Protecting quantum entanglement by repetitive measurement

Bultink, C.C.

DOI
10.4233/uuid:25a762bf-f782-4ac2-b997-c91e95605c4f

Publication date
2020

Document Version
Final published version

Citation (APA)
Bultink, C. C. (2020). *Protecting quantum entanglement by repetitive measurement*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:25a762bf-f782-4ac2-b997-c91e95605c4f

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

PROTECTING QUANTUM ENTANGLEMENT BY REPETITIVE MEASUREMENT

PROTECTING QUANTUM ENTANGLEMENT BY REPETITIVE MEASUREMENT

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates,
to be defended publicly on
Wednesday 16th of September 2020 at 10:00 o'clock

by

Cornelis Christiaan BULTINK

Master of Science in Physics,
Technische Universiteit, Delft,
born in Bergen op Zoom, the Netherlands.

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	chairperson
Prof.dr. L. DiCarlo	Delft University of Technology, promotor
Prof.dr.ir. L.M.K. Vandersypen	Delft University of Technology, promotor
Independent members:	
Prof.dr. M. Möttönen	Aalto University, Finland
Prof.dr. B. Huard	ENS de Lyon, France
Prof.dr. D.P. DiVincenzo	RWTH Aachen & Forsch. Zentr. Jülich, Germany
Prof.dr. S.D.C. Wehner	Delft University of Technology
Prof.dr. G.A. Steele	Delft University of Technology

Printed by: Gildeprint, Enschede – www.gildeprint.nl

Cover: Artistic impression of the devices and pulsed operations thereon discussed in this work.
Design by Wouter Geense,
Device design by René Vollmer, Marc Beekman and Nadia Haider.

Casimir PhD Series, Delft-Leiden 2020-24

ISBN 978-90-8593-449-3

An electronic version of this dissertation is available at `http://repository.tudelft.nl/`.

Begin anywhere.

CONTENTS

PROLOGUE: "WARNING: 5 FAULTS FOUND!"

It was around 1:00 AM on a cold winter night in 2019 when I stepped out of a bar in the historic center of Delft. Time to go home, after celebrating Suzanne van Dam's successful PhD defense. Evenings like these remind me why science is such a great endeavour. People might be working long hours to reach only tiny steps. But there is a team of people around you that understands the struggle. And, when those tiny steps lead to a big one, or occasionally even a breakthrough, celebration comes with all sorts of weird traditions. In short, it was an evening with a warm QuTech family feeling. For me personally it was also a nice change of scenery. Over the months prior, I had been working intensely on the final experiment of this thesis. In particular this week, I was finetuning day-in day-out to hunt for data sets that were fit for publication. Although the efforts of many had been finally coming together for quite some time: from design, to fabrication, to the integration of the setup; it is this final stage of an experiment that can make or break years of struggle for the whole team.

In complete ignorance of what was about to happen, I drew my phone to see if I had missed anything. Upon lighting the screen, my full attention was drawn: "9 unread messages from La Maserati", the name of the dilution refrigerator I was using. This is the gigantic and complex system that cools down the quantum chip close to absolute zero and thereby enables the control of quantum phenomena. For a second, I comforted myself, as likely the messages were caused by a failure in the pressure monitoring system. This was a routinely experienced false alarm from La Maserati that leads to a burst of automatically-generated warnings. It mostly occurred at the least convenient time to manually reset the monitoring. Opening one of the messages, however showed that the situation was more severe.

$$
\begin{aligned}
&\textit{WARNING: 5 FAULTS FOUND!}\\
&\textit{Injection pressure is unsafe!}\\
&\textit{OVC pressure is unsafe!}\\
&\textit{Still pressure is unsafely high!}\\
&\textit{Still is running hot!}\\
&\textit{3K Plate is running hot!}\\
&\textit{First responder} = \textit{Niels B}\\
&T\_3K = 19095\ mK.\\
&T\_Still = 1170\ mK.\\
&T\_MClo = 327.6\ mK.\\
&P\_5 = 1601\ mBar.\\
&P\_IVC = 0.000564\ mBar.\\
&P\_OVC = 0.006134\ mBar.
\end{aligned}
$$

Fri, Feb 01, 2019 9:37:00 PM, La Maserati

The untrained eye will note that the five warnings probably mean things are not great. The trained eye will note that each individual pressure and temperature reading is completely off. i.e. far beyond the point of getting things back on track. Reading this message thereby abruptly marked the end of the experimental work of this thesis, the part I liked most.

It was quite a shock, but no panic. Fortunately we were already sitting on high quality data. Soon, we needed to go into writing mode to push for publication anyway. It was however required to pay a direct visit to the lab to bring the system to a safe mode. It was a welcome surprise for my two travel companions, Anne-Marije and Jules, who were clearly craving a nightly adventure in the lab. After going through the standard procedures: switching off the control electronics and recovering the precious helium-3 from the internals, it was my goal to leave the fridge on its way back down towards three Kelvin. This intermediate temperature would allow us going back to base temperature (0.02 Kelvin) during the next day. However, after a few attempts, I had to conclude that something blocked the fridge from lowering its temperature again. I even tried to call my promotor Leo DiCarlo, but the calls were unanswered. What would one expect at this time? It was far passed 3:00 AM and although the pressure in the fridge had lowered, the pressure on my relationship was clearly rising; time to drive home to Rotterdam. Surely, the situation would be clearer in the morning, after some rest.

Then, when stepping out of the faculty, the night took another unexpected turn. Looking from the parking space outside, we noticed a dark figure was walking down halfway the staircase. We asked ourselves: “Who in earth would still be in the building at this point?” It was not easy to tell as the mysterious figure was obscured by the many QuTech logos that covered the window (aka ‘het douchegordijn’). Anxiously, we awaited the person to arrive. And then, the one we had all been waiting for entered the stage. With the door wide open and a loud “TADA, here am I!”, prof. dr. Leonardo DiCarlo claimed his spot. A strangely exultant appearance for this time of the day. It turned out he was so relieved to see us, because we could save him from a sixteen-kilometer bike ride. This was his daily commute to Rotterdam, for which he had been charging with a power nap in his office, after Suzannes’ party.

