Figure 4.1 are shown in Figure 4.5. The computer is responsible for preparing the experiment and executing the numerical algorithm determining the parameter values for each iteration. To do this, the computer relies on two python packages, *PycQED* for cQED-specific routines [196], and *QCoDeS* for the framework of instrument drivers [197]. Part of the preparation consists of generating and uploading a sequence of control pulses and markers to the AWG. Once an experiment starts, the AWG is responsible for all time-critical matters, including gating the readout pulses on the microwave source and triggering the data acquisition on the FPGA controller. The control pulses are generated using 4 AWG channels, 2 for the I and Q quadratures of the Gaussian component and 2 for the quadratures of the derivative component. The components are up-converted using single-sideband mixers and a constant microwave tone as a local oscillator (LO). This allows independent control over the amplitude of both pulse components, using either the AWG or the VSM. The frequency of the pulses can be changed by changing the frequency of the LO. Note that all these controls can be applied without regenerating and uploading the sequence of control pulses to the AWG.

The transmon (same as used in Chapter 5) is operated at its coherence sweetspot, with transition frequency 6.47 GHz, -315 MHz anharmonicity, relaxation time $T_1 = 22 \mu\text{s}$ and echo time $T_{2,\text{echo}} = 39 \mu\text{s}$. It is readout by interrogating its dispersively coupled resonator near its fundamental with a tone at 6.848 GHz. Readout transients are amplified at the front end of the amplification chain by a Josephson parametric amplifier operated in the non-degenerate mode, providing 14 dB of gain. The FPGA controller performs final demodulation, integration and discrimination of measurement transients and real-time calculation of ε .

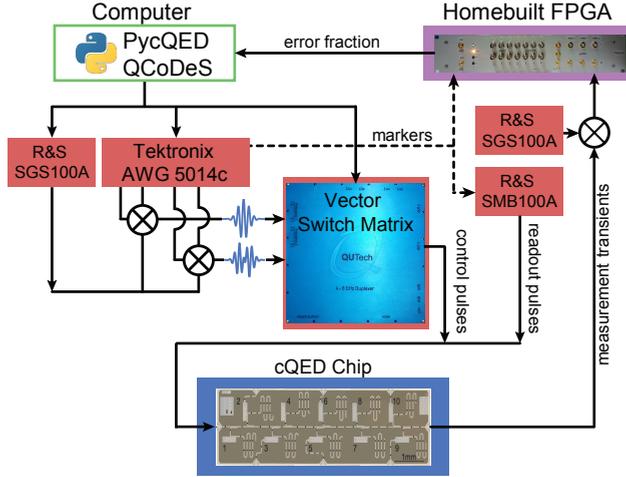


Figure 4.5.: Schematic overview of the hardware components used in the numerical tuneup.

4.5.2 Signal and noise of the restless cost function

We experimentally obtained the signal and noise of RRB presented in Figure 4.3 of the main text from 50 RRB experiments ($N = 8000$ measurement outcomes each) at each N_{Cl} (32 values) and F_{Cl} (5 values). Here, F_{Cl} was varied by changing A_{Cl} . The procedure was repeated 10 times for all settings to build up statistics. In this section, we present the derivation of the extended model used to predict these curves (Section 4.5.3), using independent measurements of qubit T_1 fluctuations performed one day apart (Section 4.5.4).

4.5.3 Modeling

We develop a model for the RRB experiment to capture both the signal and noise obtained experimentally. The standard deviation differs from that simply expected from a binomial distribution. This is hypothesized to be caused by T_1 fluctuations that are quasi-static during individual RRB experiments, but dynamic on the time scale required for 50 repetitions. We attempt to match the experimental results with a model containing T_1 and its fluctuations, a relaxation independent pulse error p_{pulse} , and a SPAM offset $p_s^{(c)}$. Independent measurements of the average and standard deviation of T_1 , and extractions of p_{pulse} and $p_s^{(c)}$ from the data in Figure 4.2(a) are used to produce the model curves in Figure 4.3.

Modeling without T_1 fluctuations

The time taken for a single-shot RRB experiment can be written $\tau_{\text{RRB}} = \tau_{\text{RO}} + \tau_{\text{Cl}} N_{\text{Cl}}$. The static time $\tau_{\text{RO}} = 4.25 \mu\text{s}$ is the readout-and-depletion time, whilst the Clifford-dependent time $\tau_{\text{Cl}} = 37.5 \text{ ns}$ is the average time it takes to perform a Clifford gate. To each of these we

can associate an error rate, making the total error rate per single-shot experiment (assuming independent error rates)

$$p_e = p_s +_p N_{\text{Cl}} \times_p p_c.$$

Here, p_s is the error contribution due to SPAM, and $p_c = 1 - F_{\text{Cl}}$ is the error contribution per Clifford. We must be careful with adding probabilities here, as two errors cancel. This is taken care of by an independent probabilistic addition $a +_p b = a + b - 2ab = a(1 - b) + b(1 - a)$, and a probabilistic multiplication $c \times_p a$ (with $c \in \mathbf{N}$). The multiplication can be defined in two equivalent ways: as multiple additions: $a +_p a +_p \dots +_p a$ (repeated c times for c a positive integer), or as a direct calculation of the probability of an odd number of errors occurring over c events with an error rate of a .

The latter construction allows for a direct simplification. We write the sum over all odd numbers n of the probability of n errors occurring, which can be counted directly via combinatorics:

$$N_{\text{Cl}} \times_p p_c = N_{\text{Cl}} p_c (1 - p_c)^{N_{\text{Cl}}-1} + \binom{N_{\text{Cl}}}{3} p_c^3 (1 - p_c)^{N_{\text{Cl}}-3} + \dots$$

This can be recognized as the odd terms from the binomial expansion of $((1 - p_c) \pm p_c)^{N_{\text{Cl}}}$, which can be singled out by canceling the even terms.

$$\begin{aligned} N_{\text{Cl}} \times_p p_c &= \frac{1}{2} \left[((1 - p_c) + p_c)^{N_{\text{Cl}}} - ((1 - p_c) - p_c)^{N_{\text{Cl}}} \right] \\ &= \frac{1}{2} \left[1 - (1 - 2p_c)^{N_{\text{Cl}}} \right], \end{aligned}$$

resulting in a final error rate

$$p_e = p_s + \frac{1}{2} [1 - (1 - 2p_c)^{N_{\text{Cl}}}] (1 - 2p_s). \quad (4.4)$$

Modeling with T_1 fluctuations

If p_s or p_c fluctuate, the error rate p_e for any given single-shot experiment is drawn from a distribution with mean $\overline{p_e}$. This in turn can be calculated assuming that p_s and p_c are drawn from a normal distribution, giving

$$\overline{p_e} = \int dp_s dp_c p_e(p_s, p_c) \frac{1}{2\pi} \exp \left(- \frac{1}{2} \begin{pmatrix} p_s & p_c \end{pmatrix} \Sigma^{-1} \begin{pmatrix} p_s \\ p_c \end{pmatrix} \right) |\Sigma|^{-1/2}.$$

Here, Σ is the covariance matrix;

$$\Sigma = \begin{pmatrix} \text{var}(p_s) & \text{covar}(p_c, p_s) \\ \text{covar}(p_c, p_s) & \text{var}(p_c) \end{pmatrix},$$

with $\overline{p_c}$ ($\text{var}(p_c)$) and $\overline{p_s}$ ($\text{var}(p_s)$) the means (variances) of p_c and p_s , respectively, and $\text{covar}(p_c, p_s)$ the covariance between p_c and p_s . The inverse of Σ can be calculated,

$$\Sigma^{-1} = \frac{1}{\text{var}(p_s)\text{var}(p_c) - \text{covar}(p_s, p_c)^2} \begin{pmatrix} \text{var}(p_c) & -\text{covar}(p_s, p_c) \\ -\text{covar}(p_s, p_c) & \text{var}(p_s) \end{pmatrix}.$$

We make the simplifying assumption that $\text{covar}(p_c, p_s) \ll \max(\text{var}(p_c), \text{var}(p_s))$, leaving us with

$$\Sigma^{-1} \approx \begin{pmatrix} \frac{1}{\text{var}(p_s)} & 0 \\ 0 & \frac{1}{\text{var}(p_c)} \end{pmatrix}.$$

From here the integral in p_s can be evaluated:

$$\bar{p}_e = \bar{p}_s + \frac{1}{2}(1 - 2\bar{p}_s) \left(1 - \int dp_e (1 - 2p_c)^{N_{\text{Cl}}} \frac{\exp\left(\frac{-(p_c - \bar{p}_c)^2}{2 \text{var}(p_c)}\right)}{\sqrt{2\pi \text{var}(p_c)}} \right).$$

In order to calculate the integral in p_c we expand in terms of powers of p_c , allowing the result to be expressed in terms of moments of the normal distribution

$$\int dp_e (1 - 2p_c)^{N_{\text{Cl}}} \frac{\exp\left(\frac{-(p_c - \bar{p}_c)^2}{2 \text{var}(p_c)}\right)}{\sqrt{2\pi \text{var}(p_c)}} = \sum_{n=1}^{N_{\text{Cl}}} \binom{N_{\text{Cl}}}{n} (-2)^n \langle p_e^n \rangle,$$

with $\langle p_e^n \rangle$ the n -th moment of the normal distribution. This may then be expanded in terms of the variance $\text{var}(p_{\text{Cl}})$ to obtain

$$\int dp_e (1 - 2p_c)^{N_{\text{Cl}}} \frac{\exp\left(\frac{-(p_c - \bar{p}_c)^2}{2 \text{var}(p_c)}\right)}{\sqrt{2\pi \text{var}(p_c)}} = \sum_{n=0}^{N_{\text{Cl}}/2} \text{var}(p_c)^n (1 - 2p_c)^{N_{\text{Cl}} - 2n} \frac{N_{\text{Cl}}!}{(N_{\text{Cl}} - 2n)! 2n!!}.$$

Here, $2n!!$ is the product of even positive numbers less than $2n$. We then approximate this to lowest order in $\text{var}(p_c)$ (observed in the experiment to be $\approx 0.01\bar{p}_c$). Note that although this term contains prefactors of N_{Cl} , it also contains prefactors of $(1 - 2p_c)^{N_{\text{Cl}}}$, which prevent it from growing in the large N_{Cl} limit. This leaves

$$\bar{p}_e = \bar{p}_s + \frac{1}{2}[1 - (1 - 2\bar{p}_c)^{N_{\text{Cl}}}] (1 - 2\bar{p}_s),$$

and

$$\begin{aligned} \text{var}(p_e) &= (1 - 2\bar{p}_c)^{2N_{\text{Cl}}} \text{var}(p_s) + N_{\text{Cl}}^2 (1 - 2\bar{p}_s)^2 (1 - 2\bar{p}_c)^{2(N_{\text{Cl}} - 1)} \text{var}(p_c) \\ &\quad + 2N_{\text{Cl}} (1 - 2\bar{p}_s) (1 - 2p_c)^{2N_{\text{Cl}} - 1} \text{covar}(p_c, p_s). \end{aligned}$$

Measurements of ε_{R} use $N = 8000$ single-shot measurement outcomes, which we assume are selected from a binomial distribution with mean $(1 - P)$. P is in turn selected from a distribution with mean \bar{p}_e and standard deviation σ_{p_e} . Let N_e be the number of erroneous measurements, given as $N_e = N\varepsilon_{\text{R}}$. In order to calculate the mean and variance in N_e , we have to calculate the first and second moments of the distribution, averaged over all P . We assume a normal distribution for P . For the first moment we obtain

$$\begin{aligned} \langle N_e \rangle &= \int_{-\infty}^{\infty} \left[\sum_{k=0}^N k \binom{N}{k} P^k (1 - P)^{N-k} \right] e^{-\frac{(P - \bar{p}_e)^2}{(2\sigma_{p_e}^2)}} \frac{1}{\sqrt{2\pi\sigma_{p_e}^2}} dP \\ &= N \int_{-\infty}^{\infty} P e^{-\frac{(P - \bar{p}_e)^2}{(2\sigma_{p_e}^2)}} \frac{1}{\sqrt{2\pi\sigma_{p_e}^2}} dP = N\bar{p}_e. \end{aligned}$$

As expected, the average number of erroneous measurements equals the total number of measurements multiplied by the average error, and is unaffected by fluctuations. For the second moment we calculate

$$\begin{aligned} \langle N_e^2 \rangle &= \int_{-\infty}^{\infty} \left[\sum_{k=0}^N k^2 \binom{N}{k} P^k (1-P)^{N-k} \right] e^{-\frac{(P-\bar{p}_e)^2}{(2\sigma_{p_e})^2}} \frac{1}{\sqrt{2\pi\sigma_{p_e}^2}} dP \\ &= \int_{-\infty}^{\infty} (NP + N(N-1)P^2) e^{-\frac{(P-\bar{p}_e)^2}{(2\sigma_{p_e})^2}} \frac{1}{\sqrt{2\pi\sigma_{p_e}^2}} dP \\ &= N\bar{p}_e + N(N-1)(\bar{p}_e^2 + \sigma_{p_e}^2). \end{aligned}$$

This leads to the final result:

$$\text{var}(\varepsilon_R) = \frac{1}{N}\bar{p}_e(1-\bar{p}_e) + \frac{N-1}{N}\text{var}(p_e). \quad (4.5)$$

The simple model without T_1 fluctuations can be recovered here by setting $\text{var}(p_e) = 0$.

Asymmetry

Due to the asymmetry of T_1 , the error rate $p_e^{(j)}$ depends on whether the qubit is in the excited or ground state during τ_{RO} . The measurement, lasting $\tau_m = 1 \mu\text{s}$, is T_1 rather than noise limited. We can approximate it by perfect state update and measurement at $\tau_b \approx 4\tau_m/7 = 0.57 \mu\text{s}$ [79], followed by a rest time $\tau_a = \tau_{\text{RO}} - \tau_b = 3.68 \mu\text{s}$ before the beginning of the next Clifford sequence. Let the system state at the point of the measurement (i.e., τ_b into the measurement time) be $|j\rangle$ with $j = 0$ or 1 . If a single error occurs during the sequence, the flipping sequence will revert the qubit to the same state $|j\rangle$ at the next measurement point. This implies that the process is biased towards states with higher error rate, and so the error rate cannot be simply averaged over that expected individually for $|0\rangle$ and $|1\rangle$. Instead, we let the population fraction of $|j\rangle$ over the experiment be f_j , and solve the steady-state rate equation for f_j :

$$f_j = p_e^{(j)} f_j + (1 - p_e^{(1-j)})(1 - f_j).$$

This leads to an error rate of

$$p_e = \frac{p_e^{(0)}(1 - p_e^{(1)}) + p_e^{(1)}(1 - p_e^{(0)})}{(1 - p_e^{(0)}) + (1 - p_e^{(1)})}. \quad (4.6)$$

The error during the RRB sequence is state independent, and so the adjustment to Equation (4.4) comes solely from the adjustment to the SPAM error:

$$p_e^{(j)} = p_s^{(j)} + \frac{1}{2}[1 - (1 - 2p_c)^{N_{\text{Cl}}}] (1 - 2p_s^{(j)}),$$

with

$$p_s^{(0)} = p_s^{(c)} + (1 - e^{-\tau_b/T_1}), \quad p_s^{(1)} = p_s^{(c)} + (1 - e^{-\tau_a/T_1})e^{-\tau_b/T_1}.$$

Here, $p_s^{(c)}$ is a small error accounting for non- T_1 SPAM. Substituting these into Equation (4.6) allows for the calculation of the error p_e as a function of p_c , N_{C1} , and T_1 . In order to calculate the standard deviation, we must then calculate the first derivative, via

$$\frac{\partial p_e}{\partial T_1} = \sum_j \frac{\partial p_e}{\partial p_e^{(j)}} \left(\frac{\partial p_e^{(j)}}{\partial p_s^{(j)}} \frac{\partial p_s^{(j)}}{\partial T_1} + \frac{\partial p_e^{(j)}}{\partial p_c} \frac{\partial p_c}{\partial T_1} \right). \quad (4.7)$$

Here, the value of $\frac{\partial p_c}{\partial T_1}$ is obtained by assuming that p_c can be split into a constant pulse error probability p_{pulse} plus a T_1 -induced error probability $p_c^{(T_1)} = 1 - F_{C1}^{(T_1)}$, with $F_{C1}^{(T_1)}$ as defined in Eq. (3).

4

4.5.4 Measurement of T_1 fluctuations

We perform repeated measurements of T_1 one day after the RRB experiments. We extract T_1 from exponential best fits to standard sliding π -pulse experiments. These measurements rely on qubit initialization by waiting. The benefit of this method is that one can measure T_1 fluctuations independently from fluctuations in residual qubit populations, gate fidelity and readout fidelity (unlike restless sequences). The downside is that one can only probe T_1 in $\Delta t = 2.0$ s intervals. We measure T_1 in $L = 234$ runs l of $M = 21$ measurements each, and calculate the single-sided power spectral density (PSD) as

$$S_{T_1}(f) = \frac{2\Delta t}{LM} \sum_{l=1}^L \left| \sum_{m=1}^M \delta T_{1,l}[m] e^{-i2\pi f m \Delta t} \right|^2,$$

where $\delta T_{1,l}[m] = T_{1,l}[m] - \frac{1}{M} \sum_{m'=1}^M T_{1,l}[m']$. We fit $S_{T_1}(f) = \alpha (f/1 \text{ Hz})^\beta$ to the experimental PSD, finding best-fit parameters $\alpha = 8.4 \cdot 10^{-13} \text{ s}^2/\text{Hz}$ and $\beta = -0.81$ (data and fit are shown in Figure 4.6). Extrapolating the PSD to higher frequencies, we can estimate the expected σ_{T_1} in the RRB experiments of Section 4.5.2, by integrating over the frequency interval bounded above by the rate of single RRB experiments ($f_u = 1/0.074$ s at low N_{C1}), and below by the acquisition time for 50 such experiments ($f_l = 1/3.7$ s). We find $\overline{T_1} = 21.6 \mu\text{s}$ and

$$\sigma_{T_1} = \left(\int_{f_l}^{f_u} S_{T_1} df \right)^{1/2} = 2.4 \pm 0.1 \mu\text{s}.$$

We estimate the uncertainty in σ_{T_1} by splitting the dataset into 6 subsets of equal length.

4.5.5 Relation to experiment

Using the measured $\overline{T_1}$, we fit Equation (4.6) to the data in Figure 4.2(a) to extract a common $p_s^{(c)} = 0.006$ and curve specific p_{pulse} . We use Equations (4.6) and (4.7) to obtain the model curves for $\overline{\Delta \varepsilon_R}$ and $\overline{\sigma_{\varepsilon_R}}$ shown in Figure 4.3 of the main text, finding good agreement with experiment.

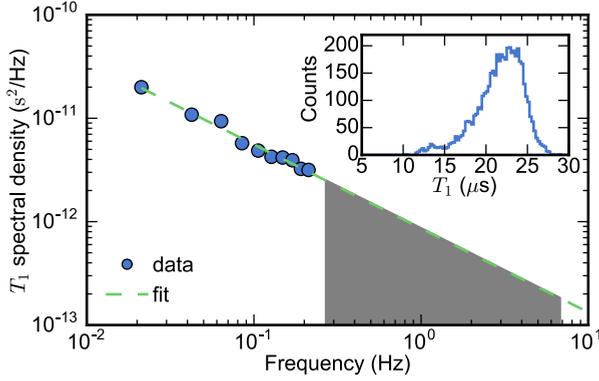


Figure 4.6.: Power spectral density of T_1 fluctuations. Main panel: measured single-sided PSD of T_1 fluctuations and best fit (see details in text). The indicated frequency range is that relevant for estimating σ_{T_1} in the RRB experiments of Section 4.5.2. Inset: Histogram of 4914 T_1 measurements. The set has $\overline{T_1} = 21.6 \mu\text{s}$.

4.5.6 Gate set tomography and randomized benchmarking fidelities

In order to compare results from GST to those acquired using CRB, the results of GST need to be converted to Clifford fidelities. GST performs a full self-consistent tomography of the gates in the set $\{I, X90, Y90, X180, Y180\}$, consisting of the identity and positive $\pi/2$ and π rotations around the x and y axes. The super-operators for the gates in the gate set are extracted from the GST data using pyGSTi [153]. These are then used to construct the 24 elements ($G_{\text{Cl}_n}^{\text{GST}}$) of the (single-qubit) Clifford group (\mathcal{G}_{Cl}) according to the decomposition of [190]. To account for the missing negative rotations in the gate set, we replace negative rotations with their positive counterparts (e.g., $-X90 \rightarrow X90$). For each of these operations, the depolarization probability is calculated as the geometric mean over all poles of the Bloch sphere $|\rho_i\rangle\rangle$ (using the super-operator formalism), of the overlap between the obtained state $G_{\text{Cl}_n}^{\text{GST}}|\rho_i\rangle\rangle$ and the target state $|\rho_t\rangle\rangle$:

$$p_n = \sqrt[n]{\prod_{\rho_i} \langle\langle \rho_t | G_{\text{Cl}_n}^{\text{GST}} | \rho_i \rangle\rangle},$$

where the target state is the state one would get if the gates were perfect:

$$|\rho_t\rangle\rangle = G_{\text{Cl}_n}^{\text{Ideal}} |\rho_i\rangle\rangle.$$

p_{Cl} is the geometric mean of the individual depolarization probabilities for all $G_{\text{Cl}_n} \in \mathcal{G}_{\text{Cl}}$ and related to F_{Cl} through $F_{\text{Cl}} = \frac{1}{2} + \frac{1}{2}p_{\text{Cl}}$.

Table 4.2 summarizes the gate fidelities found after performing the two-parameter optimization, for the four starting (A_G, A_D) conditions discussed in the main text.

	Conventional	Restless
F_I	0.99928 ± 0.00007	0.99921 ± 0.00005
F_{X90}	0.99927 ± 0.00005	0.99925 ± 0.00004
F_{X180}	0.99920 ± 0.00007	0.99910 ± 0.00005
F_{Y90}	0.99908 ± 0.00005	0.99906 ± 0.00005
F_{Y180}	0.99901 ± 0.00008	0.99891 ± 0.00005
F_{Cl}^{GST}	0.99909 ± 0.00005	0.99907 ± 0.00003
F_{Cl}	0.9991	0.9991

Table 4.2.: Measured gate fidelities in GST. Gate fidelities correspond to average gate fidelities for the four starting conditions of the two-parameter optimization as discussed in the main text.

4

4.5.7 Verification of conventional and restless tuneup

The speed, robustness and accuracy of the two- and three- parameter optimizations are tested during an 11-hour period by interleaving conventional and restless tuneups with CRB and T_1 experiments. The data summarized in Table 1 of the main text is shown in Figure 4.7. The two-parameter (three-parameter) optimization loops over 4 (8) different starting conditions as specified in the main text. The starting condition is updated after each set of conventional and restless optimizations.

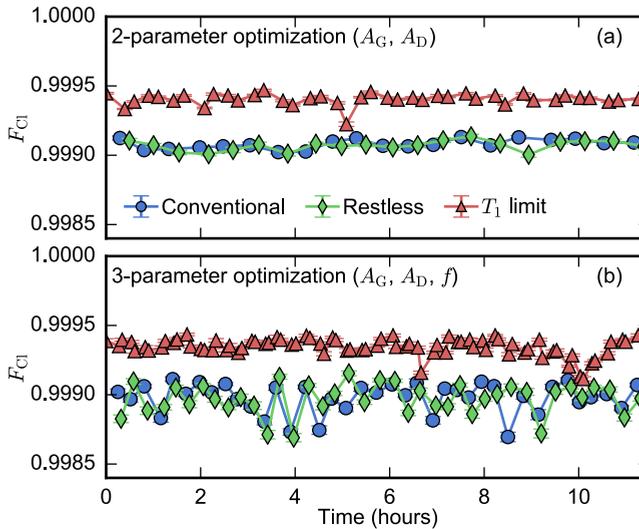
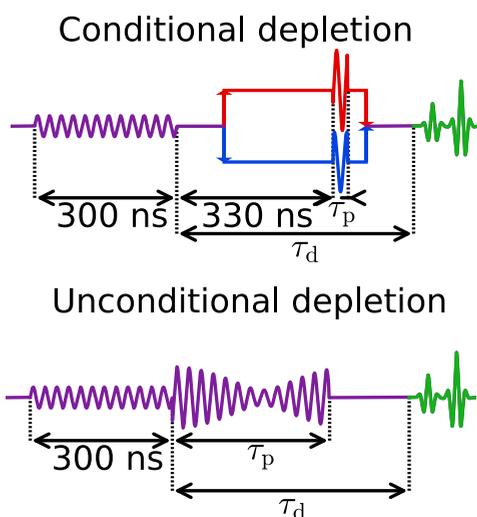


Figure 4.7.: Performance comparison of repeated restless and conventional tuneups for two parameters (a) and three parameters (b). Each iteration consists of a conventional tuneup followed by a CRB measurement of F_{Cl} , a restless tuneup followed by a CRB measurement of F_{Cl} , and a T_1 experiment to determine $F_{Cl}^{(T_1)}$. For each iteration, a new starting condition is chosen (detailed in main text) that is used for both the conventional and restless tuneup.



We present two pulse schemes to actively deplete measurement photons from a readout resonator in the nonlinear dispersive regime of circuit QED. One method uses digital feedback conditioned on the measurement outcome while the other is unconditional. In the absence of analytic forms and symmetries to exploit in this nonlinear regime, the depletion pulses are numerically optimized using the Powell method. We speed up photon depletion by more than six inverse resonator linewidths, saving ~ 1650 ns compared to depletion by waiting. We quantify the benefit by emulating an ancilla qubit performing repeated quantum parity checks in a repetition code. Fast depletion increases the mean number of cycles to a spurious error detection event from order 1 to 75 at a $1 \mu\text{s}$ cycle time.

5.1 Introduction

Many protocols in quantum information processing require interleaving qubit gates and measurements in rapid succession. For example, current experimental implementations of quantum error correction (QEC) schemes [42–44, 198–201] rely on repeated measurements of ancilla qubits to discretize and track errors in the data-carrying part of the system. Minimizing the QEC cycle time is essential to avoid buildup of errors beyond the threshold for fault tolerance.

An attractive architecture for QEC codes is circuit quantum electrodynamics (cQED) [202]. Initially implemented with superconducting qubits, this scheme has since grown to include both semiconducting [203] and hybrid qubit platforms [204, 205]. Readout in cQED involves dispersively coupling the qubit to a microwave-frequency resonator causing a qubit-state dependent shift of the fundamental resonance. This shift can be measured by injecting the resonator with a microwave photon pulse. Inversely however, resonator photons shift the qubit transition frequency (AC Stark shift [202]), leading to qubit dephasing and gate errors. To ensure photons leave the resonator before gates recommence, QEC implementations include a waiting step after measurement. During this dead time, lasting a significant part of the QEC cycle, qubits are susceptible to decoherence. Whilst many prerequisites of measurement for QEC have already been demonstrated (including frequency-multiplexed readout via a common feedline [206], the use of parametric amplifiers to improve speed and readout fidelity [207, 208] and null back-action on untargeted qubits [177]), comparatively little attention has been given to the fast depletion of measurement photons.

Two compatible approaches to accelerate photon depletion have been explored. The first increases the resonator linewidth κ while adding a Purcell filter [42, 209, 210] to avoid enhanced qubit relaxation via the Purcell effect [211]. However, increasing κ also enhances qubit dephasing (for a fixed ratio of the dispersive shift χ and κ as desired for high-fidelity readout [172, 212]) by stray photons [213, 214], introducing a compromise. The second approach actively depletes photons using a counter pulse, as recently demonstrated by McClure *et al.* [185]. This demonstration uses symmetries available when the resonator response is linear. However, reaching the single-shot readout fidelity required for QEC often involves driving the resonator deep into the nonlinear regime, where no such symmetries are available.

Here, we propose and demonstrate two methods for active photon depletion in the nonlinear dispersive regime of cQED. The first uses a homebuilt feedback controller to send one of two depletion pulses conditioned on the declared measurement outcome. The second applies a universal pulse independent of measurement outcome. We maximize readout fidelity at a measurement power two orders of magnitude larger than the power inducing the critical photon number in the resonator [202]. Missing analytic forms for this regime, we rely on numerical optimizations by Powell’s method [215] to tune up pulses, defined by two or four parameters. Both depletion methods speed up depletion by at least $\sim 1250 \text{ ns} \sim 5/\kappa$ compared to waiting. To illustrate the benefits for QEC we emulate an ancilla qubit performing parity checks [177, 216] by subjecting our qubit to repeated rounds of coherent operations and measurement. We quantify performance by extracting the mean number of rounds to an

unexpected measurement outcome (i.e. a detection event). With active depletion, we observe an increase in this mean rounds to event, $\overline{\text{RTE}}$, from 15 to 39 and reduce the cycle time to $1 \mu\text{s} \sim 4/\kappa$. By further fixing the ancilla to remain in the ground state, $\overline{\text{RTE}}$ increases to 75. Simulations [217] indicate that, when including the same intrinsic coherence for surrounding data qubits, a 5-qubit repetition code (studied in [42]) would have a logical error rate below its pseudothreshold [125].

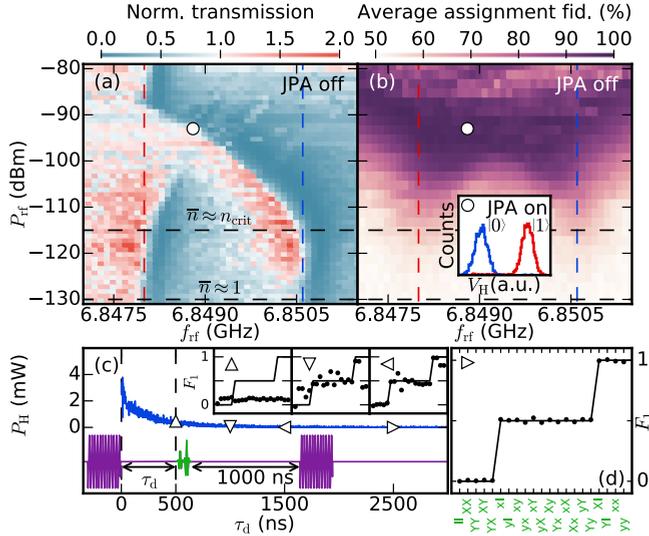


Figure 5.1.: (Color online) (a) CW feedline transmission spectroscopy as a function of incident power and frequency near the low- and high-power fundamentals of the resonator. The qubit is simultaneously driven with a weakly saturating tone. The right (left) vertical line indicates the fundamental $f_{r,|0\rangle}$ ($f_{r,|1\rangle}$) in the linear regime. The dot indicates $(P_{rf}, f_{rf}) = (-93 \text{ dBm}, 6.8488 \text{ GHz})$ used throughout the experiment. (b) Average assignment fidelity \mathcal{F}_a as a function of P_{rf} and f_{rf} ($\tau_r = 1200 \text{ ns}$, $\tau_{\text{int}} = 1500 \text{ ns}$), obtained from histograms with 4000 shots per qubit state. Inset: Turning on the JPA achieves $\mathcal{F}_a = 98.8\%$. (c) Illustration of qubits errors induced by leftover photons. At τ_d , after an initial measurement pulse ends, AllXY qubit pulse pairs are applied and a final measurement is performed 1000 ns later to measure F_1 . The transient of the decaying homodyne signal, P_H , fits $1/\kappa = 250 \pm 2 \text{ ns}$. Insets and (d): F_1 versus pulse pair for several τ_d . The ideal two-step signature is observed at $\tau_d \gtrsim 2500 \text{ ns}$.

5.2 Experimental results

5.2.1 Device characterization

We employ a 2D cQED chip containing ten transmon qubits with dedicated readout resonators, coupled to a common feedline (more details in Section 5.4.1). We focus on one qubit-resonator pair for all data presented. This qubit has frequency $f_q = 6.477 \text{ GHz}$, $T_1 = 25 \mu\text{s}$

and $T_{2,\text{echo}} = 39 \mu\text{s}$. The resonator has a low-power fundamental at $f_{r,|0\rangle} = 6.8506 \text{ GHz}$ ($f_{r,|1\rangle} = 6.8480 \text{ GHz}$) for qubit in $|0\rangle$ ($|1\rangle$), making the dispersive shift $\chi/\pi = -2.6 \text{ MHz}$. Note that this shift also corresponds to the qubit detuning per resonator photon. The fundamentals converge to the bare resonator frequency, $f_{r,\text{bare}} = 6.8478 \text{ GHz}$, at incident power $P_{\text{rf}} \gtrsim -88 \text{ dBm}$. We calibrate a single-photon power $P_{\text{rf}} = -130 \text{ dBm}$ using photon-number splitting experiments (Figure 5.7) according to [218] and a critical photon number [202] $n_{\text{crit}} = (\Delta^2/4g^2) \approx 33$ ($P_{\text{rf}} \approx -115 \text{ dBm}$) using $f_{r,|0\rangle} - f_{r,\text{bare}} = g^2/2\pi\Delta$ and $\Delta = 2\pi(f_{\text{q}} - f_{r,\text{bare}})$.

5.2.2 Measurement tune-up and the effect of leftover photons

Our first objective is to maximize the average assignment fidelity of single-shot readout,

$$\mathcal{F}_{\text{a}} = 1 - \frac{1}{2}(\epsilon_{01} + \epsilon_{10}),$$

where ϵ_{ij} is the probability of incorrectly assigning measurement result j for input state $|i\rangle$. We map \mathcal{F}_{a} as a function of the power P_{rf} and frequency f_{rf} of a measurement pulse of duration $\tau_{\text{r}} = 1200 \text{ ns}$ [Figure 5.1(b)]. \mathcal{F}_{a} is maximized at $P_{\text{rf}} = -93 \text{ dBm}$, 22 dB stronger than the n_{crit} power. The nonlinearity is evidenced by the bending of resonator lineshapes in the accompanying continuous-wave (CW) transmission spectroscopy [Figure 5.1(a)]. We make two additions to further improve \mathcal{F}_{a} . First, we turn on a Josephson parametric amplifier (JPA), providing 14 dB of gain. The improved signal-to-noise ratio allows shortening τ_{r} to 300 ns. Second, we use an optimized weight function (duration $\tau_{\text{int}} = 400 \text{ ns}$) to integrate the homodyne signal before thresholding. This weight function consists of the difference of the averaged transients for $|0\rangle$ and for $|1\rangle$ [65, 219]. These additions achieve $\mathcal{F}_{\text{a}} = 98.8\%$, with $\epsilon_{01} = 0.1\%$ and $\epsilon_{10} = 2.3\%$ [Inset, Figure 5.1(b)], limited by T_1 .

The effect of this strong measurement on coherent operations is conveniently illustrated with AllXY measurements [98, 220]. AllXY consists of 21 sequences, two pulses each [Figure 5.1(d)], applied to the qubit followed by measurement. The pulses are drawn from the set $\{I, X, Y, x, y\}$, with I the identity, and X and Y (x and y) denoting π ($\pi/2$) pulses around the x and y axis. Ideal pulses leave the qubit in $|0\rangle$ (first 5 pairs), on the equator of the Bloch sphere (next 12), and in $|1\rangle$ (final 4), producing a characteristic two-step signature in the fidelity to $|1\rangle$, F_1 [Figure 5.1(d)]. Distinct signatures reveal errors in many gate parameters [220]. Here, we apply an extra measurement pulse ending at time τ_{d} before the AllXY pulse pair to reveal the effect of leftover photons [Figure 5.1(c)]. At $\tau_{\text{d}} \sim 7/\kappa$, the characteristic signature of moderate qubit detuning is observed. At $\tau_{\text{d}} \leq 2/\kappa$, the detuning is significant with respect to the Rabi frequency of pulses, which thus barely excite the qubit.

5.2.3 AllXY as a photon detector

To find depletion pulses we rely exclusively on optimization with Powell's method and calibrate AllXY as our photon detector. We choose $\mathcal{E}_{\text{AllXY}}$ as cost function, defined as the sum of the absolute deviations from the ideal two-step result. We find experimentally that $\mathcal{E}_{\text{AllXY}} = \alpha\bar{n}(\tau_{\text{d}}) + \beta$ for average photon numbers $\bar{n} \lesssim 30$. The calibration of coefficients α and β is

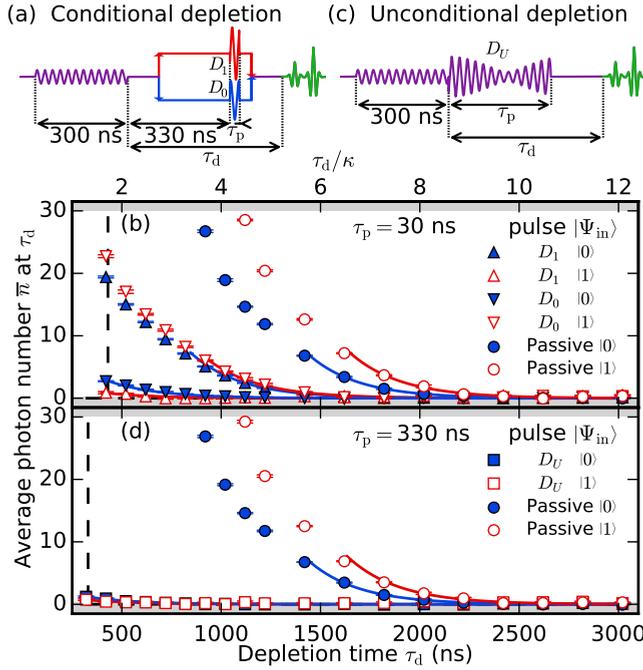


Figure 5.2.: (Color online) (a) Pulse scheme for conditional photon depletion. The controller applies a depletion pulse D_0 (at $f_{r,|0\rangle}$) or D_1 (at $f_{r,|1\rangle}$), each with separate amplitude and phase, depending on its declared measurement outcome. (b) Performance of conditional depletion. Average photon number \bar{n} as a function of τ_d for all combinations of input qubit state and depletion pulse. Compared to waiting, conditional depletion saves ~ 1250 (1800) ns for correct declaration 0 (1). (c) Pulse scheme for unconditional active depletion. The single depletion pulse D_U , immediately following the nominal measurement pulse, has four parameters corresponding to the amplitude and phase of two pulse components at $f_{r,|0\rangle}$ and $f_{r,|1\rangle}$. The summation of the two square pulse components produces the displayed beating at frequency $(f_{r,|0\rangle} - f_{r,|1\rangle})/2 = \chi/2\pi$. (d) Performance of unconditional depletion. Unconditional depletion saves ~ 1650 (1900) ns for $|0\rangle$ ($|1\rangle$). Exponential best fits (curves) to the data in the linear regime ($\bar{n} \leq 8$) give $1/\kappa = 255 \pm 5$ ns.

described in Section 5.4.2. Measurement noise limits the detector to $\delta\bar{n} \gtrsim 0.3$, providing a dynamic range of two orders of magnitude, suitable for the optimizations that follow.

5.2.4 Tune-up and comparison of two methods for active photon depletion

Our first depletion method uses a feedback controller to apply one of two depletion pulses, D_j , conditioned on the declared measurement result, $j \in \{0, 1\}$ [Figure 5.2(a)]. The pulse D_j , a square pulse of duration $\tau_p = 30$ ns, is applied at $f_{r,|j\rangle}$ by sideband modulating f_{rf} . The combined delays from round-trip signal propagation (80 ns), the augmented integration window (100 ns), and controller latency (150 ns) make D_j arrive 330 ns after the

measurement pulse ends. Each pulse is separately optimized with amplitude and phase as free parameters using a two-step procedure. We first minimize \bar{n} at $\tau_d = 1000$ ns with the qubit initialized in $|i\rangle$. This τ_d is sufficiently long to avoid saturating the detector and the sensitivity limit is reached after a few optimization rounds (further details on the optimization in Section 5.4.3). A second optimization at $\tau_d = 500$ ns further optimizes the resulting pulse and converges to $\bar{n} \sim 2.1$ (0.7) for $|0\rangle$ ($|1\rangle$), reducing τ_d by at least $5/\kappa$ compared to passive depletion [Figure 5.2(b)]. An incorrect assignment by the feedback controller leads to less effective depletion but still outperforms passive depletion.

Our second depletion method is unconditional (as in [185]), using a universal depletion pulse D_U starting immediately after the measurement pulse [Figure 5.2(c)]. To cope with the asymmetry of the nonlinear regime, we compose D_U by summing two square pulses of duration $\tau_p = 330$ ns with independent amplitude and phase at $f_{r,|0\rangle}$ and $f_{r,|1\rangle}$. These four parameters are found minimizing the sum of \bar{n} for $|0\rangle$ and $|1\rangle$, using a similar two-step procedure as for the conditional pulses (using $\tau_d = 400$ ns in the second step). This achieves $\bar{n} \sim 0.8$ (0.4) for $|0\rangle$ ($|1\rangle$) and reduces τ_d by $> 6/\kappa$ compared to passive depletion [Figure 5.2(d)]. We do not currently understand why unconditional depletion outperforms conditional depletion and why depletion for $|1\rangle$ outperforms depletion for $|0\rangle$. Numerical studies of depletion performance currently pursued outside our group [221] may soon help explain these observations and suggest other pulse parameterizations to achieve better depletion.

5.2.5 Benchmarking depletion methods with a QEC emulation: a flipping ancilla

We quantify the merits of these active depletion schemes with an experiment motivated by current efforts in quantum error correction (QEC). Specifically, we emulate an ancilla qubit undergoing the rapid succession of interleaved coherent interaction and measurement steps when performing repetitive parity checks on data qubits in a repetition code [Figure 5.3(a)]. We replace each conditional-phase gate with idling for an equivalent time (40 ns), reducing the coherent step to a 200 ns echo sequence that ideally flips the ancilla each round. As performance metric, we measure the average number of rounds to an event, $\overline{\text{RTE}}$. An event is marked by the first qubit measurement outcome deviating from the expected. Imperfections reducing $\overline{\text{RTE}}$ include qubit relaxation, dephasing and detuning during the interaction step, and measurement errors due to readout discrimination infidelity, $1 - \mathcal{F}_d$ (defined as the overlap fraction of Gaussian best fits to the single-shot readout histograms [210]).

To differentiate these sources of ancilla hardware errors, we distinguish two types of detection events, determined by the measurement outcome in the round following the first deviation (Figure 5.3(b), similar to [222]). Events of type s can result, for example, from one ancilla bit flip or from measurement errors in two consecutive rounds. In turn, events of type d can result from one measurement error or from ancilla bit flips in two consecutive rounds. Because photon-induced errors primarily lead to single bit flips, we also extract the probability of encountering an event of type s per cycle, p_s , and investigate its τ_d dependence.

Decreasing τ_d trades off T_1 -induced errors for photon-induced errors. For passive depletion, $\overline{\text{RTE}}$ is maximized to 14.6 at $\tau_d = 2200$ ns [Figure 5.3(c)]. At this optimum, depletion

occupies most of the total QEC cycle time $\tau_{\text{cycle}} = 2700$ ns. Both active depletion methods reach a higher $\overline{\text{RTE}}$ by balancing the tradeoff at lower τ_{d} . As in the optimization, we find that unconditional depletion performs best, improving the maximal $\overline{\text{RTE}}$ to 39.5 at $\tau_{\text{d}} = 700$ ns when $\bar{n} \sim 0.29$ (0.14) for $|0\rangle$ ($|1\rangle$), which reduces the optimum τ_{cycle} to 1200 ns.

The essential features of $\overline{\text{RTE}}$ for the three depletion schemes are well captured by two theory models (detailed description in Section 5.4.5). The simple model includes only qubit relaxation and non-photon-induced dephasing (calibrated using standard T_1 and $T_{2,\text{echo}}$

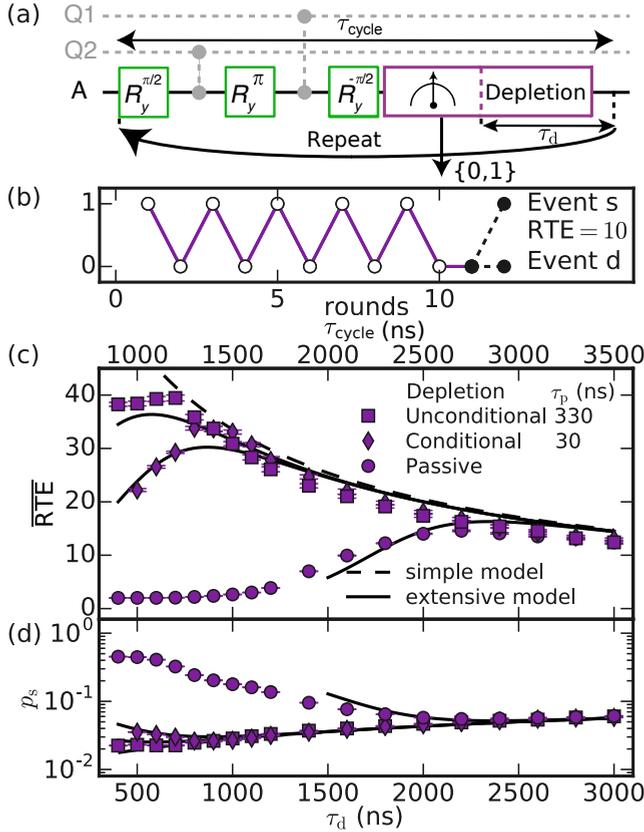


Figure 5.3.: (Color online) (a) Block diagram for parity measurements in a repetition code. The ancilla A performs an indirect measurement of the parity of data qubits q_1 and q_2 by a coherent interaction step followed by measurement. This emulation replaces conditional-phase gates by idling, reducing the coherent step to an echo sequence that ideally flips the ancilla. The measurement step is followed by a depletion step of duration τ_{d} , after which a new cycle begins. (b) Single trace of digitized measurement outcomes. The counting of rounds is ended by two types of event, s and d . (c) Average rounds to event as a function of τ_{d} . The unconditional method improves $\overline{\text{RTE}}$ by a factor 2.7 and reduces the optimum τ_{d} by a factor 2.7. (d) Per-round probability of type- s event versus τ_{d} . Added curves are obtained from the two models described in Section 5.4.5.

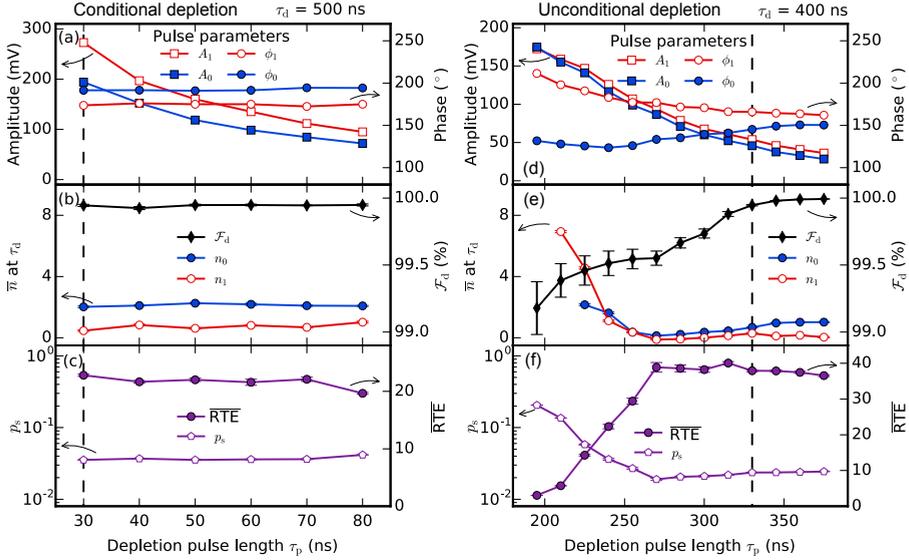


Figure 5.4.: Characterization of conditional and unconditional depletion as a function of depletion pulse length τ_p . The dashed lines indicate the pulse lengths for conditional (unconditional) depletion $\tau_p = 30$ ns ($\tau_p = 330$ ns), used in Figures 5.2, 5.3 and 5.5. All data were taken at a fixed $\tau_d = 500$ ns ($\tau_d = 400$ ns). (a) [(d)] Optimal pulse parameters after the two-step optimization protocol. (b) [(e)] Residual photon number for both qubit states and discrimination fidelity \mathcal{F}_d extracted from single shot readout histograms. (c) [(f)] Average rounds to event and per-round probability of type- s event for emulated QEC as in Figure 5.3.

measurements). The extensive model also includes photon-induced qubit dephasing and detuning during the coherent step (modeled following [223] with photon dynamics of Figure 5.2), and a measured $1 - \mathcal{F}_d = 0.1\%$ for readout. As we do not model qubit gate errors, we restrict the extensive model to $\bar{n} < 8$. The good agreement between the extensive model and experiment confirms the \bar{n} calibration and demonstrates the nondemolition character of the measurement. The conditions for nondemolition readout in the nonlinear regime have been investigated in Ref. [64].

5.2.6 Optimization of the depletion pulse length

In attempts to further shorten the depletion time we have explored depletion for various pulse lengths, finding smooth variation in optimal pulse parameters but no significant improvement of $\overline{\text{RTE}}$ (Figure 5.4). For a variety of τ_p , the optimized pulse amplitudes and phase parameters are shown, along with the residual photon number and results for multi-round QEC emulation. For conditional depletion, the optimal amplitude A_0 (A_1) of D_0 (D_1) decreases smoothly as τ_p increases, whereas the optimal phase ϕ_0 (ϕ_1) remains constant. The residual \bar{n} and readout discrimination infidelity do not show any dependence on τ_p . As expected, there is no dependence of \mathcal{F}_d on τ_p as there is no overlap between the depletion pulse and

integration window. $\overline{\text{RTE}}$ and per-round probability of type- s event for emulated QEC in the flipping configuration do not show any dependence on τ_p either. For unconditional depletion, the optimal values of the four parameters, defining the universal depletion pulse D_U , evolve smoothly as τ_p is varied. The residual \bar{n} first decreases weakly with decreasing τ_p but increases sharply for $\tau_p < 250$ ns. A smooth decrease in \mathcal{F}_d is observed for decreasing τ_p . We attribute this effect to the overlap between D_U and the measurement integration window. We note that slightly higher $\overline{\text{RTE}}$ might be achieved by implementing a short wait time between the measurement pulse and the depletion pulse to combine the lower achieved \bar{n} for $\tau_p = 270$ to 315 ns with the higher \mathcal{F}_d of the longer pulses. However, we did not explore this experimentally.

5.2.7 Benchmarking depletion methods with a QEC emulation: a non-flipping ancilla

The QEC emulations can be made more sensitive to leftover photons by harnessing the asymmetry of qubit relaxation. Specifically, we change the polarity of the final $\pi/2$ pulse ideally returning the qubit to the input state $\Psi_{\text{in}} = |0\rangle$ before measurement and depletion (results for $\Psi_{\text{in}} = |1\rangle$ are discussed in Section 5.4.4). This change removes relaxation as a source of spurious detection events. For this configuration, unconditional depletion improves $\overline{\text{RTE}}$ from 1 to 75 at a 1 μs cycle time [Figure 5.5]. For longer τ_d $\overline{\text{RTE}}$ reaches a ceiling of 168, which is set by intrinsic decoherence in the coherent step and readout discrimination infidelity. Again, unconditional depletion performs best, but the reduction of $\overline{\text{RTE}}$ at short τ_d evidences the performance limit reached by our pulses. In a QEC context, the key benefit of active depletion in this non-flipping variant will be an increase in $\overline{\text{RTE}}$ due to lower per-cycle probability of data qubit errors, afforded by reducing τ_{cycle} by $6/\kappa$. Evidently, this effect is not captured by our emulation, which is only sensitive to ancilla hardware errors. In quantum error correcting schemes, a trade-off will need to be made between shortening cycle times and increasing ancilla fidelity, especially as the different error sources contribute differently to the fidelity of an encoded logical qubit [45].

5.3 Conclusions

The RTE experiments motivate two points for discussion and outlook. First, they highlight the importance of digital feedback [63] in QEC to keep ancillas in $|0\rangle$ as much as possible (as used in a cat code [201]). Second, $\overline{\text{RTE}}$ emerges as an attractive performance metric for every element in the QEC cycle, not just the depletion. The advantage over traditional tune-up methods is the speed gained by not reinitializing in $|0\rangle$ after measurement [99] and the ability to tune without interrupting ongoing error correction [170].

In summary, we have investigated two active methods for fast photon depletion in the non-linear regime of cQED, relying on numerical optimizations to successfully outperform passive depletion by $> 6/\kappa$. Active photon depletion will find applications in quantum computing scenarios which interleave qubit measurements with coherent qubit operations. Here, we have focused on quantum error correction, emulating an ancilla qubit performing repetitive parity checks in a repetition code. Future experiments could map out the theoretically challenging

non-linear readout regime to find the optimum parameters for fast and nondemolition readout and depletion. Motivated by [64], future experiments will investigate the space of parameters (Δ , κ , g) and especially lower Δ , to pinpoint the optimal conditions for high-fidelity, nondemolition transmon readout in the nonlinear regime. Finally, combining active depletion with Purcell filtering will reduce the QEC cycle time to ~ 500 ns, sufficient to cross the error pseudothreshold in small surface codes at state-of-the-art transmon relaxation times [125].

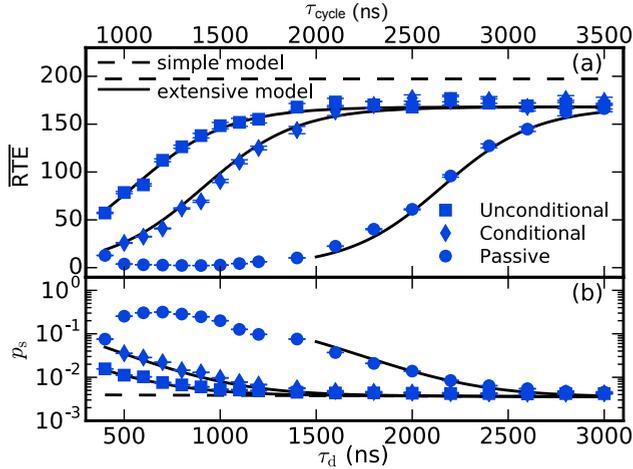


Figure 5.5.: (Color online) Emulation of repeating parity measurement for a non-flipping ancilla starting in $|0\rangle$. This variant uses the sequence of Figure 5.3(a) but with opposite polarity on the final $\pi/2$ pulse in order not to flip the ancilla. (a) $\overline{\text{RTE}}$ is no longer sensitive to qubit relaxation during τ_d and reaches a ceiling of ~ 168 set by intrinsic decoherence in the coherent step and readout discrimination infidelity. (b) Per-round probability of type- s event as a function of τ_d . Added model curves include the same calibrated errors as in Figure 5.3.

5.4 Methods

5.4.1 Experimental setup

Figure 5.6 shows the device and experimental setup, including a full wiring diagram. The chip contains ten transmon qubit-resonator pairs. All experiments presented target pair 2. The experimental setup is similar to that of previous experiments [43], but with an important addition labeled QuTech Control Box. This homebuilt controller, comprised of 4 interconnected field-programmable gate arrays (Altera Cyclone IV), has digitizing and waveform generation capabilities. The 2-channel digitizer samples with 8-bit resolution at 200 MSamples/s. The 6-channel waveform generator produces qubit and resonator pulse envelopes with 14-bit resolution at 200 MSamples/s.

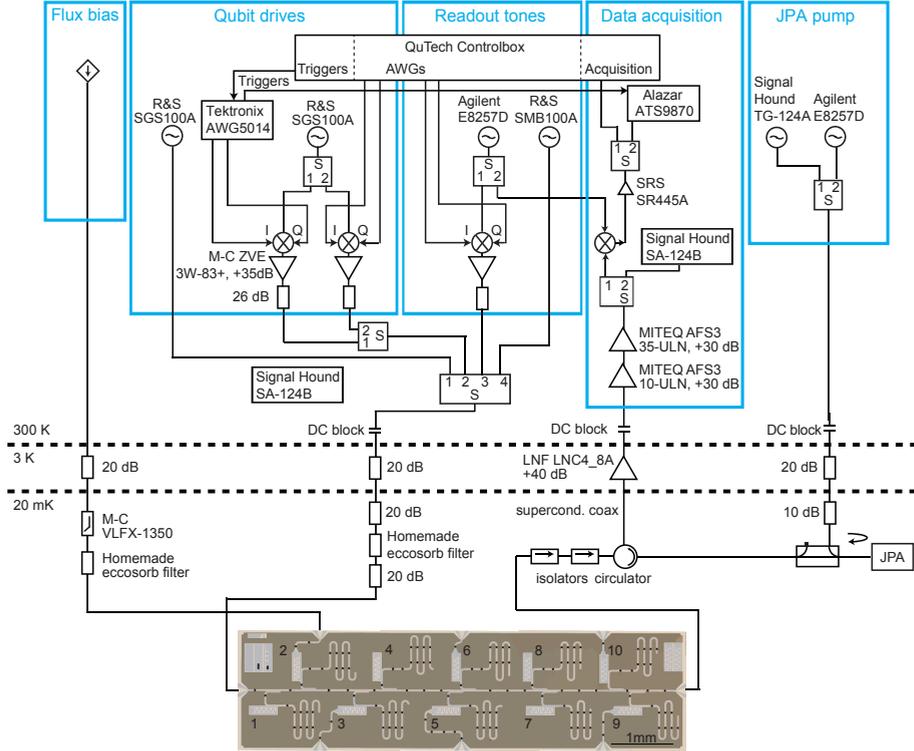


Figure 5.6.: Photograph of the cQED chip and complete wiring diagram of electronic components inside and outside the $^3\text{He}/^4\text{He}$ dilution refrigerator (Leiden Cryogenics CF-450). The chip contains ten transmon qubits individually coupled to dedicated readout resonators. All resonators couple capacitively to the common feedline traversing the chip. All data shown correspond to qubit-resonator pair 2. Dark features traversing the coplanar waveguide transmission lines are NbTiN bridges which interconnect ground planes and suppress slot-line mode propagation.

5.4.2 Photon number calibration

Figure 5.7 contains the calibration of the photon number using AllXY error ($\mathcal{E}_{\text{AllXY}}$) as a detector. $\mathcal{E}_{\text{AllXY}}$ is defined as the average absolute deviation from the ideal 2-step result in an AllXY experiment. To calibrate the detector the resonator is populated using a long (1800 ns) readout pulse with a varying pulse amplitude before measuring the AllXY. This pulse amplitude is converted to an average photon number using the single-photon power that is extracted from a photon number splitting experiment. We fit the form $\mathcal{E}_{\text{AllXY}} = \alpha \bar{n} + \beta$ to the data for each input state separately, with α and β as free parameters. The best-fit functions are used throughout the experiment to convert $\mathcal{E}_{\text{AllXY}}$ to \bar{n} .

5.4.3 Numerical optimization of depletion pulses

This paragraph further describes the optimization of depletion pulses, including the optimization ansatzes and convergence criteria. As optimization algorithm we use the implementation of Powell's method [215] in SciPy; `scipy.optimize.fmin_powell` [224].

For conditional depletion, the pulse for $|0\rangle$ ($|1\rangle$) at frequency $f_{r,|0\rangle}$ ($f_{r,|1\rangle}$) is optimized with $\mathcal{E}_{\text{AllXY}}$ as the cost function with amplitude and phase as free parameters. In the first optimization step with $\tau_d = 1000$ ns, an ansatz pulse is used with modulation envelope amplitude of $A_{0,\text{init}} = 0.035$ V ($A_{1,\text{init}} = 0.035$ V), equal to half the measurement modulation envelope amplitude, and with an initial phase of $\phi_{0,\text{init}} = 180^\circ$ ($\phi_{1,\text{init}} = 180^\circ$) with respect to the measurement pulse. After the first iteration, the phase of the pulse is varied with an initial step size of $+10^\circ$. After minimizing $\mathcal{E}_{\text{AllXY}}$ by only varying the phase, the algorithm optimizes the amplitude parameter starting with an initial step size of $+10$ mV. Then, the algorithm chooses nontrivial directions in its parameter space until one of three convergence criteria is met:

1. the iteration maximum of 300 is reached (reaching this limit indicates a failed convergence);
2. the change in both parameters is less than 0.001 times the initial step size;
3. the change in the cost function $\mathcal{E}_{\text{AllXY}}$ is less than 0.00005.

The second round of optimization at $\tau_d = 500$ ns uses the final pulse of the first optimization as its starting point and repeats the optimization algorithm with initial step sizes of 1° and $+1$ mV. Each iteration takes 12 s and each optimization step uses ~ 60 iterations to converge. The total two-step procedure takes ~ 48 minutes in total for the two pulses combined.

For unconditional depletion, the sum of $\mathcal{E}_{\text{AllXY}}$ for both input states is used as the cost function. The single 4-parameter pulse, composed by summing two square pulses at frequencies $f_{r,|0\rangle}$ and $f_{r,|1\rangle}$, is optimized starting from an ansatz pulse with amplitude and phase parameters $A_{0,\text{init}} = A_{1,\text{init}} = 0.035$ V and $\phi_{0,\text{init}} = \phi_{1,\text{init}} = 180^\circ$. Similar to the 2-parameter optimization, the algorithm starts at $\tau_d = 1000$ ns and starts the first optimization varying one parameter after the other (here, the chosen order is ϕ_0, ϕ_1, A_0, A_1). The same initial step sizes and convergence criteria are used as for conditional depletion, but now a maximum of 500 iterations is chosen. As for the conditional pulses, a second optimization round fine tunes the pulses, but because the unconditional pulse is shorter than the sum of latency and conditional pulse length, a depletion time of $\tau_d = 400$ ns is used. Each iteration takes 24 s. Each optimization step uses ~ 150 iterations to converge and the total two-step procedure takes ~ 2 hours.

5.4.4 Constant excited state QEC emulation

Figure 5.8 shows the emulated multi-round QEC for a non-flipping ancilla when the qubit is initialized in the excited state. This variant of the emulation uses the same sequence as Figure 5.5 but with the qubit initialized in $|1\rangle$. Varying τ_d , we find the optimum tradeoff between

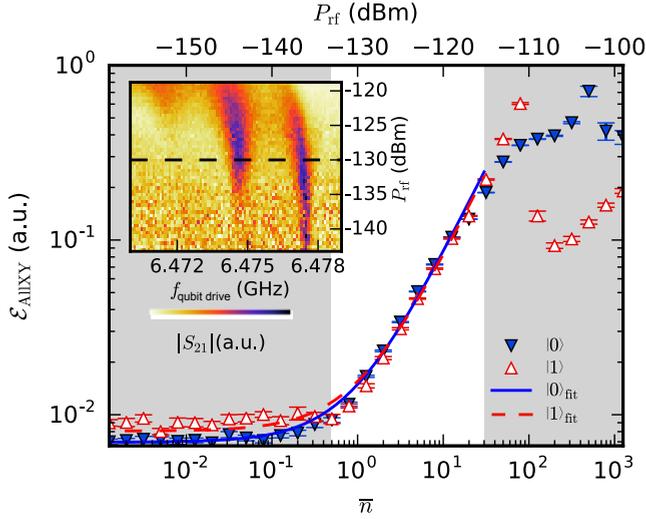


Figure 5.7.: Calibration of photon number using AllXY error. $\mathcal{E}_{\text{AllXY}}$ measured directly after a readout pulse of 1800 ns duration drives the resonator into a steady-state photon population, \bar{n} , for input states $|0\rangle$ and $|1\rangle$. The lines show a bilinear fit to the form $\mathcal{E}_{\text{AllXY}} = \alpha\bar{n} + \beta$. Inset: photon-number splitting experiment [218] used to calibrate the single-photon power level, $P_{\text{rf}} \sim -130$ dBm.

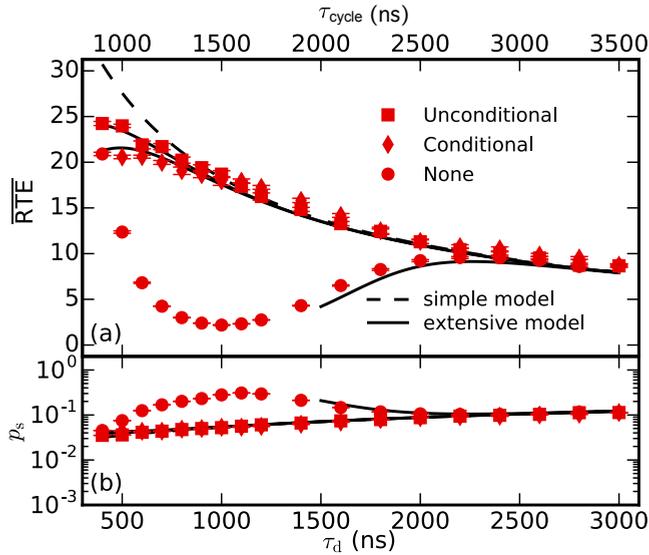


Figure 5.8.: Emulated multi-round QEC for a non-flipping ancilla in $|1\rangle$. This variant of the emulation uses the same sequence as Figure 5.5 but with the qubit initialized in $|1\rangle$. (a) Mean rounds to error detection event, $\overline{\text{RTE}}$, as a function of τ_d . (b) Per-round probability of encountering event of type s as a function of τ_d . Added curves correspond to the simple and extensive models described in Section 5.4.5.

errors induced by leftover photons and by relaxation for the three methods. Unconditional depletions performs best, increasing $\overline{\text{RTE}}$ by a factor 2.5 with respect to passive depletion. Note that passive depletion produces a spurious increase in $\overline{\text{RTE}}$ for very short τ_d . The high photon number detunes the qubit so much that qubit pulses are inoperative, causing the qubit to remain in the same state and yielding long strings of identical, expected measurement outcomes.

5.4.5 Theoretical models

We use two models to compare to data in Figures 5.3, 5.5 and 5.8 labelled simple and extensive. The simple model includes ancilla relaxation and intrinsic dephasing, providing an upper bound for the performance of the emulated multi-round QEC circuit. The extensive model further includes ancilla readout error and detuning and dephasing from the photon-induced AC Stark shift. These models use separately calibrated parameters.

The ancilla sans photon field is modeled considering amplitude and phase damping as in [2]. Single-qubit gates are approximated as 40 ns decay windows with perfect instantaneous pulses in the middle. This leads to the following scheme: $\tau_d + 20$ ns of T_1 decay, followed by a $\pi/2$ pulse, then 160 ns of $T_{2,\text{echo}}$ decay (with a π pulse in the middle), another $\pi/2$ pulse, and 20 ns of T_1 decay.

Measurement is modeled as a perfect state update S_1 , followed by a $\tau_r = 300$ ns decay window, and a second state update S_2 . The measurement signal is conditioned both on the state post- S_1 ($|\psi_i\rangle$) and post- S_2 ($|\psi_o\rangle$). If $|\psi_i\rangle = |\psi_o\rangle$ no decay occurred, and the incorrect measurement is returned with probability $1 - \mathcal{F}_d = 0.1\%$ [Figure 5.4(b)]. The only other possibility is for a single decay event (as we do not allow excitations). To zeroth order in $\tau_r/T_1 \approx 1/800$, this situation has equal probability of returning either measurement signal.

During the coherent phase, the off-diagonal elements are affected by the photon population. We model this effect following Ref. [223]:

$$\frac{d\rho^{\text{qb}}}{dt} = -i\frac{\bar{\omega}_a + B}{2}[\sigma_z, \rho^{\text{qb}}] + \gamma_1 \mathcal{D}[\sigma_-]\rho^{\text{qb}} + \frac{\gamma_{\text{CE}} + \Gamma_d}{2} \mathcal{D}[\sigma_z]\rho^{\text{qb}}. \quad (5.1)$$

Here, $\mathcal{D}[X]$ is the Lindblad operator $\mathcal{D}[X]\rho = X\rho X^\dagger - \frac{1}{2}X^\dagger X\rho - \frac{1}{2}\rho X^\dagger X$, $\gamma_1 = 1/T_1$ and γ_{CE} the pure dephasing rate [$\gamma_{\text{CE}} = (T_{2,\text{echo}})^{-1} - \frac{1}{2}T_1^{-1} = (177\mu\text{s})^{-1}$]. $\bar{\omega}_a$ is a constant rotation around the z axis of the Bloch sphere, and so is canceled by the π pulse in the coherent phase. $\Gamma_d = 2\chi\text{Im}(\alpha_0\alpha_1^*)$ is the measurement-induced dephasing, with $\alpha_{0,1}$ the qubit-state-dependent photon field amplitude and 2χ the dispersive shift per photon. This contributes a decay to the off-diagonal element of the density matrix during the coherent phase, multiplying it by

$$\exp\left[-\int \Gamma_d(t)\right], \quad (5.2)$$

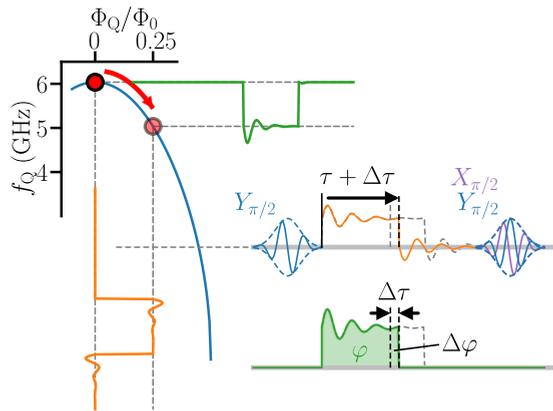
where the integral is taken over the coherent time window. $B = 2\chi\text{Re}(\alpha_0\alpha_1^*)$ is the AC Stark shift, which detunes the ancilla by an amount equal to the difference in the average photon

number over the two parts of the coherent phase. This multiplies the off-diagonal terms by a complex phase

$$\phi_{\text{Stark}} = \int_{t_A} B(t) - \int_{t_B} B(t). \quad (5.3)$$

Here, t_A and t_B are the time windows in the coherent phase on either side of the π pulse. The magnitude of the photon fields post-depletion is taken from Figure 5.2, and experiences an exponential decay at a rate that is obtained by fitting curves to the same figure. The phase difference between the fields associated with the ground and excited state grows at a rate 2χ , as extracted from Figure 5.1. As we do not model photon-induced pulse errors, we restrict our modeling to $\bar{n} < 8$, where these effects are negligible.

The experiment is simulated by storing the error-free ancilla population as a unnormalized density matrix and applying repeated cycles of the circuit. At each measurement step, the fraction of the density matrix that corresponded to an event is removed and the corresponding probability stored. The removed fraction of the density matrix is evolved for one more cycle in order to extract the event type probabilities. This is repeated until the remaining population is less than 10^{-6} .



We introduce Cryoscope, a method for sampling on-chip baseband pulses used to dynamically control qubit frequency in a quantum processor. We specifically use Cryoscope to measure the step response of the dedicated flux control lines of two-junction transmon qubits in circuit QED processors with the temporal resolution of the room-temperature arbitrary waveform generator producing the control pulses. As a first application, we iteratively improve this step response using optimized real-time digital filters to counter the linear-dynamical distortion in the control line, as needed for high-fidelity, repeatable one- and two-qubit gates based on dynamical control of qubit frequency.

6.1 Introduction

In many solid-state quantum information platforms, accurate dynamical control of qubit frequency is key to realizing single- and two-qubit gates. Common on-chip control variables include, but are not limited to, voltage on a local gate and magnetic flux through a SQUID loop. For example, voltage control is typically used for spin qubits [226–229] and gatemons [204, 230], while flux control is ubiquitous for transmon, flux and fluxonium superconducting qubits [27]. In most cases, the input control signal originates at an arbitrary waveform generator (AWG) operating at room temperature. The signal suffers linear dynamical distortions as it traverses various electrical components on the control line connecting to the quantum device, most often lying at the coldest stage of a dilution refrigerator.

If uncompensated, such distortions can have detrimental effects on gate performance, affecting fidelity and even repeatability. A salient example is the controlled-phase (CZ) gate between two transmon qubits implemented by a baseband flux pulse [26] that brings the computational state $|11\rangle$ temporarily near resonance with the non-computational state $|02\rangle$. Short-timescale distortions of the meticulously shaped flux pulse [167] can produce leakage away from the two-qubit computational subspace, leaving remnant population in $|02\rangle$. Meanwhile, long-timescale distortions make the unitary action of a flux pulse depend on the history of flux pulses applied [91, 112]. As leakage and history dependence severely limit the depth of quantum circuits that can be realized, a practical scheme for characterization and correction of pulse distortion on chip is of paramount importance.

Distortions introduced by components at room temperature (e.g., AWG bandwidth, high-pass filtering of a bias tee, skin effect in instrumentation cable) are easily characterized with a fast oscilloscope. However, distortions introduced by components inside the refrigerator (e.g., low-pass filters, impedance mismatch, skin effect in semi-rigid coaxial cable, chip packaging [231]) are generally temperature-dependent and are thus best characterized in the cold. Additionally, the on-chip response varies across devices and even between different qubits on the very same device. Evidently, the ideal strategy for characterizing pulse distortion is to use the controlled qubit itself.

A traditional method to visualize the dynamical distortion of ideally square pulses is to observe the oscillations in the excited-state population (as a function of pulse amplitude and duration) when pulsing the qubit into near resonance with another exchange-coupled qubit or a continuous drive tone. While the distortions can be gleaned from the deviation from the ideal chevron pattern [91], the inversion is challenging. More direct methods use spectroscopy [176] and Ramsey experiments [232] to measure the qubit frequency dynamics, but only during the turn-off transients following a square pulse. Most recently, a method combining continuous microwave and flux drives was developed to convert a transmon into a vector network analyzer [233] giving the frequency response of the flux control line, from which it is possible to calculate the qubit frequency dynamics for a given pulse.

In this Letter, we present Cryoscope (short for cryogenic oscilloscope), an in-situ technique using the qubit to sample control pulses of arbitrary shape at the temporal resolution of the AWG. We specifically demonstrate Cryoscope for two-junction transmon qubits, whose

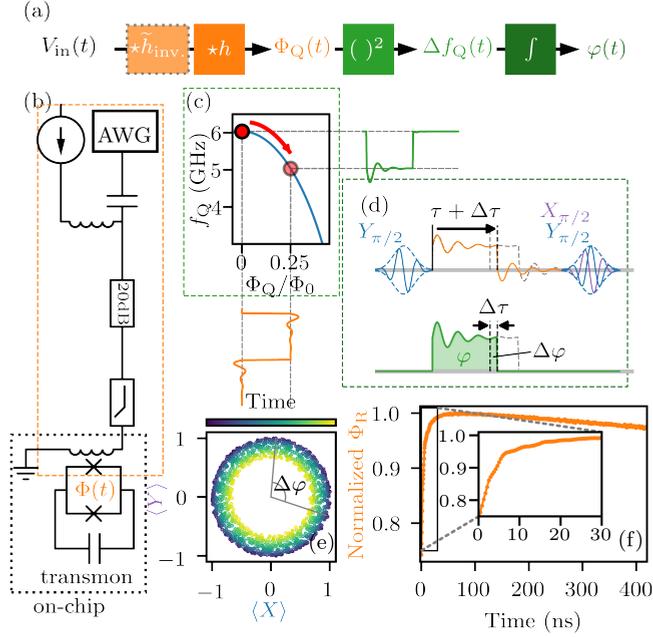


Figure 6.1.: Basic concept of Cryoscope. (a) Overview of relevant transformations involved. (b) Schematic of the control line used to control the flux Φ_Q through the transmon SQUID loop. A DC source and AWG combined at a bias tee at room temperature produce the static and dynamic components of Φ_Q . (c) When operating Cryoscope, the transmon is biased at its flux sweetspot and pulsed away only during the waiting interval between the $\pi/2$ pulses in a standard Ramsey-style experiment. (d) The difference in quantum phase $\Delta\varphi$ [shown in (e)] acquired by the qubit during Ramsey experiments with the flux pulse truncated after τ and $\tau + \Delta\tau$ provides an estimate of the instantaneous qubit detuning Δf_Q in the interval $[\tau, \tau + \Delta\tau]$, and consequently an estimate Φ_R of the instantaneous actual flux Φ_Q . The nonlinear dependence of $\Delta f_Q(\Phi_Q)$ suppresses the error produced by the difference of the two turn-off transients. (f) Reconstructed step response of the control line, normalized to maximal flux.

frequency depends quadratically (to a good approximation) on the flux through the constituent SQUID loop. However, Cryoscope is generally applicable to any system with quadratic or higher power dependence of qubit frequency on the control variable and a sweetspot where qubit frequency is at least first-order insensitive to this variable. As a first application, we use Cryoscope to iteratively measure the voltage-to-flux step response and apply predistortion corrections to the control waveforms. We predistort the waveforms digitally using finite- and infinite impulse response filters applied in real time, i.e., without precompilation of the waveform, in a manner compatible with codeword-based microarchitectures [52, 53] and feedback control. We consistently find the reconstructed step response to be within $\sim 0.1\%$ of the ideal response in several setups and devices.

6.2 Concept

The transition frequency f_Q of a two-junction transmon depends on the magnetic flux $\Phi_Q(t)$ through its SQUID loop and for symmetric junctions is given by [76]

$$f_Q(\Phi_Q) \approx \frac{1}{h} \left(\sqrt{8E_J E_C \left| \cos \left(\pi \frac{\Phi_Q}{\Phi_0} \right) \right|} - E_C \right), \quad (6.1)$$

where E_C is the charging energy, E_J is the sum of the Josephson energies of the individual junctions, Φ_0 is the flux quantum, and h is Planck's constant. In our system, the static and dynamic components of Φ_Q are produced by a DC source and an AWG, respectively, and combined at a bias tee, all at room temperature. Here, we use the DC source to null flux offsets, biasing the transmon at its maximal frequency, $f_{\max} \approx \frac{1}{h} \sqrt{8E_J E_C} - E_C$, which functions as a sweetspot with first-order insensitivity to Φ_Q . As in typical applications [61, 67, 91, 112], we use the AWG to flux pulse the transmon to detunings $\Delta f_Q(t) = f_{\max} - f_Q(\Phi_Q(t))$ up to ~ 1 GHz, corresponding to $\sim 0.25\Phi_0$.

At its core, Cryoscope is a technique using Ramsey-style experiments to obtain an estimate $\Phi_R(t)$ of the actual $\Phi_Q(t)$ produced by an AWG pulse $V_{\text{in}}(t)$. We embed the flux pulse (with varying truncation of the input) between the two $\pi/2$ pulses, which are always separated by a fixed interval T_{sep} . The first $\pi/2$ pulse (around the y axis of the Bloch sphere) prepares the qubit in the superposition state $(|0\rangle + |1\rangle)/\sqrt{2}$. An AWG pulse $V_{\text{in},\emptyset}(t)$ truncated at time τ produces a flux $\Phi_{Q,\tau}(t)$ that transforms the state to $(|0\rangle + e^{i\varphi_\tau} |1\rangle)/\sqrt{2}$, with relative quantum phase

$$\varphi_\tau/2\pi = \int_0^\tau \Delta f_Q(\Phi_{Q,\tau}(t)) dt + \int_\tau^{T_{\text{sep}}} \Delta f_Q(\Phi_{Q,\tau}(t)) dt, \quad (6.2)$$

where we explicitly separate the contributions from the flux response up to the truncation point and the subsequent turn-off transient. We complete the Ramsey experiment with two variants, one with the final $\pi/2$ rotation around y and another with it around x before measuring in order to determine the Bloch vector components $\langle X \rangle$ and $\langle Y \rangle$ from which we extract φ_τ .

We estimate $\Phi_Q(t)$ in the small time interval $[\tau, \tau + \Delta\tau]$ using the following procedure. First, we measure φ_τ and $\varphi_{\tau+\Delta\tau}$ to compute

$$\overline{\Delta f_R} \equiv \frac{\varphi_{\tau+\Delta\tau} - \varphi_\tau}{2\pi\Delta\tau} = \frac{1}{\Delta\tau} \int_\tau^{\tau+\Delta\tau} \Delta f_Q(\Phi_{Q,\tau+\Delta\tau}(t)) dt + \varepsilon, \quad (6.3)$$

which gives the average detuning $\overline{\Delta f_Q}$ during the interval, with inaccuracy

$$\varepsilon = \frac{1}{\Delta\tau} \left(\int_{\tau+\Delta\tau}^{T_{\text{sep}}} \Delta f_Q(\Phi_{Q,\tau+\Delta\tau}(t)) dt - \int_\tau^{T_{\text{sep}}} \Delta f_Q(\Phi_{Q,\tau}(t)) dt \right), \quad (6.4)$$

given by the difference in the phase contributions of the turn-off transients.

The phase contribution from the turn-off transients is suppressed because of the steep return to the first-order flux-insensitive sweetspot of the nearly quadratic $\Delta f_Q(\Phi_Q)$. Numerical simulations indicate that $|\varepsilon|/\overline{\Delta f_R} \lesssim 10^{-2}-10^{-3}$ for dynamical distortions of typically

used components [67, 91]¹. This suppression of $|\varepsilon|/\overline{\Delta f_R}$ would improve for higher order of nonlinearity in $\Delta f_Q(\Phi_Q)$.

Finally, we obtain the reconstructed $\Phi_R(t)$ by inversion of Equation (6.1). The ability of Cryoscope to reconstruct pulses of arbitrary shape is shown in Section 6.4.5 for the case of a pulse shaped as a traditional Dutch canal skyline.

We briefly discuss some technical aspects of the implementation. We set $\Delta\tau = 1/2.4 \text{ ns}$, the minimum allowed by the sampling rate of the AWG (Zurich Instruments HDAWG). The separation time T_{sep} is set 100 ns longer than the largest chosen τ to negate the need for fine timing calibrations and to be less sensitive to residual detuning during the final rotation. The phase φ_τ is determined by combining the $\langle X \rangle$ and $\langle Y \rangle$ data. Before unwrapping the phase it is demodulated using the highest frequency component of a Fourier transform of the $\langle X \rangle$ and $\langle Y \rangle$ data. A second-order Savitzky-Golay filter is then used to determine the derivative by fitting a polynomial in a small window around each data point. The estimated detuning $\overline{\Delta f_R}$ is a sum of the frequency extracted using the Savitzky-Golay filter, the demodulation frequency and, when using large flux pulse amplitudes, appropriate multiples of the 1.2 GHz Nyquist frequency. The Nyquist order can be determined by acquiring Cryoscope traces for square pulses with different amplitudes and observing when the mean frequency wraps as the pulse amplitude is increased. Because distortions can cause the instantaneous detuning to be slightly lower or higher than the mean detuning, amplitudes close to the Nyquist wrapping should be avoided.

6.2.1 Correcting distortions

As a first demonstration of Cryoscope, we measure the voltage-to-flux step response $s(t)$ of the control line. The result shown in Figure 4.1(f) reveals clear deviations from the ideal, with dynamics on timescales comparable to typical pulse durations ($\sim 40 \text{ ns}$) and much longer. These dynamics are the result of compounded linear dynamical distortions and thus can be described by convolution of the input $V_{\text{in}}(t) = V_0 \cdot u(t)$ (where $u(t)$ is the Heaviside step function) with the system impulse response h , $\Phi_Q(t) = h \star V_{\text{in}}(t)$. We furthermore assert that the system is causal so that $s(t) = 0$ for $t < 0$.

As an application of Cryoscope, we make iterative use of real-time digital filtering (available in the AWG) and Cryoscope to improve the step response. The goal of this procedure is to determine the filter $h_{\text{flt}} = \tilde{h}_{\text{inv}}$, that best inverts h such that the corrected step response $s_{\text{corr}}(t) = h_{\text{flt}} \star s(t)$ approximates $u(t)$ as close as possible.

First, several first-order infinite impulse response (IIR) filters are applied to eliminate dynamics on timescales longer than 30 ns. The IIR filters are designed to each correct a step response of the form $s(t) = g(1 + Ae^{-t/\tau_{\text{IIR}}}) \cdot u(t)$, where A is the amplitude coefficient, τ_{IIR} is the time constant of the filter and g is a gain constant. The coefficients of the filters are determined by performing a least-squares optimization of a prediction of $s_{\text{corr}}(t)$ based on a model of the IIR filters and the measured $s(t)$. Because the IIR filters are applied in real-time on the hardware, there are small differences between the ideal filter and the implementation

¹ $\frac{|\varepsilon|}{\overline{\Delta f_R}}$ can be slightly larger for certain idealized filters such as a single-pole low-pass filter..

which are taken into account in the model. We typically require 3–5 such IIR filters in order to correct $s(t)$ between 30 – 200 ns. Cryoscope is used to evaluate the corrections of the IIR filters [Figure 6.2(a)] and shows a reconstruction in which the slow dynamics are corrected to within $\sim 0.1\%$.

Next, a finite impulse response (FIR) filter is used to correct for the remaining short (< 30 ns) timescale dynamics. The FIR filter is described by 40 parameters that in turn describe the 72 coefficients (30 ns) of the filter. The values are found by minimizing the distance between the predicted signal and the ideal step response using the CMA-ES algorithm [234]. A third Cryoscope measurement is performed to test the accuracy of the corrections. This final step can be used to iteratively fine tune the FIR coefficients if required. No such iterations were required to achieve a reconstructed step response accurate to $\sim 0.1\%$ shown in Figure 6.2(a).

6.2.2 Verifying distortion corrections

To independently characterize the corrections, we perform a chevron experiment without and with the predistortions applied [Figure 6.2(b,c)]. In this experiment, two qubits (q_1 and q_0) are prepared in the $|11\rangle$ state using π pulses, a square flux pulse of varying duration and amplitude is applied to the higher frequency qubit (q_0) to tune $|11\rangle$ into (near) resonance with $|02\rangle$, the same interaction that is exploited to realize a CZ gate. With no predistortions applied [Figure 6.2(b)], the pattern of q_1 population as a function of pulse amplitude (horizontal axis) and duration (vertical axis) is visibly asymmetric – fringes on the right-hand side are more visible, and the pattern bends towards large pulse amplitudes for short pulse durations. These two features are signatures of the finite rise time of the applied pulse. In contrast, when predistortions are applied [Figure 6.2(c)], the pattern is almost perfectly left-right symmetric, both in terms of visibility and shape, indicating a near-perfect rectangular pulse. Using Cryoscope, we can predict the pulse amplitude that results in exact $|11\rangle$ – $|02\rangle$ degeneracy at every point in time. The prediction [white curve in Figure 6.2(b, c)] overlaps with the path along which the oscillations are slowest, providing an independent verification (although less quantitative) of our method.

6.2.3 Cryoscope sensitivity

Having established the ability to measure and correct distortions, we investigate the sensitivity of Cryoscope. Figure 6.3(a) presents the unprocessed measurement of $\langle X \rangle$ for three values of qubit detuning during the rectangular pulse. In all cases we observe decaying oscillations. The decay is faster the larger the pulse amplitude due to reduced coherence of the qubit further away from sweetspot. The reconstructed instantaneous flux in a 100 – 200 ns window [Figure 6.3(b,c)] fluctuates around the mean value, in a range decreasing with the amplitude of the rectangular pulse. We interpret that for larger detuning the qubit precession is faster, resulting in a larger phase acquired between subsequent time steps and consequently yielding a more accurate measurement of the instantaneous detuning relative to nearly the same sampling noise.

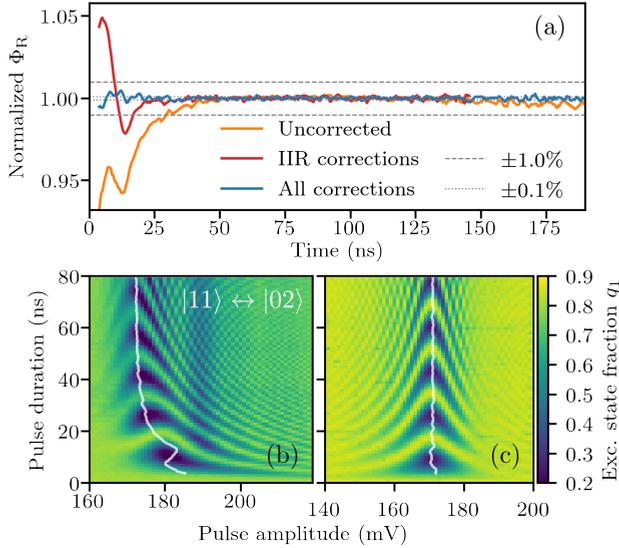


Figure 6.2.: Reconstructed step response without and with distortion corrections (for a qubit on a different device from that of Figure 4.1) normalized to flux between 40 and 125 ns. (a) Cryoscope measurements of uncorrected (orange) and corrected step responses with IIR corrections only (red) and FIR and IIR corrections (blue). (b-c) Chevron experiments without and with predistortion corrections (not corrected for readout error). The overlaid curve indicates resonance between $|11\rangle$ and $|02\rangle$, predicted using the step response reconstructed with Cryoscope. See text for details.

We define a signal-to-noise ratio to quantify the influence of dephasing and precession rate on Cryoscope sensitivity,

$$\text{SNR} = \frac{\overline{\Phi_R}}{s_{\Phi_R}}. \quad (6.5)$$

We define as signal the mean amplitude of the optimally corrected, reconstructed flux $\overline{\Phi_R}$ and as noise the standard deviation s_{Φ_R} . The SNR is experimentally determined for several time windows and amplitudes of the rectangular flux pulse [Figure 6.3(d)]. We perform 10 Cryoscope experiments for every data point to extract $\overline{\Phi_R}$ and s_{Φ_R} in the relevant time interval. In the 100 – 200 ns window, SNR increases quadratically with pulse amplitude, indicating that detuning increases, while the qubit coherence is not affected on this short timescale. In contrast, the increase of SNR is slower for the other time windows. In particular, for the 1200 – 1300 ns window, the SNR reaches a maximum for pulse amplitude $\Phi_Q \approx 0.17 \Phi_0$. The maximum indicates the configuration in which the benefit of increased precession rate balances out the drawback of the reduced qubit coherence (due to increased sensitivity to flux noise).

The SNR is also affected by acquisition and processing parameters. The former is the precision with which the qubit state can be determined, which depends on the number of

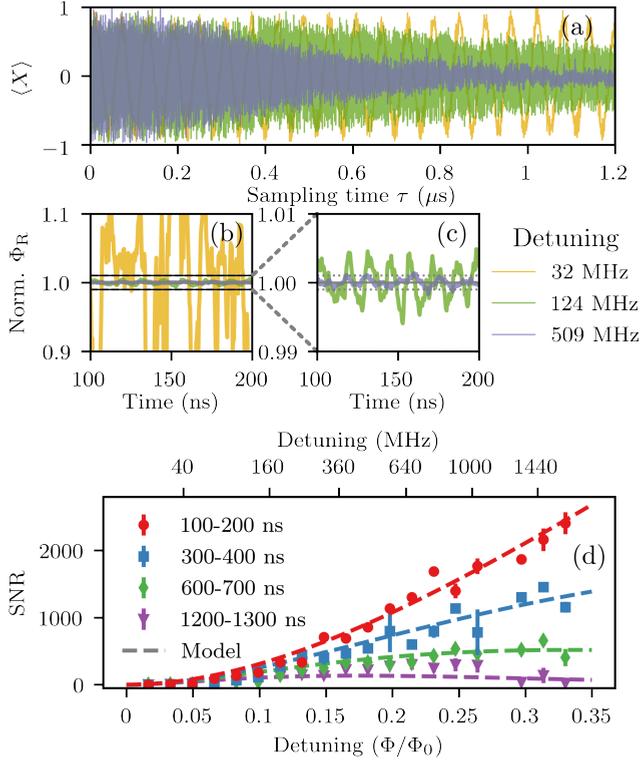


Figure 6.3.: Cryoscope signal-to-noise ratio. (a) Raw measurements of $\langle X \rangle$ for individual Cryoscope traces using different detuning. (b, c) Zoom of reconstructed signal (normalized to the mean flux). The dotted curves denote deviations of 0.1%. (d) SNR at various timescales and detunings.

averages and the readout fidelity. The latter is a matter of applied data filtering and can be adjusted depending on the temporal resolution demanded.

All these factors can be combined in a model yielding

$$\text{SNR} = c\Phi_Q^2 \exp(-(\Gamma_0 + 2a\Gamma_1\Phi_Q)t), \quad (6.6)$$

where t is the time of reconstruction, c accounts for sampling noise and filtering effects in data processing, Γ_0 is a sweetspot dephasing rate, Γ_1 quantifies the power of flux noise and the qubit detuning from sweetspot is $\Delta f_Q(\Phi_Q) = a\Phi_Q^2$. The interplay between quadratic and exponential terms in Φ_Q represents the trade-off between improved sensitivity to the shape of flux pulse versus reduced signal visibility due to dephasing. The prefactor c can be increased by averaging more or alternatively improving the readout fidelity. We fix values of a and Γ_0 based on independent measurements and perform a fit of the two-parameter model (c and Γ_1), finding a good agreement with the data [Figure 6.3(d)].

6.3 Conclusions

In conclusion, we have demonstrated a method capable of sampling on-chip flux pulses by exploiting the nonlinear flux dependence of transmon frequency. This characterization method is straightforward to use and generalizable to any qubit system with baseband control of the qubit frequency and a sweetspot with respect to the control variable. Furthermore, we have demonstrated the capability to correct distortions as demonstrated by a reconstructed step response accurate to $\sim 0.1\%$. The identified corrections were applied in real time, making the correction method compatible with an instruction-based control architecture [52, 53]. Cryoscope has already been used to tune-up fast, high-fidelity, and low-leakage CZ gates for a QEC experiment [61, 112] and parametrized iSWAP interactions in a variational quantum eigensolver [67, 74].

6.4 Methods

This section provides experimental details and derivations supporting claims made in the main text. First, we describe the experimental setup. We then discuss the limitations of the Cryoscope, showing how undesired distortions are suppressed for a typical step response and how the nonlinear response of the qubit to flux helps in reconstructing the step response. The third section details a simple model that describes the signal-to-noise ratio of the experiment. Next, we provide details on the hardware implementations of the FIR and IIR filters used to correct distortions in real time. Finally, we provide experimental data demonstrating the ability to use Cryoscope to reconstruct an arbitrary signal.

6.4.1 Device and experimental setup

The data shown in this letter were acquired using two devices mounted in different dilution refrigerators. In all experiments, a Zurich Instruments HDAWG equipped with real-time digital filters was used to generate the flux pulses. The output of the AWG was connected to the RF port of a Mini-Circuits ZFBT 6GW+ bias tee while the DC port was connected to a DC current source. The RF+DC port of the bias tee was connected to the flux control line entering the fridge. The flux control line contains a 20 dB attenuator at the 4 K stage as well as a Mini-Circuits VLFX1050 low-pass filter and a homebuilt eccosorb filter before being connected to the flux control line on the device. The control-line coaxial cables between 4 K and mixing chamber plate were superconducting (NbTi, inner and outer conductor) for one device and stainless steel (inner and outer) for the other. In all cases it was possible to correct distortions to within $\sim 0.1\%$.

6.4.2 Limitations of the Cryoscope

In this section we first investigate the accuracy of the Cryoscope for a physically motivated step response including distortions due to AWG bandwidth, bias tee, skin effect, and on-chip response. We show that the inaccuracy is small using typical distortion parameters. Next, we

investigate the effects of a single-pole low-pass filter for which the error in the reconstruction is significant on the timescale of the filter.

Our analysis is based on a numerical calculation of the acquired relative phase φ_τ , yielding a noiseless Cryoscope measurement. Specifically,

$$\varphi_\tau = 2\pi \int_0^\infty a [\Phi_s (s(t) - s(t - \tau))]^k dt, \quad (6.7)$$

where Φ_s is the amplitude of the applied square flux pulse, $s(t)$ is the step response and a parametrizes the dependence between magnetic flux Φ and qubit detuning in the quadratic approximation: $\Delta f_Q = a\Phi^k$ with $k = 2$. Because the phases are calculated in simulation it is possible to extract the contributions to $\overline{\Delta f_R}$:

$$\begin{aligned} \overline{\Delta f_R}(\tau, \Delta\tau) &\equiv \frac{\varphi_{\tau+\Delta\tau} - \varphi_\tau}{2\pi\Delta\tau} \\ &= \frac{1}{\Delta\tau} \underbrace{\int_\tau^{\tau+\Delta\tau} \Delta f_Q(\Phi_{Q,\tau+\Delta\tau}(t)) dt}_{\overline{\Delta f_Q}} \\ &\quad + \underbrace{\frac{1}{\Delta\tau} \int_{\tau+\Delta\tau}^{T_{\text{sep}}} \Delta f_Q(\Phi_{Q,\tau+\Delta\tau}(t)) dt}_{\varepsilon_{\tau+\Delta\tau, \text{Off}}} - \underbrace{\frac{1}{\Delta\tau} \int_\tau^{T_{\text{sep}}} \Delta f_Q(\Phi_{Q,\tau}(t)) dt}_{\varepsilon_{\tau, \text{Off}}}. \end{aligned} \quad (6.8)$$

Cryoscope reconstruction of a typical step response

An overview of the distortion models used can be found in Table 6.1. The response of the HDAWG is taken into account by performing a convolution with an impulse response extracted from the measured step response. This step response was measured when the HDAWG was operated in amplified mode and shown in Figure 6.4. The step response of the bias tee is modeled as a single exponential high-pass filter of the form $s(t) = e^{-t/\tau_{\text{HP}}} \cdot u(t)$ and several exponential filters of the form $s(t) = 1 + Ae^{-t/\tau_{\text{IR}}} \cdot u(t)$, where τ_{IR} and τ_{HP} are the relevant time constants, A is an amplitude coefficient and $u(t)$ is the Heaviside step function. The coefficients used are based on a measured step response for a bias tee and are the same as in Ref. [67]. We note that the coefficients are known to vary slightly between different bias tees of the same model. The skin effect is modeled according to Ref. [235] with an attenuation of $\alpha_{\text{GHz}} = 2.1$ dB at 1 GHz. The signal is filtered with a Savitsky-Golay filter in order to determine the derivative and the impulse response.