After learning the details of our nightly adventure, we went back in together to take another look at the fridge. Rather quickly (with his experience), we narrowed down the reason why the fridge would not return to three Kelvin. It was the same reason that made the system unstable in the first place: a small but sudden rise in pressure in the outer vacuum chamber. This had created a thermal bridge between the hot room and the cold inner workings. The root cause for this sudden burst of gas remains a mystery until today. But what I did learn, is that I will never forget this night. Everything came together that makes science so engaging: the pleasure of finding things out, the excitement of running an experiment that pushes the edge of what is possible, but mostly, the family feeling that binds together those who share this passion.

SUMMARY

Information processing based on the laws of quantum mechanics promises to be a revolutionary new avenue in information technology. This emerging field of quantum information processing (QIP) is however challenged by the fragile nature of the quantum bits (qubits) in which quantum information is stored and processed. An error in even a single qubit makes the quantum processor go off-track, corrupting the calculation as a whole. Therefore, the chance for an erroneous outcome increases with the number of qubits in the processor. Large-scale QIP thus hinges on the ability to correct for these errors. Classical information processing often uses error correction algorithms to identify errors by checking whether information is consistent in multiple copies. This strategy is unfortunately not applicable to QIP as quantum states cannot be copied. Moreover, direct measurements on qubits collapse their quantum states, reducing them to classical information. Fortunately, the theory of quantum error correction (QEC) overcomes these complications by encoding quantum information in entangled states of many qubits and performing parity measurements to identify errors in the system without destroying the encoded information. Implementing these codes is challenging as it requires many qubits and quick interleaving of operations and measurements. Moreover, to not introduce more errors in the system than QEC can solve for, these operations and measurements need to be of sufficient fidelity and speed.

Circuit quantum electrodynamics (cQED) is one of the most successful platforms for implementing QEC. Most notably, QEC codes with five to nine quantum bits have shown the preservation of the classical degree of freedom of the encoded information. However, QEC implementations prior to the start of this thesis, have not succeeded in preserving quantum states. This was mainly caused by the long time required for qubit readout compared to the qubit coherence time, the time during which they can hold their information. Ratios of 0.2-0.5 were achieved. In this thesis, we implement several improvements to accelerate qubit readout, avoid its unwanted back-actions on other qubits and make use of qubits with an improved coherence time. These steps improve the measurement-time to coherence-time ratio by a factor ten to 0.025-0.05. We ultimately demonstrate the benefits of these improvements by preserving an entangled state during repeated QEC over tens of error correction cycles.

In the first chapters of this thesis, we improve several aspects of repetitive readout in QEC. Chapter 1 introduces quantum computing and QEC. It provides an overview of the status of experimental work and motivates the use of cQED for this thesis. Chapter 2 provides an introduction to superconducting qubits and summarizes cQED. In this platform, superconducting qubits are coupled to a superconducting resonator which mediates between the qubit and the environment. By detuning the resonance frequency of the resonator with respect to the transition frequency of the qubit, energy exchange between the qubit and its environment is minimized; suppressing a primary source of error. Via the resulting so-called dispersive inter-

action, the resonator's resonance frequency is slightly dependent on the qubit's state. This mechanism provides an indirect method for the readout of the qubit state. The qubit state is measured by analysing the resonator's response to a pulse near its resonance frequency. Beyond this standard approach, an additional resonator (Purcell filter) can be added to increase the qubit's isolation. Different configurations are compared in this chapter and our final choice is motivated. Chapter 3 explores the use of active depletion of measurement photons after qubit measurement has been performed. We demonstrate that this technique reduces the time required for the overal readout process by a factor ~ 3. The benefit for QEC is explored by emulating repeated quantum parity checks using one qubit. The reduction in error rate was found to be between a factor ~ 2 and ~ 75 depending on the emulation. These results strongly advocate the use of active photon depletion in QEC. In Chapter 4, we use sequences of interleaved measurements and single-qubit operations to assess and optimize the operation performance. Repetitive readout allows the calibration of three main parameters. By using numerical optimization techniques, the gate is fine tuned in less than a minute, reliably achieving a gate fidelity of 0.999. This fidelity lies well beyond the threshold for QEC. In Chapter 5, we focus on the measurement efficiency and propose a method to assess and improve the efficiency of elements in the qubit readout chain. A recent breakthrough in qubit readout is the use of special superconducting amplifiers that operate with near perfection. The key performance metric for these amplifiers and the following readout chain is the quantum efficiency, which is the fraction of readout photons that effectively reaches the observer. We show that the qubit itself can be used as the ideal sensor to determine the quantum efficiency. The efficiency measurements are consistent for arbitrary readout conditions, even for measurements with the strangest dynamics. This is a key tool for the tune-up of amplifiers for optimal readout and to distinguish sources of imperfection.

In Chapter 6, we ultimately move to a multi-qubit paradigm to implement and test the improvements made in previous chapters by implementing a QEC code that stabilizes an entangled state. The improved readout topology with a dedicated Purcell filter per qubit allows fast measurement with negligible back-action on the untargeted qubits. This allows us to create entanglement by parity measurement with a high fidelity of $\sim 95\%$. Repeated parity measurements protect this entanglement from arbitrary qubit errors during > 25 parity measurements. Furthermore, we demonstrate that the same QEC measurements can be used to detect leakage out of the qubit subspace to higher-energy states. This last form of error is natively not addressed by QEC but is detrimental to quantum computing in most platforms. We demonstrate that, by applying a separate error analysis (using a hidden Markov model), we can infer this leakage while tracking standard qubit errors. This opens a new route to fault-tolerant quantum computation in the presence of qubit errors and leakage.

Chapter 7 finally discusses the implications of this work for quantum computing. The experimental results show that an architecture has been built up with all necessary components for the preservation of logical information with larger numbers of qubits. We underline this conclusion by projecting in a detailed simulation how a seventeen-qubit QEC experiment would perform, building on the results of this thesis. Its experimental realization will be the next milestone towards fault-tolerant quantum computing.