We model the effect of distortions on square pulses (truncated at time τ and $\tau + \Delta\tau$) that detune the qubit by $\Delta f_Q = 800$ MHz Figure 6.5. The contribution of the individual turn-off transients $\varepsilon_{\tau, \text{Off}}$ and $\varepsilon_{\tau+\Delta\tau, \text{Off}}$ is typically significantly larger than $\overline{\Delta f_Q}$ as it takes multiple $\Delta\tau$ for the qubit to return to the sweetspot. However, their difference ε is smaller;

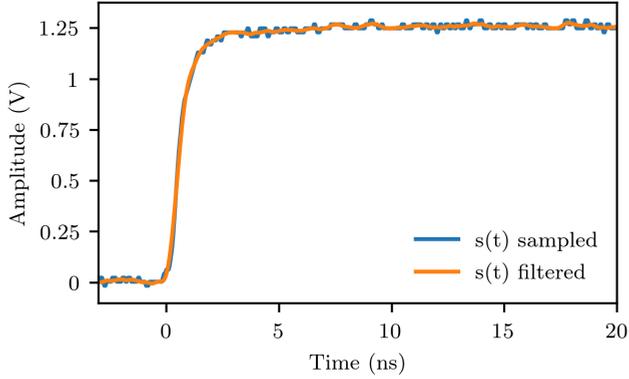


Figure 6.4.: Measured step response of the HDAWG in amplified mode, measured using a Rohde & Schwarz RTO1024 oscilloscope (blue). The signal is filtered with a Savitsky-Golay filter (orange) in order to determine the impulse response from the derivative.

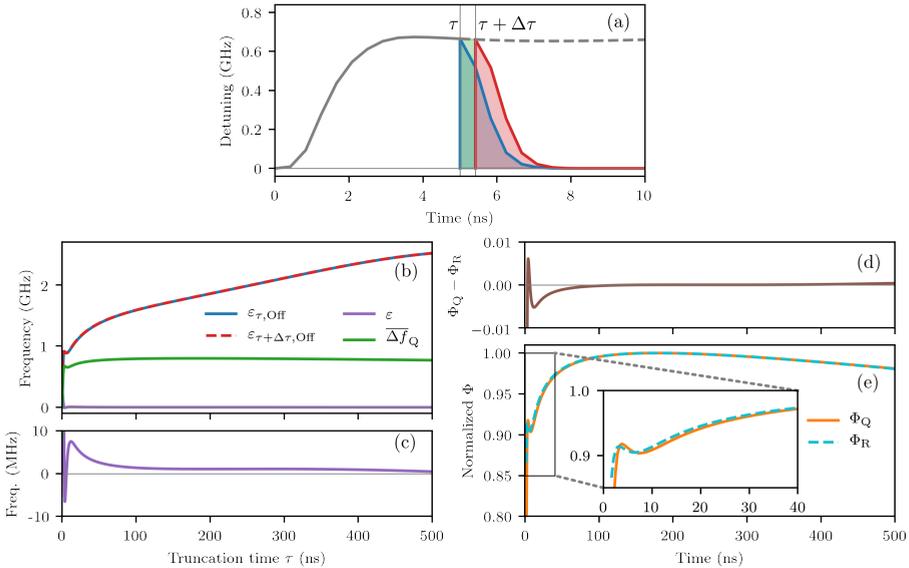


Figure 6.5.: Simulated reconstruction of a typical step response (cumulative effect of models of Table 6.1) using Cryoscope. (a) Detuning of the qubit when applying square pulses truncated at τ and $\tau + \Delta\tau$. Shaded areas illustrate the contributions to $\overline{\Delta f_R}$ from $\overline{\Delta f_Q}$ (green), $\varepsilon_{\tau, \text{Off}}$ (blue) and $\varepsilon_{\tau + \Delta\tau, \text{Off}}$ (red). (b, c) Contributions to $\overline{\Delta f_R}$ from the true detuning (green), the individual turn-off transients (blue and red), and the difference of turn-off transients (purple). (e, f) Comparison of the reconstructed flux $\Phi_R(t)$ to the true flux $\Phi_Q(t)$.

Effect	Model	Model parameters	Notes
AWG response	$\star h_{\text{AWG}}$	-	Measured step response of HDAWG in amplified mode (Figure 6.4)
Bias tee	$s(t) = (e^{-t/\tau_{\text{HP}}}) \cdot u(t)$	$\tau_{\text{HP}} = 41 \mu\text{s}$	Mini-Circuits ZFBT-6GW+
Bias tee	$s(t) = (1 + A \cdot e^{-t/\tau_{\text{TIR}}}) \cdot u(t)$	$\tau_{\text{TIR}} = 15 \mu\text{s}$ $A = 0.13$	Ref. [67]
Bias tee	$s(t) = (1 + A \cdot e^{-t/\tau_{\text{TIR}}}) \cdot u(t)$	$\tau_{\text{TIR}} = 6.4 \mu\text{s}$ $A = 0.99$	Ref. [67]
Skin effect	$s(t) = (1 - \text{erfc}(\alpha_{\text{GHz}}/21\sqrt{t})) \cdot u(t)$	$\alpha_{\text{GHz}} = 2.1 \text{ dB}$	Model according to Ref. [235]
On-chip response	$s(t) = (1 + A \cdot e^{-t/\tau_{\text{TIR}}}) \cdot u(t)$	$\tau_{\text{TIR}} = 2 \text{ ns}$ $A = 0.6$	

Table 6.1.: Overview of the distortion models used and their coefficients. Figure 6.5 illustrates the cumulative influence of all listed effects and their reconstruction using Cryoscope.

with the exception of the first few samples, $\epsilon \lesssim 8 \text{ MHz} \approx \overline{\Delta f_Q} \cdot 10^{-2}$ [Figure 6.5(b, c)]. We next use the acquired phases as input for the Cryoscope analysis to obtain the reconstructed flux $\Phi_R(t)$ and compare it to the true flux $\Phi_Q(t)$ [Figure 6.5(d, e)]. We observe a matching of $\Phi_R(t)$ to $\Phi_Q(t)$ better than 1% for $t > 3 \text{ ns}$. Note that the data shown in Figure 6.5(d, e) is normalized to the maximal flux.

Cryoscope reconstruction of a single-pole low-pass filter step response

For completeness, we demonstrate that Cryoscope may reconstruct the step response poorly for specific filters. A simple example is the single-pole low-pass filter, whose step response is

$$s_{\text{LP}}(t) = (1 - e^{-t/\tau_{\text{LP}}}) \cdot u(t), \quad (6.9)$$

where τ_{LP} is the time constant. Such a filter does not accurately represent our setup but is easy to describe analytically and therefore is a good choice to demonstrate the origin of potential errors and to show the relevance of the nonlinear qubit response to flux. We find that the reconstructed step response $s_{R,\text{LP}}(t)$ differs from $s_{\text{LP}}(t)$ by more than 1% for $t \lesssim 4\tau_{\text{LP}}$.

We also use this simple example to show that Cryoscope is more accurate for higher degrees of nonlinearity. Specifically, we calculate $s_{R,\text{LP}}(t)$ for different forms of qubit detuning on flux: $\Delta f(\Phi) = a\Phi^k$ where $k \in \mathbb{Z}^+$.

In general, the phase φ_τ (setting $T_{\text{sep}} = \infty$) expressed in terms of the impulse response $h = ds/dt$ is

$$\begin{aligned} \varphi_\tau &= 2\pi a \int_0^\infty \left[\int_0^\infty h(t-t') dt' - \int_0^\infty h(t-\tau-t') dt' \right]^k dt \\ &= 2\pi a \int_0^\tau \left[\int_0^t h(t-t') dt' \right]^k dt + 2\pi a \int_\tau^\infty \left[\int_0^\tau h(t-t') dt' \right]^k dt, \end{aligned} \quad (6.10)$$

while the reconstructed step response is given by

$$s_R(\tau) = \left(\frac{2\pi}{a} \frac{d\varphi}{d\tau} \right)^{1/k} = \left(\int_\tau^\infty \left\{ h(t-\tau) \left[\int_0^t h(t-t') dt' \right]^{k-1} \right\} dt \right)^{1/k}. \quad (6.11)$$

For the single-pole low-pass filter,

$$s_{R,\text{LP}}(t) = \left[1 - e^{-t/\tau_{\text{LP}}} \right]^{\frac{k-1}{k}} u(t) = \begin{cases} (1 - e^{-t/\tau_{\text{LP}}}) \cdot u(t) & \text{for } k \rightarrow \infty \\ u(t) & \text{for } k = 1 \end{cases} \quad (6.12)$$

We observe that, $s_{R,\text{LP}}$ matches s_{LP} in the limit $k \rightarrow \infty$. On the other hand, for $k = 1$ (i.e. linear dependence of qubit frequency on flux) the reconstruction gives $u(t)$, completely missing the dynamics.

By directly comparing Equation (6.9) to Equation (6.12) we can place an upper bound on the Cryoscope inaccuracy:

$$0 \leq s_{R,LP}(t) - s_{LP}(t) \leq \frac{1}{k} e^{-t/\tau_{LP}}. \quad (6.13)$$

The factor $1/k$ clearly shows that the nonlinear dependence on flux is essential for Cryoscope to work and that a higher power dependence increases its accuracy.

The reduced inaccuracy with higher order k can intuitively be understood by considering that ε is determined by contributions during the turn-off transients (see Equation (6.8)). In the case of a linear dependence, i.e., $k = 1$, contributions during the turn-off transients are weighted as strong as contributions during the pulse, making it impossible to reconstruct the waveforms. With a parabolic dependence, i.e., $k = 2$, the contributions during the off-transients are suppressed. The higher the order k , the more strongly ε is suppressed.

6.4.3 Cryoscope signal-to-noise ratio

Taking the expectation values $\langle X \rangle$ and $\langle Y \rangle$ in Cryoscope, denoted x and y for brevity, the probability density $p(x, y)$ of measuring (x, y) , given that the true values are (x_0, y_0) , is assumed to be normally distributed with standard deviation σ dependent on readout fidelity and number of averages,

$$p(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(x - x_0)^2 + (y - y_0)^2}{2\sigma^2}\right). \quad (6.14)$$

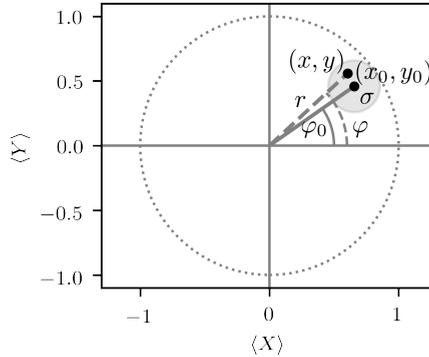


Figure 6.6.: Illustration clarifying the variables used. The probability of measuring (x, y) , given that the actual values are (x_0, y_0) , is taken to be normally distributed with standard deviation σ denoted by the grey circle.

The values x and y can be represented as an amplitude r and a phase φ . Assuming a perfect (non-distorted) square pulse, the value r is affected by the dephasing of the qubit $r_\tau = \exp(-(\tau/T_2^*(\Phi_Q))^\alpha)$ where $T_2^*(\Phi)$ is the flux dependent dephasing rate and $\alpha \in [1, 2]$. Meanwhile, the phase $\varphi_\tau = 2\pi\tau\Delta f_Q(\Phi_Q)$, where $\Delta f_Q(\Phi_Q)$ is the qubit detuning from the operating point as a function of flux-pulse amplitude. The phase is measured with

some error σ_φ that depends on r and σ . Ultimately, to reconstruct Φ_R , we are interested in the phase change between two measurements with slightly different τ , $\Delta\varphi = \varphi_{\tau+\Delta\tau} - \varphi_\tau$, with error $\sigma_{\Delta\varphi}$.

Error propagation

Without loss of generality, φ_0 can be set to 0, setting $y_0 = 0$ in Eq. (6.14). Rewriting $p(x, y)$ in spherical coordinates,

$$p(r, \varphi) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{r^2 - 2rx_0 \cos(\varphi) + x_0^2}{2\sigma^2}\right), \quad (6.15)$$

where the subscript from r_τ and φ_τ is dropped for brevity. As we are only interested in the error of the phase, we can integrate to find

$$\begin{aligned} p(\varphi) &= \int_0^\infty p(r, \varphi) r dr = \frac{1}{2\pi} \exp\left(-\frac{x_0^2}{2\sigma^2}\right) \\ &\quad + \frac{\cos(\varphi)}{\sqrt{8\pi}} \frac{x_0}{\sigma} \exp\left(-\frac{x_0^2 \sin^2(\varphi)}{2\sigma^2}\right) \left[\operatorname{erf}\left(\frac{x_0 \cos(\varphi)}{\sqrt{2}\sigma}\right) + 1 \right]. \end{aligned} \quad (6.16)$$

When the the visibility of the Ramsey oscillations x_0 is much larger than σ this simplifies to

$$p(\varphi) = \frac{\cos(\varphi)}{\sqrt{8\pi}} \frac{x_0}{\sigma} \exp\left(-\frac{x_0^2 \sin^2(\varphi)}{2\sigma^2}\right) \left[\operatorname{erf}\left(\frac{x_0 \cos(\varphi)}{\sqrt{2}\sigma}\right) + 1 \right]. \quad (6.17)$$

Because we have set $\varphi_0 = 0$ and $x_0 \ll \sigma$, the small-angle approximation ($\varphi \ll 1$) can be made, so that this simplifies further to

$$p(\varphi) = \frac{x_0}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x_0^2 \varphi^2}{2\sigma^2}\right). \quad (6.18)$$

Since the distribution is normal, we conclude that

$$\sigma_\varphi = \frac{\sigma}{x_0} \quad \text{and} \quad \sigma_{\Delta\varphi} = \frac{\sqrt{2}\sigma}{x_0}. \quad (6.19)$$

SNR formula

Ultimately, the SNR of Cryoscope is affected by the following factors:

1. Readout fidelity and averaging, captured by σ ;
2. Flux dependent qubit dephasing $T_2^*(\Phi_Q)$, affecting visibility $r = x_0 = \exp(-(\tau/T_2^*(\Phi_Q))^\alpha)$, with $1 \leq \alpha \leq 2$;
3. Rate at which the phase is acquired, proportional to $\Delta f_Q(\Phi_Q)$;
4. Filtering effects in the data processing, that affect SNR linearly.

The phase $\Delta\varphi$ in the time interval $\Delta\tau$ is

$$\Delta\varphi = 2\pi\Delta f_Q(\Phi_Q)\Delta\tau, \quad (6.20)$$

while the noise of the phase measurement is

$$\sigma_{\Delta\varphi} = \frac{\sqrt{2}\sigma}{x_0}, \quad (6.21)$$

leading to

$$\text{SNR} = c' \times \frac{2\pi \times \Delta f_Q(\Phi_Q) \times \Delta\tau \exp(-t/T_2^*(\Phi_Q)^\alpha)}{\sqrt{2}\sigma}. \quad (6.22)$$

Here, c' is a constant that accounts for filtering effects in data processing. As we cannot distinguish between σ and effects of filtering c' , and c' is unknown, we can absorb all multiplicative factors

$$\text{SNR} = c'' \times \Delta f_Q(\Phi_Q) \exp(-(\tau/T_2^*(\Phi_Q))^\alpha). \quad (6.23)$$

To evaluate the model, we use a quadratic dependence of the detuning on the flux $\Delta f_Q(\Phi_Q) = a\Phi_Q^2$. We find that this dependence matches well the experimentally measured dependence in studied range of pulse amplitudes, up to $0.33 \Phi_0$ for $a = 16.9 \frac{\text{GHz}}{\Phi_0^2}$ (Figure 6.7). Furthermore, we use $\alpha = 1$ and a dephasing rate given by

$$\frac{1}{T_2^*(\Phi_Q)} = \Gamma = \Gamma_0 + \Gamma_1 \left| \frac{d\Delta f_Q}{d\Phi_Q} \right|, \quad (6.24)$$

where Γ_0 describes the flux-independent dephasing, and Γ_1 parametrizes the contribution to dephasing due to $1/f$ flux noise [236]. Finally, in the limit of small errors:

$$\text{SNR} = c\Phi_Q^2 \exp(-(\Gamma_0 + 2a\Gamma_1\Phi_Q)t). \quad (6.25)$$

The joint fit to all data in Fig. 3 uses $\Gamma_0 = 66.7 \cdot 10^{-3} \text{ s}^{-1}$, corresponding to the measured sweetspot $T_2^* = 15 \mu\text{s}$ and yields $\Gamma_1 = 0.213 \times 10^{-3} \Phi_0$ equivalent to $\sqrt{A_\Phi} = 12 \times 10^{-6} \Phi_0$ for the single-sided flux noise power spectrum $S(f) = A_\Phi/f$ in a reasonable agreement with the typically reported values[237–240] and consistent with values measured in our group [112, 241].

6.4.4 Real-time predistortion filters

We make use of two types of digital filters to correct for distortions in real time: a finite impulse response (FIR) filter for short-timescale ($< 30 \text{ ns}$) distortions and a first-order infinite impulse response (IIR) filter.

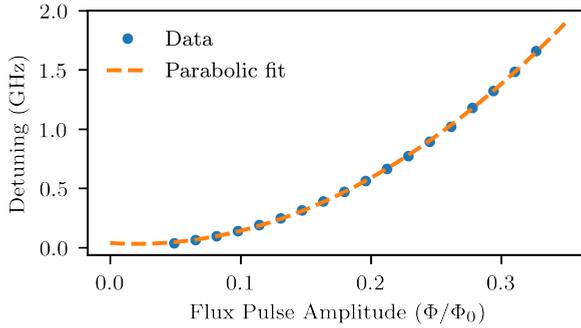


Figure 6.7.: Measured dependence of detuning on the applied flux through the SQUID loop of the studied transmon qubit.

FIR filter

A FIR filter implements a convolution with the impulse response $h_{\text{FIR}}[n] = b_n$ specified by the coefficients b_i and described by:

$$y[n] = \sum_{i=0}^N b_i x[n - i], \quad (6.26)$$

where $x[n]$ is the signal at time n at the input of the filter and $y[n]$ is the signal at the output of the filter.

The real-time FIR filter allows specifying 40 parameters to determine the first 72 coefficients b_i . The first 8 parameters directly correspond to the first 8 b_i coefficients while the remaining 32 parameters set pairs of parameters. Allowing a total of 72 coefficients b_i to be set, corresponding to a filter length of 30 ns for the AWG sampling rate of 2.4 GSa/s.

Exponential over- and under-shoot correction IIR filter

An IIR filter is slightly more complicated than a FIR filter because it includes feedback:

$$a_0 y[n] = \sum_{i=0}^N b_i x[n - i] - \sum_{i=1}^M a_i y[n - i]. \quad (6.27)$$

Here, a_i are the feedback coefficients that describe how $y[n]$ depends on values at the output of the filter at preceding times. Equation (6.27) is known as the time-domain difference equation and is a generalization of Equation (6.26).

A first-order IIR filter is implemented in hardware and intended to correct an exponential over- or undershoot in the step response. For a step response described by

$$s(t) = g(1 + Ae^{-t/\tau_{\text{IIR}}}) \cdot u(t), \quad (6.28)$$

where τ_{IIR} is the time constant, A the amplitude and g is a gain correction factor that is ignored. The filter that corrects for this effect is described by the coefficients

$$b_0 = 1 - k + k \cdot \alpha, \quad b_1 = -(1 - k) \cdot (1 - \alpha), \quad a_0 = 1 \quad \text{and} \quad a_1 = -(1 - \alpha),$$

with

$$\alpha = 1 - e^{1/f_s \tau_{\text{IIR}}(1+A)} \quad \text{and} \quad k = \begin{cases} \frac{A}{(1+A)(1-\alpha)}, & \text{if } A < 0 \\ \frac{A}{(1+A-\alpha)}, & \text{if } A \geq 0, \end{cases}$$

where f_s is the sampling rate and $a_i, b_i = 0$ for $i > 1$.

The limitations of the hardware implementation of the IIR filter can best be described using an equivalent representation of the filter. The ideal IIR filter obeys the following difference equation mapping the input samples $x[n]$ to the output samples $y[n]$

$$y[n] = (1 - k) x[n] + k u[n], \quad (6.29)$$

where $u[n]$ represents the state of the IIR filter, which is determined by the recursive difference equation, known as an exponential moving average

$$u[n] = u[n - 1] + \alpha(x[n] - u[n - 1]). \quad (6.30)$$

Implementing the recursion in Equation (6.30) directly with state-of-the-art digital signal processing hardware is infeasible due to the high sampling rate (2.4 GSa/s). Instead, the real-time filters compute the state variable $u[n]$ based on an average of 16 samples. Furthermore, the IIR filter is operated at a clock frequency of 300 MHz, which means that the state variable $u[n]$ gets updated only every 8-th sample. The down sampled $u[n]$ is combined with the input signal $x[n]$ in Equation (6.29) at the full sampling rate. These hardware approximations were taken into account when modeling the impulse response of the filter.

Modelling filters

The Python library SciPy [224] provides a function called “filter(b, a, sig)”, which applies the filter defined by the coefficient vectors “b” and “a” to the signal defined by the vector “sig”. In this work, we use this function to predict the effect of applying the real-time predistortion filters. Because the implementations of the real-time digital filters in hardware requires certain modifications to approximate the ideal filter operation there are slight differences between the real-time filters and the ideal filters. These deviations are taken into account when we predict the effect of applying a specific filter.

6.4.5 Using Cryoscope to measure arbitrary shapes

Cryoscope is capable of sampling arbitrary flux control pulses. To demonstrate this capability we have chosen a typical Amsterdam canal skyline as an example of an arbitrary shape for the flux pulse. Figure 6.8 demonstrates this capability by providing a near perfect reconstruction of the target waveform. Note that the reconstruction involves no free parameters.

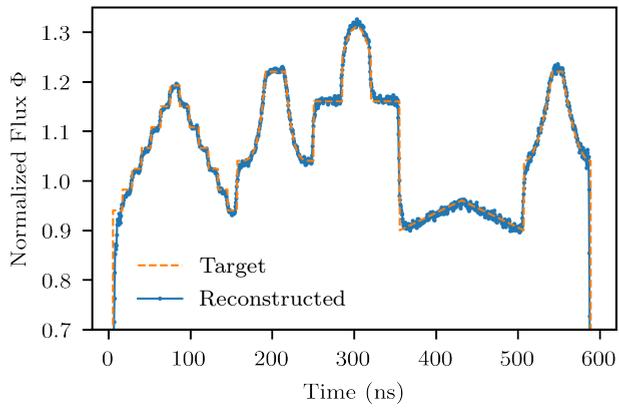
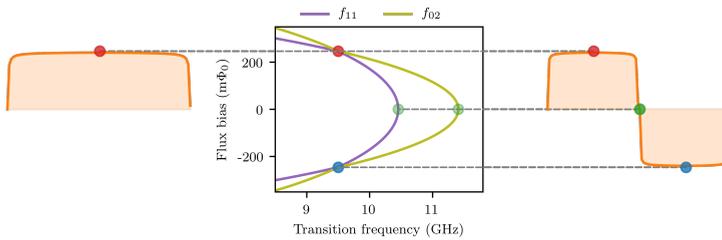


Figure 6.8.: Reconstruction of an arbitrary waveform consisting of a typical Amsterdam canal skyline.

A FAST, HIGH-FIDELITY CONDITIONAL-PHASE GATE EXPLOITING LEAKAGE INTERFERENCE IN WEAKLY ANHARMONIC SUPERCONDUCTING QUBITS



Conditional-phase (CZ) gates in transmons can be realized by flux pulsing computational states towards resonance with non-computational ones. We present a 40 ns CZ gate based on a bipolar flux pulse suppressing leakage (0.1%) by interference and approaching the speed limit set by exchange coupling. This pulse harnesses a built-in echo to enhance fidelity (99.1%) and is robust to long-timescale distortion in the flux-control line, ensuring repeatability. Numerical simulations matching experiment show that fidelity is limited by high-frequency dephasing and leakage by short-timescale distortion.

7.1 Introduction

A steady increase in qubit counts [242–245] and operation fidelities [78, 99, 117, 120, 130] allows quantum computing platforms using monolithic superconducting quantum hardware to target outstanding challenges such as quantum advantage [180, 246, 247], quantum error correction (QEC) [42, 43, 61, 248, 249], and quantum fault tolerance (QFT) [48, 182]. All of these pursuits require two-qubit gates with fidelities exceeding 99%, fueling active research.

There are three main types of two-qubit gates in use for transmon qubits, all of which harness exchange interactions between computational states ($|ij\rangle, i, j \in \{0, 1\}$) or between computational and non-computational states (i or $j \geq 2$), mediated by a coupling bus or capacitor. Cross-resonance gates [114, 117] exploit the exchange interaction between $|01\rangle$ and $|10\rangle$ using microwave-frequency transversal drives. Parametric gates [96, 120] employ radio-frequency longitudinal drives, specifically flux pulses modulating the qubit frequency, to generate sidebands of resonance between $|01\rangle$ and $|10\rangle$ for iSWAP or between $|11\rangle$ and $|02\rangle$ or $|20\rangle$ for conditional phase (CZ). The oldest approach [26, 95] uses baseband flux pulses to tune $|11\rangle$ into near resonance with $|02\rangle$ to realize CZ. Either because they explicitly use non-computational states, or because of frequency crowding and the weak transmon anharmonicity, the three approaches are vulnerable to leakage of information from the computational subspace. Leakage is very problematic in applications such as QEC, complicating the design of error decoders and/or demanding operational overhead to generate seepage [56–60], generally reducing the error thresholds for QFT. This threat has motivated the design of fast-adiabatic pulses [167] to mitigate leakage and architectural choices in qubit frequency and coupler arrangements [121] to explicitly avoid it. Surprisingly, many recent demonstrations [117, 120, 250] of two-qubit gates place emphasis on reaching or approaching 99% fidelity without separately quantifying leakage.

Although baseband flux pulsing produces the fastest two-qubit gates to date (30–45 ns), two challenges have kept it from becoming the de facto two-qubit gating method. First, because the pulse displaces one qubit 0.5–1 GHz below its flux-symmetry point, i.e., the sweetspot, the sensitivity to flux noise increases dephasing and impacts fidelity. The second challenge is non-atomicity. If uncompensated, distortions in the flux-control lines originating from limited waveform-generator bandwidth, high-pass bias tees, low-pass filters, impedance mismatches, on-chip response, etc., can make the action of a pulse depend on the history of flux pulses applied. To date, predistortion corrections have been calculated in advance, requiring prior knowledge of the timing of all the flux-pulse-based operations required by the quantum circuit, and significant waveform memory. This standard practice is incompatible with real-time determination and execution of operations, as is required for control flow and feedback in a fully programmable quantum computer [52, 53].

In this Letter, we introduce a fast (40 ns), low-leakage (0.1%), high-fidelity (99.1%), and repeatable flux-pulse-based CZ gate suitable for a full-stack quantum computer executing operations in real time on transmon-based quantum hardware. These attractive characteristics are enabled by a zero-average bipolar flux-pulsing method, nicknamed Net-Zero (NZ), which uses the $|11\rangle \leftrightarrow |02\rangle$ avoided crossing twice. Harnessing the analogy to a Mach-Zehnder in-

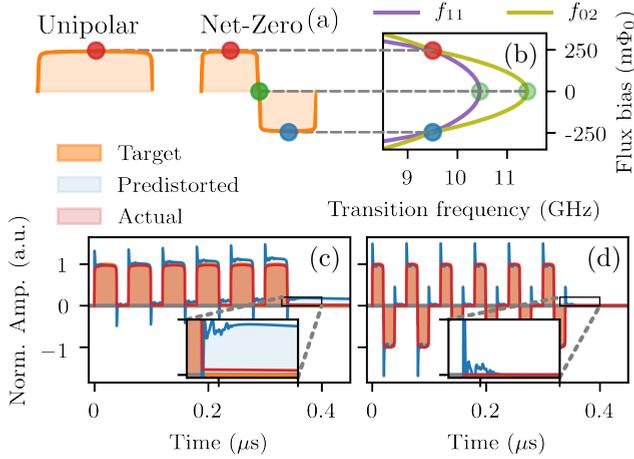


Figure 7.1.: (a) Schematic representation of unipolar and NZ pulses that tune into resonance with (b) $|11\rangle \leftrightarrow |02\rangle$ in order to perform CZ gates. Repeated applications of unipolar (c) and NZ (d) CZ pulses showing the target (orange), predistorted (blue), and actual (red) waveforms for an imperfect distortion correction. The insets in (c) and (d) show the differing accumulation in the required predistortion correction.

terferometer, NZ exploits destructive interference to minimize leakage to $|02\rangle$ while approaching the speed limit set by the exchange coupling in the two-excitation manifold. The flux symmetry of the transmon Hamiltonian makes the phases acquired by the pulsed qubit first-order insensitive to low-frequency flux noise, increasing fidelity relative to a unipolar pulse. Crucially, the zero-average characteristic makes NZ insensitive to long-timescale distortions remaining in the flux-control line after real-time pre-compensation, making the CZ gate repeatable. Detailed numerical simulations supplied with calibrated experimental parameters and direct measurement of short-timescale distortions show an excellent match to experiment, and indicate that fidelity is limited by high-frequency flux noise while leakage is dominated by remaining short-timescale distortions.

7.2 Concept

The ideal CZ gate is described by the transformation:

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\phi_{01}} & 0 & 0 \\ 0 & 0 & e^{i\phi_{10}} & 0 \\ 0 & 0 & 0 & e^{i\phi_{11}} \end{pmatrix}, \quad (7.1)$$

in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$, where the single-qubit phases ϕ_{01} and ϕ_{10} are even multiples of π and the conditional phase defined by $\phi_{2Q} = \phi_{11} - \phi_{01} - \phi_{10}$ is an odd multiple of π . A CZ gate of total duration $T_{CZ} = T_{2Q} + T_{1Q}$ can be realized in two

steps. First, a strong flux pulse on the higher frequency qubit moves $|11\rangle$ into the avoided crossing with $|02\rangle$ and back to acquire ϕ_{2Q} . Next, simultaneous weaker pulses on both qubits adjust the single-qubit phases. We compare two types of flux pulses, the (unipolar) pulse introduced in [167] and the NZ pulse [Figure 7.1(a)]. The NZ pulse consists of two back-to-back unipolar pulses of half the duration and opposite amplitude. Experiments are performed on a pair of flux-tunable transmons described in Section 7.5.1.

Because of distortions, the waveform $V_{\text{AWG}}(t)$ specified in an arbitrary waveform generator (AWG) does not result in the qubit experiencing the targeted flux $\Phi_{\text{target}}(t)$. These distortions can be described as a linear time-invariant system that transduces voltage to flux and is characterized by its impulse response $h(t)$. To measure $h(t)$ at the qubit, we employ the Cryoscope technique that we have introduced in Chapter 6. We then use it to construct an inverse filter \tilde{h}^{-1} , known as a predistortion correction, to compensate the distortions. By performing a convolution of the desired signal $\Phi_{\text{target}}(t)$ with \tilde{h}^{-1} , the qubit experiences the pulse

$$\Phi(t) = h * V_{\text{AWG}}(t) = h * (\tilde{h}^{-1} * \Phi_{\text{target}})(t). \quad (7.2)$$

The predistortion corrections are performed using a combination of real-time filters implemented in a Zurich Instruments HDAWG and a short (20 ns) FIR filter implemented offline.

By eliminating the DC component of the pulse, NZ CZ gates are resilient to long-timescale distortions [176]. Because the transmon Hamiltonian is symmetric with respect to the sweetspot, it is possible to use both positive and negative amplitudes to perform a CZ gate [Figure 7.1(b)] while satisfying the zero-average condition

$$\int_0^{T_{\text{CZ}}} \Phi_{\text{target}}(t') dt' = 0. \quad (7.3)$$

If Equation (7.3) holds, the DC component is zero and the components in the Fourier transform $\Phi_{\text{target}}(\omega)$ at frequencies $\omega \lesssim \frac{2\pi}{T_{\text{CZ}}}$ are suppressed. Writing Equation (7.2) in the Fourier domain: $\Phi(\omega) = \mathcal{H}(\omega) \cdot \tilde{\mathcal{H}}^{-1}(\omega) \cdot \Phi_{\text{target}}(\omega)$, it follows that if $\Phi_{\text{target}}(\omega)$ does not contain any components at $\omega < \frac{2\pi}{T_{\text{CZ}}}$, then $\Phi(\omega)$ does not depend on any components of $\mathcal{H}(\omega)$ at frequencies $\omega < \frac{2\pi}{T_{\text{CZ}}}$. As a consequence, the required corrections for NZ pulses do not accumulate, eliminating the need for accurate long-timescale distortion corrections and the resulting history-dependent errors [Figure 7.1(d)].

7.3 Experimental results

7.3.1 Repeatability

To measure the repeatability of CZ gates, the phase (ϕ_{01}) acquired by the pulsed qubit during a CZ gate is measured as a function of the separation time T_{Sep} between pulses (Figure 7.2). Because of the detuning from the sweetspot, a small change in amplitude during the pulse leads to a significant change in frequency. This makes the acquired phase sensitive to distortions. We observe that not correcting distortions leads to significant phase errors (~ 80 deg). Correcting distortions using a predistortion filter keeps the error small (< 10 deg) for the first

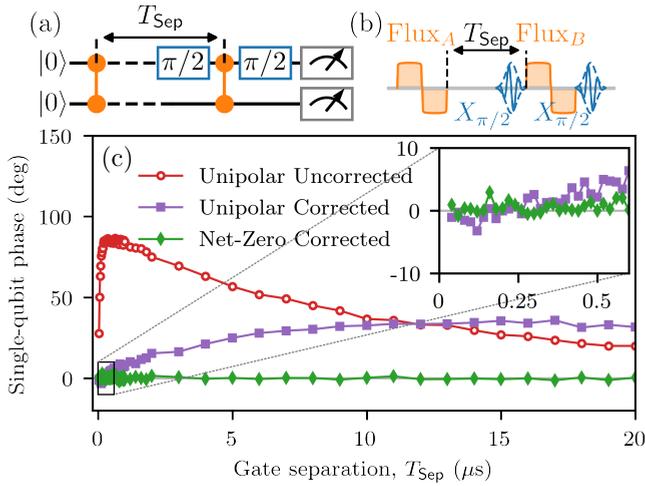


Figure 7.2.: History dependence of flux pulses. Circuit (a) and pulses (b) used to measure the phase acquired during a pulse as a function of separation time T_{Sep} to another pulse. Pulses are calibrated to correspond to CZ gates. (c) Acquired single-qubit phase for unipolar pulses without (red), and with (purple) predistortion corrections and NZ pulses with predistortion corrections (green).

500 ns but shows history-dependent behavior for longer timescales. Using NZ pulses in combination with a predistortion filter eliminates all history dependence. Hence, we conclude that NZ pulses are robust against remaining long-timescale distortions.

7

7.3.2 Echo-effect

We next investigate a built-in echo effect that provides protection against flux noise. Because the derivative of the flux arc is equal and opposite in sign at the positive and negative halves of the NZ pulse, we expect ϕ_{01} and ϕ_{2Q} to be first-order insensitive to low-frequency flux noise. As a test, we measure the dependence of ϕ_{2Q} on an applied DC flux offset for both a unipolar and NZ CZ gate [Figure 7.3]. As shown in Figure 7.3(b), ϕ_{2Q} is first-order (second-order) sensitive for a unipolar (NZ) pulse. We have also measured how the dephasing time depends on the detuning for both a square flux pulse and two half-square flux pulses with opposite sign (Section 7.5.3). We find that the dephasing rate is significantly reduced when the opposite-sign flux pulses are used, confirming that NZ pulses have a built-in echo effect.

7.3.3 Leakage

The pulse shape is intended to minimize leakage and is described by two parameters (Section 7.5.2). Parameter θ_f is a measure of the flux at the middle of the unipolar pulse, and at the middle of each half of NZ. States $|11\rangle$ and $|02\rangle$ are resonant at $\theta_f = \pi/2$. Parameter λ_2 tunes the sharpness of the pulse rise and fall. We follow [151] in defining the leakage (L_1)

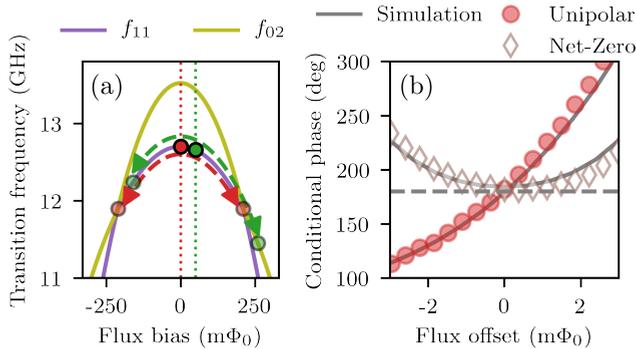


Figure 7.3.: Echo effect in NZ pulses. (a) Level diagram showing the effect of a drift in flux on a NZ pulse: a NZ pulse will move to the interaction point on both sides (red); when the bias is offset (green), one side will overshoot while the other side will undershoot the interaction point, canceling the acquired extra phase. (b) Measured dependence of conditional phase on applied DC flux offset for both NZ (diamond) and unipolar (circles) $T_{CZ} = 60$ ns pulses ($T_{2Q} = 40$ ns). Solid lines correspond to simulation (Section 7.5.3), dashed line indicates 180 deg. The unipolar (NZ) is first-order (second-order) sensitive to the applied offset.

of an operation as the average probability that a random computational state leaks out of the computational subspace.

In order to gain insight into how ϕ_{2Q} and L_1 depend on the pulse shape, we perform an experiment and compare this to simulations. The conditional oscillation experiment (Figure 7.4) consists of a Ramsey-like experiment that allows us to measure ϕ_{2Q} and estimate L_1 . This experiment measures the phase acquired during an (uncalibrated) CZ gate by the target qubit ($q_{\text{targ.}}$) while either leaving the control qubit ($q_{\text{contr.}}$) in the ground state, or adding an excitation to $q_{\text{contr.}}$. The difference between the phase acquired when $q_{\text{contr.}}$ is in $|0\rangle$ and when $q_{\text{contr.}}$ is in $|1\rangle$ gives ϕ_{2Q} . If leakage from $|11\rangle$ to $|02\rangle$ occurs, $q_{\text{contr.}}$ is in $|0\rangle$ when the second π pulse is applied, adding, instead of removing, an excitation to $q_{\text{contr.}}$. The leakage probability L_1 can be estimated as $\widetilde{L}_1 = m/2$, where m is the population difference on the control qubit between both variants of the experiment. Because of relaxation effects, \widetilde{L}_1 slightly overestimates L_1 .

The simulations model the system realistically and allow us to extract ϕ_{2Q} , L_1 and the average gate fidelity F for a single application of the gate (Section 7.5.3). The pulse is modeled as a trajectory in a two-qutrit Hamiltonian. The noise model accounts for relaxation and dephasing effects as well as the effect of remaining distortions. The latter are measured using the Cryoscope technique [225]. For the dephasing we take into account the different timescales on which flux noise acts as well as the measured dependence on the flux bias.

Both experiment and simulation show a fringe of low leakage [Figure 7.4(b,d)]. This fringe can be understood as “leakage interference” between $|11\rangle$ and $|02\rangle$ by analogy to a Mach-Zehnder interferometer (Section 7.5.6). Such analogy has been exploited in a variety of platforms [251–255] to demonstrate coherent control of a single qubit by showing Stückelberg

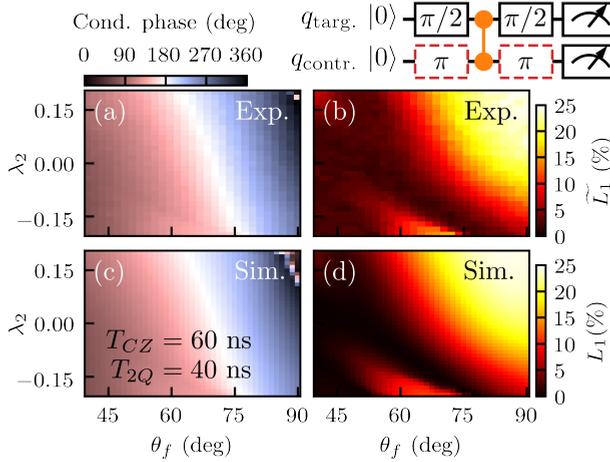


Figure 7.4.: Conditional phase (a, c) and leakage (b, d) for a $T_{CZ} = 60$ ns ($T_{2Q} = 40$ ns) NZ flux pulse as a function of pulse parameters θ_f and λ_2 for both experiment (a, b) and simulation (c, d). The conditional phase increases with θ_f and λ_2 , since both of these have the effect of making the pulse spend more time close to the interaction point. Leakage tends to increase significantly with larger values of θ_f with the exception of a diagonal fringe.

oscillations [256] as a consequence of periodic driving of the qubit into an avoided crossing. Here we pulse in-and-out of $|11\rangle \leftrightarrow |02\rangle$ twice to realize low-leakage two-qubit gates. The states $|11\rangle$ and $|02\rangle$ correspond to two paths of the interferometer. The first part of the NZ pulse (red in Figure 7.1) corresponds to the first (imbalanced) beamsplitter. In general, after the first beamsplitter most of the population remains in $|11\rangle$ but part is transferred to $|02\rangle$. Pulsing through the sweetspot (green in Figure 7.1) corresponds to the arms of the interferometer. The two paths are detuned by ~ 800 MHz, causing a phase to be acquired before the paths are recombined at the second half of the NZ pulse (blue in Figure 7.1) corresponding to the second beamsplitter. The phase difference between the two paths will cause interference that either enhances or suppresses the leakage to $|02\rangle$.

7.3.4 Performance

Given the good correspondence between experiment and simulation (Figure 7.4), we can use simulations to explore the parameter space $(\theta_f, \lambda_2, T_{2Q})$ to find the shortest T_{2Q} enabling a high-fidelity, low-leakage CZ gate. The minimum CZ gate duration is fundamentally limited by the coupling strength J_2 as the time required to acquire 180 degrees of conditional phase at the avoided crossing: $T_{2Q} \geq \frac{\pi}{J_2} = 25$ ns. We find a $T_{2Q} = 28$ ns NZ pulse using leakage interference to achieve low leakage. The use of interference is demonstrated by the fact that the corresponding half pulse displays high leakage (Section 7.5.6). We append $T_{1Q} = 12$ ns flux pulses on both qubits to correct the single-qubit phases, making the total duration of the phase-corrected CZ gate $T_{CZ} = 40$ ns. We ensure that these phase-

correction pulses satisfy Equation (7.3) and have a sufficiently low amplitude to not affect ϕ_{2Q} and L_1 significantly.

We characterize the performance of the CZ gate using an interleaved randomized benchmarking protocol [78, 146] with modifications that allow us to quantify leakage ([92, 151] and Section 7.5.7). The randomized benchmarking sequences are based on 300 random seeds. For each seed, every data point is measured 104 times. We measure an average gate fidelity $F = 99.10\% \pm 0.16\%$ and leakage $L_1 = 0.10\% \pm 0.07\%$ for the NZ pulse with $T_{CZ} = 40$ ns [Figure 7.5(a,b)]. We could not perform similar measurements for the unipolar pulse since this gate is not repeatable, as demonstrated in Figure 7.2.

7.3.5 Performance limitations

It is possible to investigate the limits to the performance of the NZ CZ using simulation and compare to the unipolar CZ, even though this is not possible in experiment since the unipolar CZ lacks the required characteristic of being repeatable. We simulate these gates for a range of different error models [Figure 7.5(c,d)]. For each we optimize over θ_f and λ_2 to find the lowest ε and the corresponding L_1 . A first observation is that the infidelity ($\varepsilon = 1 - F$) of the NZ gate does not significantly increase when the low-frequency flux-noise components are included, whereas this does affect the unipolar pulse. It appears that the difference in ε between the unipolar and NZ pulses for the full model can be attributed completely to this effect. This observation is consistent with the echo effect demonstrated in Figure 7.3. Looking at the L_1 error budgets, L_1 is limited by short-timescale distortions. This is understandable as minimizing L_1 requires the pulse to follow a precise trajectory. Distortions also increase ε through L_1 (Section 7.5.3). The simulations also indicate that dephasing causes leakage. This can be understood as dephasing effectively corresponds to an uncertainty in the energy levels. The simulated L_1 is larger than the measured L_1 . This could be explained in two ways, either the distortions are less severe than our estimate, or the simulations, only concerned with a single application of the gate, do not take into account all the relevant effects. Specifically, because the population in the leakage subspace does not completely decohere, this population can seep back into the computational subspace due to an interference effect (similar to that in the NZ pulse itself) at subsequent applications of the gate. Because the first CZ gate cannot benefit from this coherence, the simulations, which only deal with a single CZ gate, slightly overestimate the effective leakage.

7.4 Conclusions

In summary, we have demonstrated a flux-based CZ gate for transmon qubits that is fast, low-leakage, high-fidelity and repeatable. The gate is realized using a bipolar Net-Zero flux pulse that harnesses leakage interference to achieve speed while maintaining low leakage. The NZ pulse exploits the flux symmetry of the pulsed transmon to build in an echo effect on its single-qubit phase and the conditional phase, increasing fidelity relative to a unipolar pulse. Finally, the action of the NZ pulse is robust to long-timescale distortions in the flux-control line remaining after real-time pre-compensation, enabling the repeatability of the CZ gate. These

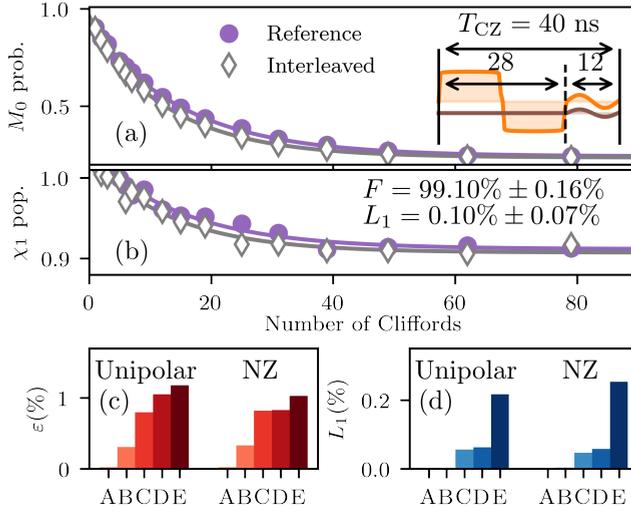


Figure 7.5.: Interleaved randomized benchmarking with leakage modification and simulated performance using different error models for a $T_{CZ} = 40$ ns NZ CZ gate ($T_{2Q} = 28$ ns), schematically shown in the diagram. (a) Survival probability M_0 of recovering $|00\rangle$ for reference and interleaved two-qubit randomized benchmarking sequence. (b) Population in the computational subspace \mathcal{X}_1 . Simulated ϵ (c) and L_1 (d) for different error models (Section 7.5.3) for $T_{CZ} = 40$ ns unipolar and NZ pulses ($T_{2Q} = 28$ ns). The error models (A to E) contain: no noise (A), relaxation (B), all Markovian noise components (C), Markovian and quasi-static flux noise components (D) and all noise components including distortions (E).

features make the realized NZ CZ gate immediately useful in high-circuit-depth applications of a full-stack quantum computer in which a controller issues operations to execute on the quantum hardware in real time. For example, current work in our group uses NZ CZ gates to stabilize two-qubit entanglement by multi-round indirect parity measurements [61]. Future work will incorporate NZ CZ gates into our scheme [121] to realize a surface-code-based logical qubit [48] with monolithic transmon-cQED quantum hardware.

7.5 Methods

This section contains detailed information on the experimental protocols and the simulations performed in this work. Section 7.5.1 provides relevant device parameters. Section 7.5.2 describes the parametrization used for the unipolar and NZ pulses. Section 7.5.3 describes the simulations in detail. Section 7.5.4 and Section 7.5.7 describe protocols used to characterize the flux pulses. Section 7.5.5 investigates the limitations of the CZ gate. Section 7.5.6 discusses the Mach-Zehnder interferometer analogy in detail.

7.5.1 Device parameters

All experiments were performed on a circuit-QED quantum chip containing three starmon-type [121] transmon qubits, labeled q_H , q_M , and q_L . Pairs q_H - q_M and q_M - q_L are coupled by separate bus resonators. Each qubit has a microwave drive line for single-qubit gating, a flux-bias line for local and ns-timescale control of the qubit frequency, and dedicated, fast readout resonators with Purcell protection for the qubits. The readout resonators are coupled to a common feedline, allowing independent readout of the three qubits by frequency multiplexing.

In this work we focus on the transmon pair q_H - q_M . We have achieved similar performance (fidelity, leakage and gate time) for the pair q_M - q_L . Relevant device parameters are given in Table 7.1.

Parameter	q_L	q_M	q_H
$\omega/2\pi$ operating point (GHz)	5.02	5.79	6.87
$\omega/2\pi$ sweetspot (GHz)	5.02	5.79	6.91
$\eta/2\pi$ (MHz)	-300	-300	-331
$J_1/2\pi$ avoided crossing (MHz)	17.2		14.3
T_1 (μs)	31.8	15.2	19.2
T_2^* operating point (μs)	14.0	14.8	3.2
T_2^E operating point (μs)	33.8	19.4	14.7
$\sim \omega_{\text{bus}}/2\pi$ (GHz)	8.5		8.5

Table 7.1.: Parameters of the three-transmon device: qubit frequency (ω), anharmonicity (η), exchange coupling between $|01\rangle$ and $|10\rangle$ (J_1), dephasing times (T_1, T_2^*, T_2^E) and bus-resonator frequency (ω_{bus}). Experiments in this work are performed with the pair q_H - q_M . q_H is operated 40 MHz below its sweetspot to minimize interaction with a spurious two-level system right at the sweetspot frequency.