SAMENVATTING

Informatieverwerking die gebaseerd is op de wetten van de kwantummechanica belooft een revolutionair nieuw hoofdstuk te worden in de informatietechnologie. Het snel ontwikkelende veld van kwantuminformatietechnologie heeft echter te kampen met de fragiliteit van de kwantumbits waarin de kwantuminformatie wordt opgeslagen en verwerkt. Een fout in een enkel kwantumbit kan de gehele kwantumprocessor laten ontsporen, waardoor de berekening in zijn geheel onbetrouwbaar wordt. Dit heeft tot gevolg dat de kans op een foute uitkomst toeneemt met het groeien van het aantal kwantumbits. Voor het uitvoeren van kwantumberekeningen op grote schaal is het kunnen corrigeren van deze fouten daarom cruciaal. Klassieke informatietechnologie maakt vaak gebruik van foutcorrectie om fouten te detecteren door kopiën van informatie te testen op consistentie. Een dergelijke aanpak is helaas niet toepasbaar op kwantumberekeningen aangezien kwantumtoestanden niet gekopieerd kunnen worden en directe metingen van kwantumbits deze reduceren tot klassieke bits. Gelukkig biedt de theorie van kwantumfoutcorrectie een oplossing. Door de kwantuminformatie van een enkele kwantumbit te encoderen in verstrengelde toestanden van een aantal kwantumbits, wordt de informatie verspreid zonder deze te hoeven kopiëren. Daarbij worden pariteitmetingen gebruikt om fouten te traceren zonder dat de onderliggende kwantumtoestanden vernietigd worden. Het implementeren van deze algoritmes is echter uitdagend omdat er veel kwantumbits voor nodig zijn en omdat kwantumoperaties en metingen in hoog tempo afgewisseld moeten worden. Daarnaast, om te voorkomen dat dit meer fouten veroorzaakt dan dat de foutcorrectie kan oplossen, dienen deze operaties en metingen voldoende nauwkeurig en snel te zijn.

Circuit kwantumelectrodynamica (cQED) is een van de vooraanstaande systemen waarop kwantumfoutcorrectie wordt geïmplementeerd. Het meest opmerkelijk zijn de implementaties waarin met vijf tot negen kwantumbits de klassieke vrijheidsgraad van de geëncodeerde informatie beschermd bleef. Desalniettemin, was het voor aanvang van dit proefschrift nog niet gelukt om kwantumtoestanden succesvol te beschermen. Dit lag hoofdzakelijk aan de lang benodigde tijd om kwantumbits uit te lezen, in verhouding tot hun coherentietijd, de tijd waarin kwantumbits hun informatie bewaren. De ratio's lagen tussen 0.2 en 0.5.

In dit proefschrift wordt een aantal verbeteringen uitgevoerd in het uitlezen van kwantumbits die leiden tot een snellere uitlezing en worden verstoringen (veroorzaakt door de meting) van andere kwantumbits voorkomen. Ook hebben de kwantumbits een langere coherentietijd dan voorheen beschikbaar was. Dit leidt tot een verbetering van de ratio uitleestijd-tot-coherentietijd tot 0.025-0.05. Ultiem komen de voordelen hiervan tot uiting door een verstrengelde toestand te beschermen gedurende tientallen herhaalde cycli van kwantumfoutcorrectie.

In de eerste hoofdstukken van dit proefschrift, worden verschillende aspecten van het herhaaldelijk uitlezen van kwantumbits verbeterd. Hoofdstuk 1 is een introductie tot de kwantumcomputer en kwantumfoutcorrectie. Het geeft een overzicht van de voorgaande experimentele resultaten en motiveert het gebruik van cQED in dit proefschrift. In Hoofdstuk 2 worden supergeleidende kwantumbits beschreven. Daarnaast biedt het een inleiding in het cQED platform waarin supergeleidende kwantumbits gekoppeld worden aan een resonator. De resonator vormt een medium tussen het kwantumbit en de buitenwereld. Door de resonantiefrequentie van de resonator verschillend te maken van die van het kwantumbit, wordt de uitwisseling van energie tussen het kwantumbit en de buitenwereld geminimaliseerd; een primaire bron van fouten wordt zo onderdrukt. De overblijvende zogenoemde dispersieve koppeling tussen kwantumbit en resonator, zorgt dat de resonantiefrequentie van de resonator licht afhankelijk is van de toestand van het kwantumbit. Dit wordt gebruikt als het uitleesmechanisme. De toestand van het kwantumbit wordt uitgelezen door de resonator met een microgolfpuls te injecteren en diens reflectie te analyseren. Buiten deze standaardaanpak, kan een tweede resonator (Purcell filter) toegevoegd worden om het kwantumbit verder te isoleren van de buitenwereld. Verschillende topologieën van resonatoren worden vergeleken en onze uiteindelijke keuze wordt onderbouwd. Hoofdstuk 3 verkent het gebruik van actieve depletie van uitleesfotonen nadat kwantumbits worden uitgelezen. We laten zien dat deze techniek de totale uitleestijd van het kwantumbit reduceert met een factor ~ 3. Het voordeel van fotondepletie wordt verder onderzocht door een herhaaldelijke kwantumpariteitmeting te emuleren met één kwantumbit. De ondervonden foutreductie met fotondepletie ligt tussen een factor ~ 2 en ~ 75 afhankelijk van de emulatie. Dit resulteert in een sterke aanbeveling voor het gebruik van fotondepletie bij foutcorrectie. In Hoofdstuk 4 gebruiken we snel afwisselende uitlezingen en één-kwantumbitoperaties om de operatienauwkeurigheid te beoordelen en te optimaliseren. Herhaaldelijk uitlezen stelt ons in staat de belangrijkste drie parameters te optimaliseren in minder dan één minuut tot een nauwkeurigheid van 0.999. In Hoofdstuk 5 onderzoeken we de efficientie van de uitleesketen en stellen een nieuwe methode voor om deze te meten en te verbeteren. Een van de grote doorbraken in het uitlezen van kwantumbits, is het gebruik van speciale supergeleidende versterkers die dicht perfectie naderen. De belangrijkste maat voor perfectie is de kwantumefficientie, die de fractie van uitleesfotonen weergeeft die de observator op een nuttige manier bereiken. We laten zien dat het kwantumbit zelf als ideale sensor gebruikt kan worden en dat traditionele methodes waarin additionele apparatuur gebruikt wordt overbodig zijn. De nieuwe meetmethode is consistent te gebruiken met arbitraire uitleescondities, zelfs wanneer een meetpuls gekozen wordt met arbitraire dynamiek. De methode biedt een belangrijk instrument om de supergeleidende versterkers te kalibreren en om verschillende bronnen van imperfectie in de uitleesketen te identificeren.