7.5.2 Flux pulse parametrization

Unipolar and NZ pulses are based on the Martinis-Geller parametrization for fast-adiabatic gates [167]. This parametrization is determined by the Hamiltonian [Equation (7.11)] projected onto a two-dimensional subspace. In the case of the CZ gate, this subspace is spanned by the states $|11\rangle$ and $|02\rangle$. The projected Hamiltonian, H_{subspace} , takes the form

$$H_{\text{subspace}} = \begin{pmatrix} \frac{\epsilon}{2} & J_2 \\ J_2 & -\frac{\epsilon}{2} \end{pmatrix}, \quad (7.4)$$

where $\epsilon = \omega_{|02\rangle} - \omega_{|11\rangle}$ is the bare detuning between $|11\rangle$ and $|02\rangle$ and J_2 is their coupling. The detuning ϵ is controlled by flux whereas J_2 is considered to be constant. We define the angle θ as

$$\theta \equiv \arctan\left(\frac{2J_2}{\epsilon}\right). \quad (7.5)$$

Note that $\theta = \pi/2$ at $\epsilon = 0$.

The waveform is expressed as a series

$$\theta(\tau(t)) = \theta_i + \sum_{j=1}^N \lambda_j \left(1 - \cos \left(\frac{2\pi \cdot j \cdot \tau(t)}{T_{2Q}} \right) \right), \quad (7.6)$$

where T_{2Q} is the pulse duration, θ_i corresponds to the detuning at the operating point and τ is proper time, which is related to real time t through $t(\tau) = \int_0^\tau d\tau' \sin(\theta(\tau'))$.

We truncate the series to $N = 2$. We make use of the relation between the angle at the middle of the unipolar pulse (θ_f) and the odd λ coefficients

$$\theta_f \equiv \theta(T_{2Q}/2) = \theta_i + 2 \sum_{j \text{ odd}}^N \lambda_j, \quad (7.7)$$

to define the entire waveform using three parameters: θ_f , λ_2 , and T_{2Q} . A NZ pulse is a sequence of two concatenated unipolar pulses, each lasting $T_{2Q}/2$ time and with the same θ_f and λ_2 .

There are a few more transformations required in order to have a waveform in terms of the flux $\Phi_{\text{target}}(t)$ [Figure 7.6]:

$$\theta(t) \mapsto \epsilon(t) \mapsto \omega_{q_H}(t) \mapsto \Phi_{\text{target}}(t). \quad (7.8)$$

The first transformation uses Equation (7.5): $\epsilon(t) = 2J_2/\tan \theta(t)$. The second one uses the fact that by definition $\epsilon(t) = \omega_{|02\rangle}(t) - \omega_{|11\rangle}(t) = \omega_{q_H}(t) + \eta_{q_H} - \omega_{q_M}$. The qubit frequency depends on flux according to the formula

$$\omega_{q_H}(\Phi) = (\omega_{q_H}^0 - \eta_{q_H}) \sqrt{\left| \cos \left(\frac{\Phi}{\Phi_0} \pi \right) \right|} + \eta_{q_H}, \quad (7.9)$$

where $\omega_{q_H}^0$ is the sweetspot frequency and η_{q_H} the anharmonicity, reported in Table 7.1. We refer to this relation between frequency and flux as the flux arc. The flux arc has been measured in the experiment and we find that it matches well with Equation (7.9). We invert Equation (7.9) to convert $\omega_{q_H}(t) \mapsto \Phi_{\text{target}}(t)$. Since $\omega_{q_H}(\Phi) = \omega_{q_H}(-\Phi)$, there is a positive and a negative solution for every value of ω_{q_H} . In the case of a unipolar pulse, we always consider the positive solution, whereas, in the case of a NZ pulse, the first and second half of the pulse use the positive and negative solutions, respectively. Changes that are clearly visible in the θ parametrization correspond to only a small change in the applied flux. This provides intuition why even a small distortion of the applied flux can have a relatively large effect on the gate quality.

7.5.3 Simulation structure

The simulations model the system, consisting of two coupled transmons, using a two-qutrit Hamiltonian. One of the two transmons, namely q_H , is actively pulsed into resonance according to the pulse parametrization described in Section 7.5.2. The simulations (Figure 7.7)

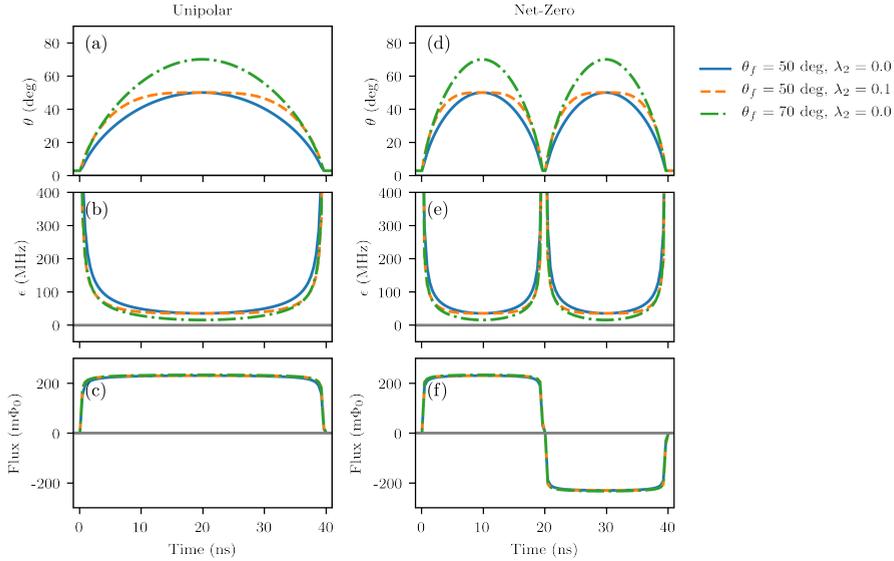


Figure 7.6.: Unipolar (a-c) and NZ pulses (d-e) represented in terms of θ (a, d), bare detuning ϵ (b, e) and flux Φ (c, f). The center of the unipolar pulse is controlled by θ_f , while λ_2 controls the sharpness of rise and fall of the pulse.

7

include distortions, relaxation and flux-dependent dephasing effects. The error model also includes a distinction between Markovian (fast) and non-Markovian (slow) noise in order to accurately model dephasing effects. The simulations are used to calculate the propagator or time-evolution superoperator, from which the quantities of interest - fidelity, leakage and conditional phase - are extracted.

System Hamiltonian

The system is composed of two transmons coupled via a bus resonator. We exclude the resonator from the model by making the assumption that it always remains in its ground state (it is excited only “virtually”). We restrict each transmon to its first three energy levels. In the dispersive regime, in the rotating-wave approximation, the Hamiltonian is given by

$$H(t) = \omega_{q_M} a_{q_M}^\dagger a_{q_M} + \frac{\eta_{q_M}}{2} (a_{q_M}^\dagger)^2 a_{q_M}^2 + \omega_{q_H}(\Phi(t)) a_{q_H}^\dagger a_{q_H} + \frac{\eta_{q_H}}{2} (a_{q_H}^\dagger)^2 a_{q_H}^2 \quad (7.10)$$

$$+ J_1(\Phi(t)) (a_{q_M} a_{q_H}^\dagger + a_{q_M}^\dagger a_{q_H}), \quad (7.11)$$

where only the higher-frequency transmon (q_H) is actively fluxed. Here a_{q_i} is the annihilation operator restricted to the first three energy levels, ω_{q_i} and η_{q_i} are the qubit frequency and anharmonicity, respectively, and J_1 is the coupling. The coupling is weakly flux-dependent since $J_1(\Phi) \approx \frac{g_{q_M} g_{q_H}}{2} (\Delta_{q_M}^{-1} + \Delta_{q_H}^{-1}(\Phi))$, with g_{q_i} the coupling of q_i to the bus resonator and $\Delta_{q_i} \approx \omega_{\text{bus}} - \omega_{q_i} \gg g_{q_i}$ given the parameters in Table 7.1. When we generate the

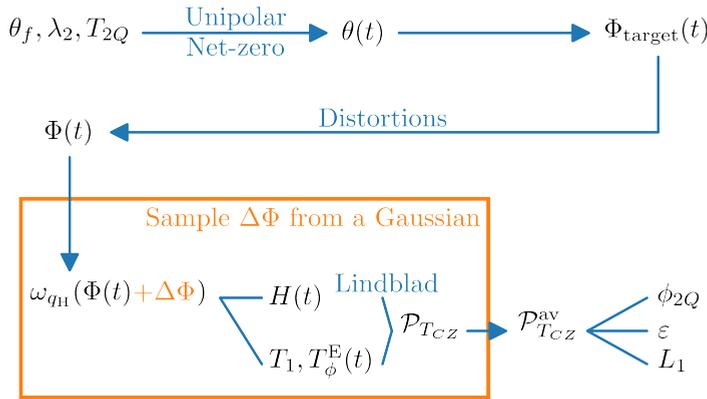


Figure 7.7.: The parameters θ_f , λ_2 and the gate time T_{2Q} determine either a unipolar pulse or a NZ pulse in terms of $\theta(t)$, see Equation (7.6). $\theta(t)$ is converted into $\Phi_{\text{target}}(t)$ through various transformations described in Section 7.5.2. Pulse distortions are applied by convolution to compute $\Phi(t)$ experienced by the qubit. The solution of the Lindblad equation is the time-evolution superoperator \mathcal{P}_{TCZ} . Averaging over a Gaussian distribution for the quasi-static flux bias $\Delta\Phi$, we obtain the average superoperator $\mathcal{P}_{TCZ}^{\text{av}}$. From that any quantity of interest can be computed, in particular the conditional phase ϕ_{2Q} , the average gate infidelity ε and the leakage L_1 .

flux pulse according to Section 7.5.2, we consider $J_2 = \sqrt{2}J_1$ to be constant and J_2 equal to its measured value at the $|11\rangle \leftrightarrow |02\rangle$ avoided crossing, whereas in the simulations we take into account the dependence of J_1 and J_2 on Φ .

Distortions

The flux pulse at the qubit is subject to distortions altering the shape of the waveform as experienced by the qubit. Distortions are described as a linear time-invariant system fully characterized by the impulse response h of the system. We best compensate such distortions by predistorting the desired pulse $\Phi_{\text{target}}(t)$ with an impulse response \tilde{h}^{-1} designed to invert h . Then, the actual pulse $\Phi(t)$ experienced by the qubit is given by

$$\Phi(t) = (h * V_{\text{AWG}})(t) = (h * (\tilde{h}^{-1} * \Phi_{\text{target}}))(t) = ((\tilde{h}^{-1} * h) * \Phi_{\text{target}})(t), \quad (7.12)$$

where $*$ denotes convolution. The distortions remaining after applying \tilde{h}^{-1} are determined by measuring the step response $s(t) = \int_0^t dt' (\tilde{h}^{-1} * h)(t')$ (Figure 7.8) using the Cryoscope technique [225]. The impulse response extracted from these data is used to distort the pulses in simulations.

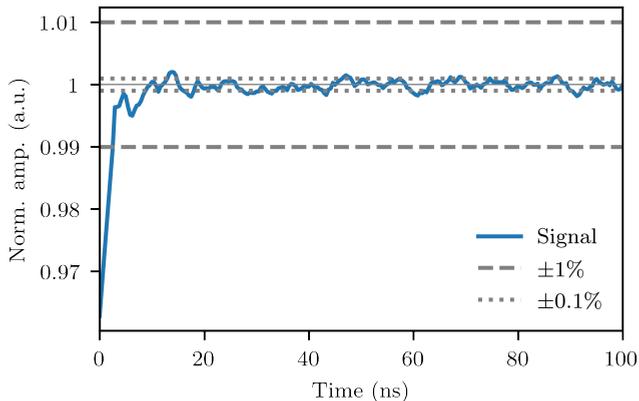


Figure 7.8.: Step response at the qubit after applying distortion corrections, measured using the Cryoscope technique [225]. The impulse response extracted from this experiment is used to distort the pulses in the simulations. In the case of perfect distortion corrections, the normalized amplitude would have value 1 for all times larger than zero.

Noise model

There are two major error sources in superconducting qubits: relaxation and flux noise. The latter has a power spectral density $S_f \sim A/f$, where f is frequency and \sqrt{A} is a constant of the order of $10 \mu\Phi_0$, with Φ_0 the flux quantum. S_f contains both high-frequency and low-frequency components: we phenomenologically distinguish high and low frequencies depending on whether they are larger or smaller than $1/T_{CZ}$. Relaxation and high-frequency flux-noise components are Markovian noise processes since they act on a timescale shorter than the gate time. On the other hand, the low-frequency flux-noise components determine a non-Markovian noise process, since they induce correlations across different gates.

We perform two experiments to quantify the strength of the dephasing affecting q_H : a Ram-Z and an Echo-Z experiment [Figure 7.9]. In these experiments, the dephasing times T_{2,q_H}^* (Φ) and T_{2,q_H}^E (Φ), respectively, at different flux sensitivities $\frac{1}{2\pi} \frac{\partial \omega_{q_H}}{\partial \Phi}$ are measured while applying a flux pulse. In the Ram-Z experiment, this flux pulse is square. In the Echo-Z experiment, the flux pulse consists of two square half pulses that detune the qubit by the same amount in magnitude but with opposite-sign sensitivity. We perform these experiments for a range of fluxes. The experimental data for q_H is represented in Figure 7.9. On the other hand, the static qubit q_M is always operated at the sweetspot. Therefore, we only use the measured Ramsey and Echo dephasing times at the sweetspot, reported in Table 7.1. The relaxation times T_{1,q_H} and T_{1,q_M} , are also reported in this table.

We assume that the low-frequency flux-noise components are echoed out in an Echo-Z experiment. In other words, we assume that T_{1,q_i} , T_{2,q_i}^E (Φ) quantify the strength of the Markovian noise. On the other hand, we assume that T_{1,q_i} , T_{2,q_i}^* (Φ) quantify the strength of the overall noise (both Markovian and non-Markovian). The strength of the non-Markovian noise alone cannot be extracted directly from the experiment. However, in the following we

explain the model that we use fitting the experimental data (Figure 7.9). In this way we can simulate separately both the Markovian and non-Markovian noise, and obtain a realistic simulation of the system.

Model of Markovian noise.

A Markovian evolution is modeled with the Lindblad equation

$$\dot{\rho}(t) = -i[H(t), \rho(t)] + \sum_{j,q_i} \left(c_{j,q_i}(t) \rho(t) c_{j,q_i}^\dagger(t) - \frac{1}{2} \{c_{j,q_i}^\dagger(t) c_{j,q_i}(t), \rho(t)\} \right) =: \mathcal{L}_t(\rho(t)), \quad (7.13)$$

where \mathcal{L}_t is the time-dependent Lindbladian defined by the Hamiltonian [Equation (7.11)] and by the jump operators $\{c_{j,q_i}(t)\}$ specified in Equations (7.14) and (7.16) to (7.18) below.

To model relaxation, we use the jump operator

$$c_{0,q_i} = \sqrt{\frac{1}{T_{1,q_i}}} a_{q_i}. \quad (7.14)$$

To model pure dephasing, we first define a pure-dephasing time

$$T_{\phi,q_i}^E(\Phi) = \left(\frac{1}{T_{2,q_i}^E(\Phi)} - \frac{1}{2T_{1,q_i}} \right)^{-1}, \quad (7.15)$$

Ignoring relaxation-induced dephasing in this paragraph, the coherence $\langle 0 | \rho_{q_i}(\Phi) | 1 \rangle$ decays as $e^{-t/T_{\phi,q_i}^E(\Phi)}$, where ρ_{q_i} is the qutrit reduced density matrix. In Figure 7.9 we see that the decay rates have a linear dependence on the flux sensitivity. Ignoring the anharmonicity, the frequency of the $|2\rangle$ state is twice the frequency of the $|1\rangle$ state, therefore, the sensitivity of the $|2\rangle$ state is twice as high. Given these two observations, we assume that $\langle 0 | \rho_{q_i}(\Phi) | 2 \rangle \propto e^{-t/(T_{\phi,q_i}^E(\Phi)/2)}$ and $\langle 1 | \rho_{q_i}(\Phi) | 2 \rangle \propto e^{-t/T_{\phi,q_i}^E(\Phi)}$. We find that such decay rates can be realized by the following jump operators

$$c_{1,q_i}(\Phi(t)) = \sqrt{\frac{8}{9T_{\phi,q_i}^E(\Phi(t))}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}_{q_i}, \quad (7.16)$$

$$c_{2,q_i}(\Phi(t)) = \sqrt{\frac{2}{9T_{\phi,q_i}^E(\Phi(t))}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}_{q_i}, \quad (7.17)$$

$$c_{3,q_i}(\Phi(t)) = \sqrt{\frac{2}{9T_{\phi,q_i}^E(\Phi(t))}} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}_{q_i}. \quad (7.18)$$

Instead, if one would use only

$$c'_{1,q_i}(\Phi(t)) = \sqrt{\frac{2}{T_{\phi,q_i}^E(\Phi(t))}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}_{q_i}, \quad (7.19)$$

which produces the same Lindbladian as $c''_{1,q_i}(\Phi(t)) = \sqrt{2/T_{\phi,q_i}^E(\Phi(t))} a_{q_i}^\dagger a_{q_i}$, then one would get $\langle 0|\rho_{q_i}(\Phi)|2\rangle \propto e^{-t/(T_{\phi,q_i}^E(\Phi)/4)}$ and $\langle 1|\rho_{q_i}(\Phi)|2\rangle \propto e^{-t/T_{\phi,q_i}^E(\Phi)}$. This means that Equation (7.19) would be the correct modeling if the decay rates in Figure 7.9 would depend quadratically on the sensitivity, but they do not.

The formal solution of Equation (7.13) is given by

$$\rho(t) = \mathcal{T} e^{\int_0^t dt' \mathcal{L}_{t'}} (\rho(0)), \quad (7.20)$$

where \mathcal{T} is the time-ordering operator. We call $\mathcal{P}_{TCZ} := \mathcal{T} e^{\int_0^{T_{2Q}} dt' \mathcal{L}_{t'}}$ the propagator or time-evolution superoperator, evaluated up to the gate time T_{CZ} , which includes an idling time T_{1Q} to account for the noise during the single-qubit phase correction pulses. The propagator \mathcal{P}_{TCZ} can be computed by solving the differential Equation (7.13), or as

$$\mathcal{P}_{TCZ} \simeq e^{\delta t \mathcal{L}_{TCZ - \delta t}} e^{\delta t \mathcal{L}_{TCZ - 2\delta t}} \dots e^{\delta t \mathcal{L}_{2\delta t}} e^{\delta t \mathcal{L}_{\delta t}} e^{\delta t \mathcal{L}_0}, \quad (7.21)$$

for a sufficiently small δt . In the simulations we use $\delta t = 0.1$ ns. In the Liouville representation, this equation is a product of matrices. We find that this method is an order of magnitude faster than using the qutip [257] differential equation solver.

Model of non-Markovian noise.

We model the low-frequency flux-noise components as quasi-static. Since the static qubit q_M is always operated at the sweetspot, where the sensitivity to flux noise is zero, we apply this model only to q_H . We assume that the qubit experiences a random, fixed flux offset $\Delta\Phi$ during the execution of a gate, but that $\Delta\Phi$ varies across different gates. For $\Delta\Phi \ll 1$, the effect of such offset on the pulse trajectory can be approximated at first order as $\omega_{q_H}(\Phi(t) + \Delta\Phi) \approx \omega_{q_H}(\Phi(t)) + \frac{\partial \omega_{q_H}(\Phi(t))}{\partial \Phi} \Delta\Phi$, where $\frac{1}{2\pi} \frac{\partial \omega_{q_H}(\Phi)}{\partial \Phi}$ is the flux sensitivity. Using Equation (7.9) we can see that $\frac{\partial \omega_{q_H}(\Phi)}{\partial \Phi} = -\frac{\partial \omega_{q_H}(-\Phi)}{\partial \Phi}$. In the case of a NZ pulse, this implies that first-order frequency variations in the first half of the pulse are canceled by an equal and opposite variation in the second half, resulting in an echo effect.

We take the probability distribution p_σ of $\Delta\Phi$ to be Gaussian $p_\sigma(\Delta\Phi) = e^{-(\Delta\Phi)^2/(2\sigma^2)}/(\sqrt{2\pi}\sigma)$, where σ is the standard deviation of the Gaussian. Averaging over this distribution, we get the final propagator

$$\mathcal{P}_{TCZ}^{\text{av}} = \int_{-\infty}^{+\infty} d(\Delta\Phi) p_\sigma(\Delta\Phi) \cdot \mathcal{P}_{TCZ}(\Delta\Phi), \quad (7.22)$$

which gives the time evolution including all the noise sources in the model, both Markovian and non-Markovian.

The standard deviation σ is not directly measured in the experiment. Instead, we fit this model to the experiment simulating a Ram-Z and Echo-Z experiment for q_H (Figure 7.9). We vary the value of σ while keeping the Markovian noise model described above fixed. We find that the value $\sigma = 55 \mu\Phi_0$ best fits both the Ram-Z and Echo-Z data at the same time. This is the value we use in all the simulations in this paper.

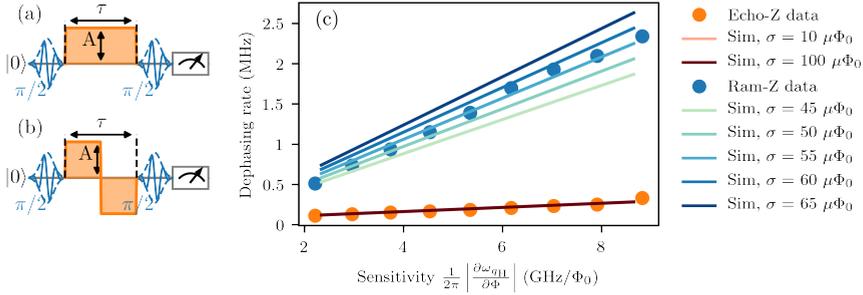


Figure 7.9.: Comparison of experimental data and simulation (c) for the Ram-Z (a) and Echo-Z (b) experiments. In the Ram-Z (Echo-Z) experiment, the dephasing time is measured using a (two-half) square flux-pulse(s). All simulated curves include the effects of both the Markovian and non-Markovian noise. Only the strength of the non-Markovian noise [Equation (7.22)], quantified by σ , is varied, while the strength of the Markovian noise, quantified by T_{1,q_H} and $T_{\phi,q_H}^E(\Phi)$, is kept fixed. We see that the value $\sigma = 55 \mu\Phi_0$ best fits the Ram-Z data. It fits the Echo-Z data as well, given that the simulated curves are equal even for σ 's that differ by an order of magnitude. This agrees with the intuition that the non-Markovian noise is echoed-out in an Echo-Z experiment.

Quantities of interest

To quantify the quality of the CZ gate, we are interested in computing the conditional phase, the leakage and the average gate fidelity from the propagator $\mathcal{P}_{TCZ}^{\text{av}}$. In the following, we summarize their definitions for a generic superoperator \mathcal{P} .

We call \mathcal{X}_1 the computational subspace, spanned by the 2-qubit energy levels $|00\rangle, |01\rangle, |10\rangle$ and $|11\rangle$ at the operating point. The phases acquired by those states under the action of \mathcal{P} are computed as

$$e^{i\phi_{ij}} = \frac{\langle ij | \mathcal{P}(|ij\rangle \langle 00|) |00\rangle}{|\langle ij | \mathcal{P}(|ij\rangle \langle 00|) |00\rangle|}, \quad (7.23)$$

where $i, j \in \{0, 1\}$. If \mathcal{P} is unitary, that is, $\mathcal{P}(\rho) = U\rho U^\dagger$ for some unitary U , then Equation (7.23) reduces to $e^{i\phi_{ij}} = \frac{\langle ij | U | ij \rangle}{|\langle ij | U | ij \rangle|}$, and, if U is diagonal, then we simply have $U |ij\rangle = e^{i\phi_{ij}} |ij\rangle$. The phase ϕ_{00} of the ground state can be set to 0. The single-qubit phases are given by ϕ_{01} and ϕ_{10} . The conditional phase ϕ_{2Q} is defined as the phase acquired by the target qubit conditional on the state of the control qubit and it is given by

$$\phi_{2Q} = \phi_{11} - \phi_{10} - \phi_{01}. \quad (7.24)$$

Note that ϕ_{2Q} is invariant under single-qubit Z rotations.

We follow the definitions in [151] for leakage, seepage and average gate fidelity. The leakage of a superoperator \mathcal{P} is defined as

$$\begin{aligned} L_1 &= 1 - \int_{\psi_1 \in \mathcal{X}_1} d\psi_1 \operatorname{Tr}_{\mathcal{X}_1} \left(\mathcal{P}(|\psi_1\rangle\langle\psi_1|) \right) \\ &= 1 - \frac{1}{\dim \mathcal{X}_1} \sum_{i,j \in \{0,1\}} \operatorname{Tr}_{\mathcal{X}_1} \left(\mathcal{P}(|ij\rangle\langle ij|) \right). \end{aligned} \quad (7.25)$$

The quantity L_1 represents the average probability that a random computational state leaks out of \mathcal{X}_1 .

The seepage of a superoperator \mathcal{P} is defined as

$$L_2 = 1 - \int_{\psi_2 \in \mathcal{X}_2} d\psi_2 \operatorname{Tr}_{\mathcal{X}_2} \left(\mathcal{P}(|\psi_2\rangle\langle\psi_2|) \right), \quad (7.26)$$

where \mathcal{X}_2 is the leakage subspace.

The average gate fidelity, evaluated in the computational subspace, between \mathcal{P} and a target unitary U is defined as

$$\begin{aligned} F &= \int_{\psi_1 \in \mathcal{X}_1} d\psi_1 \langle \psi_1 | U^\dagger \mathcal{P}(|\psi_1\rangle\langle\psi_1|) U | \psi_1 \rangle \\ &= \frac{\dim \mathcal{X}_1 (1 - L_1) + \sum_k \left| \operatorname{Tr}_{\mathcal{X}_1} (U^\dagger A_k) \right|^2}{\dim \mathcal{X}_1 (\dim \mathcal{X}_1 + 1)}, \end{aligned} \quad (7.27)$$

where the $\{A_k\}$ are the Kraus operators of \mathcal{P} . The average gate infidelity is defined as $\varepsilon = 1 - F$. We can see from Equation (7.27) that F is affected by L_1 . There are two contributions: one is explicit in the first term at the numerator, the other is implicit in the second term and is due to the fact that the Kraus operators of a leaky superoperator are in general different from the ones of a non-leaky superoperator. For a two-qubit gate, the explicit contribution to ε is equal to $L_1/5$, whereas the implicit one is evaluated numerically.

7.5.4 Conditional oscillation experiment

The conditional oscillation experiment (Figure 7.10) can be used to measure the single-qubit phases (ϕ_{01} and ϕ_{10}) and the conditional phase (ϕ_{2Q}), and to estimate the leakage (L_1) defined in Equation (7.25). In the conditional oscillation experiment, two variants of the same experiment are performed. In the first variant (Off), the target qubit ($q_{\text{targ.}}$) is rotated onto the equator of the Bloch sphere by a $\pi/2$ pulse and the control qubit ($q_{\text{contr.}}$) is left in the ground state. After that, a flux pulse is applied that is intended to perform a CZ gate. A recovery $\pi/2$ rotation, performed around an axis in the equatorial plane forming an angle ϕ with the X axis, is applied to $q_{\text{targ.}}$ before measuring the state of both qubits simultaneously. In the second variant (On), $q_{\text{contr.}}$ is rotated into the excited state before applying the CZ gate. Then, $q_{\text{contr.}}$ is pulsed back to the ground state before measuring both qubits.

The conditional phase ϕ_{2Q} can be extracted directly from the phase of the oscillations and corresponds to the difference in phase between the oscillations (Figure 7.10). The single-qubit phase ϕ_{10} (ϕ_{01}) can be measured by letting q_M (q_H) take the role of $q_{\text{targ.}}$ and correspond directly to the measured phase of $q_{\text{targ.}}$ in the Off variant.

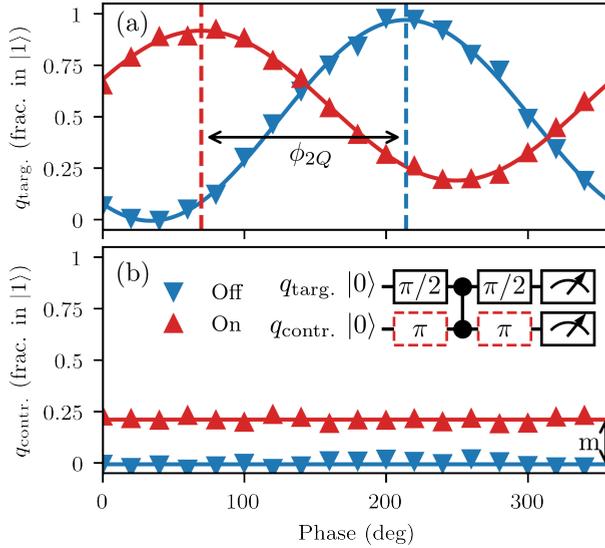


Figure 7.10.: The conditional oscillation experiment described in Section 7.5.4.

The quantity denoted by m in Figure 7.10 is called the missing fraction. In the idealized case in which there is no noise and no leakage to other levels, we calculate $L_1 = m_{\text{idealized}}/2$. We see numerically that such relation approximately holds in the complete modeling with noise. Therefore, we define a leakage estimator $\tilde{L}_1 = m/2$, where m is the measured value. Due to relaxation effects, \tilde{L}_1 generally overestimates L_1 . The advantage of estimating the leakage with \tilde{L}_1 rather than with a randomized benchmarking experiment (Section 7.5.7) is that it is much faster. In this way we can quickly acquire a scan of the leakage landscape to find pulse parameters giving a low-leakage CZ gate. Further characterization is then carried out with randomized benchmarking.

7.5.5 Optimal performance

Using simulations, it is possible to find the optimal parameters (θ_f and λ_2) for a given T_{2Q} in order to perform a CZ gate. We optimize over the infidelity ε . In Figure 7.11, the minimal infidelity ε and the corresponding leakage L_1 are shown as a function of T_{2Q} . Contrary to all the other figures in this paper, the simulations shown in Figure 7.11 do not include the effect of distortions. The shortest duration for which a NZ pulse with low leakage and high fidelity can be performed is $T_{2Q} = 28$ ns, close to the speed limit of $T_{2Q} = 25$ ns, set by the interaction strength. The difference in minimal infidelity between the unipolar and the NZ pulse is attributed to the built-in echo effect that makes the NZ pulse resilient to low-frequency flux-noise components. Unipolar pulses with good performance could in principle be realized slightly faster ($T_{2Q} = 26$ ns) than NZ pulses, due to the fact that NZ needs ~ 2 ns to sweep from one avoided crossing to the other in the middle of the pulse, during which no conditional phase is accumulated. We remark that we can study the performance of

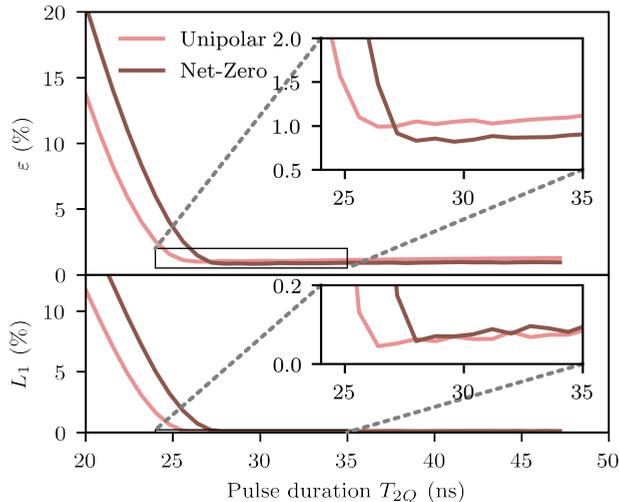


Figure 7.11.: Minimal infidelity (ε), optimized over θ_f and λ_2 for a fixed T_{2Q} ($T_{1Q} = 12$ ns for all T_{2Q}), and leakage (L_1) evaluated at the minimal infidelity. Contrary to all other figures in this paper, the simulations shown here do not include distortions because we want to quantify the intrinsic optimal performance of unipolar and NZ pulses against Markovian and non-Markovian noise. We see that both ε and L_1 decrease fast approaching the speed limit $\pi/J_2 \sim 25$ ns. Then NZ achieves lower infidelity and we can attribute this to the echo effect. We can use these simulations to find that the minimal T_{2Q} to realize a high-fidelity, low-leakage NZ pulse is $T_{2Q} = 28$ ns.

7

a single application of the unipolar CZ gate in simulation, but that this is not representative of the performance in the experiment since the unipolar pulse is not repeatable as demonstrated in the main text.

The simulated landscape of the shortest duration ($T_{2Q} = 28$ ns) high-fidelity low-leakage NZ pulse is compared to experiment in Figure 7.12. There is a relatively large region of low leakage at high θ_f (90-130 deg) that can be found in both simulation and experiment. The $T_{2Q} = 28$ ns pulses described in the main text are operated at the marked point ($\theta_f = 125$ deg, $\lambda_2 = -0.1$).

7.5.6 Net-Zero pulses as a Mach-Zehnder interferometer

To better understand the working of a NZ pulse, it is helpful to draw an analogy to a Mach-Zehnder interferometer [251–256]. In a NZ pulse, the trajectory first approaches the $|11\rangle \leftrightarrow |02\rangle$ avoided crossing at positive flux amplitude, then it sweeps through the sweetspot, and it finally goes in and out of the $|11\rangle \leftrightarrow |02\rangle$ avoided crossing at negative flux amplitude. We argue that those three parts of the pulse correspond respectively to an (unbalanced) beamsplitter, to the arms of an interferometer, and to another (identical) beamsplitter. We make a few idealizations in this analysis. Namely, we ignore the weak coupling to other states and

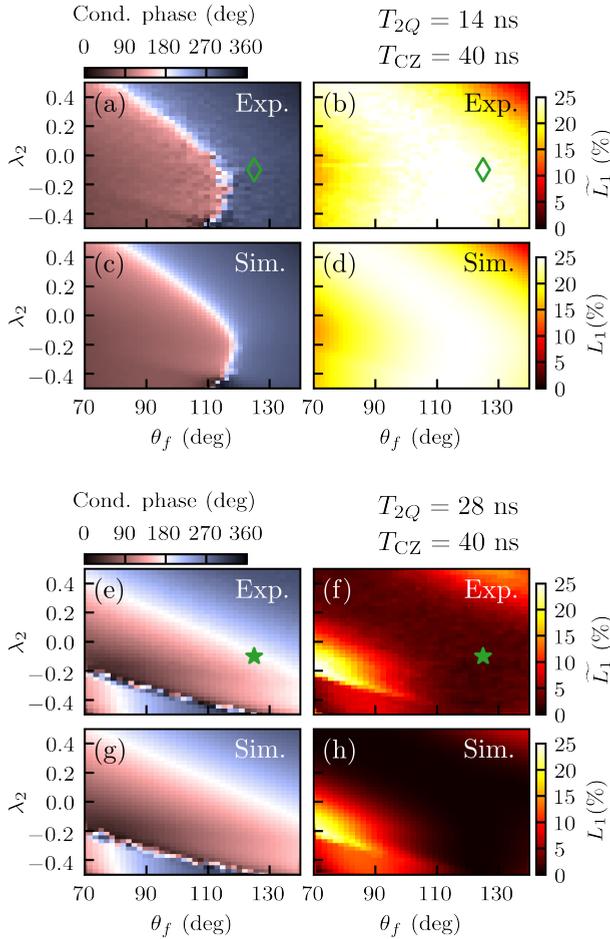


Figure 7.12.: Matching of experimental (a,b,e,f) and simulated (c,d,g,h) landscapes of conditional phase and leakage as a function of the parameters θ_f and λ_2 of a $T_{2Q} = 14$ ns unipolar (a,b,c,d) pulse and of a $T_{2Q} = 28$ ns NZ (e,f,g,h) pulse. The $T_{2Q} = 28$ ns NZ pulse consists of two concatenated $T_{2Q} = 14$ ns unipolar pulses with opposite polarity. Phase corrections are appended to get a total length $T_{CZ} = 40$ ns. We find that the matching is excellent in both cases. The star (green) marks the point ($\theta_f = 125$ deg, $\lambda_2 = -0.1$) used in the interleaved randomized benchmarking experiment described in the main text. A transparent diamond (green) marks the corresponding point for the $T_{2Q} = 14$ ns unipolar pulse. Given that the $T_{2Q} = 14$ ns unipolar pulse does not show regions of low leakage, we conclude that the broad area of low leakage for the $T_{2Q} = 28$ ns NZ pulse is a fringe of destructive leakage interference. We have verified this also by varying the interference condition and observing this fringe move across the landscape, similarly to Figure 7.14 and as described in Section 7.5.6.

we consider a purely unitary process. Moreover, there is not a clear-cut separation between the beamsplitters, where the qubits are strongly coupled, and the arms of the interferometer, where they are effectively uncoupled. However, since the sweep in the middle is very fast, for the sake of this model it does not really matter where the line is drawn.

In general, a unipolar pulse has the following effect on the $|11\rangle$ state

$$|11\rangle \mapsto e^{i\phi_{2Q}^{\text{half}}} \sqrt{1-\alpha^2} |11\rangle + \alpha |02\rangle, \quad (7.28)$$

where $\alpha \in \mathbb{R}$ and $\alpha^2 = 4L_1^{\text{half}}$ (assuming no leakage to other states). In other words, during the first half of a NZ pulse, $|11\rangle$ acquires a certain conditional phase ϕ_{2Q}^{half} and it can also leak to $|02\rangle$, for example if the parameters of the pulse are not properly chosen or if the pulse is too short.

Unitarity implies that $|02\rangle \mapsto \alpha |11\rangle - e^{-i\phi_{2Q}^{\text{half}}} \sqrt{1-\alpha^2} |02\rangle$. Overall, modulo a global phase, this amounts to the unitary

$$B_1 = \begin{pmatrix} e^{i\phi_{2Q}^{\text{half}}} \sqrt{1-\alpha^2} & \alpha \\ \alpha & -e^{-i\phi_{2Q}^{\text{half}}} \sqrt{1-\alpha^2} \end{pmatrix}, \quad (7.29)$$

which is a beamsplitter that also imparts a conditional phase.

During the sweep across the sweetspot, $|11\rangle$ and $|02\rangle$ quickly acquire a relative phase φ due to the large energy gap between them (~ 800 MHz). We can formalize this with the unitary

$$P_\varphi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\varphi} \end{pmatrix}, \quad (7.30)$$

which is a phase shifter.

The second beamsplitter, B_2 , is equal to B_1 due to the symmetry of the pulse. The total evolution is given by

$$B_2 P_\varphi B_1 = B_1 P_\varphi B_1 = \begin{pmatrix} e^{i2\phi_{2Q}^{\text{half}}} \left((1-\alpha^2) + \alpha^2 e^{i\tilde{\varphi}} \right) & \alpha \sqrt{1-\alpha^2} e^{i\phi_{2Q}^{\text{half}}} (1 - e^{i\tilde{\varphi}}) \\ \alpha \sqrt{1-\alpha^2} e^{i\phi_{2Q}^{\text{half}}} (1 - e^{i\tilde{\varphi}}) & \alpha^2 + (1-\alpha^2) e^{i\tilde{\varphi}} \end{pmatrix}, \quad (7.31)$$

where $\tilde{\varphi} := \varphi - 2\phi_{2Q}^{\text{half}}$. We are interested in the first matrix element because it gives the leakage L_1^{NZ} and conditional phase ϕ_{2Q}^{NZ} at the end of a NZ pulse. Explicitly

$$L_1^{\text{NZ}} = (\alpha^4 + (1-\alpha^2)^2 + 2\alpha^2(1-\alpha^2)\cos\tilde{\varphi})/4, \quad (7.32)$$

$$\phi_{2Q}^{\text{NZ}} = 2\phi_{2Q}^{\text{half}} + \arctan\left(\frac{\alpha^2 \sin\tilde{\varphi}}{(1-\alpha^2) + \alpha^2 \cos\tilde{\varphi}}\right). \quad (7.33)$$

There are two cases in which L_1^{NZ} can be made zero. The first one is when $\alpha^2 = 0$. This is when the half pulse has zero leakage in the first place. We refer to this case as the adiabatic condition. The second case is when $\alpha^2 \neq 0$ but $\tilde{\varphi} = (2k+1)\pi$, with k an integer. We refer to this second case as the interference condition. We point out that, in either case, the second term in Equation (7.33) is zero, which implies that $\phi_{2Q}^{\text{NZ}} = 2\phi_{2Q}^{\text{half}}$ whenever $L_1^{\text{NZ}} = 0$. As a

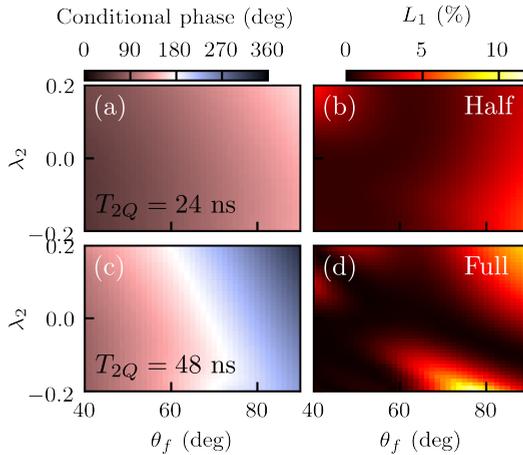


Figure 7.13.: Simulation of conditional phase and leakage landscapes as a function of the parameters θ_f , λ_2 of a half (a,b) and full (c,d) $T_{2Q} = 48$ ns NZ pulse ($T_{CZ} = 60$ ns). The half pulse consists of only the first part of the NZ pulse, which is effectively a $T_{2Q} = 24$ ns unipolar pulse ($T_{CZ} = 60$ ns). Naively one may expect both the conditional phase and the leakage of the full pulse to be approximately twice that of the half-pulse. However, this is not the case for the leakage. In (b) we see a low-leakage area due to the adiabaticity of the pulse. We find this low-leakage area in (d) as well. However, an interference fringe is visible that does not occur for the half pulse.

consequence, the speed limit to do a NZ CZ with low leakage is the same as for the unipolar pulse (π/J_2). We also note that if $L_1^{\text{half}} = \alpha^2/4$ is large and if L_1^{NZ} is low, it follows that the latter must result from destructive interference of leakage.

It is possible to explore both the adiabatic and interference conditions for low leakage in the simulations (Figure 7.13). When performing a $T_{2Q} = 24$ ns unipolar (Half NZ) pulse, only the adiabatic condition can be used to achieve a low leakage. This condition is visible as the dark region in [Figure 7.13(b)]. When simulating a $T_{2Q} = 48$ ns (Full) NZ pulse, a low-leakage fringe is visible [Figure 7.13(d)] corresponding to the interference condition.

The position of the interference fringe should depend on the time between the two halves of the pulse. This can be explored by adding a buffer time Δt between the two halves of the pulse in simulation. For a $T_{2Q} = 40$ ns + Δt pulse, the fringe can be seen to move over the leakage landscape (Figure 7.14). The period corresponds to the expected period of $\sim 1/800$ MHz = 1.25 ns.

7.5.7 Leakage modification for randomized benchmarking

Leakage out of the computational subspace is determined using the protocol introduced in [151], which constitutes a modification of the randomized benchmarking protocol.

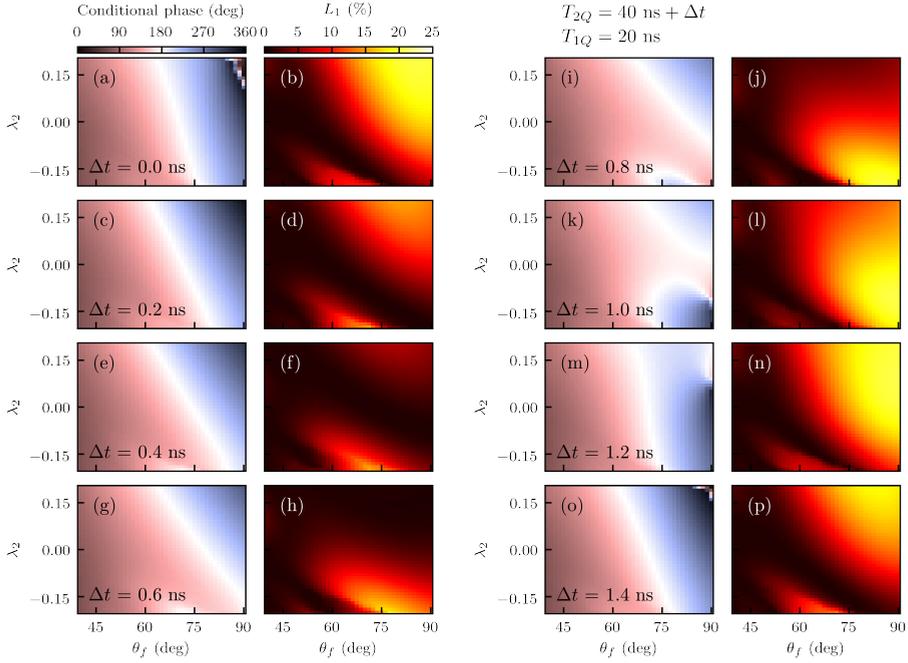


Figure 7.14.: Moving interference fringes. To observe the effect of changing the length of the arms of the interferometer, a buffer (Δt) is added between the first and second part of the strong NZ pulse ($T_{2Q} = 40 \text{ ns} + \Delta t$, $T_{1Q} = 20 \text{ ns}$) in simulation. The low-leakage fringe can clearly be seen to move over the landscape.

To determine the populations in the ground (g), first-excited (e), and second-excited (f) states we follow the procedure described in [92]. In this procedure, a given experiment is performed in two different variants: once in the normal way, giving signal S_I , and once with a π pulse on the $g - e$ transition appended at the end of the sequence just before the measurement, giving signal S_X . When the respective reference signals V_0 , V_1 , and V_2 of a transmon qubit prepared in the g , e and f state are known, the respective populations of the g and e states, P_0 and P_1 , can be extracted using

$$\begin{bmatrix} V_0 - V_2 & V_1 - V_2 \\ V_1 - V_2 & V_0 - V_2 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} = \begin{bmatrix} S_I - V_2 \\ S_X - V_2 \end{bmatrix}, \quad (7.34)$$

under the assumption that higher-excited levels are unpopulated (in other words, $P_0 + P_1 + P_2 = 1$, where P_2 is the population in the f state).

Following [151], we fit the population $P_{\mathcal{X}_1}$ in the computational subspace \mathcal{X}_1 to a single exponential

$$P_{\mathcal{X}_1}(N_{\text{Cl.}}) = A + B\lambda_1^{N_{\text{Cl.}}}, \quad (7.35)$$

where N_{Cl} is the number of Cliffords. The average leakage (L_1) and seepage (L_2) rates [Equations (7.25) and (7.26)] per Clifford can then be estimated as

$$L_1^{\text{Cl.}} = (1 - A)(1 - \lambda_1), \quad (7.36)$$

$$L_2^{\text{Cl.}} = A(1 - \lambda_1). \quad (7.37)$$

Using the fitted value of λ_1 , the survival probability M_0 is then fitted to a double exponential of the form

$$M_0(N_{\text{Cl.}}) = A_0 + B_0\lambda_1^{N_{\text{Cl.}}} + C_0\lambda_2^{N_{\text{Cl.}}}. \quad (7.38)$$

The average gate infidelity per Clifford $\varepsilon^{\text{Cl.}}$ is given by

$$\varepsilon^{\text{Cl.}} = 1 - \frac{1}{d_1} [(d_1 - 1)\lambda_2 + 1 - L_1], \quad (7.39)$$

with $d_1 = \dim \mathcal{X}_1$. We note that if the leakage is weak ($\lambda_1 \ll \lambda_2$ and $B \ll A$), this reduces to the conventional randomized benchmarking formula. We refer to this experiment as the reference sequence.

This method is used in combination with interleaved randomized benchmarking [146] to extract the average gate infidelity (ε^{CZ}) and leakage (L_1^{CZ}) per CZ gate

$$\varepsilon^{\text{CZ}} = 1 - \frac{1 - \varepsilon^{\text{Int.}}}{1 - \varepsilon^{\text{Cl.}}}, \quad (7.40)$$

$$L_1^{\text{CZ}} = 1 - \frac{1 - L_1^{\text{Int.}}}{1 - L_1^{\text{Cl.}}}, \quad (7.41)$$

where $\varepsilon^{\text{Int.}}$ ($L_1^{\text{Int.}}$) stands for the average gate fidelity (leakage) in the interleaved sequence of the interleaved randomized benchmarking experiment.

Last year's Nature paper should be this year's calibration.

— John Martinis [258]

This chapter discusses the software that is required to control a multi-qubit processor. We start by discussing the requirements for experiment control software and introduce PycQED, an open-source software platform for controlling quantum computing experiments. We discuss both the design philosophy of PycQED as well as the core concepts of this framework and their limitations. We next discuss automated characterization and calibration. We introduce dependency graphs as a useful abstraction, the mock qubit and emulation as an essential development tool and give examples of the kind of system-level analysis that is required to make use of the characterization and calibration data.