In Hoofdstuk 6 gaan we over tot het gebruik van meer kwantumbits om de verbeteringen in voorgaande hoofdstukken samen te laten komen in een implementatie van foutcorrectie. De vernieuwde uitleestopologie en methodes zorgen ervoor dat we individuele kwantumbits snel kunnen uitlezen met minimale verstoringen van de overige kwantumbits. Dit stelt ons in staat kwantumbits te verstrengelen door middel van een pariteitmeting met een betrouw-

baarheid van 0.95. Herhaalde pariteitmetingen helpen ons de verstrengeling te beschermen tegen arbitraire fouten in de onderliggende kwantumbits. Daarnaast introduceren we in dit experiment het gebruik van dezelfde foutcorrectiemetingen om de lekkage van kwantumbits naar hogere energietoestanden te detecteren. Deze laatste vorm van fouten worden niet standaard aangepakt in kwantumfoutcorrectie, maar zijn desalniettemin desastreus voor kwantumcomputers. We laten zien dat door middel van een losstaande foutanalyse met een hidden Markov model de lekkage kunnen oppikken tijdens het standaard foutcorrectieprotocol. Dit opent een nieuwe weg in fouttolerantie in de aanwezigheid van zowel fouten in het kwantumbit als lekkage naar andere energietoestanden.

Hoofdstuk 7 beschouwt de implicaties van dit proefschrift voor kwantumcomputers. De resultaten laten zien dat een architectuur met alle benodigde componenten is opgebouwd voor foutcorrectie met meer kwantumbits. Deze conclusie wordt onderbouwd met gedetailleerde simulaties van een foutcorrectie-algoritme met zeventien kwantumbits. De experimentele verwezenlijking hiervan zal de volgende mijlpaal zijn op weg naar fouttolerante kwantumcomputers.

1 INTRODUCTION

As far as the laws of mathematics refer to reality, they are not certain; and as far as they are certain, they do not refer to reality

Albert Einstein

Shut up and calculate

Richard Feynman

1.1 Quantum mechanics: from an oddity to a potential powerful resource

At the brink of quantum mechanics (QM) in the early 20th century, scientists were utterly puzzled whether it could be a valid description of the physical world. The central phenomena (i) superposition (a particle's ability to be in multiple states at the same time), (ii) the idea that the act of a measurement probabilistically forces particles into 'classical' states and (iii) entanglement (a shared state amongst particles leading to instantaneous back-action at arbitrary distance) were unacceptable to many as these phenomena do not align with our daily observations of the macroscopic world. Although many experimental tests confirm quantum mechanics to be an accurate description of reality, its interpretation remains a topic of debate to date. Nowadays, however, the main focus of research in quantum mechanics has shifted to unlocking its potential as a technical and computational resource. In 1982, R. P. Feynman described that the simulation of quantum mechanical behavior (like for instance chemical reactions) is inefficient on a classical computer and that simulating them on a quantum system is a more efficient approach [1]; as they both behave quantum mechanically. Hence, the first concept for a quantum computer was born. A few years later, further ideas sprung up to use quantum computers for general information processing tasks (which do not have a quantum mechanical description). Examples include Shor's algorithm [2] to factorize large numbers and Grover's algorithm [3] to search large, unstructured databases of information. More recently, the list of algorithms is growing with ways of solving sets of linear equations [4]. What makes these 'quantum' algorithms so interesting over 'classical' approaches, is that the required resources (time and hardware) grow exponentially on a classical computer whereas

they scale polynomially for the quantum computer. In short, problems that take billions of years with all current supercomputers combined, become practically solvable with a quantum computer.

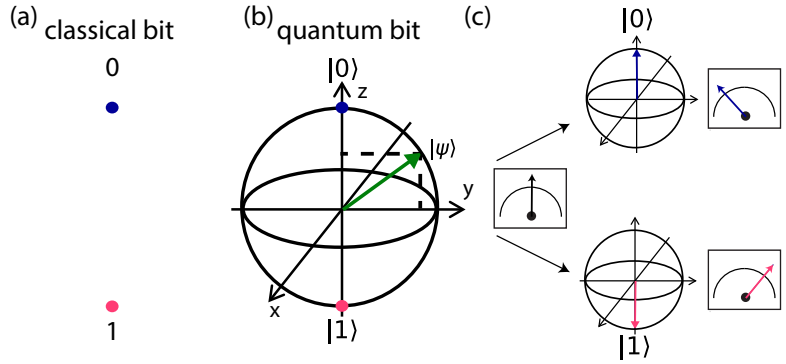


Figure 1.1: Graphical representations of a classical bit, a quantum bit and state collapse under measurement **(a)** A classical bit is limited to two values, '0' and '1'. **(b)** Qubit representation on the Bloch sphere. The unit vector on the Bloch sphere represents the qubit state $|\psi\rangle$ which can have any position on the sphere. The north pole represents $|0\rangle$, the south pole $|1\rangle$ and all other positions on the sphere represent superposition states. **(c)** Projective state measurement. When a projective measurement is performed along $\hat{z}$, the probability to measure $|0\rangle$ is $P_{|0\rangle} = |\alpha|^2 = (z+1)/2$ and the probability to measure $|1\rangle$ is $P_{|1\rangle} = |\beta|^2 = 1-(z+1)/2$. After measurement, the qubit is left in the measured state.

1.2 The basics of a quantum computer and criteria for a good one

To understand the difference between classical and quantum computers, we zoom into the most fundamental level of the information that is processed. In a classical computer, the fundamental building blocks for storing and processing information are called bits. Bits are limited to have only two possible values, 0 or 1 [Fig. 1.1(a)]. The fundamental building blocks of quantum computers are quantum bits, or qubits. Just as their classical counterparts, qubits are described by two basis states, $|0\rangle$ and $|1\rangle$. However, the crucial difference is rooted in three quantum mechanical phenomena as already briefly described above: superposition, the unavoidable perturbative effect of measurement, and entanglement.

i Superposition denotes the ability of quantum systems to be in more than one of the 'classical' states at the same time. For a qubit, this means that it is not limited to be just $|0\rangle$ or $|1\rangle$, but it can also be $|0\rangle$ and $|1\rangle$ at the same time; hence, a superposition state. Mathematically, the state $|\psi\rangle$ of a single qubit is written down as a linear combination of the two basis states

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,,$$

with complex numbers α and β which satisfy the normalization condition $|\alpha|^2+|\beta|^2=1$. All possible qubit states can be visualized on the surface of a sphere: the Bloch sphere [Fig. 1.1(b)] [5]. The two poles are the two basis states: north, $|0\rangle$ and south, $|1\rangle$.

ii A remarkable property of a quantum state is that it is impossible to examine it without affecting it, except for specific states, known as the eigenstates. For instance, when a qubit is measured along the $\hat{z}$ axis, the outcome will be one of the measurement's eigenstates: $|0\rangle$ and $|1\rangle$. The coefficients α and β determine the probability of finding the qubit in $|0\rangle$ with $|\alpha|^2$ and $|1\rangle$ with $|\beta|^2$ [Fig. 1.1(c)], which is known as the Born rule [6]. The widely accepted Copenhagen interpretation tells us that the act of measurement itself forces the qubit into one of the two basis states. This feature is known as the collapse of the wave function, which more generally describes the transition between a continuum of probable outcomes, to a discrete eigenstate of the measurement. After a measurement is performed, the qubit is left in the eigenstate. Subsequent measurements (following the Born rule) then lead to identical outcomes.

iii Entanglement is the possibility of multiple elements to share a superposition state. An example of a two-qubit entangled state is the Bell state

$$|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}.$$

Just as for the single-qubit case, a two-qubit measurement collapses the state to an eigenstate. For a measurement along at least one of the qubit's $\hat{z}$, these are: $|00\rangle$ or $|11\rangle$, with equal probability, $P_{|00\rangle} = P_{|11\rangle} = 0.5$. The weirdness of an entangled qubit pair is revealed by measuring an individual qubit and correlating its outcome to a measurement of the other. The measurement of the first qubit directly affects the second qubit and this interaction is independent of the distance between the two qubits. This instantaneous back-action, which provenly happens faster than the speed of light [7] was the crucial unacceptable feature of QM for scientists like A. Einstein [8].

As mentioned, most current research is not directly aimed to make sense of these phenomena, but like this thesis, focuses on their implementation for the realization of quantum technology. This realization starts with making qubits with the highest possible quality. Just as classical bits can be implemented in many physical systems: holes in a punch cards, the reflectivity of a cd track, or currents through a transistor, we can implement qubits in many different physical systems. Qubit implementations that are under study include: nuclear spins, electron spins, ions, optical systems and electrical circuits [9]. All of these systems have their own advantages and disadvantages with regard to the processing of quantum information. To objectively examine and compare these systems, D. DiVincenzo proposed a list of criteria [10]. The criteria: (i) A physical system on which qubits are well-defined two-level systems and which can scale up to many qubits. (ii) The ability to initialize the qubits in a well-defined

state. (iii) Qubits should contain information for sufficient time to perform subsequent computational steps. (iv) Qubits should be both controllable individually with single-qubit control and jointly to create entanglement. (v) Qubits should be measurable individually and selectively.

1.3 Overcoming the fragility of quantum information processing

The most fundamental road block for quantum information processing seems to be the fragile nature of quantum information. First, this fragility is reflected in the limited ability of qubits to hold their information due to spurious interactions with their environment. The loss of information in this way is known as decoherence. Second, qubits are essentially analog devices. Single-qubit operations can, for instance, be represented on the Bloch sphere as rotations around an arbitrary axis. Inevitable miscalibrations (even the tiniest) in these qubit operations, eventually make a quantum computer go off-track, leading to an incorrect outcome. Naturally, the effect of both of these sources of error increases with the the number of qubits and computational steps involved in a calculation. There are two important pathways to address this.

The first is to simply improve the qubit coherence time (the typical time for a decoherence event to have happened) and the quality of single- and two-qubit operations. Following this path, the first quantum computers are being built that can outperform classical computers on some tasks, which are specifically designed to showcase the advantage of quantum computers over classical computers. For this mile stone, known as quantum supremacy or quantum advantage, approximately 50 qubits are required to run for 40 computational steps with an error per operation of ~ 0.003 [11]. At the time of writing, the first experimental implementation has reached this point [12]. The following era, during which the first useful calculations will be explored on noisy and smallish devices, is hence referred to as the noisy intermediate-scale quantum-computing (NISQ) era [13]. This pathway may however not be followed indefinitely as the chance for calculation errors keeps increasing with the system size and the number of computational steps.

The second, more rigorous path to battle fragility, is to empower quantum computers with the ability to correct for errors during the processing of information. To this end, several Quantum Error Correction (QEC) schemes [14–18] were invented in the mid 90's. These schemes make quantum information processing tolerant to errors, 'fault tolerant', by encoding the information of a single 'logical' qubit on a larger number of 'physical' qubits. Because of the added redundancy, individual errors in a part of the system, can be measured and corrected without compromising the encoded information on the logical qubit. Crucially, the tolerance of these algorithms to combinations of errors improves when the number of physical qubits increases. This means that, as opposed to the NISQ approach, arbitrarily low error rates can be achieved, when scaling to larger and larger quantum computers. Following this approach, more heavy computational tasks will come in reach, provided that we can make quantum computers with enough qubits. For instance, applying Shor's algorithm to crack a 2048-bit RSA encryption code would require an error rate of 0.001 per operation using ~ 20 million qubits [19, 20] over a period of 8 hours.