Controlling a quantum computing system is an often underestimated challenge. When taking a top-down approach, such as the full-stack view on quantum computing (Section 2.1), one is led to believe that the control problem is limited to defining the right instruction set and compilation of an algorithm. Ensuring that the instructions correspond to the right physical operation would then be a low-level routine that belongs in the domain of analog control electronics. Quantum computers, however, are different in that these calibrations are significantly more complex. Calibrations are typically physics experiments that require both low-level access to control parameters and measured signals, as well as programs and data analysis that would traditionally be defined at the top layer of the stack.

In this chapter, we give an overview of the software required to control a quantum system. We start by giving an overview of the basic lab infrastructure based on PycQED [196], the software platform we developed for this purpose. We first discuss the design philosophy of PycQED before going into the core concepts of the framework. In the next section, we take a look at automation of the characterization and calibration process. When scaling up to hundreds of qubits, the current paradigms start to break down. In the final section, we focus on the limitations of our current approach and outline the developments we anticipate.

8.1 Basic lab infrastructure – design philosophy

In the first place, the software infrastructure must support the needs of an experimental physics lab. These needs revolve around the ability to rapidly perform a wide variety of small scale experiments while tracking the state of the system and allowing the user, an experimental physicist, to continuously monitor what is going on.

PycQED [196] was developed to provide the tools required to control (increasingly) complex experiments in the domain of superconducting quantum computing. PycQED is an open source¹ Python platform for quantum computing built on top of QCoDeS [197] that contains all basic functionality that one expects from an experiment control package and is used for a wide variety of experiments in multiple labs. It is also very much a package that was developed in an academic lab, which means that it suffers from a characteristic lack of documentation, dead or abandoned code, and limited test coverage. This makes PycQED well suited for a case study on what works well and what does not. PycQED was developed as an open-source package with three design goals in mind. PycQED had to be extensible, easy to use, and automatable.

8.1.1 Extensibility

Here, extensible means that it is possible to extend the functionality of PycQED to support experiments and use-cases that we did not anticipate when designing the initial version of the software. Extensibility is reflected in the code having modular functionality that can be reused, and if required, replaced with newer versions or combined in different ways. Having a modular architecture implies a standardization of certain core-concepts that can be reused.

¹Released under the MIT license.

8.1.2 Usability

Easy to use means that a student should be able to run a basic experiment with minimal training. As a consequence, the code should be based around a small set of core concepts that once understood should allow the user to, at least in principle, express any experiment using these concepts. Easy to use also implies that a library of standard experiments should be provided, both to serve as an example for more advanced experiments as well as to quickly start running experiments. Finally, because debugging setups is such a common activity in experimental physics, it is extra important that the code itself is readable and avoids complex constructs. Ease of use is a particularly important requirement as the goal of an experimental physics lab is to do science in the first place, and to only develop software if this supports the primary goal.

8.1.3 Automatability

To automate data acquisition, it is essential to close the loop between experiments and analysis. Closing this loop is significantly easier if all of the experiment control software is written in a single programming language such as python. Although this may seem obvious, it is still surprisingly common for labs to rely on a patchwork of distinct tools for different tasks such as analysis, instrument monitoring, waveform generation, etc., all implemented in different programming languages and tools such as LabView, Mathematica, Igor, and MATLAB². This patchwork of tools is a natural consequence of post-docs and PhD students taking tools they are familiar with and gluing them together under the pressure of short-term scientific results.

To facilitate the automation of experiments, there must be minimal barriers between daily use of the software, consisting of defining basic experiments and performing basic analysis, and advanced usage, such as the automation of experiment routines and setting up advanced analyses. To reduce the barriers between basic and advanced usage, the user is expected to control experiments through a command-line interface (CLI) such as a console or Jupyter notebook [259] instead of through a graphical user interface (GUI). The GUI that is present is meant only for monitoring purposes and includes live visualizations of the running experiment, the status of the instruments, and data browsing. The advantage of this minimal GUI philosophy is that the concepts and code users encounter when controlling their experiments is the same as that used in the measurement libraries. This makes it easier to debug setups when something is not working as expected and makes it possible to contribute the measurement code directly to the group repository, or even a community repository, with minimal changes. The drawback of a minimal GUI is of course that the entry barrier for new users is increased slightly as they need basic programming knowledge.

8.1.4 Open source

Besides the arguments for open science in general [260], there are important practical considerations why experiment control software frameworks should be open source, even in highly

²In our lab, a patchwork of different tools centered around LabView was used before the adoption of PycQED.

competitive academic or commercial settings. These arguments are debuggability, vendor lock-in, and standardization.

When performing experiments it is essential to be able to distinguish physical phenomena from artifacts caused by electronics or programming errors. To be able to effectively debug an experimental setup, it is important that the user can find out exactly what a function or method does. Although documentation can describe what a function or method is supposed to do, it cannot show how this is done. To understand what is going on, it is important to have access to the source code, and, ideally, be able to modify the code for debugging purposes.

The second argument is more commercial in nature. Due to the nature of experimental physics, a lot of projects build on preceding work, creating a lock-in both in terms of experiment code that needs to be rewritten when changing software platforms as well as a lock-in in terms of the training required before one can effectively use a new platform. There are several risks associated with vendor lock-in. First of all, it is possible for the license cost to change significantly. Due to the lock-in, it is often cheaper to pay these increasing costs rather than to invest in an alternative platform. A second risk relates to the priorities of the vendor. A project can be abandoned or change its priorities because of commercial interests. For instance, the party that is leading the project can go out of business, be acquired by a commercial party with different interests³, or simply lose interest in further developing a project. Although a similar risk of diverging priorities exists for open-source projects, this risk is limited to future developments as one can always branch away from the main project.

In addition, it makes sense for academic and private parties to collaborate on the parts of the system that are common as this can reduce costs and improve quality. Working on a common framework for controlling experiments can save on development costs, and in the long run on training costs as people moving between organizations are already familiar with the tools. An additional benefit is that it can result in superior software because of the additional manpower, and the different perspectives provided by the participating parties.

There are of course, also arguments against developing open-source software. For example, the latest experiments and calibration routines are often seen as a competitive advantage and can show competitors what a group is working on, increasing the risk of being scooped. Although I disagree from an open-science perspective, one has to acknowledge the reality of academic and commercial incentives. If one is concerned about sharing unique capabilities, it is possible to separate the common software framework from the latest measurement and calibration routines in different repositories with different licenses. A detailed discussion of different open-source models in science and how to maintain open-source projects is beyond the scope of this thesis but certainly warrants further discussion.

8.2 Basic lab infrastructure – core concepts

PycQED [196] is based on several core concepts. By understanding how these concepts work and how they can be generalized, it should be possible to understand how all of the mea-

³As an example, Labber, a company making experiment control software recently got acquired by Keysight. Although hard to predict, it is unlikely that this will be positive for future compatibility with electronics from different vendors.

surement code works. These concepts are **Instruments and Parameters**, the **Measurement Control, Data, and Analysis**. In this section, we focus on the concept behind the framework, not the implementation. When we do refer to the implementation, this serves as an example to illustrate a point.

8.2.1 Instruments and Parameters

An important task of experiment control software is to track the state of the system. For this, PycQED relies on the Instrument and Parameter classes from QCoDeS [197].

A Parameter is an object that represents a state variable of the system. It contains meta-data for units, labels, and a docstring⁴, as well as methods for setting and getting values, validating the input, and returning the last known value. An important difference with a regular variable is that assigning and retrieving values requires explicit calls to the set and get methods. Although some consider this *unpythonic*⁵ this is done with good reason. Although a parameter essentially represents a variable, one cannot simply assign values to it. Setting or getting a parameter involves actively performing certain actions that (typically) take a non-negligible amount of time, something that should be reflected in how users interact with parameters.

An Instrument is a container for parameters. An Instrument typically corresponds directly to a physical instrument, providing access to the functionality of the hardware through the parameters. By controlling the hardware through this interface, it is possible to track the state of all parameters in the software and provide a similar interface for all instruments regardless of the manufacturer. Logging of parameters is provided through the snapshot method which provides the last known value of all parameters in an instrument. We return to the snapshot when discussing **data**.

Within PycQED, an Instrument does not have to correspond to a physical piece of equipment. A meta instrument is an instrument that can contain references to other instruments but to the user acts like a regular instrument. Examples of meta instruments include a composite AWG, which contains references to n c -channel AWGs but acts as a single $n \cdot c$ channel AWG to the user, or the flux control, which allows setting the flux on individual qubits in units of Φ_0 and controls a multi-channel current source under the hood while taking a flux-crosstalk correction into account.

It can be convenient to use an instrument to track system parameters that do not correspond to a hardware setting. One of the most powerful abstractions of PycQED is that of the qubit and device objects. These instruments are used to store system parameters such as the qubit frequency or the amplitude required to perform a π -pulse and contain methods to configure the underlying hardware as well as methods to perform basic experiments based on the values of these parameters. The experiments and calibration routines defined in the

⁴A docstring is a short string describing what a segment of code does.

⁵Code is considered pythonic when it exploits features of the Python language to produce code that is clear, concise and maintainable. The pythonic way of handling set methods is to overload the assignment operator '='. This aspect of pythonic is expressed in the statement: "There should be one – and preferably only one – obvious way to do it." – The Zen of Python [261]. See also [262] for a short discussion on what pythonic is.

qubit and device objects are part of the libraries used for characterization and calibration. In Section 8.3 we discuss how to automate this process.

8.2.2 Measurement Control

At the core of PycQED is the Measurement Control (MC). The MC is in charge of the data-acquisition loop and is based on the notion that every experiment consists of the same three elements: some parameter(s) is/are varied, some other parameter is measured, the data is stored to disk, and a basic analysis of the data is performed.

Although these elements can be encapsulated in a simple for loop, there are some benefits to standardizing this through a more rigid interface such as the MC. By enforcing a standardized interface, it is possible to standardize data storage, including all the relevant metadata such as units, labels, and instrument settings, which is essential for reusing analysis code. By standardizing the interface for the data-acquisition loop, it is also possible to provide an always-on real-time visualization of the running experiment, a feature that is essential when operating an experiment.

Running a basic experiment using the MC is simple. Listing 8.1 shows a heterodyne spectroscopy experiment. The parameter to vary is set as the `sweep_function`, the `sweep_points` define the values to loop over, and the `detector_function` defines the quantity to measure⁶.

```

1 import numpy as np
2 from pycqed.analysis import measurement_analysis as ma
3 # A simple heterodyne spectroscopy experiment
4 MC.set_sweep_function(source.frequency)
5 MC.set_sweep_points(np.arange(4.65e9, 1e6, 4.85e9))
6 MC.set_detector_function(HeterodyneDetector())
7 MC.run(name='Spectroscopy') # start the experiment
8 ma.MeasurementAnalysis() # perform a standard analysis

```

Listing 8.1.: A basic spectroscopy experiment expressed using the Measurement Control.

One of the downsides of standardization is, of course, that one is less flexible. It might seem paradoxical then, that the real power of the MC is in how it deals with more advanced experiments. There are several smaller problems, that in themselves are not difficult to solve, but do have multiple valid solutions.

A common experiment is to measure a heat map of a quantity y_0 in which a parameter x_0 is varied over n values in an inner loop, and another parameter x_1 is varied over m values in an outer loop. There are two reasonable options on how to save the data: The acquired data can be stored as an $n \cdot m$ array with two smaller arrays of length n and m to denote the

⁶Both `sweep_function` and `detector_function` are technically misnomers as neither is a function. Instead, these are objects that have attributes specifying the labels and units as well as methods to set and or get these quantities. A better naming would be e.g., `Settables` and `Gettables`.

```

1 from pycqed.measurement.optimization import nelder_mead
2 from pycqed.analysis import measurement_analysis as ma
3 # Three parameter restless tuneup
4 G_amp = VSM.mod1_ch1_gaussian_amp
5 D_amp = VSM.mod1_ch1_derivative_amp
6 freq = mw_source1.frequency
7 # define options for the optimizer
8 ad_func_pars = {'adaptive_function': nelder_mead,
9                 'x0': [1, 0, 5.235e9], # the initial values
10                'initial_step': [0.05, 0.05, 1e6],
11                'minimize': True}
12 # configure the MC and perform the experiment
13 MC.set_sweep_functions([G_amp, D_amp, freq])
14 MC.set_adaptive_function_pars(ad_func_pars)
15 MC.set_detector_function(restless_cost_function)
16 MC.run(name='restless tuneup', mode='adaptive')
17 ma.OptimizationAnalysis(label='restless')

```

Listing 8.2.: Restless tuneup, an adaptive loop expressed using the Measurement Control.

setpoints, or the data can be saved as three columns of length $n \cdot m$ denoting the values of x_0 , x_1 , and y_0 , which is compatible with higher dimensional sweeps and non-grid-like heat maps. Both choices have their merits. The power of MC is that it has already made this choice⁷, so that the user does not have to solve this problem again, and analysis and visualization code can make assumptions on how the data is formatted.

A useful characteristic of the MC interface is that there is a hard separation between the parameters being varied, the values over which they are varied, and the quantity being measured. The utility of this abstraction is most evident in the adaptive loop, which was essential in achieving the results of Chapter 4. In *restless tuneup of high-fidelity qubit gates*, emphasis is put on how a calibration routine can be described as an optimization problem and how to define a proper cost function for such an optimization. Listing 8.2 shows an optimization over three parameters using the Nelder-Mead [191] algorithm based on the cost function of Chapter 4. This construct can be used to perform an optimization over any set of parameters, with an optimization algorithm of choice, while minimizing a custom cost function of choice. An optimization algorithm does not have to minimize or maximize a measured quantity, we have also used adaptive-sampling algorithms [263] that maximize the expected information gain per acquired data point.

Many experiments follow the structure of the MC but cannot be controlled by a simple python for loop (mostly) due to hardware constraints. An example of such an experiment is a

⁷PycQED stores the data as individual columns corresponding to the different quantities.

VNA sweep, in which a vector network analyzer measures the S parameters of a system for a set of frequencies. This experiment is virtually identical to a frequency sweep using a spectrum analyzer⁸. In the first example, one has to configure the VNA and trigger it, the hardware is in control of the data acquisition loop. In the second example the frequency of a microwave source is set and the transmission is measured on the spectrum analyzer in a for loop controlled by the MC. For both experiments, the data should be stored in the same way, and the experiment should be set up in the same way. To achieve this, the MC makes a distinction between soft and hard experiments. In soft experiments, the MC is in control of the data-acquisition, while for hard experiments, the MC is not in control of the data-acquisition loop. For hard experiments, the sweep and detector functions contain additional prepare methods that ensure the hardware is properly configured before triggering the experiment. Because hardware generally returns data in its particular format, the hard detector functions also take care of reshaping the data to be compatible with the MC. Allowing the user to use the same interface for these kinds of experiments while taking care of properly configuring the hardware makes it significantly easier to perform complex experiments.

The separation between sweep and detector functions is also useful in hardware controlled experiments. As an example, consider the randomized benchmarking experiment with leakage modification featured in Chapter 7. In this example, the sweep function contains a method to program the randomized benchmarking sequence onto the control hardware, the sweep points determine the number of Cliffords for each data point, and the detector function determines the acquisition mode. Depending on the analysis one wants to perform on the RB data, it might suffice to save only the average outcome for each Clifford, but if one wants to perform advanced analyses on the RB data, one should save the individual shots as complex I,Q values. The separation between sweep and detector functions allows one to replace the detector function based on the desired acquisition mode, while the same detector functions can also be used for other time-domain experiments such as a simple T_1 experiment.

8.2.3 Data

To understand the data paradigm of PycQED it is useful to define what is meant by an experiment, a dataset and (instrument) metadata. As an *experiment*, we take any process in which some parameters are varied while observing others in a controlled manner. According to this definition, any run of the **Measurement Control** is an experiment. All data relating to an experiment is stored in an *experiment container* consisting of the dataset, the metadata specifying the configuration of the setup, as well as graphs and other analytical output relating to the experiment. In PycQED, the experiment container is a folder in the operating system with a unique, and readable identifier based on the date and time (see Figure 8.1). The *dataset* contains the results of the experiment, it roughly corresponds to a table of the experimental data with labels indicating the parameter names and units. The experimental *metadata* specifies the configuration of the setup and contains a snapshot of all parameters of all instruments at

⁸A spectrum analyzer does not include phase information.

the start of the experiment. The dataset and the experimental metadata are implemented as a single HDF5 file⁹.

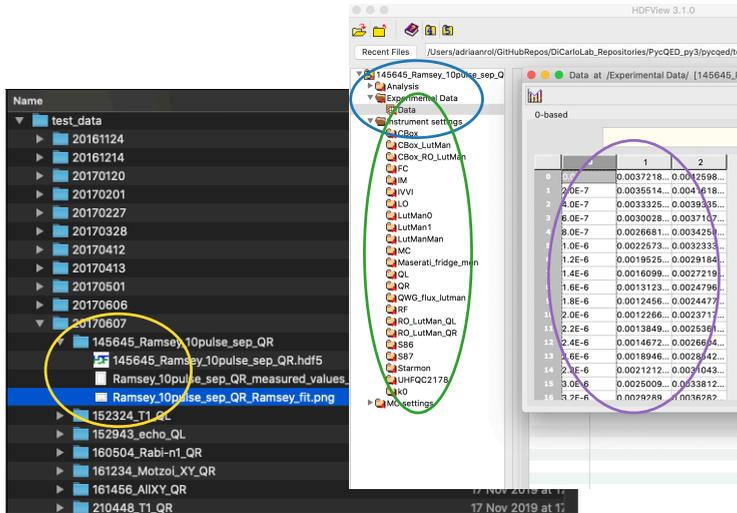


Figure 8.1.: The PycQED data format saves every experiment in its own folder using the date and time as a unique identifier (yellow). The files are stored in HDF5, a hierarchical data format containing multiple groups (blue). A snapshot of all instruments setting (green) is stored alongside the experimental data (purple).

This data paradigm, based on having a single file or folder per experiment, is widespread across the solid-state physics community and a good fit for the requirements of conventional solid-state and small-scale quantum computing experiments. The folder structure is easy to understand and saving the analysis results, such as graphs and fits, next to the data makes it possible to meaningfully browse through the data folder. Because all experiment data is stored in a single experiment folder, it is straightforward to share data with a collaborator.

The limitations of this paradigm become evident when considering slightly more complex experiments. Take for example the experiments described in Yan *et al.* [172] and Luthi *et al.* [241] in which relaxation and dephasing rates are investigated as a function of flux bias. For every T_1 and T_2 data point several experiments need to be performed: find the resonator frequency, calibrate the π -pulse amplitude using a Rabi experiment, and measure the relaxation and dephasing times. The analysis of this experiment requires extracting, and filtering data from many different datasets¹⁰. In this particular example, one would create a for loop that iterates over all experiments with "T1" or "T2" in the filename and extracts the flux-bias and frequency settings from the metadata, and the coherence time from the analysis result. One would then apply some filtering to eliminate invalid data. Although there is nothing pre-

⁹The Hierarchical Data Format version 5 (HDF5), is an open source file format that supports large, complex, heterogeneous data.

¹⁰Another example would be a measurement of DC-flux crosstalk, in which multiple qubit spectroscopy experiments are combined to determine the sensitivity to different flux-bias sources. What complicates this experiment is that multiple qubits are often measured simultaneously.

venting this type of analysis, the focus of the data paradigm on an experiment as a single atomic unit makes it challenging to deal with relations between experiments, leaving the user with the manual task of extracting this knowledge¹¹. It should be clear that this approach does not scale to larger and more complex systems where relations between experiments become more important.

An often proposed solution to these problems is to implement the data storage using a database management system (DBMS). Using a DBMS based on SQL, or on a NoSQL alternative, allows the use of a structured language for querying the database, network access to the database, and mature tools for data handling. Additionally, these database technologies are designed to scale. However, there are several disadvantages to using a DBMS that may explain the lack of adoption in the experimental physics community. DBMSes tend to be more complex to work with than simply writing and reading to a file on disk and require specialized expertise to set up and maintain. The users need to learn a specialized querying language to interact with the database and face constraints designed to protect the data integrity when multiple users or processes are accessing the database. A consequence of this increased complexity is that the data becomes less portable, forcing the user to revert to storing data in text files when working offline or sharing data with a collaborator. On top of that, the tools to browse the data, and look at the results of previously performed analyses, are often lacking, partially because it is not clear where to store the outcome of an analysis. And finally, if the data paradigm is the same, a query used to select a specific dataset is asking the database the same question whether it is expressed in (No)SQL or python¹².

It should be clear then, that simply taking the existing data paradigm and implementing it in a DBMS will not work. For a DBMS system to be adopted, it needs to take into account both the technical aspects of database design that provide scalability, as well as provide the utilities that allow the user to browse the data, manipulate and store the data locally, store analysis results, and share data with collaborators. A proper DBMS solution takes this a step further and rethinks the data paradigm to deal with datasets that contain data from multiple experiments natively, handles the system configuration metadata in a smarter way, and supports tracking of derived device level metrics such as coherence times and fidelities.

8.2.4 Analysis

All experiments produce data that needs to be analyzed. Most analyses involve one or more of the following steps. First, the data is extracted and some basic data processing is performed such as reshaping the data or filtering out invalid data points. Next, the data is fit to some model in order to determine certain quantities of interest. Finally, one or more figures are made and the figures and quantities of interest are saved. These steps are encoded in the `run_analysis()` method of the `BaseDataAnalysis` class in PycQED. Individual analy-

¹¹Because of the manual nature of this task, it is challenging to find relations that one is not explicitly looking for.

¹²The (No)SQL query will probably have a better performance and more consistent syntax than the python version.

ses inherit from this base class¹³ and implement their own initialization and data processing methods as well as specifying the fits to be performed and the figures to be made.

The advantage of this class based structure is that it provides a clear template for new analyses that helps users when implementing their own analyses. The template ensures that the interfaces of the analysis object are uniform, and storing of the analysis results such as figures and fits are in the right location. There are two downsides to this analysis framework, not all analyses follow the steps from `run_analysis()`, and not all analyses are based on a single data file.

Although the base analysis proscribes the steps of an analysis, it is quite simple to also override `run_analysis()` and add more steps to this method. It still makes sense to separate the analysis itself into multiple functions that are called from this method as it helps in preventing the analysis code turning into spaghetti. Besides structuring the analysis code into multiple methods it is good practice to also put reusable code into separate functions.

The second limitation is not inherent to the analysis framework, but relates to the data paradigm described in Section 8.2.3. As a consequence of the data paradigm, which is focused on a single file per experiment, it is difficult to perform analyses which require data from multiple files. The `BaseDataAnalysis` contains methods to help with this data extraction, but these methods are not documented well and can be cumbersome to use. Another consequence of this focus on individual files is that it is often not clear where the figures and quantities extracted from these analyses should be stored.

Looking further into the future, one can imagine that it is desirable to run some analyses in real-time to, e.g., serve as a sub-routine in a larger experiment or to provide a live dashboard. These types of analyses would not fit in the current framework.

8.2.5 Pulse sequencing

Although beyond the scope of PycQED and of this thesis, pulse sequencing is an important part of controlling a prototype quantum computer and a subject in itself. To perform a basic time-domain experiment such as a T_1 experiment, or a more complex quantum algorithm, one needs to program pulses in an AWG. Pulse sequencing is the task of determining which pulse to play when and on what channel. For our purposes, it is useful to give a brief overview of the tools available for this task. We make a distinction between first-generation, second-generation and third-generation pulse sequencing software.

With first-generation pulse sequencing, we refer to code that was used to program long numerical waveforms on hardware such as the Tektronix AWG5014. This code was typically developed by PhD students and in general very hardware and experiment specific. Because of the hardware-specific nature of these tools, they require the user to be intimately familiar with the details of the hardware. An advantage of these tools is that they are incredibly flexible in the programs that can be specified, allowing, for example, sample-specific timing control

¹³Within PycQED there exist two base analysis classes, a legacy version, `MeasurementAnalysis` located in `pycqed.analysis.measurement_analysis` and `BaseDataAnalysis` located in `pycqed.analysis_v2.base_analysis`. This section discusses the concepts behind the second version of the analysis module.

and numerical predistortion filters taking into account the full history of an AWG channel. Examples of first-generation pulse-sequencing software are the Mathematica notebooks used in DiCarlo lab up to 2015 and qtlab pulsar developed in the Hanson Lab.

The second-generation pulse sequencing software was developed based on the realization that specifying quantum programs as long numerical waveforms does not scale. Second-generation pulse sequencing software focuses strongly on the circuit model of quantum computing and typically attempts to hide details relating to the hardware implementation from the user. Most of the quantum languages, such as cQASM [264], eQASM [53], openQASM/QISKIT [265, 266] CirQ [75], projectQ [55, 267], OpenQL [268], and Quill [269] can be considered second-generation pulse sequencing software. The codeword based QuSurf architecture featured in this thesis is a good example of a second-generation quantum control system. We recommend Fu [134] for an extensive discussion on eQASM and the instruction set that forms a central part of sequencing in the QuSurf architecture.

Third-generation pulse sequencing software is based on the idea that it is useful to provide access to the details of the control pulse to the user. Third-generation pulse sequencing software tries to combine the flexibility of first-generation sequencing with the scalability of second-generation sequencing. This is particularly relevant as most of the experiments that are performed are calibrations, not algorithms, and these require pulse level control. Although it seems very similar to first-generation sequencing software, the implementation is generally a lot more mature as well as being to a large extent hardware agnostic. I consider pulse sequencing software third generation if it has the capability to, at least in principle, take advantage of hardware features that allow more efficient sequencing. Examples of third-generation pulse sequencing software include QISKIT Open Pulse [270], QGL [271], and QuPulse [272].

8.3 Automated characterization and calibration

Due to variations in the fabrication process all qubits need to be individually characterized and calibrated before the quantum processor can be operated as a quantum computer. The goal of the characterization process¹⁴ is to learn properties of the system such as resonator quality factors, qubit frequencies, crosstalk, gate fidelities, etc. This information can be used to improve the device design and fabrication. This task is challenging because system parameters can fluctuate over time (Section 3.4.2), depend on each other, and suffer from crosstalk (Section 3.5). For small processors (≤ 7 qubits) the characterization and calibration is typically done manually. In this semi-manual procedure, common experiments such as a qubit spectroscopy or a Rabi experiment are automated while the human operator updates settings based on the measurements and determines what experiments to perform next. Quantities of interest, such as the maximum qubit frequency, crosstalk or gate fidelity, are collected and summarized in a digital notebook.

There are several problems with this semi-manual approach. First of all, it is very labor intense, fully characterizing a single 7-qubit device can take up several weeks. Second, because the user decides what experiments to perform, the procedures tend to be slightly

¹⁴Characterization protocols are discussed extensively in Chapter 3.

different for each device, depending on the experience of the operator and changing priorities. Because the procedure is subtly different each time, it is hard to determine how long the process takes and what the optimal procedure is. Third, because the data collected during these processes is summarized in digital notebooks it requires considerable extra effort to compare metrics across different devices. Most important of these limitations is that this approach does not scale.

In this section we discuss **graph-based tuneup** using AutoDepGraph [273], **system emulation** and **system-level analysis**. **Graph-based tuneup** is a framework to combine different calibration experiments and automate the characterization and calibration process. **System emulation** is an essential part of developing and maintaining automated calibration routines. In **system-level analysis** we discuss how to make use of the data collected during the calibration process. We do not discuss individual characterization and calibration routines here as these have been extensively covered in this thesis. Of special note is the use of optimal control theory to construct control pulses. We recommend the lecture notes by Wilhelm *et al.* [274] for an introduction into the subject.

8.3.1 Graph-based tuneup

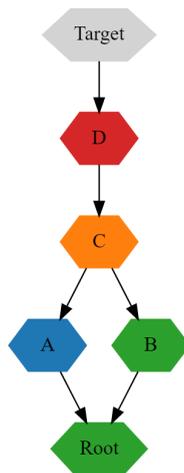


Figure 8.2.: Example of a directed acyclic graph used for calibrations. The color of each node indicates its state: *active* (blue), *good* (green), *needs calibration* (orange), *unknown* (gray) and *bad* (red). Edges determine the order of calibration.

AutoDepGraph [273] is an open source¹⁵ framework for calibrations that is strongly inspired by Kelly *et al.* [275] where the use of a *directed acyclic graph* (DAG) is proposed to encode the calibration problem. Encoding calibrations in a DAG later proved to be essential in demonstrating quantum supremacy [36]. AutoDepGraph featured prominently in the Bachelor project of van Abswoude [276] where it was used to automate single-qubit char-

¹⁵Released under the MIT license.

acterization. In this framework, each calibration step corresponds to a *node* in a DAG. The nodes have a *state* attribute indicated by color in the example graph in Figure 8.2. Directed edges visualized by arrows indicate the dependencies of the nodes. By calling a node, the node is calibrated. To do this, the logic within each node will first check if the parents are in a good state and (recursively) calibrate them if this is not the case. If all parents are in a good state, the node will calibrate itself.

As an example, consider the Rabi node in Figure 8.3. To calibrate the π -pulse amplitude, one needs to know the qubit frequency, have a working readout and calibrate the mixers. The Rabi node in Figure 8.3 depends directly on the mixer calibrations and the qubit frequency, and indirectly, through multiple spectroscopy nodes, on the readout. After ensuring the parents nodes are in a good state, the node has access to a check and a calibrate method. The calibrate method performs an experiment, analyzes the data, and updates the parameters. In the case of a π -pulse calibration, the experiment would be to measure a Rabi oscillation, the analysis would fit a cosine to the data which is then used to update the pulse-amplitude parameter. A check is a quick experiment to determine if (re)calibration is required. In the example of the π -pulse amplitude, measuring a few data points of the Rabi oscillation and assessing that these correspond to the expected result often suffices to declare the system "in spec" and conclude that no calibrations are required.

Thinking about calibrations in terms of a DAG can be rather abstract and one might wonder what the point is. The DAG itself does not provide the automation, this still happens within the automated experiment routines. Rather, the DAG is the glue that connects different experiments together. The framework proposes a clear definition and interface of check and calibration experiments. This makes it possible to work on individual experiments in isolation, greatly improving the modularity and testability of the code.

Another advantage of encoding calibrations in a DAG is that it frames the problem in a manner that facilitates thinking and discussing. If you, the reader, are an expert in circuit QED experiments, Figure 8.3 will most likely trigger several questions. Why does the Rabi calibration require the qubit to be at the sweetspot? Why does the Rabi depend on mixer skewness if it is only used to measure coherence times? Most likely, these questions did not occur when reading the preceding paragraph where we stated that the Rabi calibration depends on the mixer calibrations, the qubit frequency and a readout calibration. Besides questions on the structure of the graph itself, it also raises questions on calibrations that defy the existing graph structure. How does one include a multiplexed T_1 experiment that measures the relaxation time of multiple qubits? Is it possible to replace multiple nodes of qubit calibration with a single adaptive calibration? Using a DAG makes it significantly easier to have discussions about the system-level characterizations and converge on an optimal, and uniform, calibration procedure.

8.3.2 System emulation

One of the most challenging tasks in automating calibrations is the development of the measurement code. Because a live setup is typically required for testing, code development is

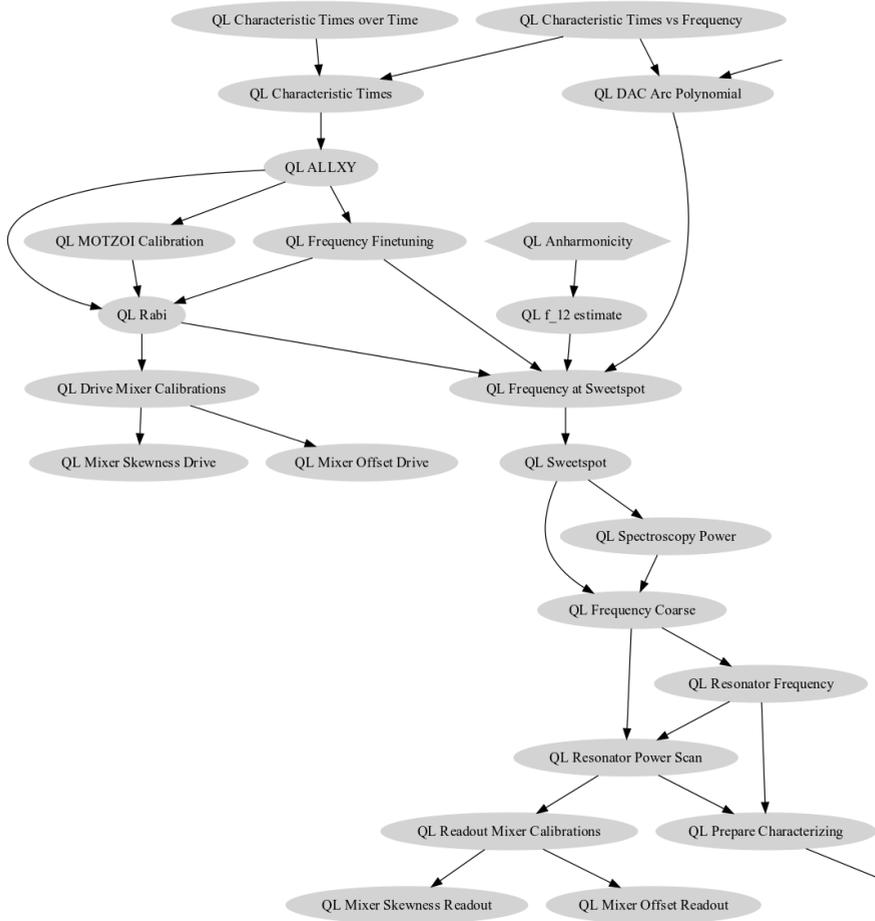


Figure 8.3.: Example calibration graph for single-qubit calibrations, part of a larger graph. Node labels denote individual calibration experiments, QL is short for "Qubit Left". Graph from [276], figure generated using the real-time calibration monitor of AutoDepGraph [273].

time consuming and performing (automated) regression tests is difficult. The solution to this problem is to create a system emulator. The goal of this emulator is not to accurately model (simulate) the system but to qualitatively emulate the system behavior. An emulation of an experiment can be significantly faster than doing it on a live setup, for a typical time-domain experiment (e.g., a Rabi) emulation takes a fraction of a second where the same experiment takes on the order of minutes on a live setup. An emulation can also be easily modified to cover edge cases or model effects that are not fully understood yet. This is important as it is often difficult or undesired to reproduce certain failure modes in a live setup. A side benefit of developing an emulator together with the measurement code is that one is forced to explicitly model how the system should behave. This improves the understanding of the system and the quality of the code.

As part of our efforts to automate qubit calibrations in PycQED, we have created mock qubit and mock device objects. The readout signal in the mock transmon is based on a simple analytical model describing the resonator response. The measured transmission S_{21} at a drive frequency f is described by [277]:

$$S_{21} = A \left(1 + \alpha \frac{f - f_r}{f_r} \right) \left(1 - \frac{Q_l |Q_e| e^{i\theta}}{1 + 2iQ_l \frac{f - f_r}{f_r}} \right) e^{i(\phi_v f + \phi_0)}, \quad (8.1)$$

where A is the transmission amplitude away from resonance, f_r is the qubit-state dependent resonance frequency and α allows for a linear variation in the overall transmission chain in the frequency range around the resonator. Q_l (Q_e) is the loaded (extrinsic) quality factor. ϕ_v and ϕ_0 relate to propagation delays to and from the sample. A simple if statement takes care of using different configurations to describe the low-power, non-linear and high-power regimes. From this model, the signal levels for both the $|0\rangle$ and $|1\rangle$ states are determined. If the populations in the $|0\rangle$ and $|1\rangle$ states are known, the experimental data can be faked. In the same way that the readout is emulated based on a simple analytical model, the populations in the $|0\rangle$ and $|1\rangle$ states can either be based on a circuit level simulation or on a simple analytical model. A Rabi oscillation, for example, can be based on a cosine function whose amplitude and phase depend on the amplitude and detuning of the driving pulse. All parameters in these models can be configured to emulate different scenarios. When these parameters are kept hidden from the user, and with the addition of a configurable noise level, it is possible to construct a fairly sophisticated emulation of the qubit characterization process. This can be used both for educational purposes¹⁶ and to perform integration testing.

8.3.3 System-level analysis

8

The final ingredient in automated system characterization and calibration is the analysis. Where the analysis described in Section 8.2.4 focuses on a single experiment, the analysis here is on a system level. The data from the characterization process can be used to provide insight on the performance, compare metrics or statistics between different devices and calibrations or to improve the characterization process itself.

The first and most obvious use of the characterization data is to summarize the device performance. The simplest way of doing this is to collect this information in a large standardized table. Tables are useful when one needs to provide an overview of multiple individual values. However, before looking into the table in detail a visualization is often a better way to get a quick overview of the system. Figure 8.4(a) shows a useful system level visualization that combines the device topology with a quantity of interest such as, in this case, the gate infidelity. This visualization can help in understanding the performance of a device as it makes it easy to relate the numbers to the device topology, and when using consistent color scales, make comparisons between different quantities on the same device or to make comparisons between devices. Similar visualizations are used in Figure 3.8, Fig. 4 of [50] and Figs. S20-S24 of [36].

¹⁶An example notebook for basic qubit characterization using the mock-transmon as an emulator can be found in the PycQED [196] repository under examples/Controlling a Transmock setup.ipynb.

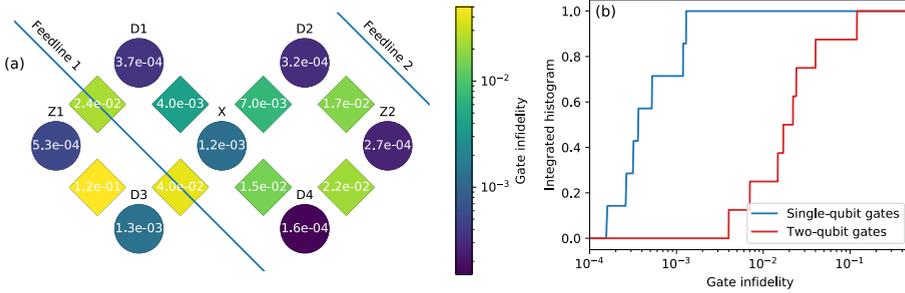


Figure 8.4.: Two different visualizations of gate infidelity for single-qubit and two-qubit gates on the Chimaera device. Performance is characterized using randomized benchmarking protocols A.1.1 and A.1.2. The visualization in (a) combines the device topology with the gate performance, circles correspond to single-qubit gates and diamonds to two-qubit gates. (b) shows the same data in a cumulative histogram.

Where the visualization of Figure 8.4(a) is useful to identify, understand, and compare individual error sources, it is not well suited to assess system performance. To quantify the performance of different calibration schemes and techniques or to compare performance between devices, it is more useful to look at the distribution of errors [Figure 8.4(b)]. A distribution can convey more information than the derived statistical quantities (mean, standard deviation, etc.). Distributions of errors are used in, e.g., Corcoles *et al.* [50] to compare performance of different devices and in Arute *et al.* [36] to assess the impact of different calibration routines.

A third kind of analysis that one might want to perform is not about the device but rather about the system itself. A calibration protocol should ideally be fast, robust, and accurate. By storing metadata on the performance of the DAG, such as the time it takes to execute a node, how often a node needs to be recalibrated, etc., it is possible to identify what calibrations are most in need of improvement.

The largest impediment to performing these kinds of analysis is the data paradigm discussed in Section 8.2.3. The data paradigm does not make it impossible to perform these kinds of analysis, but it does make it difficult to explore the data and perform these analyses on the fly. The three kinds of analysis considered above on the qubit level, the system level, and the process level, expose the requirements for a next-generation data paradigm more explicitly. In addition to the dataset, it needs to provide a good framework to handle experiment metadata, device metadata, quantities of interest extracted from analyzing an experiment, and calibration metadata.

With a proper data infrastructure in place it is relatively straightforward to create standardized reports of based on the calibration or provide a dashboard that displays the device performance in real-time. Looking to the future, I suspect that having this data infrastructure can enable more advanced analyses of data from multiple devices and identify relations between parameters that we have not thought of.

8.4 Summary

In this chapter we have given an overview of the software required to control a multi-qubit system.

Experiment control software should be based on a few core concepts so that it is easy to understand and implement, and have minimal barriers between basic and advanced use to facilitate automation of experiments by the users. The concepts of instruments, parameters and a measurement loop are well established and have fairly mature implementations. Although there is a widespread data paradigm based on a single file per experiment, this paradigm can be limiting when performing more complex experiments. There is no clear solution to this problem and these limitations can make it difficult to perform analyses involving data from multiple experiments. These problems become more evident when moving to larger systems.

When scaling to larger systems it is essential to automate the characterization and calibration procedures. A DAG is a useful abstraction to combine individual routines and give insight in the system level calibrations. To develop and maintain the calibration routines it is useful to simultaneously develop a system emulator as it enables regression tests and enhances the understanding of the system. To make full use of the automated characterization system level analyses have to be performed. There are several visualizations being adopted by the community to give insight into the system level performance. The largest impediment to these system level analyses is currently the data paradigm.

The future is not something to predict. The future is something to build.

— Franco Ongaro (ESA)

In this chapter, we first summarize the conclusions of each chapter of this thesis. Based on the results obtained during the completion of this work and developments in the field, we reflect on our approach to building a prototype quantum computer. We end by providing an outlook on the challenges involved in scaling to a kilo-qubit sized system.

9.1 Conclusions

The aim of this thesis was to address the challenges in controlling a prototype quantum computer based on superconducting transmon qubits. This section summarizes the main results and conclusions of each chapter.

- Chapter 1 introduces quantum computing, discussing the quantum mechanical concepts of superposition, entanglement and measurements and how these can be used to perform computations. Explicit attention is given to the notion that qubits are fragile and the concept of quantum error correction.
- In the first part of [building a superconducting quantum system](#) the [the full-stack quantum computer](#) computer is introduced as a vision of what a quantum computer can look like. Some limitations that break the abstraction layers of the full-stack quantum computer are highlighted. We conclude that it is too early to precisely define the responsibilities of each layer. To deal with the resulting uncertainty we propose an approach to building quantum computers consisting of three parts: vision, focus and slack.
- In the second part of Chapter 2 the [QuSurf architecture](#) is introduced, our design for a prototype quantum computer capable of realizing quantum fault tolerance. Key in the QuSurf architecture is the choice for a flux-pulsing based two-qubit gate. The flux dance that has been designed to work with this type of two-qubit gate uses a copy-pastable 8-qubit unit cell to enable a scalable surface code. The quantum device is controlled using the QuSurf control system, a centralized architecture based on codewords that trigger control pulses corresponding to individual operations. The QuSurf architecture is realized in the Quantum Infinity, a small-scale prototype quantum computer.
- In [assessing the performance of superconducting quantum processors](#) we discuss two approaches to characterizing quantum systems, the system agnostic *black-box* approach and the ad-hoc *experimentalist's approach*. A combination of these approaches is used to give an overview of the dominant error sources in the QuSurf architecture. We conclude that the error rates of all operations in isolation are at or below the fault-tolerance threshold. However, crosstalk errors are problematic, both from a performance perspective as well as from a characterization perspective. Crosstalk currently limits performance in larger devices. We anticipate that parameter drift will become relevant as systems become larger, calibration times increase, and error rates go down.
- We address single-qubit gates in Chapter 4. Here we introduce restless tuneup, a calibration protocol that provides a tenfold speedup over traditional methods to robustly achieve a coherence-limited error rate of $\epsilon_g = 5.3 \cdot 10^{-4}$ per single-qubit gate within one minute. This speedup is achieved by avoiding qubit initialization by relaxation and instead utilizing correlations between the outcomes of consecutive QND measurements. This method extends to higher gate fidelities due to a near-constant signal-to-noise ratio and can be modified for calibrations of other operations such as two-qubit gates.

- In Chapter 5 we address the duration of the measurement operation, one of the dominant error sources in the error-correcting cycle of the surface code. By using active photon depletion we speed up photon depletion by more than six inverse resonator linewidths, reducing the time before operations on a qubit can recommence after a measurement. The benefit of the depletion protocol is evaluated by emulating an ancilla qubit undergoing repeated parity checks in a repetition code.
- In Chapter 6 we introduce Cryoscope, an in-situ technique which uses a qubit to accurately sample the flux pulses used to dynamically control its frequency. We use Cryoscope to measure and correct the step response of the dedicated flux-control lines of transmon qubits as needed for high-fidelity two-qubit gates.
- In Chapter 7 we introduce Net Zero, a new type of conditional-phase gate for transmon qubits providing several key improvements over standard flux-pulsing-based versions. The Net-Zero gate uses “leakage interference” to minimize leakage to non-computational states. The zero-average, bipolar shape of the pulse makes the gate robust to long timescale distortions in the flux control line and additionally provides an echo effect. We demonstrate a state-of-the-art repeatable conditional-phase gate of duration 40 ns achieving an error rate of $\epsilon_{CZ} = 9 \cdot 10^{-3}$ and $L_1 = 1.0 \cdot 10^{-3}$.
- In Chapter 8 we have given an overview of the software required to control a multi-qubit system. We have describing the design philosophy and the core concepts of PycQED, an open-source software package we developed for this purpose. When scaling to larger systems it is essential to automate the characterization and calibration process. We have introduced dependency graphs as a useful abstraction and system emulation as an essential development tool for automating this process and have ended the chapter by providing an outlook on the system-level analysis required to scale to larger systems.

9.2 Reflections on the QuSurf architecture

The QuSurf architecture was conceived several years ago [83, 121] and has remained largely unchanged, a testament to its success. Nonetheless, were we to design the system today we would make two important changes. First, we would abandon the VSM pulse broadcasting scheme and modify the control electronics to take advantage of this change. And second, we would introduce tunable couplers.

Abandoning the VSM pulse broadcasting scheme

To make economical use of control hardware the QuSurf architecture makes use of selective broadcasting of common control pulses tailored to individual (same-frequency) qubits, saving on the amount of AWGs, mixers and microwave sources required for microwave control [92]. Because the error correcting cycle of the surface code requires the same single-qubit gates on multiple qubits at the same time this introduces no compilation overhead.

Although this concept looks very attractive from a resource perspective, there are some drawbacks. To be able to use common control pulses next-nearest neighbor qubits are tuned to identical frequencies. The coupling between these same-frequency qubits is not negligible (see Table 3.5) resulting in a crosstalk error than can be eliminated by slightly detuning the qubits. Second, the qubit sweetspot is often different from the target frequency because of fabrication uncertainties resulting in reduced coherence under operation. Both of these limitations can be resolved by bypassing the VSM and using frequency bands instead of discrete frequencies. Because these problems are severe, this is already being done in current generation devices. Although these problems (frequency targeting and residual exchange coupling) can be addressed in the near future there are other reasons to abandon the VSM concept.

The requirement that control pulses for different qubits have to be common introduces overhead¹ and increases the complexity of compilation. This requirement also makes it impossible to perform virtual Z gates (Section 2.2.1). To address this limitation, the codeword protocol of the QuSurf architecture needs to be expanded slightly as well. By allowing codewords to update the phase of all subsequent pulses for a specific qubit² it is possible to perform virtual Z gates. This feature is currently being implemented in the latest version of the Zurich Instruments HDAWG.

The use of a codeword-based paradigm imposes several other limitations that can be worked around, such as the limited gate set (addressed in [66]) or the inability the precompile distortion corrections (addressed in Chapters 6 and 7), or have relatively limited impact (the 20 ns time grid). Operating in this paradigm has taught us a lot about interfacing with the higher levels of the quantum computing stack and has allowed us to develop a quantum instruction set [52, 53]. As such, I would argue that this simplification of the system has been a successful one. Nonetheless, as discussed in Section 2.1.1 there are limitations that make this concept less suited for NISQ [33] applications.

Introducing tunable couplers

We have seen in Chapter 3 that crosstalk due to residual coupling is the dominant error source in the Chimaera-S7 device. Even if all qubit frequencies and coupling strengths are on target this effect is not negligible. The natural solution to this problem is to replace the coupling buses with tunable couplers [104–106]. However, this greatly increases the system complexity and adds at least one extra control line per coupler, corresponding to two extra control lines per qubit on an infinite 2D grid. There are several arguments why adding tunable couplers is nonetheless a good idea.

An important problem that is not discussed very often is how to mitigate the impact of two-level systems (TLS) [94, 173, 278, 279] on multi-qubit devices. TLS cause a significant reduction in qubit coherence at specific frequencies when they couple to a qubit. Currently the three most

¹With the exception of highly structured sequences such as the error correcting cycle.

²Note that this is different from simply triggering a predefined waveform as it affects the shape of all subsequent operations. This conceptually simple change has implications for the technical implementation.

popular mitigation strategies to deal with TLS present in the system are: 1. wait for the TLS to disappear or drift to another frequency, 2. detune the qubit from the TLS, 3. thermal cycle the fridge and hope the TLS disappears. For small ($\lesssim 7$ qubits) devices it is feasible to hope for a TLS to disappear or perform a thermal cycle. However, for larger devices the odds of having no TLS affecting any qubit become increasingly slim so that one is forced to detune the qubits. Having tunable couplers removes the residual coupling as a constraint when trying to optimize the trade offs between the different error mechanisms [36]. Investigating the physical nature of TLS, exploring qubit designs and fabrication recipes that reduce the presence of TLS, and alternate mitigation strategies are areas of active research [94, 278, 279] that will prove vital for superconducting quantum computing.