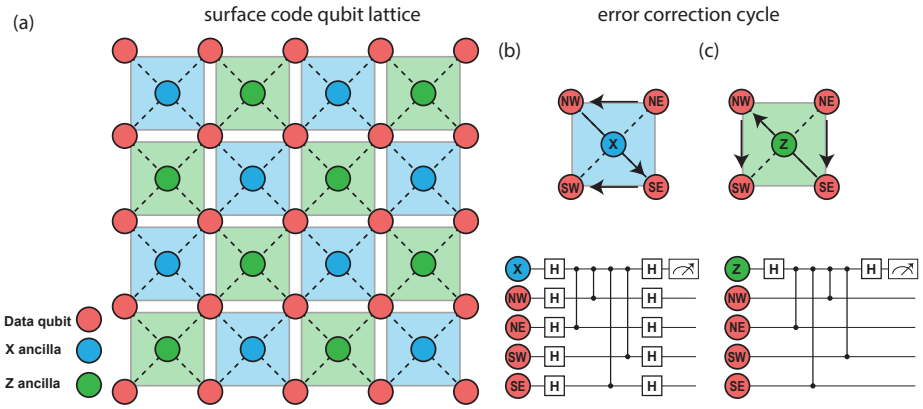


Figure 1.2: (a) Lattice of the surface code with data qubits (red), Z-ancillas (green) and X-ancillas (blue). (b) [(c)] Gate and measurement sequence for a single cycle of X-parity [X-parity] measurement of 4 surrounding data qubits. Squares with the letter H represent single-qubit operations, vertical lines connecting dots represent two-qubit operations. Rectangles with a measurement symbol represent measurements. Image obtained from Ref. [21] with minor modifications.

In particular, QEC implementations of the surface code [17, 19] are momentarily of specific interest as this scheme requires only nearest-neighbour interactions between qubits. Moreover, this scheme allows for relatively high error thresholds on the order of 1% per qubit operation or measurement. In the surface code, a logical qubit is encoded in a two-dimensional lattice of data qubits. The lattice of data qubits is interleaved with ancillary qubits [Fig. 1.2(a)] of two flavours, X and Z, which are used to compare up to four data qubits via parity measurements. These detect erroneous rotations around the qubit's z and x axis, respectively [Fig. 1.2(b, c)]. By repeated performance of these parity measurements, physical errors are projected and signaled to an error decoder. This decoder then matches the obtained parity outcomes to the most likely underlying physical errors [22]. These physical errors can finally be corrected in post-processing or corrections can be applied in real time to restore the original state.

1.4 Superconducting qubits: a promising platform for fault tolerance

The work presented in this thesis is performed exclusively with superconducting qubits. Especially, since the invention of the superconducting transmon qubit (transmission line shunted plasma oscillation qubit) in 2007 [23], this platform has become one of the leading platforms within solid-state quantum information processing. Solid-state systems have the potential advantage of being easily produced at large scale (DiVincenzo critereon i) with currently available lithographic patterning techniques. Within the cQED platform, transmons outstand for their relatively long coherence time [24] over which they preserve quantum information (DiVincenzo critereon iii). A few years after the first demonstrations of multi-qubit algorhithms [25–

29], the fidelities have improved and error rates have been achieved in single-qubit and two-qubit gates [30] with errors, and qubit measurements [31–33] at or below the threshold for the most forgiving QEC schemes.

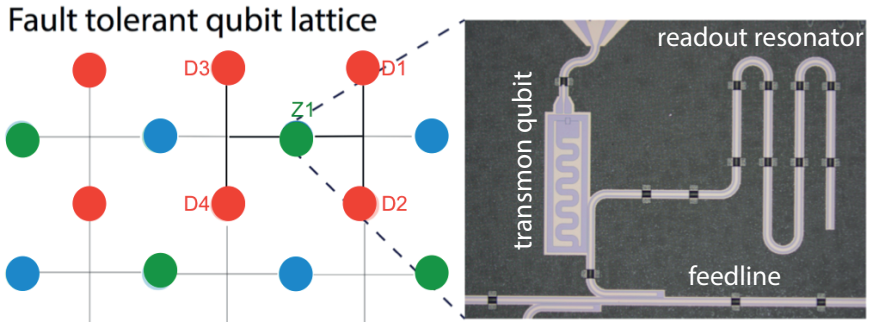


Figure 1.3: A transmon qubit in a surface-code layout. Photograph of a transmon and readout resonator, which is coupled to a feedline. Via the feedline, a readout pulse is injected towards the readout resonator and by analysing the reflected pulse, the qubit state is inferred (the imaged transmon is used in Chs. 3 and 4)

1.5 Towards logical qubits with superconducting circuits

The state-of-the art QEC experiments with transmons performed before this thesis contained either a one-dimensional strand of the surface code using five [34] or nine qubits [35], or a four-qubit square patch [36]. The work in this thesis is directly aimed at realizing a key milestone in quantum computation. Namely, the preservation of a logical qubit in a scalable hardware architecture. This joined project between TU Delft, ETH Zurich, TNO and Zurich Instruments, aims for this goal using transmon qubits in a Surface-code layout with 17 qubits as its first natural step [37] [Fig. 1.4(a)]. The nine data qubits in this device allow protection against any single error in either its qubit operation or measurement and its logical qubit is expected to have a longer lifetime than its constituent physical qubits with currently achieved experimental performance [38]. Furthermore, extending the lattice to five by five data qubits (49 qubits in total), is expected to lead an order of magnitude lower logical error rate. This lowering of the error rate with increased lattice size is referred to as being below the threshold for fault tolerance.