Because two-qubit gates in the QuSurf architecture rely on the on-off ratio of static coupling, the transmons involved need to detune far from the sweetspot during these gates. This makes the qubits sensitive to small changes in flux, such as those from flux noise causing dephasing and those caused by remnant distortions. In fact, this limits performance of the Net-Zero two-qubit gate (Section 7.3.5). When using tunable couplers it is possible to operate the qubits at frequencies that are closer together, reducing the sensitivity to these effects [280]. Although this introduces an extra control parameter in the calibration problem, the extra degree of freedom can be used to decrease leakage [68]³ or to perform parametrized two-qubit gates [67, 280]. Although the exact parametrization of two-qubit gates using tunable couplers is different, one would still use the Cryoscope (Chapter 6) to correct distortion effects and modify the pulse shapes using Net-Zero (Chapter 7) to ensure repeatability. A complementary technique to reduce the sensitivity to flux that can be pursued in parallel is to use SQUIDs with asymmetric Josephson junctions [90, 91].

Based on these arguments: the elimination of residual ZZ crosstalk, the ability to detune from unwanted TLS, the reduced sensitivity to flux during two-qubit gates, and the simplification of parametrized two-qubit gates, adding tunable couplers seems worth the increased system complexity.

9.3 Towards a kilo-qubit system

Where the preceding section provided a reflection on our chosen architecture and provided an outlook on the changes we would make to scale to 50 – 100 qubits. Here we provide an outlook on what we believe will be the main challenges in scaling to a useful kilo-qubit (kQb) sized system. We frame these challenges in terms of three big problems: **the application problem**, **the fabrication problem** and **the calibration problem**.

³Strictly speaking the dependence of the coupling strength on the qubit frequency can also be used to achieve the relevant resonance condition but this is simpler using a tunable coupler.

9.3.1 The application problem

For the last several years the community has focused on the number of qubits n in a device as the single most important performance metric. If the power of quantum computing scales exponentially with the number of qubits, then the number of (logical) qubits is the most important performance metric. However, while classical bits can be approximated as error free and ideal, this is a gross oversimplification for qubits. When considering physical error rates on the order of $\epsilon \sim 10^{-4}$ one requires 10^4 physical qubits per logical qubit to achieve a (close to ideal) logical error rate of $\epsilon_L \sim 10^{-15}$ [72, 281]. Current estimates suggest that one would require around 20 million qubits to factor a number that is too large to tackle using classical algorithms [282].

A 20 million qubit quantum processor is, at this point in time, inconceivably large and as such not specific enough to be useful as a vision (see Section 2.1.2). Nevertheless, small quantum computers consisting of ~ 50 (physical) qubits are already too large to simulate and can outperform even the largest supercomputers on specifically constructed problems [36]. This suggests that it may be possible that these pre-error correction systems, known as “Noisy Intermediate Scale Quantum” (NISQ) [33] systems, will be able to outperform classical computers on certain *useful* applications. This introduces the first big challenge:

Problem 1 (The Application Problem)

Are there applications where it is advantageous to use a NISQ quantum computer?

Although it is at this point not clear what these applications will be, it seems likely that the answer lies in Feynman’s original idea [1]; using quantum systems to simulate quantum systems⁴.

To give a meaningful answer to this question, it is important to know how powerful a potential NISQ system can be. To describe the power of a quantum computer, one has to take into account not only the number of qubits n , but also the number of operations that can be performed, typically expressed using the circuit depth d . A metric that combines these two is the quantum volume V_Q [34, 35, 283]. The quantum volume is defined as $\log_2(V_Q) = \operatorname{argmax}_n \min(n, d(n))$ ⁵. This definition loosely coincides with the complexity of classically simulating model circuits. In this definition, the circuit depth d corresponds to the number of circuit layers that can be executed before (on average) a single error occurs. A circuit layer corresponds to a combination of arbitrary two-qubit operations between disjoint pairs of qubits. It is possible to estimate d as $d \approx 1/\epsilon_{1\text{step}} = 1/n\epsilon_{\text{eff}}$, where the effective error rate ϵ_{eff} is the average error rate per two-qubit operation. Limitations in connectivity, parallelism and gate set introduce an overhead in the physical implementation of a circuit layer so that in general $\epsilon_{\text{eff}} \geq \epsilon$, where ϵ is the average error rate for physical operations.

At current error rates of $\epsilon \sim 10^{-3}$ this implies that there is no power in having a device larger than $n \sim 31$ qubits (Figure 9.1), which can be efficiently simulated. This seems to contradict the results of Arute *et al.* [36], in which a $n = 53$ qubit device with $\epsilon \sim 10^{-3}$ was

⁴Expressed in the opening quote of this thesis (Chapter 1)

⁵The definition of V_Q is different in [34, 283] and [35]. Here we use the definition from [35].

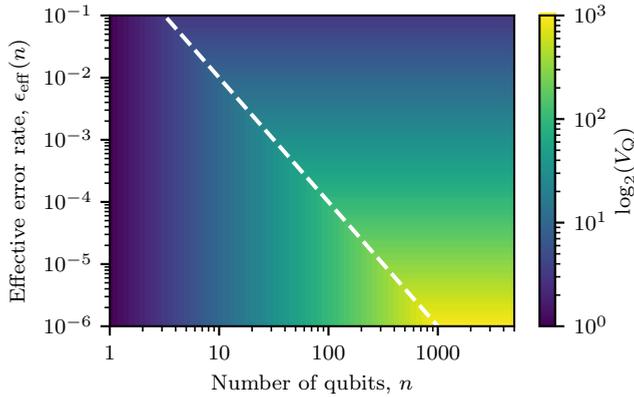


Figure 9.1.: Quantum volume V_Q as a function of the effective error rate ϵ_{eff} and number of qubits n . Dashed line indicates the point where the n and ϵ_{eff} contribute equally to V_Q . Figure based on [34] using the definition for V_Q of [35].

used to perform a computation that could not be simulated in a reasonable amount of time. This can be understood by looking closer at the quantum volume. The quantum volume was designed as a binary metric: can a device run an algorithm? For many algorithms, a single error indicates failure, however, for other applications, such as sampling a distribution (as is done in [36]), one can tolerate a limited amount of errors simply by averaging. From this we can infer that NISQ applications will have to be able to tolerate a limited amount of errors, either because of the nature of the application (e.g., sampling a distribution) or by using error mitigation techniques.

Another limitation of the quantum volume is that it quantifies the ability to run circuits of equal width and depth. However, the computational power of short-depth circuits is not yet fully understood, and it can be argued, that even short-depth circuits lie beyond the reach of classical computing [34, 284, 285]. As such it is likely that potential NISQ applications will be short depth to limit the amount of errors that accumulate. At current error rates of $\epsilon \sim 10^{-3}$ a 1 kQb system is right at the point where one can still execute a single layer without a single error occurring. If one considers using a fraction of the qubits as ancillas for error mitigation and uses an algorithm that is somewhat robust to the remaining errors, the kQb processor is the largest scale NISQ device that is interesting at current error rates.

Here, we have focused on the quantum volume as analyzing the limitations of this metric can be instructive in understanding some characteristics of potential NISQ algorithms. Recent works on *volumetric benchmarks* have expanded upon the quantum volume [286, 287], addressing both the circuit width depth trade-off and the binary nature of the quantum volume.

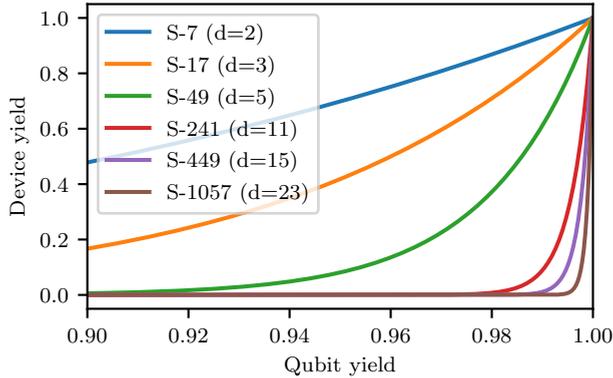


Figure 9.2.: Probability of producing a working device versus the yield of individual qubits. Different lines correspond to different size surface code devices, code distance denoted in brackets.

9.3.2 The fabrication problem

Having set a target for the size of the system (1 kQb) and the performance ($\epsilon < 10^{-3}$) one has to design and fabricate devices capable of reaching this target. Qubit coherence ($\sim 50 \mu\text{s}$, see Chapter 3) should be sufficient to reach $\epsilon < 10^{-3}$ for every operation. More challenging will be to scale up the design to ~ 1 kQb while maintaining these levels of performance. We call this the fabrication problem:

Problem 2 (The fabrication problem)

Can we design and fabricate large devices that consistently have a high yield and coherence?

Although the current surface code design (excluding tunable couplers) is copy pastable (Figure 2.6), that does not mean that it is scalable in practice. If we define an individual qubit to be working if 1. all control lines are working, 2. coherence is larger than a specified target (e.g. $50 \mu\text{s}$) and 3. the relevant parameters (charging and Josephson energies, coupling to coupling bus/tunable couplers, readout resonator parameters etc.) are within a specified tolerance, we can take this probability to predict device yield. If we take a simplistic model in which device yield, defined as all qubits working, is simply the product of the individual qubit yields (Figure 9.2) the odds of producing a working kQb device are increasingly small even at an exceptional yield of 99% per qubit.

To tackle this problem one needs to either become robust against *missing* qubits at the algorithm level, which falls in the domain of the *application problem*, or find a way to increase device yield for a given qubit yield. A promising concept would be to link together smaller devices within the same fridge. Although the odds of producing a single monolithic kQb are vanishingly small, one can increase the probability by combining multiple smaller patches, which have a reasonable yield, and replacing only the patches that do not work. Existing flip-

chip architectures [123], on which the readout resonators, Purcell filters and coupling buses are on a different chip than the qubits, can be seen as a prototype of this technique as they effectively link together different devices. It is only a small step to imagine using the coupling plane to connect qubits on different (but adjacent) chips. Note that what we envision is subtly different from the chip-to-chip entanglement discussed in [288] which is more powerful. This proposal does not require long-distances as it only attempts to create modularity.

Related to the question of yield is the question of size: does a kQb processor fit in a fridge? Although transmon qubits ($\sim 400 \mu\text{m}^2$) are often seen as large in comparison to e.g., spin-qubits or dopant-based qubits, the processor sizes are not limited by the qubit size, but rather by the size of the I/O [122]. The footprint of a single VIA is currently 1 mm^2 ⁶, a single transmon (including tunable couplers) requires 4.2 control lines⁷ putting the total footprint at $\sim 5 \text{ mm}^2$ per qubit. Let us assume that we can translate this footprint into a pitch of 2.5 mm in which case a kQb processor would be about 6.4 cm^2 . As such 1 kQb would fit on a 100 mm wafer (with a surface area of 7.8 cm^2).

Although this back of the envelope calculation indicates that a kQb processor would be about the size of a single 100 mm wafer, it also highlights the importance of the interconnect size. Where it is possible to reduce the on-chip footprint to about 1 mm^2 for each interconnect, regular SMA connectors have a diameter of 7.9 mm resulting in a footprint of about 1 cm^2 per connector. At about 4.2 lines per qubit this would mean that a kQb processor would require a solid $70 \times 70 \text{ cm}$ block of SMA connectors. Even when using the smaller SMP connectors one would still require $35 \times 35 \text{ cm}$ of connectors. As cable dimensions are typically significantly smaller than the connector sizes a natural solution includes the cabling in the sample mount. In this way, the signal integrity can be preserved while the fan out can be taken care of elsewhere.

Not only the footprint of the lines is relevant but also the heat load. The heat load consists of two contributions, a passive contribution coming from the fact that there is a conducting line connecting the sample to room temperature, and an active contribution consisting of power dissipation happening in the line. Attenuating the power of signals intended for the qubit is required to manage the noise temperature of the signals. For a system up to ~ 100 qubits the heat load can be managed by using standard cable technologies and attenuators [289]. To reduce the active contribution one can consider using directional couplers that transmit only part of the signal while sending the return signal to a higher temperature stage where more cooling power is available. The passive contribution to the heat load can be reduced by using specialized cable technologies. A promising approach to is to use microwave striplines etched on a flexible substrate to produce cables with lower thermal conductivity and a smaller form factor [290]. Because of the reduced form factor, these cabling technologies are a natural candidate for integrating in the sample mount mentioned in the preceding paragraph.

At this point it is unclear if better interconnects and cabling technologies will be sufficient to realize a kQb device. There are several techniques that can be used to reduce the number

⁶And can be scaled down to 0.4 mm^2 [122].

⁷1 microwave drive, 1 flux bias, 2 coupler bias and a feedline input and output shared by ~ 10 qubits.

of lines by a constant factor. The concept of dedicated drive-lines per qubit can be dropped in favor of a frequency multiplexing scheme in which several qubits (~ 5) operated at different frequencies share a drive line. These changes however do not address how the number of lines scales (linearly) but only change the prefactor. At some point one has to consider Rent's rule [291]. To change the scaling of control one has to find multiplexing schemes for all types of control (microwave, flux, measurement) similar to the VSM scheme [92, 121] for microwave pulses. The constraints imposed by such a scheme will have significant consequences for how it can execute algorithms and furthermore requires exquisite control over device fabrication. As such we do not expect such a scheme to be viable in the near future.

Until now we have glossed over one of the most important aspects of the fabrication problem; coherence. Qubit performance is inherently limited by coherence and although we now regularly measure coherence times $T_1 > 50 \mu\text{s}$ this is not yet a reliable occurrence. It will be a large challenge in itself to better understand the what is limiting coherence and to reliably fabricate high coherence devices. Achieving high coherence will be especially challenging as significant changes to the design are required such as the addition of tunable couplers and the connection of different sub patches. All of these changes have the potential to impact coherence.

9.3.3 The calibration problem

The last of our three big problems ties everything together. After finding a useful application for a kQb sized NISQ system, and designing and fabricating a processor, the system should be operated. Due to variations in the fabrication process all qubits need to be individually characterized and calibrated before the system can be operated as a quantum computer. This task is challenging because system parameters can fluctuate over time (Section 3.4.2), depend on each other, and suffer from crosstalk (Section 3.5). As calibration techniques have been a central theme in this thesis (see e.g., Chapters 4 to 7 and Section 8.3), it should be clear that calibrating the system is more challenging than executing algorithms. We call this challenge the calibration problem:

Problem 3 (The calibration problem)

Can we characterize and calibrate a kQb sized system in a reasonable amount of time?

Solving the calibration problem will be key in resolving the fabrication problem. To achieve a high yield and coherence, one needs to understand how changes in design and fabrication affect the system. By connecting the automated calibration, that is required to operate the system, to a database infrastructure, it is possible to close the loop between design, fabrication, and characterization shown in Figure 4.5. The system level analyses and visualizations discussed in Section 8.3.3 will be essential in effectively closing this loop. Just as important as having the infrastructure in place is knowing how to efficiently characterize the system. To this end, new characterization protocols that combine the best of the the experimentalist's approach and the the black-box approach will be needed. For a more extensive discussion on the calibration problem, we refer the reader back to Section 8.3.

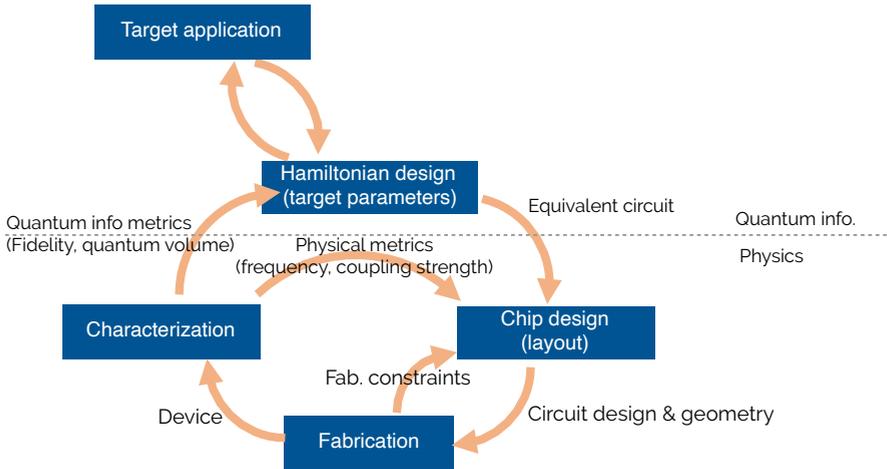


Figure 9.3.: The engineering cycle of quantum device development consists of several steps. A target application influences the chip design in which an equivalent circuit and its target parameters are determined. This design serves as the input for a second design step in which the circuit design and geometry of the device is determined, taking into account constraints of the fabrication process. After fabricating the device, the system is characterized resulting in knowledge about the system. This knowledge is then used to iterate on the design steps.

Even after careful calibration and characterization of the system, one question remains: How can one trust the output of a quantum computer? Normally, for small systems ($\lesssim 40$ qubits) one would simulate the algorithm and verify the output. However, for a kQb processor this is not possible. There are several ways of dealing with this problem. One option is to use test circuits with a known solution, if the quantum computer correctly reproduces these results then it can also be trusted for another algorithm. In another approach, used by Arute *et al.* [36], the individual error rates of all operations are carefully measured and used to predict the outcome of several test circuits. From here, the complexity of the test circuits is gradually increased until it is no longer possible to simulate the system. Because the experiment continued to perform as expected, trusting the final result was transformed from a leap of faith into a small and logical step.

This appendix provides details on the protocols used by to determine performance metrics. The appendix contains five sections pertaining to different kinds of errors. The first section details protocols to determine gate errors, the second section errors during idle operations, the third section describes reset and initialization errors, the fourth section is about measurement characterization, the fifth is about crosstalk and the final section is about other kinds of errors such as leakage.

For errors of the crosstalk type, it is usually required to perform an architecture-specific experiment to isolate the different effects. If these effects describe particular (usually coherent) multi-qubit processes, we convert them to the form of Pauli error rates per qubit and per operation when reporting an error rate.

In reporting the metrics, we choose for the duration τ_I of an idle gate the duration of the longest gate, namely that of CZ gates, $\tau_I = \tau_{CZ} = 40$ ns.

Timescale	Value
Single-qubit gate time τ_{1Q}	20 ns
Two-qubit gate time $\tau_{2Q} = \tau_{CZ}$	40 ns

A.1 Gate errors

A.1.1 Single-qubit randomized benchmarking

Single-qubit randomized benchmarking is performed by constructing sequences of n random Cliffords for different sequence lengths N . A total of 200 randomizations (seeds) are used for each length n . Cliffords are decomposed into physical operations ($\pm X90$, $\pm Y90$, $X180$, $Y180$, requiring an average of 1.875 operations per Clifford) as described in [99]. The resulting outcomes are fitted with an exponential of the form:

$$F_{|0\rangle} = Ap^N + B. \quad (\text{A.1})$$

The error rate per Clifford is related to the RB decay parameter through:

$$\epsilon_{\text{Cl}} = 1 - \left(p + \frac{1-p}{2} \right). \quad (\text{A.2})$$

The reported error rate is per gate and can be extracted using:

$$\epsilon_g = 1 - (1 - \epsilon_{\text{Cl}})^{1/n_g}, \quad (\text{A.3})$$

Clifford type	Frac. of Cliff. group	Decomposition	Avg. # of CZ	Avg. # SG
Single-qubit	1/8	SC	0	3.75
CNOT-like	1/8	SC-CZ-SC	1	6.25
iSWAP-like	3/8	SC-CZ-SC-CZ-SC	2	9.42
SWAP-like	1/8	SC-CZ-SC-CZ- -SC-CZ-SC	3	8.75
Average			1.5	8.25

Table A.1.: Decomposition of different subsets of two-qubit Cliffords into alternating simultaneous single-qubit Cliffords (SC) and CZ gates (not all possible SC should be chosen at all positions, see Ref. [78] for details). The single-qubit Clifford gates are further decomposed into 1-3 single qubit rotations (SG), following Ref. [190].

where $n_g = 1.875$ is the average number of gates per Clifford. This protocol can be supplemented with a leakage modification to extract the leakage per gate. See the section on "Leakage modification" below.

A.1.2 Two-qubit randomized benchmarking with leakage modification

For 2RB, we follow the description in Ref. [78]. We randomly select k strings of N elements from the 11520-element Clifford group on two qubits, and then append an $N+1$ th recovery Clifford that inverts the action of each string. Each Clifford is then further expanded into a string of gates from our gate set ($\pm X90$, $\pm Y90$, $X180$, $Y180$, CZ), by first decomposing into strings of simultaneous single-qubit Clifford (SC) operations and CZ gates, and then expanding the SC into single-qubit rotations (see Appendix A.1.2). We note that this decomposition is not optimal, for instance, the element of the Clifford group representing the CZ gate is in fact decomposed into the string " $Z90$, CZ , $-Z90$ ".

We compile the sequences to run on our device using the OpenQL compiler, which means that single-qubit gates on different qubits are scheduled simultaneously whenever possible. We measure the "survival" probability of measuring "00" averaged over all k strings and for different lengths N , and fit an exponential decay, giving the error rate per Clifford.

By assuming all errors to be due to the two-qubit CZ gate, an upper bound to the error per CZ gate can be found using

$$\epsilon_{CZ} = 1 - (1 - \epsilon_{Cl})^{1/n_{CZ}}, \quad (\text{A.4})$$

where $n_{CZ} = 1.5$ is the average number of CZ gates per Clifford.

This protocol can be supplemented with a leakage modification to extract the leakage per gate. See the section on "Leakage modification" below.

A.1.3 Two-qubit interleaved randomized benchmarking

Interleaved two-qubit randomized benchmarking (2IRB) can be used to characterize the performance of a CZ gate. By both interleaving a CZ gate between each Clifford in two-qubit randomized benchmarking (A.1.2) and performing the non-interleaved two-qubit randomized benchmarking experiment. The difference in error rates for the conventional (ϵ_{CZ}) and interleaved sequences (ϵ_{Int}) correspond to the error rate per interleaved (CZ) gate [146]:

$$\epsilon_{CZ} = 1 - \frac{1 - \epsilon_{Int}}{1 - \epsilon_{Cl}}. \quad (\text{A.5})$$

A.1.4 Single-qubit and two-qubit gate-set tomography

Gate set tomography (GST) is performed using the pyGSTi package. For the error rate ϵ_{GST} , only the process infidelity on a per-gate basis is reported:

$$\epsilon_{GST} = 1 - F_{pro}. \quad (\text{A.6})$$

The germ length is provided in the notes for the method. For more details, pyGSTi reports are available upon request.

A.2 Idle error rates

A.2.1 Idle decay

To characterize errors during idling the following protocol is followed:

1. The qubit is prepared in $|0\rangle$ by idling several T_1 .
2. The qubit is measured to post-select on the "0" outcome.
3. The qubit is initialized into (a) $|0\rangle$, (b) $|1\rangle$, or (c) $|+\rangle$ using a single-qubit gate.
4. Idle for a time $N\tau_I$.
5. Apply a pre-measurement gate to rotate to the Z basis.
6. Measure in Z basis, repeat to obtain average outcome $P_i(N)$
7. The decay curves will be polluted by SPAM errors, so will be fitted to the model described below.

A basic decay model is used that includes an offset to account for SPAM errors:

$$P(N) = A_0 e^{-\frac{N}{N_1} - \frac{N^2}{N_2}} + B_0. \quad (\text{A.7})$$

In most cases, the data is compatible with $N_2 = \infty$. For consistency, we report the error rates as

$$\epsilon_I = 1/N_1 \quad (\text{A.8})$$

per τ_I . For case (c), a single echo pulse can be included in the middle of the idle period.

A.3 Measurement errors

A measurement M of a state $|\psi\rangle_i$ returns an outcome $m \in \{0, 1\}$ and leaves the system in a state $|\phi\rangle_o$. There are two types of errors that have to be characterized for measurements, errors in the state assignment and deviations from a perfect quantum nondemolition (QND) measurement.

Measurement histograms

Before defining a threshold, each instance (or shot) of a measurement will return an arbitrary complex number. To extract information about the qubit state from this, these outcomes are consolidated in a histogram. The possible outcomes of a good qubit measurement will form two groups, each with a Gaussian distribution associated with the qubit being in either $|0\rangle$ or $|1\rangle$. Based on these histograms, the data is typically rotated and a threshold is defined. The measurement M will then return $m = 0$ when the value is below this threshold and $m = 1$ when it is above.

A.3.1 Measurement assignment errors

We define as the average assignment error $\epsilon_{\text{avg.ass.}}$ as the average probability (for ground and excited state) that the measurement result does not correlates with the input state:

$$\epsilon_{\text{avg.ass.}} = \frac{P(m = 0 \mid |1\rangle_i) + P(m = 1 \mid |0\rangle_i)}{2}, \quad (\text{A.9})$$

and correspondingly the average assignment fidelity:

$$F_{\text{avg.ass.}} = 1 - \epsilon_{\text{avg.ass.}}. \quad (\text{A.10})$$

In the single-shot readout experiment the system is repeatedly prepared in $|0\rangle$ and $|1\rangle$ followed by a measurement. $\epsilon_{\text{avg.ass.}}$ can be determined by correlating the declared states with the prepared states.

A.3.2 Measurement QNDness

The QNDness of a strong measurement (which completely dephases a qubit) can be expressed using classical probabilities. We define the relaxation (excitation) during measurement $\epsilon_{\text{msmt. rel.}}$ ($\epsilon_{\text{msmt. exc.}}$)¹ as the probability of the system being in the $|0\rangle_o$ ($|1\rangle_o$) state after the measurement when the system was in the $|1\rangle_i$ ($|0\rangle_i$):

$$\epsilon_{\text{msmt. rel.}} = P(|0\rangle_o \mid |1\rangle_i), \quad (\text{A.11})$$

$$\epsilon_{\text{msmt. exc.}} = P(|1\rangle_o \mid |0\rangle_i) \quad (\text{A.12})$$

¹Note that this is not equal to the measurement *induced* relaxation (excitation). The measurement induced relaxation (excitation) is the difference between the relaxation (excitation) during measurement and the relaxation (excitation) one would experience when no operation (idle) is performed for the same duration.

We define the QND error ϵ_{QND} as the average probability (for ground and excited state) that the state after the measurement does not correlates with the input state:

$$\epsilon_{\text{QND}} = \frac{\epsilon_{\text{msmt. rel.}} + \epsilon_{\text{msmt. exc.}}}{2}. \tag{A.13}$$

The measurement butterfly

A measurement M_1 can be characterized using the butterfly experiment consisting of three consecutive measurements M_0, M_1 and M_2 (shown in Figure A.1). A qubit q_i is first prepared in $|0\rangle$ with high fidelity by combining initialization through relaxation with post selection on m_0 . Next, a π pulse is or is not applied to prepare the system in $|\psi\rangle_i \in \{|0\rangle, |1\rangle\}$ followed by two measurements M_1 and M_2 . By repeating this experiment it is possible to estimate eight conditional probabilities:

$$P(m_2, m_1 \mid |\psi\rangle_i). \tag{A.14}$$

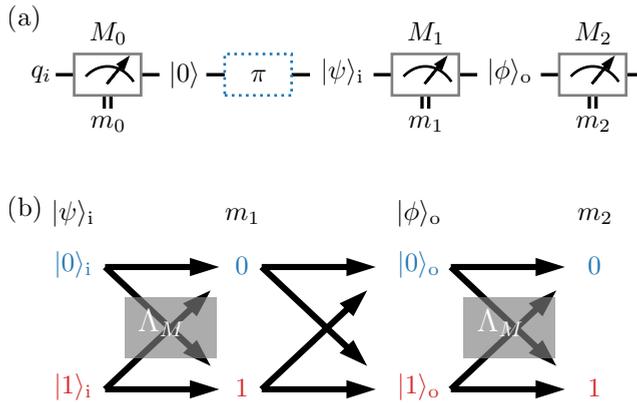


Figure A.1.: (a) Circuit diagram for the butterfly experiment to characterize measurement M_1 on qubit q_i . Post selection based on an initial measurement M_0 is used to initialize in $|0\rangle$. (b) Correlations between m_1 and $|\psi\rangle_i$ are described by Λ_M

In order to estimate the conditional probabilities $P(|\phi\rangle_o \mid |\psi\rangle_i)$ required to determine the QNDness of the measurement we first determine the matrix Λ_M that describes the relation between the $P(m_1)$ and $|\psi\rangle_i$:

$$\Lambda_M P(|\psi\rangle_i) = P(m_1), \tag{A.15}$$

where

$$\Lambda_M = \begin{bmatrix} P(m_1 = 0 \mid |0\rangle_i) & P(m_1 = 0 \mid |1\rangle_i) \\ P(m_1 = 1 \mid |0\rangle_i) & P(m_1 = 1 \mid |1\rangle_i) \end{bmatrix}. \quad (\text{A.16})$$

Note that $\epsilon_{\text{avg. ass.}}$ can be determined based on the off-diagonal elements of this matrix.

Because M_2 and M_1 correspond to the same (measurement) operation $\Lambda_M P(|\phi\rangle_o) = P(m_2)$ [see also Figure A.1(b)]. So that $P(|\phi\rangle_o)$ can be determined by inverting Λ_M :

$$\Lambda_M^{-1} P(m_2) = P(|\phi\rangle_o). \quad (\text{A.17})$$

Combining this we get

$$\begin{bmatrix} P(|0\rangle_o, m_1 \mid |\psi\rangle_i) \\ P(|1\rangle_o, m_1 \mid |\psi\rangle_i) \end{bmatrix} = \Lambda_M^{-1} \begin{bmatrix} P(m_2 = 0, m_1 \mid |\psi\rangle_i) \\ P(m_2 = 1, m_1 \mid |\psi\rangle_i) \end{bmatrix}, \quad (\text{A.18})$$

and for completeness:

$$\epsilon_{\text{msmt. rel.}} = P(|0\rangle_o, m_1 = 0 \mid |1\rangle_i) + P(|0\rangle_o, m_1 = 1 \mid |1\rangle_i), \quad (\text{A.19})$$

$$\epsilon_{\text{msmt. exc.}} = P(|1\rangle_o, m_1 = 0 \mid |0\rangle_i) + P(|1\rangle_o, m_1 = 1 \mid |0\rangle_i). \quad (\text{A.20})$$

A.4 Reset and initialization errors

Qubit initialization can either be done deterministically or probabilistically. Deterministic methods include measurement and feedback and pumping scheme. Probabilistic schemes involve an initialization measurement (M_0) prior to the algorithm and post-selecting experiment runs conditioned on the outcome m_0 .

A.4.1 Reset and initialization error

The remaining initialization error of any of these schemes can be extracted by following M_0 with an additional measurement M_1 and count the fraction of outcomes that indicate an unwanted outcome. For ground state preparation this error is defined as the probability of being in the $|1\rangle$ state after the first measurement:

$$\epsilon_{\text{in}} = P(|1\rangle). \quad (\text{A.21})$$

A.5 Crosstalk errors: Idle crosstalk

A.5.1 Residual-ZZ interaction

The coupling of neighboring transmons q_i and q_j by a coupling bus is used to implement CZ gates by flux-detuning the higher-frequency qubit so that the two-qubit states $|11\rangle$ and $|20\rangle$ are in resonance. At the idle frequencies, the system is ideally far detuned from that point, but leads to a residual idle Hamiltonian of the form

$$H = \zeta_{ij} |11\rangle \langle 11| = -\frac{\zeta_{ij}}{2} (1 - Z_i - Z_j + Z_i Z_j). \quad (\text{A.22})$$

Following this definition

$$\zeta_{ij} = E_{11} - E_{01} - E_{10} \quad (\text{A.23})$$

where E_{kl} corresponds to the energy of the state with k (l) excitations on qubit i (j). Since the interaction is always on, the single-qubit terms can be taken as a renormalization of the qubit frequency. Per operation, we thus obtain a coherent correlated ZZ error with angle

$$\theta = \frac{\zeta_{ij} \tau I}{4}. \quad (\text{A.24})$$

The interaction strength corresponds to the difference in frequency $\Delta\omega_i^{(j)}$ of Ramsey oscillations of q_i with q_j in state $|1\rangle$ and $|0\rangle$.

$$\Delta\omega_i^{(j)} = (E_{11} - E_{01}) - (E_{10} - E_{00}), \quad (\text{A.25})$$

$$\Delta\omega_i^{(j)} = (\zeta_{ij} + E_{10} + E_{01} - E_{01}) - (E_{10}), \quad (\text{A.26})$$

$$\Delta\omega_i^{(j)} = \zeta_{ij}. \quad (\text{A.27})$$

The frequency difference is identical to the residual coupling $\Delta\omega_i^{(j)} = \zeta_{ij}$.

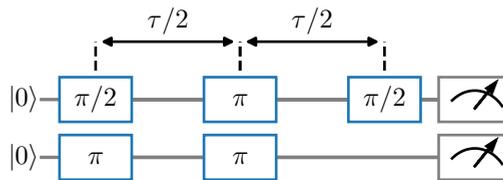


Figure A.2.: Circuit diagram for the residual-ZZ-echo experiment.

This frequency difference is measured through the residual-ZZ-echo experiment (Figure A.2). An echo experiment over a time τ is performed on q_i while an excitation is added and subsequently removed from q_j . In the first arm of the echo experiment, q_i will acquire a phase $\varphi_A = (\omega_i + \zeta_{ij}) \cdot \tau/2$ which is partially canceled by the phase acquired in the second arm

$\varphi_B = (-\omega_i) \cdot \tau/2$ resulting in an oscillation with frequency $\zeta_{ij}/2$ in the measured signal of q_i :

$$\varphi = \varphi_A + \varphi_B, \quad (\text{A.28})$$

$$\varphi = (\omega_i + \zeta_{ij}) \cdot \tau/2 + (-\omega_i) \cdot \tau/2, \quad (\text{A.29})$$

$$\varphi = (\omega_i + \zeta_{ij} - \omega_i) \cdot \tau/2, \quad (\text{A.30})$$

$$\varphi = (\zeta_{ij}/2) \cdot \tau. \quad (\text{A.31})$$

Note the factor 2.

For coherent errors, we report the error rate as,

$$\epsilon_{ZZ} = \theta^2 = \left(\frac{\zeta_{ij}\tau_I}{4} \right)^2. \quad (\text{A.32})$$

A.5.2 Residual exchange interaction

Some next-nearest neighboring qubits in the QuSurf architecture are targeted to the same frequency. The residual coupling between these pairs leads to an exchange interaction [92].

$$H = J_R |10\rangle \langle 01| + h.c. \quad (\text{A.33})$$

The result is a coherent bit-flip error for both qubits with angle per idle gate

$$\theta = J_R \tau_I, \quad (\text{A.34})$$

and corresponding error

$$\epsilon_{XX} = \theta^2 = (J_R \tau_I)^2. \quad (\text{A.35})$$

To determine J_R , a T_1 relaxation experiment is performed on qubit q_i . If $J_R \cdot T_1 \gg 1$, an oscillation will be visible in the T_1 relaxation experiment. The frequency of this oscillation corresponds to J_R [92].

A.6 Crosstalk errors: Weight-1 operation crosstalk

A.6.1 Microwave cross driving

This describes the effect of spill-over of a microwave signal to a same-frequency qubit. As a metric that isolates this effect, we measure the microwave isolation for each pair of qubit q_i and microwave drive line D_j , defined as the pulse amplitude A_{ij}^{180} required for performing

an X180 operation on q_i through D_j , normalized by the required amplitude when using the correct driveline D_i , expressed in power reduction:

$$M_{ij} = 20 \log_{10} \left(\frac{A_{ij}^{180}}{A_{ii}^{180}} \right) [\text{dB}]. \quad (\text{A.36})$$

The result is a coherent rotation around an unknown axis in the $X - Y$ plane, with angle

$$\theta = \pi \frac{A_{ij}^{180}}{A_{ii}^{180}}. \quad (\text{A.37})$$

The error rate is then given by

$$\epsilon_{\text{CD}_{ij}} = \theta^2 = \left(\pi \frac{A_{ij}^{180}}{A_{ii}^{180}} \right)^2. \quad (\text{A.38})$$

A.6.2 Simultaneous single-qubit RB

Perform a single-qubit RB on qubit q_i as specified in A.1.1 in order to determine ϵ_{g_i} . Perform a second round of single-qubit RB on qubit q_i , while simultaneously performing a randomized benchmarking sequence on another qubit q_j to determine $\epsilon_{g_i}^{(j)}$ [149]. The error due to crosstalk between qubits q_i and q_j is then given by the difference in error rates:

$$\epsilon_{\text{mwX}_{ij}} = \epsilon_{g_i} - \epsilon_{g_i}^{(j)}. \quad (\text{A.39})$$

A.6.3 Measurement-induced dephasing

The measurement pulse applied to a qubit q_j can lead to parasitic measurement of another qubit q_i on the same feedline [130, 177]. The effect is quantified by the integrated measurement-induced dephasing [102, 130, 178], with the corresponding phase-flip error per measurement

$$\epsilon_{\phi_{\text{M}}} = \frac{1}{2} \left(1 - e^{-\overline{\Gamma_{m,ij}} \tau_{\text{m}}} \right). \quad (\text{A.40})$$

The measurement-induced dephasing time $\overline{\Gamma_{m,ij}}$ is measured by performing a Ramsey experiment on qubit i while applying a measurement pulse to qubit j .

This dephasing only leads to an operation error if q_i was not intended to be measured. In the QuSurf architecture, it is thus only required to reduce this error if q_i and q_j are of different type (i.e., not both X ancillas, Z ancillas, or data qubits).

A.6.4 Measurement detection crosstalk

The measurement detection crosstalk describes the dependence of the measurement outcome of qubit q_i on the state of qubit q_j . We quantify this dependence by first defining the average cross-measurement fidelity [130]:

$$F_{ij} = \langle -P(e_i | I_j) - P(g_i | \pi_j) \rangle, \quad (\text{A.41})$$

where e_i (g_i) denotes the assignment of q_i to the excited (ground) state, π_j (I_j) denotes the preparation with (without) a π pulse on q_j and the average $\langle \cdot \rangle$ is taken over assignment (preparation) of all qubits but i (j). These probabilities are extracted for all combinations of qubits ($i \neq j$). For the device we obtain an overall average measurement detection crosstalk error of

$$\epsilon_{\text{mdX}} = \langle 1 - |F_{ij}| \rangle. \quad (\text{A.42})$$

A.7 Other error sources

A.7.1 Leakage

The main method for quantifying leakage during gates is the leakage modification for randomized benchmarking. This is described in the sections on randomized benchmarking.

Leakage modification for randomized benchmarking

In order to determine the leakage rate out of the computational subspace, the protocol from [151] is used. In order to use this method, the randomized benchmarking experiment needs to be modified to determine the populations in the computational subspace.

To determine the populations in the ground (g), first-excited (e), and second-excited (f) states we follow the procedure described in [92]. In this procedure, a given experiment is performed in two different variants: once as normal giving signal SI , and once with a π pulse on the ge -transition appended at the end of the sequence just before measuring, giving signal SX . When the respective reference signals V_0 , V_1 , and V_2 of a transmon qubit prepared in the g , e and f state are known, the respective populations of the g and e states, P_0 and P_1 , can be extracted using:

$$\begin{bmatrix} V_0 - V_2 & V_1 - V_2 \\ V_1 - V_2 & V_0 - V_2 \end{bmatrix} \begin{bmatrix} P_0 \\ P_1 \end{bmatrix} = \begin{bmatrix} SI - V_2 \\ SX - V_2 \end{bmatrix} \quad (\text{A.43})$$

under the assumption that even higher-excited levels are unpopulated (i.e., $P_0 + P_1 + P_2 = 1$, where P_2 is the population in the f -state). Following [151], we fit the population of the computational subspace X_1 , to a single exponential:

$$X_1(N_{CL}) = A + B\lambda_1^{N_{CL}}. \quad (\text{A.44})$$

The average leakage (L_1) and seepage (L_2) rates per Clifford can then be extracted as:

$$L_1 = (1 - A)(1 - \lambda_1), \quad (\text{A.45})$$

$$L_2 L_2 = A(1 - \lambda_1). \quad (\text{A.46})$$

Using the fitted value of λ_1 , the survival probability M_0 is then fitted to a double exponential of the form:

$$M_0(N_{CL}) = A_0 + B_0 \lambda_1^{N_{CL}} + C_0 \lambda_2^{N_{CL}}. \quad (\text{A.47})$$

The average gate infidelity $\overline{\epsilon_{X_1}}$ of the gate set is given by:

$$\overline{\epsilon_{X_1}} = 1 - \frac{1}{d_1} [(d_1 - 1)\lambda_2 + 1 - L_1]. \quad (\text{A.48})$$

Note that if leakage is weak ($\lambda_1 \ll \lambda_2$ and $B \ll A$), then this reduces to the conventional RB formula. Reported error rates $\epsilon_{X_{1,g}}$ and leakage rates ϵ_L are per gate and are converted using

$$\epsilon_{X_{1,g}} = 1 - (1 - \overline{\epsilon_{X_1}})^{ng} \quad \text{and} \quad (\text{A.49})$$

$$\epsilon_L = 1 - (1 - L_1)^{ng}, \quad (\text{A.50})$$

in the same way as in methods [A.1.1](#) and [A.1.2](#).

A.7.2 Non Markovianity

There are currently no standardized protocols in use to characterize the Markovianity of gates. Non Markovianity is typically detected as a model violation or deviation from the expected results in another characterization method such as [A.1.4](#).

ACKNOWLEDGEMENTS

Like any project, a thesis does not write itself, and I could not have gotten this far without the help and support of many people. I would like to use these final pages to express my gratitude and thanks to you, the amazing people whom I have had the pleasure to work with during these years.

First of all, **Leo**. Without you, there would not be a superconducting quantum system to control. You first showed me your ideas for the Surface-17, or "Ninja-star" as we like to call it, when I was looking for a PhD. The idea of actually building this system resonated very strongly with me. It has been amazing to have contributed to this vision and to have seen it grow from an idea on a slide to a system in the lab. I have come to know you as a very passionate and inspiring teacher and scientist. Your scientific standards have made me more precise and aware, especially during moments in which I otherwise might have hand-waved my way out of things. And although we sometimes disagree on how to approach certain problems, your focus has balanced out my long-term bias and helped me to critically reflect on many of my assumptions. I hope to have (scientific) discussions and meantime IPAs with you for many years to come. Thank you, Leo.

Lieven, thank you for being my second promotor and the active role you have taken in the writing of this thesis. Your detailed feedback has motivated me to expand on several sections. I hope that some of the ideas expressed in this thesis can be applied to your research on spin systems and wish you the best of luck in your new position as director of research at QuTech.

I would like to thank all committee members for taking the time to review my thesis and oppose me in the defense ceremony. Professor Devoret, **Michel**, it is an honor to have you on my committee. I do apologize for the time difference; I wish we could have hosted you in Delft to have more discussions and make it more worthwhile for you. **Robin**, it has been a pleasure working with you and your team. I have learned a lot about QCVV and more fundamental subjects during our post site visit and conference discussions. It is safe to say that without you, Chapter 3 would not have been written. I hope to be the annoying experimentalist complaining about "exotic" errors from reality for many conferences to come. **Frank**, thank you for your help in realizing my first first-author paper. Your efforts in making optimal control theory work on physical systems and your work on the European quantum community set you apart from the rest. **Carmina**, it has been a pleasure to work with you and the other members of the Quantum & Computer Engineering department. You bring a different viewpoint to a room filled with physicists. **Barbara**, you have been an inspiration to work with. The breadth of your knowledge and interests, both within and outside of physics, never ceases to amaze me. As one of the few theorists active on both quantum algorithms and quantum error correction, as

well as on the details of device physics and control pulses, I am certain that if there will be a breakthrough in NISQ quantum computing, you will be involved. In the meantime, I hope to continue to read your blog posts. As an experimentalist trying to build a quantum computer, thank you for trying to get us to do so. **Gary**, your guidance has been invaluable, thank you for providing a more fundamentally interesting alternative to the road-map focused QuTech. I admire your efforts in making science a better and more open place. Also, you have the best coffee in the building.

To my paronyms, **Niels** and **Tom**, we started this together five years ago and I am happy that we are finishing it together as well. **Niels**, as my academic twin brother we started within a month of each other as gates guy and measurement guy. Although we started out working on the same project, our ways of working were very different. Where I tried to build a system that could withstand the ages, you recklessly copy pasted code that made the experiment run today. You then controlled this experiment with the efficiency and discipline of a machine. You taught me when to respect the duct tape, and when just good enough is far superior to perfect. In hindsight, our different way of working is probably exactly why we were such an effective team. Being in academia also meant that we traveled a lot, I have particularly good memories of our first IARPA tech meeting in Zürich, the LA March meeting (and a certain party), and our trip to China. Thank you for these memories. I hope we can continue to complement each other, now as commercial neighbors in the E-wing. **Tom**, you joined us a few months later to help us simulate some experiments, not knowing how many late nights we would cause you. However, the challenges of the experiment didn't phase you as you found the time to write the equivalent of *three* theses in your PhD, something which never ceases to amaze. Besides the many discussions on error correction, algorithms and experiments, the things I remember best are our trips. Somewhere in our first year we drove down to Switzerland to go skiing for a weekend. Because I was, predictably, stuck in the lab you offered to help out with the experiment. Although this did not accelerate my work, we did end up in the mountains where I dragged you down all the off-piste and black slopes I could find. Failing to understand the British/Aussie/Kiwi politeness I continued to do so on the subsequent skiing trips until you finally dragged me up a mountain in Colorado last year. Thank you for being an awesome mate.

To all my lab mates and to the people I have worked with over the years, thank you for the support and joy you brought to my work. There is no way I can do justice to all our interactions in these short pages.

Chris, having started well before me, you were the rebellious older brother. Your lobbying for proper control software is what made my PhD possible. It is in those early days of PycQED that the foundation for our friendship was formed. I fondly remember sleeping on your couch after a late session in the lab before flying to Copenhagen to present PycQED for the first time. Your knowledge of the literature combined with your creativity meant that you were often the most active contributor to any scientific discussion in the lab and *the* most active contributor to the QuTech blog. Although this may have defocused your efforts a bit, I am convinced that it is the basis for an outstanding scientific career. It has certainly inspired me to read more

papers and look for links in the literature. Similarly defining for my PhD were the weekly jam sessions at your place. You have achieved what no one else could, you got me to play the guitar and sing. **Gijs**, during our brief overlap you coached us when setting up PycQED and taught me the basics of IQ modulation. **Stefano**, as the senior post-doc you were the rigorous methodical worker. I hope life is well in the bay area. **Serwan**, as Chris' master student you were heavily involved in the creation of PycQED and fork-fights with Florian. **Flo**, I'm sorry for bothering you with git and breaking your experiment. I hope your recent spin adventures are enjoyable. I probably shouldn't say this, but you can hit me up for a Lüthi Meister anytime. I wish you all the best in Portland. **Nathan**, it is amazing what you built up in such a short amount of time while still having the patience to help us out, explain things, and teach us. I am sure you will have a very successful group in Sydney. **Xiang**, as the representative of EWI you were put in an impossible position having to balance the often conflicting scientific interests of two different disciplines. You have been the bridge that made the collaboration possible and one of the main reasons we could give our yearly demo to Intel. I really enjoyed it when you showed your favorite Chinese restaurant in Delft and when you cooked for us at the end of your PhD. The spicy nightmare was way too pleasant, but maybe that is for the better. **Ramiro**, you have been a chaotic sunshine of energy in the lab which sticks to a Latin clock. Your work on variational algorithms has always had my interest. It was a pleasure finally working with you directly during the last year of my thesis. I'm looking forward to seeing you have a dance-off at your defense party when the world is hopefully open again. **Brian**, you are one of the smartest and kindest people I have ever met. Your skill in running experiments while at the same time being our in house theory department was truly unique (and a bit intimidating). I wish you all the best in Finland. **Jacob**, your ability to systematically debug a system has helped me out more than once, but your ability to put things in perspective helped me the most. In your role as the old wise man, you have kept the lab together in the toughest of times. **Gerco**, good luck with freelancing. **Xavi**, as Tom's first PhD student, but more importantly, as my former master student and all-time theory collaborator, you have a special place in our lab. I most enjoyed your bleak world view and debating politics and science with you for hours. Good job on moving out of the city. Let's go for a ride again soon. **Filip**, it is great to see that you have found your passion again. **Thijs**, I love your chill. Good luck in continuing the work on the nanowire transmons. **Ale**, during my time in Delft you have been synonymous with fabrication. One of the things I love about you is that you live and breathe what you love. Besides your ability to conjure up magic in the cleanroom, you also know how to party. Let's go build that quantum computer! **Nands**, you have been a special kind of constant during my time in Delft. It is great to finally see you get the time to work on your own projects. I hope it'll lead to nice publications and a beautiful thesis, you deserve it. **Christos**, thank you for your work in the cleanroom and that blog post. **Nadia and Marc**, thank you for relentlessly iterating the designs of the devices. **Duije**, thank you for your work on the VSM, QWG and all the other microwave engineering projects that affected my work. It was a pleasure working with you. **Wouter**, your design work on the QWG and CC is impressive. I remember that I once ran into an issue with the QWG and the feature that was required to resolve my issue was already implemented. **Martin**, thank you for taking the

work on the CC to the next level. I am really enjoying our technical discussions. **Jules**, I really enjoyed the discussions during the many late working sessions and dinners at the hangout. I look forward to continuing that tradition in the E-wing. **Miguel**, you have some big shoes to fill. I am very impressed by how quickly you have grown into the bridge between computer engineering and physics. Good luck in pushing the Quantum Inspire to the next level. **Jordy**, I am continuously impressed by your ability to estimate how long a task will take in combination with your ability to deliver and always go home on time. **Raymond, Raymond, and Marijn**, thank you for having an open door policy for all the technical issues and questions we can come up with. Without you, QuTech would not be what it is today. **Matt** you are amazing at what you do. I love how knowledgeable and passionate you are about cooling. The lab is lucky to have you. **Stan**, you exceeded all expectations in your work on the Quantum Infinity. Your artwork shining on the cover of this thesis is testament to that. **Slava & Boris**, thank you for continuing quantumsim and modeling all those errors that do not fit into a circuit diagram. **Slava**, I love your dark sense of humor and your knowledge of the human psyche. **Jorge & Hany** it's good to see you settling in and working on the next generation of projects. **Kian** it was great having you rejoin the lab, if even only for a little while. I wish you all the best in team Diamond.