The operations to perform QEC on these devices can be roughly divided in single-qubit gates, two-qubit gates and qubit measurement [Fig. 1.2(b, c)]. The work presented in this thesis has focused primarily on the improvement of repeated readout of qubits. Qubit readout with transmons is performed by dispersively coupling the qubit to a microwave-frequency resonator [Fig. 1.3]. This causes the fundamental resonance of the resonator to be slightly dependent on the qubit state. This frequency shift is probed by applying a microwave pulsed tone to the resonator. The first reason why a focus on readout is important, is that the fidelity of the ancilla readout has to be within a certain threshold for the surface code to function in

the first place. At the start of this project, the error correction cycle time was dominated by the time required for measurement and the time required for photons to leave the readout resonator post measurement. To avoid a build-up of errors during this error correction cycle, the cycle time has to be decreased. Thirdly, the readout pulse which is used to perform readout on the ancilla qubits, can induce errors on other qubits. This is important for QEC, since we need to preserve quantum information during multiple measurement rounds. To reach the conditions for fault tolerance, also improvements in gates are essential. Therefore, parallel, but often overlapping work in Ref. [39] focused on the necessary improvements in single-qubit and two-qubit gates. Most notably, the improvements in fidelity and repeatability in two-qubit gates are a key ingredient. The last and most complex experiment described in this thesis benefits from all the advancements in both readout and gates, allowing us to reach state-of-the-art perfomance in a three-qubit quantum error correction experiment.

1.6 Thesis overview

This thesis focuses on the development of fast superconducting qubit readout for multi-round protocols like quantum error correction and fast gate tuneup. **Chapter two** introduces the reader to the traditional concepts of readout in cQED. In **chapter three** we reduce the measurement cycle time by actively removing photons from the readout resonators after measurement by counter driving the resonators post measurement. We show that this allows reducing the total time for an error correction cycle by a factor three. In **chapter four** we use sequences of interleaved measurements and single-qubit gates as a matter of assessing and tuning the gate performance. Repetitive readout allows us to tune up these gates to reach their performance limit in less than a minute. In **chapter five** we assess and improve the efficiency of all elements in a qubit readout chain. A key recent improvement in qubit readout is the use of special superconducting amplifiers. The key metric to determine how well these amplifiers and the rest of the amplification chain functions is the quantum efficiency (the fraction of readout photons that effectively reaches the observer). We show that the qubit itself is the ideal sensor and that efficiency measurements are consistent for arbitrary readout conditions. This is an important tool to tuneup the amplifiers for optimally fast readout and to distinguish different sources of imperfection. In **chapter six** we perform a three-qubit QEC experiment for which we have redesigned the on-chip readout topology. The improved readout topology allows fast measurement with negligible back-action of the untargeted qubits. This allows us to create entanglement by parity measurement with a high fidelity. We demonstrate the use the repeated parity measurement outcomes to not only protect this entanglement from X and Z errors but also from qubit leakage. This last form of errors is natively not addressed by QEC codes. In this chapter we however demonstrate that by a separate error analysis (using a hidden Markov model) we can infer this leakage while tracking standard qubit errors, thereby opening a new route to fault-tolerant quantum computation in the presence of qubit errors and leakage.

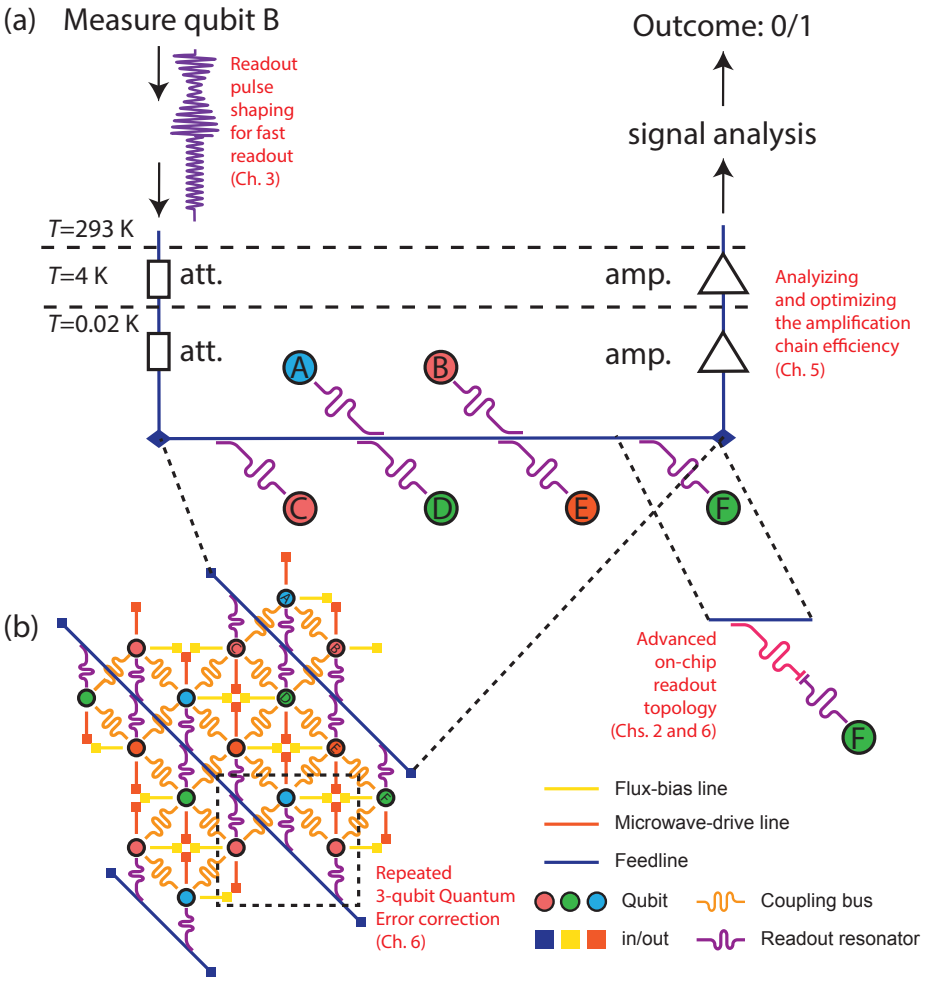
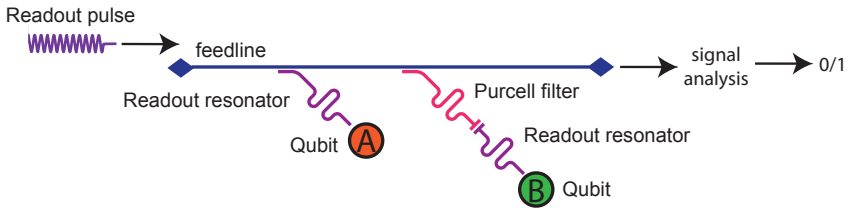