Bart, you were my first student. Your work on adaptive experiments has lasted throughout my PhD. **Sjoerd**, thank you for translating GST to experiment. **Livio**, it was great working with you. Thank you for the bouldering, for your work on Cryoscope, and for the first version of AutoDepGraph. **Rene**, thank you for designing the Purcell devices, they were instrumental in the results we achieved. **Luc**, the joy and energy you brought to the lab was inspiring. **Timo**, it was great to finally realize the dependency graph-based calibrations. **Victor**, you have combined your desire to program and simulate things with the ability to run experiments. Congratulations on taking the ideas of Net-Zero to the next level. **Berend**, you were a great help in the final stages of the Intel demo. I wish you the best of luck in finding a PhD position.

Francesco, thank you for the amazing work on simulating the Net-Zero pulses. Finishing on the ArXiv submission during the March Meeting was great teamwork. I will miss dropping by your office for some theory smalltalk. **Corsin**, thank you for turning my simple dataset into a great paper. **Lucas**, you were the first theorist I worked with, who could have thought that an idea over Pizza with Leo could turn into a paper.

The QuTech staff and leadership for transforming the Quantum Transport department into QuTech. **Leo K**, thank you starting the transformation and painting the original vision. **Ronald, Kees & Charlotte**, thank you for continuing the professionalization of QuTech. **GarreIt**, you have led the TNO efforts within QuTech and successfully bridged the culture gap between the engineers and scientists. Without you there would not be such a thing as a TNO quantum engineer. I look forward to working closely with you for the coming years. **Jeremy**, I know you as the project manager on the IARPA project, the new TNO quantum technology department is in good hands with you. **Freek**e thank you for the many coffees and discussions about the human aspects in science. Your push for the soft aspects will make the quantum community a healthier and more sustainable place. I look forward to being an active part of your quantum

ecosystem. **Leonie, Heera, Aletta, Julia** thank you for representing QuTech to the outside world. **Marja**, thank you for taking care of all my silly logistics questions, and listening to me when times were tough. **Jenny**, thanks for the coffee, it is always a pleasure. **Joanna**, thank you for processing TIM when I submitted too late. **Chantal, Yuki, Ingrid, Jolijn, Anna, Csilla, Grazia and all staff of QuTech**, thank you for taking care of all the things we scientists are not good at. **Jason, Olaf, Mark, Jelle, Remco, Siebe and all technical staff**, thank you for keeping the lab running.

The Quantum Inspire team for first assisting with the Quantum Infinity demo for Intel and later building the Quantum Inspire and taking it to the next level. **Jorrit**, your passion made you one of the best project managers I've ever seen in action. **Koen M**, I have always admired your technical vision and understanding of the relevant problems. **Kelvin**, it was a pleasure working with you on the Quantum Infinity demo. **Pieter, Nora, Thorsten, Richard, Luc, Martin, Mike, and all others who contributed to this project**, thank you for making this possible. **Garrealt**, thank you for taking the Quantum Inspire across the finish line.

Nader, Imran & Hans thank you for creating OpenQL. We have used it every single day. **Leon**, I wish you all the best at Duke. **Koen B & Carmina** thank you for your leading the efforts in the Quantum & Computer Engineering lab. I have learned a lot from all of you.

Dave, as our resident Intel engineer, you were there to balance out Ale. It was great having you here these years, but the thing I remember most fondly is our college days :). Thank you for having us in Portland and showing us the Beer O'Clock, I'm sure Ramiro is jealous. **Roman**, thank you for being the other side of the fab team. Those mojitos in LA sure packed a punch, muchos gracias. **Chris Z, Shavi, Sonika, Gian, Anne M**, it was a pleasure trying to realize your algorithms on a working quantum processor. **Jim & Jim**, thank you for your leadership over the Intel Quantum efforts. **Jeannete, Gian, Lester and all others at Intel**, thank you for a successful collaboration.

Niels H and Yves, thank you for helping us out all those times we succeeded to abuse your equipment. **Adrian, Sadik, Jana, Jan, and the greater Zurich Instruments team**, thank you for supporting our research.

Kenny, thanks for being my counterpart on the T&E team. I look forward to continuing our discussions on leakage, measurement GST, and life. **Erik**, thanks for pushing PyGSTi, it is an amazing tool. Swimming in the Limmat in Zürich was a great way to meet. **Timothy**, thank you for answering all my questions on benchmarking. **Robin and Kevin**, thank you for leading an amazing effort at the QPL.

Anton, teaching in your programming course was one of the highlights of my year. I admire your commitment to open software and science. I hope to continue to use and contribute to the tools you've created. **Joe W**, it was great to discuss code and software design with you. **Bas N**, adaptive is amazing. **Dave W, Alan, Alex, Guen, Giulio, Mikhail, Jens & William**, thank you for creating QCoDeS and continuing to support it.

Jonas, thank you for being my local benchmarking expert and taking the lead in setting up the blog. **James**, you are my favorite (lowland) Scot. **Suus**, the pasta party was great. **Suus**,

James, and Jonas, it was a pleasure to annoy everyone to write posts for us with you. **Anne-Marije, Tim, Guan & Matteo**, it was great to carry on this tradition with you. I am confident the blog is in capable hands. **Anne-Marije**, I get excited by the energy you share in all your initiatives. **Willemijn**, I always enjoyed going to the other side of the lab to distract you with some smalltalk. I wish you all the best in wrapping up your thesis. **Damaz**, I loved our discussions on code and small hackathons. **Jaap**, when is the second part of that blog post coming? **Suus & Jelmer**, thank you for organizing the running events. You made me a healthier person. **Victoria**, you are an amazing dancer and theorist. I hope you have a need for a computer capable of factoring 15 soon, so I can hook it up to your quantum internet. **Norbert**, you have been a great climbing buddy. Your success as a scientist was even more impressive, given how well you have your life in order. I hope you are enjoying life as a professional. **Nodar**, thank you for keeping QuTech free of demons and of course providing the Inspire with a silicon chip. **Sal**, thank you for your outside view on things going on in QuTech. We're going to build a quantum industry! **Remon**, your guitar skills are amazing. **Will**, thank you for carrying on Chris' tradition. The jam sessions were a highlight in my PhD. **Sara**, thank you for showing me the new optics labs. **Alessandro C**, thank you for helping with the residual coupling. I hope to use your designs in the future. **Mohamed, Joeri, Daniel, Axel, Kenneth, Maarten, David van W, David van D, Vincent, Gertjan, Bas D, Nico H, Sophie, Brian P, Stephan, Luca, Marta, Jeremy R, Mark, Arian, Sjaak, Floor, Xiao, G-Flo, Patrick, and all current and former students and post-docs**, it is you who make QuTech what it is. Thank you.

Tim & Julia, thank you for teaching me how to run an experiment. It was in part the great experience working for you that led me to pursue a PhD. **Tim**, congratulations on building your own group. I still think the carbon spin control is absolutely beautiful. **Ronald**, thank you for creating the environment and setting up a great team, first for doing my Bachelor and Master theses, and later through your leadership of QuTech. **Hannes, Wolfgang, Bas H, Suus, Norbert, Stefan, Machiel, Anais, and all the other members of team Diamond** I had the pleasure of working with. Thank you for teaching me the basics and inspiring me to pursue a PhD.

Sven & Gabriele, it was in your lab that I could experience what PhD life was like in a way that is impossible in any other way. **Jan, Joost, Joe, Arjan, Giuseppe, Chunming, and everyone else at CQC2T**, thank you for showing me an amazing time in Sydney and inspiring me to pursue a PhD.

Thank you to my parents, **Maarten & Irene, my friends & my family**, for being there when I pushed myself too hard, for showing me a mirror when I needed it, and for providing me with a perspective on what is important in life. You know who you are. Thank you.

Michiel Adriaan RoL

30–10–1989 Born in Amsterdam, The Netherlands.

Education

1994–2002 Willemspark School, Amsterdam

2002–2008 Vossius Gymnasium, Amsterdam

2008-2011 Bachelor of Science in Applied Physics

Delft University of Technology, Delft, the Netherlands

BSc. Thesis: Resonant Excitation and Detection of NV-centers
in Diamond

Supervisors: Dr. H. Bernien and Prof. dr. ir. R. Hanson

2012-2014 Master of Science in Applied Physics

Delft University of Technology, Delft, the Netherlands

Internship: Optoelectric Detection of Dopants in Silicon Nan-
odevices

Supervisors: Dr. ir. G. G. de Boo and Prof. dr. S. Rogge

MSc. Thesis: Parity Measurements on Weakly Coupled Carbon
Spins in Diamond

Supervisors: Dr. ir. J. C. Cramer, Dr. ir. T. H. Taminiau and
Prof. dr. ir. R. Hanson

2015-2020 Ph.D. in Applied Physics

Delft University of Technology, Delft, the Netherlands

Thesis: Control for programmable superconducting quan-
tum systems

Promotors: Prof. dr. L. DiCarlo and Prof. dr. ir. L. M. K. Van-
dersypen

Professional Experience

- 2012–2014 De Kleine Consultant, Amsterdam, the Netherlands
Consultant, Delft branch (2012)
Director of Internal Affairs, Executive Board (2013–2014)

Awards

- 2018 Zurich Instruments Pioneer Award, (1500 CHF prize money)
2017 Best paper Award-IEEE/ACM microarchitecture, 2017 (MICRO-50)
2017 HiPEAC Paper award at the International Symposium on Microarchitec-
ture.

LIST OF PUBLICATIONS

14. N. Khammassi, I. Ashraf, J. v. Someren, R. Nane, A. M. Krol, **M. A. Rol**, L. Lao, K. Bertels, and C. G. Almudever *OpenQL : A portable quantum programming framework for quantum accelerators.*, arXiv: 2005.13283, (2020).
13. C. C. Bultink, T. E. O'Brien, R. Vollmer, N. Muthusubramanian, M. W. Beekman, **M. A. Rol**, X. Fu, B. M. Tarasinski, V. Ostroukh, B. Varbanov, A. Bruno, L. DiCarlo, *Protecting quantum entanglement from qubit errors and leakage via repetitive parity measurements*, Science advances **6** (12), (2020).
12. X. Zou, S. P. Premaratne, **M. A. Rol**, S. Johri, V. Ostroukh, D. J. Michalak, R. Caudillo, J. S. Clarke, L. DiCarlo, A. Y. Matsuura *Enhancing a Near-Term Quantum Accelerator's Instruction Set Architecture for Material Science Applications*, IEEE Transactions on Quantum Engineering (2020).
11. **M. A. Rol**, L. Ciorciaro, F. K. Malinowski, B. M. Tarasinski, Y. Salathe, N. Haandbaek, J. Sedivy, C. C. Bultink, L. DiCarlo, *Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor*, Applied Physics Letters **116**, 054001 (2020).
10. **M. A. Rol**, F. Battistel, F. K. Malinowski, C. C. Bultink, B. M. Tarasinski, R. Vollmer, N. Haider, N. Muthusubramanian, A. Bruno, B. M. Terhal, L. DiCarlo, *Fast, High-Fidelity Conditional-Phase Gate Exploiting Leakage Interference in Weakly Anharmonic Superconducting Qubits*, Physical Review Letters **123**, 120502 (2019).
9. R. Sagastizabal, X. Bonet-Monroig, M. Singh, **M. A. Rol**, C. C. Bultink, X. Fu, C. H. Price, V. P. Ostroukh, N. Muthusubramanian, A. Bruno, M. Beekman, N. Haider, T. E. O'Brien, L. DiCarlo, *Error Mitigation by Symmetry Verification on a Variational Quantum Eigensolver*, Phys. Rev. A, **100**, 010302(R), (2019).
8. X. Fu, L. Rieseboos, **M. A. Rol**, J. van Straten, J. van Someren, N. Khammassi, I. Ashraf, R. F. L. Vermeulen, V. Newsum, K. L. Loh, J. C. de Sterke, W. J. Vlothuizen, R. N. Schouten, C. G. Almudever, L. DiCarlo, K. Bertels, *eQASM: An Executable Quantum Instruction Set Architecture*, Proceedings of 25th IEEE International Symposium on High-Performance Computer Architecture (HPCA 19) (2019),
7. C. Pfister, **M. A. Rol**, A. Mantri, M. Tomamichel, S. Wehner, *Capacity estimation and verification of quantum channels with arbitrarily correlated errors* Nature Communications **7**, 27 (2018).
6. F. Luthi, T. Stavenga, O.W. Enzing, A. Bruno, C. Dickel, N. K. Langford, **M. A. Rol**, T. S. Jespersen, J. Nygård, P. Krogstrup, L. DiCarlo, *Evolution of nanowire transmon qubits and their coherence in a magnetic field*, Physical Review Letters **120**, 100502 (2018).
5. X. Fu, **M. A. Rol**, C. C. Bultink, J. van Someren, N. Khammassi, I. Ashraf, R. F. L. Vermeulen, J. C. de Sterke, W. J. Vlothuizen, R. N. Schouten, C. G. Almudever, L. DiCarlo and K. Bertels, *A Microarchitecture for a Superconducting Quantum Processor*, IEEE Micro, **38** 40-47 (2018). [Top Picks from the 2017 Computer Architecture Conferences]
4. **M. A. Rol**, C. C. Bultink, T. E. O'Brien, S. R. de Jong, L. S. Theis, X. Fu, F. Luthi, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, D. Deurloo, R. N. Schouten, F. K. Wilhelm and L. DiCarlo, *Restless tuneup of high-fidelity qubit gates*, Physical Review Applied **7**, 041001 (2017).
3. C. C. Bultink, **M. A. Rol**, T. E. O'Brien, X. Fu, B. C. S. Dikken, C. Dickel, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, R. N. Schouten and L. DiCarlo, *Active Resonator Reset in the Nonlinear Dispersive Regime of Circuit QED*, Physical Review Applied **6**, 034008 (2016).

2. S. Asaad, C. Dickel, N. K. Langford, S. Poletto, A. Bruno, **M. A. Rol**, D. Deurloo, L. DiCarlo, *Independent, extensible control of same-frequency superconducting qubits by selective broadcasting*, npj Quantum Information **2**, 16029 (2016).
1. J. Cramer, N. Kalb, **M. A. Rol**, B. Hensen, M. S. Blok, M. Markham, D. J. Twitchen, R. Hanson, T. H. Taminiau, *Repeated quantum error correction on a continuously encoded qubit by real-time feedback*, Nature Communications **7**, 11526 (2016).

REFERENCES

- [1] R. P. FEYNMAN. *Simulating physics with computers*. International Journal of Theoretical Physics, **21** (6-7), 467–488, 1982. (see pages: [1](#), [158](#))
- [2] M. A. NIELSEN and I. L. CHUANG. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, 2000. (see pages: [1](#), [6](#), [7](#), [11](#), [32](#), [34](#), and [86](#))
- [3] M. BORN. *Zur quantenmechanik der stoßvorgänge*. Zeitschrift für Physik, **37**, 863–867, 1926. (see page: [3](#))
- [4] A. EINSTEIN, B. PODOLSKY, and N. ROSEN. *Can quantum-mechanical description of physical reality be considered complete?* Phys. Rev., **47**, 777–780, 1935. (see page: [3](#))
- [5] J. S. BELL. *On the Einstein-Podolsky-Rosen paradox*. Physics, **1**, 195–200, 1964. (see page: [3](#))
- [6] S. J. FREEDMAN and J. F. CLAUSER. *Experimental test of local hidden-variable theories*. Phys. Rev. Lett., **28**, 938–941, 1972. (see page: [3](#))
- [7] A. ASPECT, P. GRANGIER, and G. ROGER. *Experimental tests of realistic local theories via Bell's theorem*. Phys. Rev. Lett., **47**, 460–463, 1981.
- [8] G. WEIHS, T. JENNEWAIN, C. SIMON, H. WEINFURTER, and A. ZEILINGER. *Violation of bell's inequality under strict einstein locality conditions*. Phys. Rev. Lett., **81**, 5039–5043, 1998.
- [9] B. HENSEN, H. BERNIEN, A. E. DRÉAU, A. REISERER, N. KALB, M. S. BLOK, J. RUITENBERG, R. F. L. VERMEULEN, R. N. SCHOUTEN, C. ABELLÁN, W. AMAYA, V. PRUNERI, M. W. MITCHELL, et al. *Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres*. Nature, **526**, 682, 2015. (see page: [3](#))
- [10] H. EVERETT. *"relative state" formulation of quantum mechanics*. Rev. Mod. Phys., **29**, 454–462, 1957. (see page: [3](#))
- [11] K. SVOZIL. *Physical (a)causality: Determinism, randomness and uncaused events*, chapter "Shut Up and Calculate", pages 47–49. Springer International Publishing, Cham, 2018. (see page: [3](#))
- [12] J. PRESKILL. *Quantum computing and the entanglement frontier*. arXiv:1203.5813, 2012. (see page: [4](#))
- [13] A. HOLEVO. *Bounds for the quantity of information transmitted by a quantum communication channel*. Problemy Peredachi Informatsii, **9**, 3–11, 1973. (see page: [4](#))
- [14] P. W. SHOR. *Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer*. SIAM Journal on Computing, **26**, 1484, 1997. (see page: [4](#))
- [15] M. REIHER, N. WIEBE, K. M. SVORE, D. WECKER, and M. TROYER. *Elucidating reaction mechanisms on quantum computers*. Proceedings of the National Academy of Sciences, **114** (29), 7555–7560, 2017. (see page: [4](#))
- [16] J. BIAMONTE, P. WITTEK, N. PANCOTTI, P. REBENTROST, N. WIEBE, and S. LLOYD. *Quantum machine learning*. Nature, **549** (7671), 195–202, 2017. (see page: [4](#))
- [17] D. WECKER, M. B. HASTINGS, N. WIEBE, B. K. CLARK, C. NAYAK, and M. TROYER. *Solving strongly correlated electron models on a quantum computer*. Phys. Rev. A, **92**, 062318, 2015. (see page: [4](#))

- [18] W. H. ZUREK. *Decoherence and the transition from quantum to classical – revisited*. arXiv:0306072, 2003. (see page: 4)
- [19] I. L. CHUANG, L. M. K. VANDERSYPEN, X. ZHOU, D. W. LEUNG, and S. LLOYD. *Experimental realization of a quantum algorithm*. *Nature*, **393** (6681), 143–146, 1998. (see page: 4)
- [20] L. M. K. VANDERSYPEN, M. STEFFEN, G. BREYTA, C. S. YANNONI, M. H. SHERWOOD, and I. L. CHUANG. *Experimental realization of Shor’s quantum factoring algorithm*. *Nature*, **414**, 883–887, 2001.
- [21] L. M. K. VANDERSYPEN and I. L. CHUANG. *NMR techniques for quantum control and computation*. *Rev. Mod. Phys.*, **76**, 1037–1069, 2005. (see pages: 4, 16, and 31)
- [22] S. GULDE, M. RIEBE, G. P. T. LANCASTER, C. BECHER, J. ESCHNER, H. HÄFFNER, F. SCHMIDT-KALER, I. L. CHUANG, and R. BLATT. *Implementation of the Deutsch-Jozsa algorithm on an ion-trap quantum computer*. *Nature*, **421**, 48–50, 2003. (see page: 4)
- [23] C. D. BRUZEWICZ, J. CHIAVERINI, R. MCCONNELL, and J. M. SAGE. *Trapped-ion quantum computing: Progress and challenges*. *Applied Physics Reviews*, **6** (2), 021 314, 2019. (see page: 4)
- [24] P. G. KWIAT, J. R. MITCHELL, P. D. D. SCHWINDT, and A. G. WHITE. Grover’s search algorithm: an optical approach. *J. Mod. Opt.*, **47**, 257–266, 2000. (see page: 4)
- [25] S. H. TAN and P. P. ROHDE. *The resurgence of the linear optics quantum interferometer — recent advances & applications*. *Reviews in Physics*, **4**, 100 030, 2019. (see page: 4)
- [26] L. DICARLO, J. M. CHOW, J. M. GAMBETTA, L. S. BISHOP, B. R. JOHNSON, D. I. SCHUSTER, J. MAJER, A. BLAIS, L. FRUNZIO, S. M. GIRVIN, and R. J. SCHOELKOPF. *Demonstration of two-qubit algorithms with a superconducting quantum processor*. *Nature*, **460**, 240, 2009. (see pages: 5, 16, 17, 19, 21, 37, 90, and 110)
- [27] M. KJAERGAARD, M. E. SCHWARTZ, J. BRAUMÜLLER, P. KRANTZ, J. I. J. WANG, S. GUSTAVSSON, and W. D. OLIVER. *Superconducting qubits: Current state of play*. ArXiv:1905.13641, 2019. (see pages: 19, 90)
- [28] P. KRANTZ, M. KJAERGAARD, F. YAN, T. P. ORLANDO, S. GUSTAVSSON, and W. D. OLIVER. *A quantum engineer’s guide to superconducting qubits*. *Appl. Phys. Rev.*, **6** (2), 021 318, 2019. (see pages: 5, 19)
- [29] T. VAN DER SAR, Z. H. WANG, M. S. BLOK, H. BERNIEN, T. H. TAMINIAU, D. M. TOYLI, D. A. LIDAR, D. D. AWSCHALOM, R. HANSON, and V. V. DOBROVITSKI. *Decoherence-protected quantum gates for a hybrid solid-state spin register*. *Nature*, **484** (7392), 82–86, 2012. (see page: 5)
- [30] M. W. DOHERTY, N. B. MANSON, P. DELANEY, F. JELEZKO, J. WRACHTRUP, and L. C. HOLLENBERG. *The nitrogen-vacancy colour centre in diamond*. *Physics Reports*, **528** (1), 1–45, 2013. (see page: 5)
- [31] T. F. WATSON, S. G. J. PHILIPS, E. KAWAKAMI, D. R. WARD, P. SCARLINO, M. VELDHORST, D. E. SAVAGE, M. G. LAGALLY, M. FRIESEN, S. N. COPPERSMITH, M. A. ERIKSSON, and L. M. K. VANDERSYPEN. *A programmable two-qubit quantum processor in silicon*. *Nature*, **555** (7698), 633–637, 2018. (see page: 5)
- [32] F. A. ZWANENBURG, A. S. DZURAK, A. MORELLO, M. Y. SIMMONS, L. C. L. HOLLENBERG, G. KLIMECK, S. ROGGE, S. N. COPPERSMITH, and M. A. ERIKSSON. *Silicon quantum electronics*. *Rev. Mod. Phys.*, **85**, 961–1019, 2013. (see page: 5)
- [33] J. PRESKILL. *Quantum Computing in the NISQ era and beyond*. *Quantum*, **2**, 79, 2018. (see pages: 5, 156, and 158)
- [34] N. MOLL, P. BARKOUTSOS, L. S. BISHOP, J. M. CHOW, A. CROSS, D. J. EGGER, S. FILIPP, A. FUHRER, J. M. GAMBETTA, and M. GANZHORN. *Quantum optimization using variational algorithms on near-term quantum devices*. *Quantum Science and Technology*, **3** (3), 030 503, 2018. (see pages: 158, 159)

- [35] A. W. CROSS, L. S. BISHOP, S. SHELDON, P. D. NATION, and J. M. GAMBETTA. *Validating quantum computers using randomized model circuits*. Phys. Rev. A, **100**, 032328, 2019. (see pages: 5, 158, and 159)
- [36] F. ARUTE, K. ARYA, R. BABBUSH, D. BACON, J. C. BARDIN, R. BARENDS, R. BISWAS, S. BOIXO, F. G. S. L. BRANDAO, D. A. BUELL, B. BURKETT, Y. CHEN, Z. CHEN, *et al.* *Quantum supremacy using a programmable superconducting processor*. Nature, **574** (7779), 505–510, 2019. (see pages: 5, 14, 16, 55, 147, 150, 151, 157, 158, 159, and 163)
- [37] A. PERUZZO, J. R. MCCLEAN, P. SHADBOLT, M. YUNG, X. ZHOU, P. J. LOVE, A. ASPURU-GUZIK, and J. L. O'BRIEN. *A variational eigenvalue solver on a photonic quantum processor*. Nat. Commun., **5**, 4213, 2014. (see pages: 5, 12)
- [38] D. WECKER, M. B. HASTINGS, and M. TROYER. *Progress towards practical quantum variational algorithms*. Phys. Rev. A, **92**, 042303, 2015.
- [39] P. J. J. O'MALLEY, R. BABBUSH, I. D. KIVLICHAN, J. ROMERO, J. R. MCCLEAN, R. BARENDS, J. KELLY, P. ROUSHAN, A. TRANTER, N. DING, B. CAMPBELL, Y. CHEN, Z. CHEN, *et al.* *Scalable quantum simulation of molecular energies*. Phys. Rev. X, **6**, 031007, 2016. (see pages: 5, 12)
- [40] J. L. PARK. *The concept of transition in quantum mechanics*. Foundations of Physics, **1** (1), 23–33, 1970. (see page: 5)
- [41] W. K. WOOTTERS and W. H. ZUREK. *A single quantum cannot be cloned*. Nature, **246** (5429), 170–170, 1973. (see page: 5)
- [42] J. KELLY, R. BARENDS, A. G. FOWLER, A. MEGRANT, E. JEFFREY, T. WHITE, D. SANK, J. MUTUS, B. CAMPBELL, Y. CHEN, B. CHIARO, A. DUNSWORTH, I.-C. HOI, *et al.* *State preservation by repetitive error detection in a superconducting quantum circuit*. Nature, **519** (7541), 66–69, 2015. (see pages: 5, 11, 38, 74, 75, and 110)
- [43] D. RISTÈ, S. POLETTI, M. Z. HUANG, A. BRUNO, V. VESTERINEN, O. P. SAIRA, and L. DICARLO. *Detecting bit-flip errors in a logical qubit using stabilizer measurements*. Nat. Commun., **6**, 6983, 2015. (see pages: 23, 82, and 110)
- [44] J. CRAMER, N. KALB, M. A. ROL, B. HENSEN, M. MARKHAM, D. J. TWITCHEN, R. HANSON, and T. H. TAMINIAU. *Repeated quantum error correction on a continuously encoded qubit by real-time feedback*. Nat. Commun., **5**, 11526, 2016. (see pages: 5, 74)
- [45] B. M. TERHAL. *Quantum error correction for quantum memories*. Rev. Mod. Phys., **87**, 307–346, 2015. (see pages: 6, 7, and 81)
- [46] S. B. BRAVYI and A. Y. KITAEV. *Quantum codes on a lattice with boundary*. arXiv:quant-ph/9811052, 1998. (see pages: 6, 19)
- [47] E. DENNIS, A. KITAEV, A. LANDAHL, and J. PRESKILL. *Topological quantum memory*. Journal of Mathematical Physics, **43** (4452), 2002.
- [48] A. G. FOWLER, M. MARIANTONI, J. M. MARTINIS, and A. N. CLELAND. *Surface codes: Towards practical large-scale quantum computation*. Phys. Rev. A, **86**, 032324, 2012. (see pages: 6, 7, 19, 23, 58, 110, and 117)
- [49] X. FU, L. RIESEBOS, L. LAO, C. G. ALMUDEVER, F. SEBASTIANO, R. VERSLUIS, E. CHARBON, and K. BERTELS. *A heterogeneous quantum computer architecture*. In *Proceedings of the ACM International Conference on Computing Frontiers*, CF '16, pages 323–330. ACM, New York, NY, USA, 2016. (see page: 9)
- [50] A. D. CORCOLES, A. KANDALA, A. JAVADI-ABHARI, D. T. MCCLURE, A. W. CROSS, K. TEMME, P. D. NATION, M. STEFFEN, and J. M. GAMBETTA. *Challenges and opportunities of near-term quantum computing systems*. ArXiv:1910.02894, 2019. (see pages: 9, 150, and 151)

- [51] K. SVORE, M. ROETTELER, A. GELLER, M. TROYER, J. AZARIAH, C. GRANADE, B. HEIM, V. KLIUCHNIKOV, M. MYKHAILOVA, and A. PAZ. *Q#: Enabling scalable quantum computing and development with a high-level dsl*. Proceedings of the Real World Domain Specific Languages Workshop 2018 on - RWDSL2018, 2018. (see page: 9)
- [52] X. FU, M. A. ROL, C. C. BULTINK, J. VAN SOMEREN, N. KHAMMASSI, I. ASHRAF, R. F. L. VERMEULEN, J. C. DE STERKE, W. J. VLOTHUIZEN, R. N. SCHOUTEN, C. G. ALMUDEVER, L. DICARLO, and K. BERTELS. *An experimental microarchitecture for a superconducting quantum processor*. In *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO-50 '17, pages 813–825. ACM, New York, NY, USA, 2017. (see pages: 9, 24, 91, 97, 110, and 156)
- [53] X. FU, L. RIESEBOS, M. A. ROL, J. VAN STRATEN, J. VAN SOMEREN, N. KHAMMASSI, I. ASHRAF, R. F. L. VERMEULEN, V. NEWSUM, K. K. L. LOH, J. C. DE STERKE, W. J. VLOTHUIZEN, R. N. SCHOUTEN, *et al.* *eQASM: An executable quantum instruction set architecture*. In *Proceedings of 25th IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 224–237. IEEE, 2019. (see pages: 9, 24, 25, 91, 97, 110, 146, and 156)
- [54] N. KHAMMASSI, I. ASHRAF, J. V. SOMEREN, R. NANE, A. M. KROL, M. A. ROL, L. LAO, K. BERTELS, and C. G. ALMUDEVER. *Openql : A portable quantum programming framework for quantum accelerators*. arXiv:2005.13283, 2020. (see pages: 9, 25)
- [55] D. S. STEIGER, T. HÄNER, and M. TROYER. *ProjectQ: an open source software framework for quantum computing*. *Quantum*, 2, 49, 2018. (see pages: 9, 146)
- [56] P. ALIFERIS and B. M. TERHAL. *Fault-tolerant quantum computation for local leakage faults*. *Quantum Info. Comput.*, 7 (1), 139–156, 2007. (see pages: 11, 110)
- [57] A. G. FOWLER. *Coping with qubit leakage in topological codes*. *Phys. Rev. A*, 88, 042308, 2013. (see page: 39)
- [58] J. GHOSH, A. G. FOWLER, J. M. MARTINIS, and M. R. GELLER. *Understanding the effects of leakage in superconducting quantum-error-detection circuits*. *Phys. Rev. A*, 88, 062329, 2013.
- [59] J. GHOSH and A. G. FOWLER. *Leakage-resilient approach to fault-tolerant quantum computing with superconducting elements*. *Phys. Rev. A*, 91, 020302, 2015. (see page: 39)
- [60] M. SUCHARA, A. W. CROSS, and J. M. GAMBETTA. *Leakage suppression in the toric code*. *Quantum Info. Comput.*, 15 (11-12), 997–1016, 2015. (see pages: 11, 110)
- [61] C. C. BULTINK, T. E. O'BRIEN, R. VOLLMER, N. MUTHUSUBRAMANIAN, M. W. BEEKMAN, M. A. ROL, X. FU, B. TARASINSKI, V. OSTROUKH, B. VARBANOV, A. BRUNO, and L. DICARLO. *Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements*. *Science Advances*, 6 (12), 2020. (see pages: 26, 37, 38, 39, 44, 52, 92, 97, 110, and 117)
- [62] B. VARBANOV, F. BATTISTEL, B. M. TARASINSKI, V. P. OSTROUKH, T. E. O'BRIEN, L. DICARLO, and B. M. TERHAL. *Characterization and detection of leakage in simulations of a transmon-qubit surface-17 architecture*. arXiv:2002.07119, 2020. (see pages: 11, 23, and 39)
- [63] D. RISTÈ, C. C. BULTINK, K. W. LEHNERT, and L. DICARLO. *Feedback control of a solid-state qubit using high-fidelity projective measurement*. *Phys. Rev. Lett.*, 109, 240502, 2012. (see pages: 11, 37, 39, 40, 58, and 81)
- [64] D. SANK, Z. CHEN, M. KHEZRI, J. KELLY, R. BARENDS, B. CAMPBELL, Y. CHEN, B. CHIARO, A. DUNSWORTH, A. FOWLER, E. JEFFREY, E. LUCERO, A. MEGRANT, *et al.* *Measurement-induced state transitions in a superconducting qubit: Beyond the rotating wave approximation*. *Phys. Rev. Lett.*, 117, 190503, 2016. (see pages: 11, 16, 38, 80, and 82)
- [65] E. MAGESAN, J. M. GAMBETTA, A. D. CÔRCOLES, and J. M. CHOW. *Machine learning for discriminating quantum measurement trajectories and improving readout*. *Phys. Rev. Lett.*, 114, 200501, 2015. (see pages: 11, 76)

- [66] X. ZOU, S.P.PREMARATNE, M.A.ROL, S.JOHRI, V.OSTROUKH, D.J.MICHALAK, R.CAUDILLO, J.S.CLARKE, L. DICARLO, and A. Y. MATSUURA. *Enhancing a near-term quantum accelerator's instruction set architecture for material science applications*. IEEE Transactions on Quantum Engineering, **1**, 1–7, 2020. (see pages: 12, 25, and 156)
- [67] R. SAGASTIZABAL, X. BONET-MONROIG, M. SINGH, M. A. ROL, C. C. BULTINK, X. FU, C. H. PRICE, V. P. OSTROUKH, N. MUTHUSUBRAMANIAN, A. BRUNO, M. BEEKMAN, N. HAIDER, T. E. O'BRIEN, *et al.* *Error Mitigation by Symmetry Verification on a Variational Quantum Eigensolver*. Phys. Rev. A, **100**, 010302(R), 2019. (see pages: 12, 25, 92, 93, 97, 98, 100, and 157)
- [68] R. BARENDS, C. M. QUINTANA, A. G. PETUKHOV, Y. CHEN, D. KAFRI, K. KECHEDZHI, R. COLLINS, O. NAAMAN, S. BOIXO, F. ARUTE, K. ARYA, D. BUELL, B. BURKETT, *et al.* *Diabatic gates for frequency-tunable superconducting qubits*. Phys. Rev. Lett., **123**, 210501, 2019. (see pages: 12, 37, and 157)
- [69] G. G. GUERRESCHI and M. SMELYANSKIY. *Practical optimization for hybrid quantum-classical algorithms*. arXiv:1701.01450, 2017. 1701.01450. (see page: 12)
- [70] J. ROMERO, R. BABBUSH, J. R. MCCLEAN, C. HEMPEL, P. J. LOVE, and A. ASPURU-GUZI. *Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz*. Quantum Science and Technology, **4** (1), 014008, 2018. (see page: 12)
- [71] I. KASSALA and A. ASPURU-GUZI. *Quantum algorithm for molecular properties and geometry optimization*. J. Chem. Phys., **131**, 224102, 2009. (see page: 12)
- [72] S. MCARDLE, S. ENDO, A. ASPURU-GUZI, S. BENJAMIN, and X. YUAN. *Quantum computational chemistry*. ArXiv:1808.10402, 2018. (see page: 158)
- [73] A. ROGGERO and J. CARLSON. *Dynamic linear response quantum algorithm*. Physical Review C, **100** (3), 2019.
- [74] T. E. O'BRIEN, B. SENJEAN, R. SAGASTIZABAL, X. BONET-MONROIG, A. DUTKIEWICZ, F. BUDA, L. DICARLO, and L. VISSCHER. *Calculating energy derivatives for quantum chemistry on a quantum computer*. arXiv:1905.03742, 2019. (see pages: 12, 97)
- [75] C. GIDNEY *et al.* *Cirq*, 2018. (see pages: 12, 146)
- [76] J. KOCH, T. M. YU, J. GAMBETTA, A. A. HOUCK, D. I. SCHUSTER, J. MAJER, A. BLAIS, M. H. DEVORET, S. M. GIRVIN, and R. J. SCHOELKOPF. *Charge-insensitive qubit design derived from the Cooper pair box*. Phys. Rev. A, **76**, 042319, 2007. (see pages: 14, 15, 16, and 92)
- [77] J. A. SCHREIER, A. A. HOUCK, J. KOCH, D. I. SCHUSTER, B. R. JOHNSON, J. M. CHOW, J. M. GAMBETTA, J. MAJER, L. FRUNZIO, M. H. DEVORET, S. M. GIRVIN, and R. J. SCHOELKOPF. *Suppressing charge noise decoherence in superconducting charge qubits*. Phys. Rev. B, **77**, 180502, 2008. (see page: 14)
- [78] R. BARENDS, J. KELLY, A. MEGRANT, A. VEITIA, D. SANK, E. JEFFREY, T. C. WHITE, J. MUTUS, A. G. FOWLER, B. CAMPBELL, Y. CHEN, Z. CHEN, B. CHIARO, *et al.* *Superconducting quantum circuits at the surface code threshold for fault tolerance*. Nature, **508** (7497), 500, 2014. (see pages: 14, 16, 19, 43, 65, 110, 116, and 166)
- [79] T. E. O'BRIEN, B. M. TARASINSKI, and L. DICARLO. *Density-matrix simulation of small surface codes under current and projected experimental noise*. npj Quantum Information, **3** (39), 2017. (see pages: 14, 23, 28, 37, 43, and 69)
- [80] IBMQ. *IBM Quantum Experience*, 2016. (see page: 14)
- [81] C. RIGETTI. *Introducing rigetti quantum cloud services*, 2018.
- [82] D. T. MCCLURE and J. GAMBETTA. *Quantum computation center opens*, 2019. (see page: 14)

- [83] L. DICARLO. Proposal for IARPA LogiQ, not publicly available, 2015. (see pages: 14, 155)
- [84] B. D. JOSEPHSON. *Possible new effects in superconductive tunnelling*. Phys. Lett., 1 (7), 251–253, 1962. (see page: 14)
- [85] B. D. JOSEPHSON. *Coupled superconductors*. Rev. Mod. Phys., 36, 216–220, 1964. (see page: 14)
- [86] V. BOUCHIAT, D. VION, P. JOYEZ, D. ESTEVE, and M. H. DEVORET. *Quantum coherence with a single Cooper pair*. Phys. Scr., T76, 165, 1998. (see page: 15)
- [87] Y. NAKAMURA, Y. PASHKIN, and J. TSAI. *Coherent control of macroscopic quantum states in a single-Cooper-pair box*. Nature, 398, 786, 1999. (see page: 15)
- [88] J. GAMBETTA and S. SHELDON. *Cramming more power into a quantum device*, 2019. (see page: 15)
- [89] M. TINKHAM. *Introduction to superconductivity*. Dover Publications, New York, 2nd edition, 2004. (see page: 15)
- [90] M. HUTCHINGS, J. B. HERTZBERG, Y. LIU, N. T. BRONN, G. A. KEEFE, M. BRINK, J. M. CHOW, and B. PLOURDE. *Tunable superconducting qubits with flux-independent coherence*. Phys. Rev. Appl., 8 (4), 044003, 2017. (see pages: 16, 157)
- [91] N. K. LANGFORD, R. SAGASTIZABAL, M. KOUNALAKIS, C. DICKEL, A. BRUNO, F. LUTHI, D. THOEN, A. ENDO, and L. DICARLO. *Experimentally simulating the dynamics of quantum light and matter at deep-strong coupling*. Nat. Commun., 8, 1715, 2017. (see pages: 16, 43, 44, 90, 92, 93, and 157)
- [92] S. ASAAD, C. DICKEL, S. POLETTA, A. BRUNO, N. K. LANGFORD, M. A. ROL, D. DEURLOO, and L. DICARLO. *Independent, extensible control of same-frequency superconducting qubits by selective broadcasting*. npj Quantum Inf., 2, 16029, 2016. (see pages: 16, 23, 24, 37, 47, 60, 61, 116, 132, 155, 162, 172, and 174)
- [93] J. M. CHOW, J. M. GAMBETTA, A. W. CROSS, S. T. MERKEL, C. RIGETTI, and M. STEFFEN. *Microwave-activated conditional-phase gate for superconducting qubits*. New Journal of Physics, 15, 2013. (see pages: 16, 20)
- [94] P. V. KLIMOV, J. KELLY, Z. CHEN, M. NEELEY, A. MEGRANT, B. BURKETT, R. BARENDS, K. ARYA, B. CHIARO, Y. CHEN, A. DUNSWORTH, A. FOWLER, B. FOXEN, *et al.* *Fluctuations of energy-relaxation times in superconducting qubits*. Phys. Rev. Lett., 121, 090502, 2018. (see pages: 16, 30, 42, 156, and 157)
- [95] F. W. STRAUCH, P. R. JOHNSON, A. J. DRAGT, C. J. LOBB, J. R. ANDERSON, and F. C. WELLSTOOD. *Quantum logic gates for coupled superconducting phase qubits*. Phys. Rev. Lett., 91, 167005, 2003. (see pages: 16, 19, 21, 37, 65, and 110)
- [96] S. A. CALDWELL, N. DIDIER, C. A. RYAN, E. A. SETE, A. HUDSON, P. KARALEKAS, R. MANENTI, M. P. DA SILVA, R. SINCLAIR, E. ACALA, N. ALIDOUST, J. ANGELES, A. BESTWICK, *et al.* *Parametrically activated entangling gates using transmon qubits*. Phys. Rev. Applied, 10, 034050, 2018. (see pages: 16, 21, and 110)
- [97] F. MOTZOI, J. M. GAMBETTA, P. REBENTROST, and F. K. WILHELM. *Simple pulses for elimination of leakage in weakly nonlinear qubits*. Phys. Rev. Lett., 103, 110501, 2009. (see pages: 16, 35, 37, and 59)
- [98] J. M. CHOW, L. DICARLO, J. M. GAMBETTA, F. MOTZOI, L. FRUNZIO, S. M. GIRVIN, and R. J. SCHOELKOPF. *Optimized driving of superconducting artificial atoms for improved single-qubit gates*. Phys. Rev. A, 82, 040305, 2010. (see pages: 16, 35, 60, and 76)
- [99] M. A. ROL, C. C. BULTINK, T. E. O'BRIEN, S. R. DE JONG, L. S. THEIS, X. FU, F. LUTHI, R. F. L. VERMEULEN, J. C. DE STERKE, A. BRUNO, D. DEURLOO, R. N. SCHOUTEN, F. K. WILHELM, *et al.* *Restless tuneup of high-fidelity qubit gates*. Phys. Rev. Applied, 7, 041001, 2017. (see pages: 16, 30, 42, 57, 81, 110, and 165)

- [100] D. C. MCKAY, C. J. WOOD, S. SHELDON, J. M. CHOW, and J. M. GAMBETTA. *Efficient z gates for quantum computing*. Phys. Rev. A, **96**, 022330, 2017. (see pages: 16, 37)
- [101] M. BOISSONNEAULT, J. M. GAMBETTA, and A. BLAIS. *Dispersive regime of circuit qed: Photon-dependent qubit dephasing and relaxation rates*. Phys. Rev. A, **79**, 013819, 2009. (see pages: 16, 38)
- [102] J. GAMBETTA, A. BLAIS, D. I. SCHUSTER, A. WALLRAFF, L. FRUNZIO, J. MAJER, M. H. DEVORET, S. M. GIRVIN, and R. J. SCHOELKOPF. *Qubit-photon interactions in a cavity: Measurement-induced dephasing and number splitting*. Phys. Rev. A, **74**, 042318, 2006. (see pages: 16, 50, 52, and 173)
- [103] R. BARENDTS, J. KELLY, A. MEGRANT, D. SANK, E. JEFFREY, Y. CHEN, Y. YIN, B. CHIARO, J. MUTUS, C. NEILL, P. O'MALLEY, P. ROUSHAN, J. WENNER, *et al.* *Coherent Josephson qubit suitable for scalable quantum integrated circuits*. Phys. Rev. Lett., **111**, 080502, 2013. (see page: 17)
- [104] Y. CHEN, C. NEILL, P. ROUSHAN, N. LEUNG, M. FANG, R. BARENDTS, J. KELLY, B. CAMPBELL, Z. CHEN, B. CHIARO, A. DUNSWORTH, E. JEFFREY, A. MEGRANT, *et al.* *Qubit architecture with high coherence and fast tunable coupling*. Phys. Rev. Lett., **113**, 220502, 2014. (see pages: 17, 19, and 156)
- [105] M. KOUNALAKIS, C. DICKEL, A. BRUNO, N. K. LANGFORD, and G. A. STEELE. *Tuneable hopping and nonlinear cross-Kerr interactions in a high-coherence superconducting circuit*. npj Quantum Information, **4** (1), 1–7, 2018.
- [106] D. C. MCKAY, S. FILIPP, A. MEZZACAPO, E. MAGESAN, J. M. CHOW, and J. M. GAMBETTA. *Universal gate for fixed-frequency qubits via a tunable bus*. Phys. Rev. Applied, **6**, 064007, 2016. (see pages: 17, 21, and 156)
- [107] X. GU, A. F. KOCKUM, A. MIRANOWICZ, Y. XI LIU, and F. NORI. *Microwave photonics with superconducting quantum circuits*. Physics Reports, **718-719**, 1 – 102, 2017. Microwave photonics with superconducting quantum circuits. (see page: 19)
- [108] A. BLAIS, A. L. GRIMSMO, S. M. GIRVIN, and A. WALLRAFF. *Circuit quantum electrodynamics*. ArXiv:2005.12667, 2020. (see page: 19)
- [109] D. S. WANG, A. G. FOWLER, and L. C. L. HOLLENBERG. *Surface code quantum computing with error rates over 1%*. Phys. Rev. A, **83**, 020302, 2011. (see pages: 19, 23)
- [110] A. G. FOWLER, A. C. WHITESIDE, and L. C. L. HOLLENBERG. *Towards practical classical processing for the surface code*. Phys. Rev. Lett., **108**, 180501, 2012. (see pages: 19, 23)
- [111] M. NEELEY, R. C. BIALCZAK, M. LENANDER, E. LUCERO, M. MARIANTONI, A. D. O'CONNELL, D. SANK, H. WANG, M. WEIDES, J. WENNER, Y. YIN, T. YAMAMOTO, A. N. CLELAND, *et al.* *Generation of three-qubit entangled states using superconducting phase qubits*. Nature, **467** (7315), 570, 2010. (see page: 19)
- [112] M. A. ROL, F. BATTISTEL, F. K. MALINOWSKI, C. C. BULTINK, B. M. TARASINSKI, R. VOLLMER, N. HAIDER, N. MUTHUSUBRAMANIAN, A. BRUNO, B. M. TERHAL, and L. DICARLO. *Fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits*. Phys. Rev. Lett., **123**, 120502, 2019. (see pages: 19, 34, 37, 38, 90, 92, 97, and 104)
- [113] C. RIGETTI and M. DEVORET. *Fully microwave-tunable universal gates in superconducting qubits with linear couplings and fixed transition frequencies*. Phys. Rev. B, **81**, 134507, 2010. (see page: 20)
- [114] J. M. CHOW, A. D. CÓRCOLES, J. M. GAMBETTA, C. RIGETTI, B. R. JOHNSON, J. A. SMOLIN, J. R. ROZEN, G. A. KEEFE, M. B. ROTHWELL, M. B. KETCHEN, and M. STEFFEN. *Simple all-microwave entangling gate for fixed-frequency superconducting qubits*. Phys. Rev. Lett., **107**, 080502, 2011. (see page: 110)
- [115] S. POLETTI, J. M. GAMBETTA, S. T. MERKEL, J. A. SMOLIN, J. M. CHOW, A. D. CÓRCOLES, G. A. KEEFE, M. B. ROTHWELL, J. R. ROZEN, D. W. ABRAHAM, C. RIGETTI, and M. STEFFEN. *Entanglement of two superconducting qubits in a waveguide cavity via monochromatic two-photon excitation*. Phys. Rev. Lett., **109**, 240505, 2012.

- [116] H. PAIK, A. MEZZACAPO, M. SANDBERG, D. T. MCCLURE, B. ABDO, A. D. CÓRCOLES, O. DIAL, D. F. BOGORIN, B. L. T. PLOURDE, M. STEFFEN, A. W. CROSS, J. M. GAMBETTA, and J. M. CHOW. *Experimental demonstration of a resonator-induced phase gate in a multiqubit circuit-qed system*. Phys. Rev. Lett., **117**, 250502, 2016.
- [117] S. SHELDON, E. MAGESAN, J. M. CHOW, and J. M. GAMBETTA. *Procedure for systematically tuning up cross-talk in the cross-resonance gate*. Physical Review A, **93**, 060302, 2016. (see pages: 20, 110)
- [118] R. SCHUTJENS, F. A. DAGGA, D. J. EGGER, and F. K. WILHELM. *Single-qubit gates in frequency-crowded transmon systems*. Phys. Rev. A, **88**, 052330, 2013. (see page: 20)
- [119] M. BRINK, J. M. CHOW, J. HERTZBERG, E. MAGESAN, and S. ROSENBLATT. *Device challenges for near term superconducting quantum processors: frequency collisions*. In *2018 IEEE International Electron Devices Meeting (IEDM)*, pages 6.1.1–6.1.3. 2018. (see page: 20)
- [120] S. S. HONG, A. T. PAPAGEORGE, P. SIVARAJAH, G. CROSSMAN, N. DIDIER, A. M. POLLORENO, E. A. SETE, S. W. TURKOWSKI, M. P. DA SILVA, and B. R. JOHNSON. *Demonstration of a parametrically activated entangling gate protected from flux noise*. Phys. Rev. A, **101**, 012302, 2020. (see pages: 21, 110)
- [121] R. VERSLUIS, S. POLETO, N. KHAMMASSI, B. M. TARASINSKI, N. HAIDER, D. J. MICHALAK, A. BRUNO, K. BERTELS, and L. DICARLO. *Scalable quantum circuit and control for a superconducting surface code*. Phys. Rev. Appl., **8**, 034021, 2017. (see pages: 22, 23, 28, 35, 46, 110, 117, 118, 155, and 162)
- [122] A. BRUNO, S. POLETO, N. HAIDER, and L. DICARLO. *X48.00004: Extensible circuit QED processor architecture with vertical I/O*. APS March Meeting 2016. (see pages: 22, 161)
- [123] D. ROSENBERG, D. KIM, R. DAS, D. YOST, S. GUSTAVSSON, D. HOVER, P. KRANTZ, A. MELVILLE, L. RACZ, G. O. SAMACH, S. J. WEBER, F. YAN, J. L. YODER, *et al.* *3D integrated superconducting qubits*. npj Quantum Information, **3** (1), 1–4, 2017. (see pages: 22, 161)
- [124] N. T. BRONN, V. P. ADIGA, S. B. OLIVADESE, X. WU, J. M. CHOW, and D. P. PAPPAS. *High coherence plane breaking packaging for superconducting qubits*. Quantum Science and Technology, **3** (2), 2018. (see page: 22)
- [125] Y. TOMITA and K. M. SVORE. *Low-distance surface codes under realistic quantum noise*. Phys. Rev. A, **90**, 062320, 2014. (see pages: 23, 58, 75, and 82)
- [126] B. M. TARASINSKI, V. P. OSTROUKH, X. BONET-MONROIG, T. E. O'BRIEN, and B. VARBANOV. *quantumsim*. <https://gitlab.com/quantumsim>, 2016. (see page: 23)
- [127] J. Y. MUTUS, T. C. WHITE, R. BARENDS, Y. CHEN, Z. CHEN, B. CHIARO, A. DUNSWORTH, E. JEFFREY, J. KELLY, A. MEGRANT, C. NEILL, P. J. J. O'MALLEY, P. ROUSHAN, *et al.* *Strong environmental coupling in a Josephson parametric amplifier*. Appl. Phys. Lett., **104** (26), 263513, 2014. (see page: 23)
- [128] A. ROY, L. JIANG, A. D. STONE, and M. DEVORET. *Remote entanglement by coherent multiplication of concurrent quantum signals*. Phys. Rev. Lett., **115**, 150503, 2015.
- [129] C. MACKLIN, K. O'BRIEN, D. HOVER, M. E. SCHWARTZ, V. BOLKHOVSKY, X. ZHANG, W. D. OLIVER, and I. SIDDIQI. *A near-quantum-limited josephson traveling-wave parametric amplifier*. Science, **350** (6258), 307–310, 2015. (see page: 23)
- [130] J. HEINSOO, C. K. ANDERSEN, A. REMM, S. KRINNER, T. WALTER, Y. SALATHÉ, S. GASPARINETTI, J.-C. BESSE, A. POTOČNIK, A. WALLRAFF, and C. EICHLER. *Rapid high-fidelity multiplexed readout of superconducting qubits*. Phys. Rev. Appl., **10**, 034040, 2018. (see pages: 23, 52, 54, 110, 173, and 174)
- [131] C. C. BULTINK. *Protecting quantum entanglement by repetitive measurement*. PhD Dissertation, Delft University of Technology, 2020. (see page: 23)

- [132] B. M. TARASINSKI, T. E. O'BRIEN, and L. DICARLO. The QuSurf logical qubit and its simulated performance with current physical-qubit parameters. Year-2 Update of LogiQ Feasibility Study, not publicly available. (see pages: 23, 37)
- [133] B. VARBANOV, B. M. TARASINSKI, V. P. OSTROUKH, C. C. BULTINK, and L. DICARLO. The QuSurf logical qubit and its simulated performance with current physical-qubit parameters. M38 Update of LogiQ Feasibility Study, not publicly available. (see page: 23)
- [134] X. FU. *Quantum control architecture: Bridging the gap between quantum software and hardware*. PhD Dissertation, Delft University of Technology, 2018. (see pages: 25, 146)
- [135] J. HELSEN, F. BATTISTEL, and B. M. TERHAL. *Spectral quantum tomography*. arXiv:1904.00177, 2019. 1904.00177. (see page: 25)
- [136] DICARLOLAB and QUDEV. M30 reports: Crosstalk and other error sources in the QuSurf architecture. LogiQ Feasibility Study, not publicly available, 2018. (see page: 27)
- [137] M. A. ROL. *Metrics for fault tolerance, an experimentalist's approach*. In . Presented at the first "Assessing the Performance of Quantum Computers" (APQC) conference in Estes Park Colorado, 2019. (see page: 27)
- [138] J. WALLMAN, S. FLAMMIA, and I. HINCKS. *Quantum characterization, verification, and validation*. In *Oxford Research Encyclopedia of Physics*. Oxford University Press. (see page: 28)
- [139] R. BLUME-KOHOUT, J. K. GAMBLE, E. NIELSEN, K. RUDINGER, J. MIZRAHI, K. FORTIER, and P. MAUNZ. *Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography*. Nat. Commun., (14485), 2017. (see pages: 29, 34, and 58)
- [140] M. SAROVAR, T. PROCTOR, K. RUDINGER, K. YOUNG, E. NIELSEN, and R. BLUME-KOHOUT. *Detecting crosstalk errors in quantum information processors*. ArXiv:1908.09855, 2019. (see pages: 29, 35, 44, and 45)
- [141] E. MAGESAN, J. M. GAMBETTA, and J. EMERSON. *Scalable and robust randomized benchmarking of quantum processes*. Phys. Rev. Lett., **106**, 180504, 2011. (see pages: 29, 34, and 60)
- [142] E. MAGESAN, J. M. GAMBETTA, and J. EMERSON. *Characterizing quantum gates via randomized benchmarking*. Phys. Rev. A, **85**, 042311, 2012. (see pages: 29, 34, and 60)
- [143] R. BLUME-KOHOUT, J. K. GAMBLE, E. NIELSEN, J. MIZRAHI, J. D. STERK, and P. MAUNZ. *Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit*. arXiv:1310.4492, 2013. (see pages: 29, 34)
- [144] J. HELSEN. *Quantum information in the real world, diagnosing and correcting errors in practical quantum devices*. PhD Dissertation, Delft University of Technology, 2019. (see page: 29)
- [145] C. MÜLLER, J. LISENFELD, A. SHNIRMAN, and S. POLETTO. *Interacting two-level defects as sources of fluctuating high-frequency noise in superconducting circuits*. Phys. Rev. B, **92**, 035442, 2015. (see pages: 30, 42, and 62)
- [146] E. MAGESAN, J. M. GAMBETTA, B. R. JOHNSON, C. A. RYAN, J. M. CHOW, S. T. MERKEL, M. P. DA SILVA, G. A. KEEFE, M. B. ROTHWELL, T. A. OHKI, M. B. KETCHEN, and M. STEFFEN. *Efficient measurement of quantum gate error by interleaved randomized benchmarking*. Phys. Rev. Lett., **109**, 080505, 2012. (see pages: 30, 34, 116, 133, and 167)
- [147] J. HELSEN, X. XUE, L. M. K. VANDERSYPEN, and S. WEHNER. *A new class of efficient randomized benchmarking protocols*. npj Quantum Information, **5** (1), 71, 2019. (see page: 30)
- [148] X. XUE, T. F. WATSON, J. HELSEN, D. R. WARD, D. E. SAVAGE, M. G. LAGALLY, S. N. COPPERSMITH, M. A. ERIKSSON, S. WEHNER, and L. M. K. VANDERSYPEN. *Benchmarking gate fidelities in a si/sige two-qubit device*. arXiv:1811.04002, 2018. (see page: 30)

- [149] J. M. GAMBETTA, A. D. CÓRCOLES, S. T. MERKEL, B. R. JOHNSON, J. A. SMOLIN, J. M. CHOW, C. A. RYAN, C. RIGETTI, S. POLETTI, T. A. OHKI, M. B. KETCHEN, and M. STEFFEN. *Characterization of addressability by simultaneous randomized benchmarking*. Phys. Rev. Lett., **109**, 240504, 2012. (see pages: 30, 173)
- [150] J. J. WALLMAN, M. BARNHILL, and J. EMERSON. *Robust characterization of leakage errors*. New Journal of Physics, **18** (4), 043021, 2016. (see page: 30)
- [151] C. J. WOOD and J. M. GAMBETTA. *Quantification and characterization of leakage errors*. Phys. Rev. A, **97**, 032306, 2018. (see pages: 30, 36, 113, 116, 126, 131, 132, and 174)
- [152] T. PROCTOR, M. REVELLE, E. NIELSEN, K. RUDINGER, D. LOBSEY, P. MAUNZ, R. BLUME-KOHOUT, and K. YOUNG. *Detecting, tracking, and eliminating drift in quantum information processors*. ArXiv:1907.13608, 2019. (see pages: 30, 41, and 54)
- [153] E. NIELSEN, T. SCHOLTEN, K. RUDINGER, and J. GROSS. *pyGSTi: Version 0.9.1 beta*, 2016. (see pages: 30, 55, and 71)
- [154] E. NIELSEN, K. RUDINGER, T. PROCTOR, A. RUSSO, K. YOUNG, and R. BLUME-KOHOUT. *Probing quantum processor performance with pyGSTi*. Quantum Science and Technology, **5** (4), 044002, 2020. (see pages: 30, 55)
- [155] S. GUSTAVSSON, F. YAN, G. CATELANI, J. BYLANDER, A. KAMAL, J. BIRENBAUM, D. HOVER, D. ROSENBERG, G. SAMACH, A. P. SEARS, S. J. WEBER, J. L. YODER, J. CLARKE, *et al.* *Suppressing relaxation in superconducting qubits by quasiparticle pumping*. Science, **354** (6319), 1573–1577, 2016. <http://science.sciencemag.org/content/354/6319/1573.full.pdf>. (see page: 31)
- [156] A. KITAEV, A. SHEN, and M. VYALYI. *Classical and quantum computation*. American Mathematical Society, 2002. (see page: 32)
- [157] Y. R. SANDERS, J. J. WALLMAN, and B. C. SANDERS. *Bounding quantum gate error rate based on reported average fidelity*. New Journal of Physics, **18** (1), 012002, 2015. (see page: 33)
- [158] J. WALLMAN, C. GRANADE, R. HARPER, and S. T. FLAMMIA. *Estimating the coherence of noise*. New Journal of Physics, **17** (11), 113020, 2015. (see pages: 33, 34)
- [159] G. FENG, J. J. WALLMAN, B. BUONACORSI, F. H. CHO, D. K. PARK, T. XIN, D. LU, J. BAUGH, and R. LAFLAMME. *Estimating the coherence of noise in quantum control of a solid-state qubit*. Phys. Rev. Lett., **117** (26), 2016.
- [160] B. DIRKSE, J. HELSEN, and S. WEHNER. *Efficient unitarity randomized benchmarking of few-qubit clifford gates*. Phys. Rev. A, **99**, 012315, 2019. (see pages: 33, 34)
- [161] E. B. DAVIES and J. T. LEWIS. *An operational approach to quantum probability*. Communications in Mathematical Physics, **17** (3), 239–260, 1970. (see page: 35)
- [162] K. RUDINGER, T. PROCTOR, E. NIELSEN, G. RIBEILL, M. WARE, L. GOVIA, T. A. OHKI, K. YOUNG, and R. BLUME-KOHOUT. *Characterizing mid-circuit measurements with a new form of gate set tomography part 1: Theory*. APS March Meeting 2020, 2020. (see page: 35)
- [163] M. A. NIELSEN. *A simple formula for the average gate fidelity of a quantum dynamical operation*. Physics Letters A, **303** (4), 249–252, 2002. (see page: 35)
- [164] M. A. ROL, B. M. TARASINSKI, C. C. BULTINK, and L. DICARLO. QuSurf metrics report september 2018 for IARPA LogiQ, not publicly available, 2018. (see page: 37)
- [165] LogiQ monthly technical report, May 2018 (M26TR). Not publicly available. (see page: 37)
- [166] Z. CHEN, J. KELLY, C. QUINTANA, R. BARENDT, B. CAMPBELL, Y. CHEN, B. CHIARO, A. DUNSWORTH, A. G. FOWLER, E. LUCERO, E. JEFFREY, A. MEGRANT, J. MUTUS, *et al.* *Measuring and suppressing quantum state leakage in a superconducting qubit*. Phys. Rev. Lett., **116**, 020501, 2016. (see pages: 37, 60, and 61)

- [167] J. M. MARTINIS and M. R. GELLER. *Fast adiabatic qubit gates using only σ_z control*. Phys. Rev. A, **90**, 022307, 2014. (see pages: 37, 90, 110, 112, and 118)
- [168] P. MAGNARD, P. KURPIERS, B. ROYER, T. WALTER, J.-C. BESSE, S. GASPARINETTI, M. PECHAL, J. HEINSOO, S. STORZ, A. BLAIS, and A. WALLRAFF. *Fast and unconditional all-microwave reset of a superconducting qubit*. Phys. Rev. Lett., **121**, 060502, 2018. (see pages: 39, 40)
- [169] J. KELLY, R. BARENDS, B. CAMPBELL, Y. CHEN, Z. CHEN, B. CHIARO, A. DUNSWORTH, A. G. FOWLER, I.-C. HOI, E. JEFFREY, A. MEGRANT, J. MUTUS, C. NEILL, et al. *Optimal quantum control using randomized benchmarking*. Phys. Rev. Lett., **112**, 240504, 2014. (see pages: 42, 58, 60, and 64)
- [170] J. KELLY, R. BARENDS, A. G. FOWLER, A. MEGRANT, E. JEFFREY, T. C. WHITE, D. SANK, J. Y. MUTUS, B. CAMPBELL, Y. CHEN, Z. CHEN, B. CHIARO, A. DUNSWORTH, et al. *Scalable in situ qubit calibration during repetitive error detection*. Phys. Rev. A, **94**, 032321, 2016. (see pages: 42, 58, and 81)
- [171] D. RISTÈ, C. C. BULTINK, M. J. TIGGELMAN, R. N. SCHOUTEN, K. W. LEHNERT, and L. DICARLO. *Millisecond charge-parity fluctuations and induced decoherence in a superconducting transmon qubit*. Nat. Commun., **4**, 1913, 2013. (see page: 42)
- [172] F. YAN, S. GUSTAVSSON, A. KAMAL, J. BIRENBAUM, A. P. SEARS, D. HOVER, T. J. GUDMUNDSEN, D. ROSENBERG, G. SAMACH, S. WEBER, and ET AL. *The flux qubit revisited to enhance coherence and reproducibility*. Nat. Commun., **7** (1), 2016. (see pages: 74, 143)
- [173] J. J. BURNETT, A. BENGTSSON, M. SCIGLIUZZO, D. NIEPCE, M. KUDRA, P. DELSING, and J. BYLANDER. *Decoherence benchmarking of superconducting qubits*. npj Quantum Information, **5** (1), 54, 2019. (see pages: 42, 156)
- [174] S. T. SPITZ, B. M. TARASINSKI, C. W. J. BEENAKKER, and T. E. O'BRIEN. *Adaptive weight estimator for quantum error correction in a time-dependent environment*. Advanced Quantum Technologies, **1** (1), 1800012. (see page: 42)
- [175] D. I. SCHUSTER, A. WALLRAFF, A. BLAIS, L. FRUNZIO, R.-S. HUANG, J. MAJER, S. M. GIRVIN, and R. J. SCHOELKOPF. *ac Stark shift and dephasing of a superconducting qubit strongly coupled to a cavity field*. Phys. Rev. Lett., **94**, 123602, 2005. (see page: 50)
- [176] B. R. JOHNSON. *Controlling photons in superconducting electrical circuits*. PhD Dissertation, Yale University, 2011. (see pages: 51, 90, and 112)
- [177] O.-P. SAIRA, J. P. GROEN, J. CRAMER, M. MERETSKA, G. DE LANGE, and L. DICARLO. *Entanglement genesis by ancilla-based parity measurement in 2D circuit QED*. Phys. Rev. Lett., **112**, 070502, 2014. (see pages: 52, 74, and 173)
- [178] C. C. BULTINK, M. A. ROL, T. E. O'BRIEN, X. FU, B. C. S. DIKKEN, C. DICKEL, R. F. L. VERMEULEN, J. C. DE STERKE, A. BRUNO, R. N. SCHOUTEN, and L. DICARLO. *Active resonator reset in the nonlinear dispersive regime of circuit QED*. Phys. Rev. Appl., **6**, 034008, 2016. (see pages: 52, 58, 59, 62, 73, and 173)
- [179] P.-L. DALLAIRE-DEMERS and F. K. WILHELM. *Quantum gates and architecture for the quantum simulation of the fermi-hubbard model*. Phys. Rev. A, **94**, 062304, 2016. (see page: 60)
- [180] S. BOIXO, S. V. ISAKOV, V. N. SMELYANSKIY, R. BABBUSH, N. DING, Z. JIANG, M. J. BREMNER, J. M. MARTINIS, and H. NEVEN. *Characterizing quantum supremacy in near-term devices*. Nat. Phys., **14** (6), 595–600, 2018. (see pages: 58, 110)
- [181] C. HORSMAN, A. G. FOWLER, S. DEVITT, and R. V. METER. *Surface code quantum computing by lattice surgery*. New J. Phys., **14** (12), 123011, 2012. (see page: 58)
- [182] J. M. MARTINIS. *Qubit metrology for building a fault-tolerant quantum computer*. npj Quantum Inf., **1**, 15005, 2015. (see pages: 58, 110)

- [183] B. E. ANDERSON, H. SOSA-MARTINEZ, C. A. RIOFRÍO, I. H. DEUTSCH, and P. S. JESSEN. *Accurate and robust unitary transformations of a high-dimensional quantum system*. Phys. Rev. Lett., **114**, 240401, 2015. (see page: 58)
- [184] D. J. EGGER and F. K. WILHELM. *Adaptive hybrid optimal quantum control for imprecisely characterized systems*. Phys. Rev. Lett., **112**, 240503, 2014. (see page: 58)
- [185] D. T. MCCLURE, H. PAIK, L. S. BISHOP, M. STEFFEN, J. M. CHOW, and J. M. GAMBETTA. *Rapid driven reset of a qubit readout resonator*. Phys. Rev. Appl., **5**, 011001, 2016. (see pages: 74, 78)
- [186] P. CERFONTAINE, T. BOTZEM, S. S. HUMPOHL, D. SCHUH, D. BOUGEARD, and H. BLUHM. *Feedback-tuned noise-resilient gates for encoded spin qubits*. arXiv:1606.01897, 2016. (see page: 58)
- [187] M. H. DEVORET and R. J. SCHOELKOPF. *Superconducting circuits for quantum information: An outlook*. Science, **339** (6124), 1169–1174, 2013. (see page: 58)
- [188] C. WANG, C. AXLINE, Y. Y. GAO, T. BRECHT, Y. CHU, L. FRUNZIO, M. H. DEVORET, and R. J. SCHOELKOPF. *Surface participation and dielectric loss in superconducting qubits*. Appl. Phys. Lett., **107** (16), 162601, 2015. (see page: 58)
- [189] D. RISTÈ and L. DICARLO. *Digital feedback in superconducting quantum circuits*. In *Superconducting devices in quantum optics*. Springer, 2015. (see page: 58)
- [190] J. M. EPSTEIN, A. W. CROSS, E. MAGESAN, and J. M. GAMBETTA. *Investigating the limits of randomized benchmarking protocols*. Phys. Rev. A, **89**, 062321, 2014. (see pages: 60, 71, and 166)
- [191] J. A. NELDER and R. MEAD. *A simplex method for function minimization*. The Computer Journal, **7** (4), 308–313, 1965. (see pages: 63, 141)
- [192] E. Magesan, private communication. (see page: 64)
- [193] S. J. GLASER, U. BOSCAIN, T. CALARCO, C. P. KOCH, W. KÖCKENBERGER, R. KOSLOFF, I. KUPROV, B. LUY, S. SCHIRMER, T. SCHULTE-HERBRÜGGEN, D. SUGNY, and F. K. WILHELM. *Training schrödinger's cat: quantum optimal control*. The European Physical Journal D, **69** (12), 279, 2015. (see page: 65)
- [194] S. MACHNES, E. ASSÉMAT, D. TANNOR, and F. K. WILHELM. *Tunable, flexible, and efficient optimization of control pulses for practical qubits*. Phys. Rev. Lett., **120** (15), 2018. (see page: 65)
- [195] L. DICARLO, M. D. REED, L. SUN, B. R. JOHNSON, J. M. CHOW, J. M. GAMBETTA, L. FRUNZIO, S. M. GIRVIN, M. H. DEVORET, and R. J. SCHOELKOPF. *Preparation and measurement of three-qubit entanglement in a superconducting circuit*. Nature, **467**, 574, 2010. (see page: 65)
- [196] M. A. ROL, C. DICKEL, S. ASAAD, C. C. BULTINK, R. SAGASTIZABAL, N. K. L. LANGFORD, G. DE LANGE, B. C. S. DIKKEN, X. FU, S. R. DE JONG, and F. LUTHI. *PycQED*, 2016. (see pages: 65, 136, 138, and 150)
- [197] A. JOHNSON, G. UNGARETTI, *et al.* *QCoDeS*, 2016. (see pages: 65, 136, and 139)
- [198] M. D. REED, L. DICARLO, S. E. NIGG, L. SUN, L. FRUNZIO, S. M. GIRVIN, and R. J. SCHOELKOPF. *Realization of three-qubit quantum error correction with superconducting circuits*. Nature, **482**, 382, 2012. (see page: 74)
- [199] D. NIGG, M. MÜLLER, E. A. MARTINEZ, P. SCHINDLER, M. HENNRICH, T. MONZ, M. A. MARTINDELGADO, and R. BLATT. *Quantum computations on a topologically encoded qubit*. Science, **345** (6194), 302–305, 2014.
- [200] A. D. CÔRCOLES, E. MAGESAN, S. J. SRINIVASAN, A. W. CROSS, M. STEFFEN, J. M. GAMBETTA, and J. M. CHOW. *Demonstration of a quantum error detection code using a square lattice of four superconducting qubits*. Nat. Commun., **6**, 6979, 2015.

- [201] N. OFEK, A. PETRENKO, R. HEERES, P. REINHOLD, Z. LEGHTAS, B. VLASTAKIS, Y. LIU, L. FRUNZIO, S. M. GIRVIN, L. JIANG, M. MIRRAHIMI, M. H. DEVORET, and R. J. SCHOELKOPF. *Extending the lifetime of a quantum bit with error correction in superconducting circuits*. Nature, **536**, 441, 2016. (see pages: 74, 81)
- [202] A. BLAIS, R.-S. HUANG, A. WALLRAFF, S. M. GIRVIN, and R. J. SCHOELKOPF. *Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation*. Phys. Rev. A, **69**, 062320, 2004. (see pages: 74, 76)
- [203] K. D. PETERSSON, L. W. MCFAY, M. D. SCHROER, M. JUNG, J. M. TAYLOR, A. A. HOUCK, and J. R. PETTA. *Circuit quantum electrodynamics with a spin qubit*. Nature, **490** (7420), 380–3, 2012. (see page: 74)
- [204] T. W. LARSEN, K. D. PETERSSON, F. KUEMMETH, T. S. JESPERSEN, P. KROGSTROP, J. NYGÅRD, and C. M. MARCUS. *Semiconductor-nanowire-based superconducting qubit*. Phys. Rev. Lett., **115**, 127001, 2015. (see pages: 74, 90)
- [205] G. DE LANGE, B. VAN HECK, A. BRUNO, D. J. VAN WOERKOM, A. GERESDI, S. R. PLISSARD, E. P. A. M. BAKKERS, A. R. AKHMEROV, and L. DICARLO. *Realization of microwave quantum circuits using hybrid superconducting-semiconducting nanowire josephson elements*. Phys. Rev. Lett., **115**, 127002, 2015. (see page: 74)
- [206] J. P. GROEN, D. RISTÈ, L. TORNBORG, J. CRAMER, P. C. DE GROOT, T. PICOT, G. JOHANSSON, and L. DICARLO. *Partial-measurement backaction and nonclassical weak values in a superconducting circuit*. Phys. Rev. Lett., **111**, 090506. (see page: 74)
- [207] J. E. JOHNSON, C. MACKLIN, D. H. SLICHTER, R. VIJAY, E. B. WEINGARTEN, J. CLARKE, and I. SIDDIQI. *Heralded state preparation in a superconducting qubit*. Phys. Rev. Lett., **109**, 050506, 2012. (see page: 74)
- [208] D. RISTÈ, J. G. VAN LEEUWEN, H.-S. KU, K. W. LEHNERT, and L. DICARLO. *Initialization by measurement of a superconducting quantum bit circuit*. Phys. Rev. Lett., **109**, 050507, 2012. (see page: 74)
- [209] M. D. REED, B. R. JOHNSON, A. A. HOUCK, L. DICARLO, J. M. CHOW, D. I. SCHUSTER, L. FRUNZIO, and R. J. SCHOELKOPF. *Fast reset and suppressing spontaneous emission of a superconducting qubit*. Appl. Phys. Lett., **96** (20), 203110, 2010. (see page: 74)
- [210] E. JEFFREY, D. SANK, J. Y. MUTUS, T. C. WHITE, J. KELLY, R. BARENDS, Y. CHEN, Z. CHEN, B. CHIARO, A. DUNSWORTH, A. MEGRANT, P. J. J. O'MALLEY, C. NEILL, *et al.* *Fast accurate state measurement with superconducting qubits*. Phys. Rev. Lett., **112**, 190504, 2014. (see pages: 74, 78)
- [211] A. A. HOUCK, J. A. SCHREIER, B. R. JOHNSON, J. M. CHOW, J. KOCH, J. M. GAMBETTA, D. I. SCHUSTER, L. FRUNZIO, M. H. DEVORET, S. M. GIRVIN, and R. J. SCHOELKOPF. *Controlling the spontaneous emission of a superconducting transmon qubit*. Phys. Rev. Lett., **101**, 080502 (pages 4), 2008. (see page: 74)
- [212] J. GAMBETTA, A. BLAIS, M. BOISSONNEAULT, A. A. HOUCK, D. I. SCHUSTER, and S. M. GIRVIN. *Quantum trajectory approach to circuit QED: Quantum jumps and the Zeno effect*. Phys. Rev. A, **77** (1), 012112, 2008. (see page: 74)
- [213] A. P. SEARS, A. PETRENKO, G. CATELANI, L. SUN, H. PAIK, G. KIRCHMAIR, L. FRUNZIO, L. I. GLAZMAN, S. M. GIRVIN, and R. J. SCHOELKOPF. *Photon shot noise dephasing in the strong-dispersive limit of circuit QED*. Phys. Rev. B, **86**, 180504, 2012. (see page: 74)
- [214] X. Y. JIN, A. KAMAL, A. P. SEARS, T. GUDMUNDSEN, D. HOVER, J. MILOSHI, R. SLATTERY, F. YAN, J. YODER, T. P. ORLANDO, S. GUSTAVSSON, and W. D. OLIVER. *Thermal and residual excited-state population in a 3d transmon qubit*. Phys. Rev. Lett., **114**, 240501, 2015. (see page: 74)
- [215] M. J. D. POWELL. *An efficient method for finding the minimum of a function of several variables without calculating derivatives*. The Computer Journal, **7** (2), 155–162, 1964. (see pages: 74, 84)

- [216] J. M. CHOW, J. M. GAMBETTA, E. MAGESAN, D. W. ABRAHAM, A. W. CROSS, B. R. JOHNSON, N. A. MASLUK, C. A. RYAN, J. A. SMOLIN, S. J. SRINIVASAN, and M. STEFFEN. *Implementing a strand of a scalable fault-tolerant quantum computing fabric*. Nat. Commun., **5**, 4015, 2014. (see page: 74)
- [217] T. O'BRIEN, B.M.TARASINSKI, C.C.BULTINK, M.A.ROL, R.VERSLUIS, S.POLETTO, and L. DICARLO. IARPA LogiQ M6 report, QuSurf logical qubit and its simulated performance with current physical-qubit parameters, not publicly available, 2016. (see page: 75)
- [218] D. I. SCHUSTER, A. A. HOUCK, J. A. SCHREIER, A. WALLRAFF, J. M. GAMBETTA, A. BLAIS, L. FRUNZIO, J. MAJER, M. H. DEVORET, S. M. GIVIN, and R. J. SCHOELKOPF. *Resolving photon number states in a superconducting circuit*. Nature, **445**, 515, 2007. (see pages: 76, 85)
- [219] C. A. RYAN, B. R. JOHNSON, J. M. GAMBETTA, J. M. CHOW, M. P. DA SILVA, O. E. DIAL, and T. A. OHKI. *Tomography via correlation of noisy measurement records*. Phys. Rev. A, **91**, 022118, 2015. (see page: 76)
- [220] M. REED. *Entanglement and quantum error correction with superconducting qubits*. PhD Dissertation, Yale University, 2013. (see page: 76)
- [221] S. BOUTIN, C. K. ANDERSEN, J. VENKATRAMAN, and A. BLAIS. *Fast resonator reset in circuit qed using open quantum system optimal control*. APS March Meeting 2016. (see page: 78)
- [222] A. G. FOWLER, D. SANK, J. KELLY, R. BARENDS, and J. M. MARTINIS. *Scalable extraction of error models from the output of error detection circuits*. arXiv:1405.1454, 2014. (see page: 78)
- [223] A. FRISK KOCKUM, L. TORNBERG, and G. JOHANSSON. *Undoing measurement-induced dephasing in circuit QED*. Phys. Rev. A, **85**, 052318, 2012. (see pages: 80, 86)
- [224] P. VIRTANEN, R. GOMMERS, T. E. OLIPHANT, M. HABERLAND, T. REDDY, D. COURNAPEAU, E. BUROVSKI, P. PETERSON, W. WECKESSER, J. BRIGHT, S. J. VAN DER WALT, M. BRETT, J. WILSON, et al. *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*. Nature Methods, **17**, 261–272, 2020. (see pages: 84, 106)
- [225] M. A. ROL, L. CIORCIARO, F. K. MALINOWSKI, B. M. TARASINSKI, R. E. SAGASTIZABAL, C. C. BULTINK, Y. SALATHE, N. HAANDBAEK, J. SEDIVY, and L. DICARLO. *Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor*. Applied Physics Letters, **116** (5), 054001, 2020. (see pages: 89, 109, 114, 121, and 122)
- [226] S. FOLETTI, H. BLUHM, D. MAHALU, V. UMANSKY, and A. YACOBY. *Universal quantum control of two-electron spin quantum bits using dynamic nuclear polarization*. Nat. Phys., **5** (12), 903, 2009. (see page: 90)
- [227] J. MEDFORD, J. BEIL, J. TAYLOR, S. BARTLETT, A. DOHERTY, E. RASHBA, D. DIVINCENZO, H. LU, A. GOSSARD, and C. M. MARCUS. *Self-consistent measurement and state tomography of an exchange-only spin qubit*. Nat. Nanotechnol., **8** (9), 654, 2013.
- [228] A. LAUCHT, J. T. MUHONEN, F. A. MOHIYADDIN, R. KALRA, J. P. DEHOLLAIN, S. FREER, F. E. HUDSON, M. VELDHORST, R. RAHMAN, G. KLIMECK, K. M. ITOH, D. N. JAMIESON, J. C. MCCALLUM, et al. *Electrically controlling single-spin qubits in a continuous microwave field*. Sci. Adv., **1** (3), 2015.
- [229] M. VELDHORST, C. H. YANG, J. C. C. HWANG, W. HUANG, J. P. DEHOLLAIN, J. T. MUHONEN, S. SIMMONS, A. LAUCHT, F. E. HUDSON, K. M. ITOH, A. MORELLO, and A. S. DZURAK. *A two-qubit logic gate in silicon*. Nature, **526** (7573), 410–414, 2015. (see page: 90)
- [230] L. CASPARIS, T. W. LARSEN, M. S. OLSEN, F. KUEMMETH, P. KROGSTRUP, J. NYGÅRD, K. D. PETERSSON, and C. M. MARCUS. *Gatemon Benchmarking and Two-Qubit Operations*. Phys. Rev. Lett., **116** (15), 150505, 2016. (see page: 90)
- [231] B. FOXEN, J. MUTUS, E. LUCERO, E. JEFFREY, D. SANK, R. BARENDS, K. ARYA, B. BURKETT, Y. CHEN, Z. CHEN, B. CHIARO, A. DUNSWORTH, A. FOWLER, et al. *High speed flux sampling for tunable superconducting qubits with an embedded cryogenic transducer*. arXiv:1808.09612, 2018. (see page: 90)

- [232] J. KELLY. *Fault-tolerant superconducting qubits*. PhD Dissertation, University of California Santa Barbara, 2015. (see page: 90)
- [233] M. JERGER, A. KULIKOV, Z. VASSELLI, and A. FEDOROV. *In situ characterization of qubit control lines: A qubit as a vector network analyzer*. Phys. Rev. Lett., **123**, 150501, 2019. (see page: 90)
- [234] N. HANSEN. *Benchmarking the nelder-mead downhill simplex algorithm with many local restarts*. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, GECCO '09, pages 2403–2408. ACM, New York, NY, USA, 2009. (see page: 94)
- [235] R. WIGINGTON and N. NAHMAN. *Transient analysis of coaxial cables considering skin effect*. Proc. IRE, **45** (2), 166–174, 1957. (see pages: 98, 100)
- [236] J. M. MARTINIS, S. NAM, J. AUMENTADO, K. M. LANG, and C. URBINA. *Decoherence of a superconducting qubit due to bias noise*. Phys. Rev. B, **67** (9), 094510, 2003. (see page: 104)
- [237] J. BYLANDER, S. GUSTAVSSON, F. YAN, F. YOSHIHARA, K. HARRABI, G. FITCH, D. G. CORY, Y. NAKAMURA, J.-S. TSAI, and W. D. OLIVER. *Noise spectroscopy through dynamical decoupling with a superconducting flux qubit*. Nat. Phys., **7**, 565, 2011. (see page: 104)
- [238] S. M. ANTON, J. S. BIRENBAUM, S. R. O'KELLEY, V. BOLKHOVSKY, D. A. BRAJE, G. FITCH, M. NEELEY, G. C. HILTON, H.-M. CHO, K. D. IRWIN, F. C. WELLSTOOD, W. D. OLIVER, A. SHNIRMAN, *et al.* *Magnetic flux noise in dc squids: Temperature and geometry dependence*. Phys. Rev. Lett., **110**, 147002, 2013.
- [239] P. KUMAR, S. SENDELBACH, M. BECK, J. FREELAND, Z. WANG, H. WANG, C. Y. CLARE, R. WU, D. PAPPAS, and R. McDERMOTT. *Origin and reduction of 1/f magnetic flux noise in superconducting devices*. Phys. Rev. Appl., **6** (4), 041001, 2016.
- [240] C. M. QUINTANA, Y. CHEN, D. SANK, A. G. PETUKHOV, T. C. WHITE, D. KAFRI, B. CHIARO, A. MEGRANT, R. BARENDS, B. CAMPBELL, Z. CHEN, A. DUNSWORTH, A. G. FOWLER, *et al.* *Observation of classical-quantum crossover of 1/f flux noise and its paramagnetic temperature dependence*. Phys. Rev. Lett., **118**, 057702, 2017. (see page: 104)
- [241] F. LUTHI, T. STAVENGA, O. ENZING, A. BRUNO, C. DICKEL, N. LANGFORD, M. A. ROL, T. S. JESPERSEN, J. NYGÅRD, P. KROGSTROP, and L. DICARLO. *Evolution of nanowire transmon qubits and their coherence in a magnetic field*. Phys. Rev. Lett., **120** (10), 100502, 2018. (see pages: 104, 143)
- [242] J. S. OTTERBACH, R. MANENTI, N. ALIDOUST, A. BESTWICK, M. BLOCK, B. BLOOM, S. CALDWELL, N. DIDIER, E. SCHUYLER FRIED, S. HONG, P. KARALEKAS, C. B. OSBORN, A. PAPA-GEORGE, *et al.* *Unsupervised Machine Learning on a Hybrid Quantum Computer*. arXiv:1712.05771, 2017. (see page: 110)
- [243] W. KNIGHT. *IBM raises the bar with a 50-qubit quantum computer*. <https://www.technologyreview.com/s/609451/ibm-raises-the-bar-with-a-50-qubit-quantum-computer/>.
- [244] J. KELLY. *A preview of Bristlecone, Google's new quantum processor*. <https://ai.googleblog.com/2018/03/a-preview-of-bristlecone-googles-new.html>.
- [245] INTEL. *The future of quantum computing is counted in qubits*. <https://newsroom.intel.com/news/future-quantum-computing-counted-qubits/>. (see page: 110)
- [246] C. NEILL, P. ROUSHAN, K. KECHEDZHI, S. BOIXO, S. V. ISAKOV, V. SMELYANSKIY, A. MEGRANT, B. CHIARO, A. DUNSWORTH, K. ARYA, R. BARENDS, B. BURKETT, Y. CHEN, *et al.* *A blueprint for demonstrating quantum supremacy with superconducting qubits*. Science, **360** (6385), 195–199, 2018. (see page: 110)
- [247] S. BRAVYI, D. GOSSET, and R. KÖNIG. *Quantum advantage with shallow circuits*. Science, **362** (6412), 308–311, 2018. (see page: 110)

- [248] M. TAKITA, A. D. CÓRCOLES, E. MAGESAN, B. ABDO, M. BRINK, A. CROSS, J. M. CHOW, and J. M. GAMBETTA. *Demonstration of weight-four parity measurements in the surface code architecture*. Phys. Rev. Lett., **117**, 210 505, 2016. (see page: 110)
- [249] C. ANDERSEN, A. REMM, S. BALASIU, S. KRINNER, J. HEINSOO, J. BESSE, M. GABUREAC, A. WALLRAFF, and C. EICHLER. *Entanglement stabilization using parity detection and real-time feedback in superconducting circuits*. arXiv:1902.06946, 2019. (see page: 110)
- [250] Y. WANG, Y. LI, Z. YIN, and B. ZENG. *16-qubit ibm universal quantum computer can be fully entangled*. npj Quantum Inf., **4**, 46, 2018. (see page: 110)
- [251] W. D. OLIVER, Y. YU, J. C. LEE, K. K. BERGGREN, L. S. LEVITOV, and T. P. ORLANDO. *Mach-zehnder interferometry in a strongly driven superconducting qubit*. Science, **310** (5754), 1653–1657, 2005. (see pages: 114, 128)
- [252] M. SILLANPÄÄ, T. LEHTINEN, A. PAILA, Y. MAKHLIN, and P. HAKONEN. *Continuous-time monitoring of landau-zener interference in a cooper-pair box*. Phys. Rev. Lett., **96**, 187 002, 2006.
- [253] J. R. PETTA, H. LU, and A. C. GOSSARD. *A coherent beam splitter for electronic spin states*. Science, **327** (5966), 669–672, 2010.
- [254] A. CHATTERJEE, S. N. SHEVCHENKO, S. BARRAUD, R. M. OTXOA, F. NORI, J. J. L. MORTON, and M. F. GONZALEZ-ZALBA. *A silicon-based single-electron interferometer coupled to a fermionic sea*. Phys. Rev. B, **97**, 045 405, 2018.
- [255] J. ZHOU, P. HUANG, Q. ZHANG, Z. WANG, T. TAN, X. XU, F. SHI, X. RONG, S. ASHHAB, and J. DU. *Observation of time-domain rabi oscillations in the landau-zener regime with a single electronic spin*. Phys. Rev. Lett., **112**, 010 503, 2014. (see page: 114)
- [256] S. SHEVCHENKO, S. ASHHAB, and F. NORI. *Landau–zener–stückelberg interferometry*. Physics Reports, **492** (1), 1 – 30, 2010. (see pages: 115, 128)
- [257] J. JOHANSSON, P. NATION, and F. NORI. *QuTiP 2: A Python framework for the dynamics of open quantum systems*. Computer Physics Communications, **184**, 1234 – 1240, 2013. (see page: 124)
- [258] Confirmed by J.M. Martinis in personal communication. (see page: 135)
- [259] *Project jupyter*. (see page: 137)
- [260] A. BANDYOPADHYAY. *How open source approach is impacting science*, 2018. (see page: 137)
- [261] T. PETERS. *The Zen of Python*. PEP 20, 2004. (see page: 139)
- [262] M. FAASSEN. *What is pythonic?*, 2005. (see page: 139)
- [263] B. NIJHOLT, J. WESTON, J. HOOFWIJK, and A. AKHMEROV. *Adaptive: parallel active learning of mathematical functions*, 2019. (see page: 141)
- [264] N. KHAMMASSI, G. G. GUERRESCHI, I. ASHRAF, J. W. HOGABOAM, C. G. ALMUDEVER, and K. BERTELS. *cqasm v1.0: Towards a common quantum assembly language*. arXiv:1805.09607, 2018. (see page: 146)
- [265] A. W. CROSS, L. S. BISHOP, J. A. SMOLIN, and J. M. GAMBETTA. *Open quantum assembly language*. arXiv:1707.03429, 2017. (see page: 146)
- [266] H. ABRAHAM, I. Y. AKHALWAYA, G. ALEKSANDROWICZ, T. ALEXANDER, G. ALEXANDROWICS, E. ARBEL, A. ASFAW, C. AZAUSTRE, P. BARKOUTSOS, G. BARRON, L. BELLO, Y. BEN-HAIM, D. BEVENIUS, et al. *Qiskit: An open-source framework for quantum computing*, 2019. (see page: 146)
- [267] T. HÄNER, D. S. STEIGER, K. SVORE, and M. TROYER. *A software methodology for compiling quantum programs*. Quantum Science and Technology, **3** (2), 020 501, 2018. (see page: 146)

- [268] N. KHAMMASSI, I. ASHRAF, J. V. SOMEREN, R. NANE, A. M. KROL, M. A. ROL, L. LAO, K. BERTELS, and C. G. ALMUDEVER. *Openql: A portable quantum programming framework for quantum accelerators*. arXiv:2005.13283, 2020. (see page: 146)
- [269] R. S. SMITH, M. J. CURTIS, and W. J. ZENG. *A practical quantum instruction set architecture*. ArXiv:1608.03355, 2016. (see page: 146)
- [270] D. C. MCKAY, T. ALEXANDER, L. BELLO, M. J. BIERCUK, L. BISHOP, J. CHEN, J. M. CHOW, A. D. CÓRCOLES, D. EGGER, S. FILIPP, J. GOMEZ, M. HUSH, A. JAVADI-ABHARI, *et al.* *Qiskit backend specifications for openqasm and openpulse experiments*. ArXiv:1809.03452, 2018. (see page: 146)
- [271] C. A. RYAN, B. R. JOHNSON, D. RISTÈ, and G. J. RIBEILL. *QGL*, 2019. (see page: 146)
- [272] P. CERFONTAINE, S. HUMPOHL, L. PREDIGER, P. BETHKE, E. KAMMERLOHER, L. SCHREIBER, S. MEYER, B. RUMPE, and H. BLUHM. *B08.00003: qupulse: A quantum computing pulse parametrization and sequencing framework*. APS March Meeting 2020, 2020. (see page: 146)
- [273] M. A. ROL, L. GIORCIARO, and P. EENDEBAK. *AutoDepGraph*, 2017. (see pages: 147, 149)
- [274] F. K. WILHELM, S. KIRCHHOFF, S. MACHNES, N. WITTLER, and D. SUGNY. *An introduction into optimal control for quantum technologies*. ArXiv:2003.10132, 2020. (see page: 147)
- [275] J. KELLY, P. O'MALLEY, M. NEELEY, H. NEVEN, and J. M. MARTINIS. *Physical qubit calibration on a directed acyclic graph*. ArXiv:1803.03226, 2018. (see page: 147)
- [276] T. R. VAN ABSWOUE. Graph-based automation of transmon characterization. BEP report, Delft University of Technology, 2019. (see pages: 147, 149)
- [277] A. BRUNO, G. DE LANGE, S. ASAAD, K. L. VAN DER ENDEN, N. K. LANGFORD, and L. DICARLO. *Reducing intrinsic loss in superconducting resonators by surface treatment and deep etching of silicon substrates*. Appl. Phys. Lett., **106**, 182601, 2015. (see page: 150)
- [278] A. BILMES, A. MEGRANT, P. KLIMOV, G. WEISS, J. M. MARTINIS, A. V. USTINOV, and J. LISENFELD. *Resolving the positions of defects in superconducting quantum bits*. ArXiv:1911.08246, 2019. (see pages: 156, 157)
- [279] J. LISENFELD, A. BILMES, A. MEGRANT, R. BARENDS, J. KELLY, P. KLIMOV, G. WEISS, J. M. MARTINIS, and A. V. USTINOV. *Electric field spectroscopy of material defects in transmon qubits*. npj Quantum Information, **5** (1), 2019. (see pages: 156, 157)
- [280] B. FOXEN, C. NEILL, A. DUNSWORTH, P. ROUSHAN, B. CHIARO, A. MEGRANT, J. KELLY, Z. CHEN, K. SATZINGER, R. BARENDS, F. ARUTE, K. ARYA, R. BABBUSH, *et al.* *Demonstrating a continuous set of two-qubit gates for near-term quantum algorithms*. ArXiv:2001.08343, 2020. 2001.08343. (see page: 157)
- [281] E. T. CAMPBELL, B. M. TERHAL, and C. VUILLOT. *Roads towards fault-tolerant universal quantum computation*. Nature, **549** (7671), 172–179, 2017. (see page: 158)
- [282] C. GIDNEY and M. EKERÅ. *How to factor 2048 bit rsa integers in 8 hours using 20 million noisy qubits*. ArXiv:1905.09749, 2019. 1905.09749. (see page: 158)
- [283] L. BISHOP, S. BRAVYI, A. W. CROSS, J. M. GAMBETTA, and J. A. SMOLIN. *Quantum volume*. In . 2017. (see page: 158)
- [284] B. M. TERHAL and D. P. DIVINCENZO. *Adaptive quantum computation, constant depth quantum circuits and arthur-merlin games*. ArXiv:0205133, 2002. (see page: 159)
- [285] E. FARHI and A. W. HARROW. *Quantum supremacy through the quantum approximate optimization algorithm*, 2016. (see page: 159)
- [286] R. BLUME-KOHOOT and K. C. YOUNG. *A volumetric framework for quantum computer benchmarks*. ArXiv:1904.05546, 2019. (see page: 159)

- [287] T. PROCTOR, K. RUDINGER, K. YOUNG, E. NIELSEN, and R. BLUME-KOHOUT. *Measuring the capabilities of quantum computers*. ArXiv:2008.11294, 2020. (see page: 159)
- [288] C. DICKEL. *Scalability and modularity for transmon-based quantum processors*. PhD Dissertation, Delft University of Technology, 2018. (see page: 161)
- [289] S. KRINNER, S. STORZ, P. KURPIERS, P. MAGNARD, J. HEINSOO, R. KELLER, J. LÜTOLF, C. EICHLER, and A. WALLRAFF. *Engineering cryogenic setups for 100-qubit scale superconducting circuit systems*. EPJ Quantum Technology, **6** (1), 2019. (see page: 161)
- [290] S. BOSMAN, D. KUITENBROUWER, W. BOS, K. VERMEULEN, K. LINDEBERG, R. SORGEDRAGER, V. THINEY, and J. KAMMHubER. *V26.00012 : Scaling the input/output architecture of quantum processors to kqbit, and beyond, size in the nisq era*. APS March Meeting 2019. (see page: 161)
- [291] D. FRANKE, J. CLARKE, L. VANDERSYPEN, and M. VELDHORST. *Rent's rule and extensibility in quantum computing*. Microprocessors and Microsystems, **67**, 1 – 7, 2019. (see page: 162)


```

123
124
125 def randomized_benchmarking_sequence(
126     n_cl: int,
127     desired_net_cl: int = 0,
128     number_of_qubits: int = 1,
129     max_clifford_idx: int = 11520,
130     interleaving_cl: int = None,
131     seed: int=None):
132

```

"""
133 Generate a randomized benchmarking sequence for the one or two qubit clifford group.
134 """