Figure 1.4: Schematic of qubit readout and the related advances presented in this thesis. (a) Schematic for qubit readout for one feedline in a Surface-17 device. For readout, each qubit (colored dots) is coupled to an individual readout resonator, which in turn is connected to a feedline. This coupling causes the resonator's resonance frequency to be slightly dependent on the qubit state. Upon a readout instruction, room-temperature equipment generates a readout pulse. The pulse is guided by coaxial lines into the crygenic environment through various attenuators at different temperature stages. The pulse enters the quantum processor and is guided via a measurement feedline. Depending on the frequency of the pulse, the pulse interacts with a particular resonator which distorts the pulse with dependency on the qubit state. The transmitted and distorted pulse is amplified in various stages and finally analyzed at room-temperature. (b) Circuit diagram of a Surface-17 chip. Qubits are controlled and measured via various inputs and outputs on the chip. Single-qubit gates are applied through microwave-drive lines. Two-qubit gates are activated via flux-bias lines and mediated by coupling busses. Red text highlight the different topics covered in this thesis and directs to the relevant chapters.

1.7 Quantum computation: an emerging industry

The progress in quantum computation, and especially superconducting qubits has led to a new industry branch to be formed by both global enterprises (IBM, Google, Intel, Alibaba) and large-scale quantum computation startups (Rigetti computing, ionQ). Together with governmental and academic institutes, a global race has set off to build the first quantum computer that can perform tasks faster than classical computers. Next to this rivalry for the best performance, several commercial initiatives have sprung up to make prototype quantum computers available on the cloud to users worldwide. The first launch of the IBM quantum experience in 2016 contained a 5-qubit processor. These technology demonstrators seem to boost research in several ways. First, it has become an important educative tool for quantum information courses world wide. Second, they have opened experimental quantum computation to an enormous group of scientists, which has led to 72 experimental publications by researchers who have never had to be close to the experimental setups. At the time of writing the IBM demonstrator has been upgraded to 16 qubits alongside with their competitor Rigetti computing, who has launched a similar demonstrator. New announcements by other institutes proofs that this is just the beginning. IonQ has announced launching an 11-qubit demonstrator using trapped-ion based qubits. QuTech has launched a cloud service which will run on both electron-spin qubits in quantum dots as well as transmon-qubit processors.

2 THE TRANSMON QUBIT AND ITS DISPERSIVE READOUT



Superconducting transmon qubits are promising building blocks for fault-tolerant quantum computers. In this chapter, the transmon qubit is introduced with its equivalent electrical circuit. For readout, we introduce the platform of circuit Quantum Electrodynamics (cQED), in which the qubit is coupled to a readout resonator (a harmonic oscillator). Due to this coupling, the readout resonator's resonance frequency is slightly dependent on the transmon state. So, by probing this resonance frequency with a readout pulse (via a feedline) and analyzing the transmitted signal, the qubit state may be inferred. The high achievable readout fidelity and the non-demolition nature of this readout are key contributions to the success of superconducting qubits. However, for the realization of fault-tolerant quantum computing, it is additionally of key interest that readout be performed as fast as possible. Readout can be sped up by naively enlarging the coupling between qubit and resonator. However, this has the negative side-effect of creating additional energy loss in the transmon via the resonator due to the Purcell effect; creating a compromise. To avoid this tradeoff between readout time and energy loss, we implement an additional filtering resonator (the Purcell filter) which blocks energy at the qubit frequency, while transmitting energy at the readout resonator frequency. This advanced readout scheme allows speeding up the readout, while at the same time reducing loss due to the Purcell effect by orders of magnitude.

2.1 The transmon qubit: an artificial atom to store quantum information

As for many 'classical' carriers of information, quantum information is often encoded in energy states of a system. Man-made quantum circuits like quantum dots and transmon qubits are often called artificial atoms as their discrete energy levels are reminiscent of atomic spectra. As information is represented in these energy states, the ability to preserve them directly sets the ability for containing information. In quantum computing, the loss of information is captured by several time constants. The relaxation time, T_1, determines how fast a qubit relaxes to its lowest energy state (usually $|0\rangle$). The decoherence time T_2 is the typical time constant by which the relative phase of a superposition state (in the Bloch sphere) becomes random. In the evolution of superconducting qubits to the current state of affairs, these numbers are of primary interest.

2.1.1 A simple circuit

The transmon consists of two superconducting islands which couple to each other via a Josephson tunnel junction [40] through which charge (in the form of Cooper pairs) can tunnel from one island to the other. In addition, there is a capacitance between the two islands. The transmon can be modeled as a non-linear LC resonator, consisting of a capacitor and a non-linear inductor (formed by the Josephson junction) [Figure 2.2(a)]. The capacitance in this simple circuit is in reality formed by a sum of capacitances $C = C_\mathrm{s} + C_\mathrm{J} + C_\mathrm{g}$ with C_s, the capacitance of a shunting capacitor, C_J the capacitance in the Josephson junction and C_g an effective contribution by capacitive coupling of either island to the ground plane near the circuit with $C_\mathrm{g} = C_\mathrm{g1} \parallel C_\mathrm{g2}$. In total, this gives the system a charging (coulomb) energy of $E_\mathrm{C} = (e)^2/2C$ and a Josephson energy E_J (set by the junction alone). Using the cooper-pair number imbalance $\hat{n}$ and the superconducting phase between the two islands $\hat{\phi}$ we write the Hamiltonian [41]

$$\hat{H} = 4E_\mathrm{C}\left(\hat{n} - n_\mathrm{g}\right)^2 - E_\mathrm{J}\cos(\hat{\phi}), \tag{2.1}$$

with n_g, an offset charge caused by nearby charged particles leading to a quasi-static potential between the islands.

2.1.2 A fine balance between coherence and addressability

The transmon qubit is a modification of the Cooper-pair box (CPB) [43]. In the CPB, the eigenstates are charge states, defined by the static number of Cooper pairs that have tunneled across the junction, raising the energy of the circuit for every tunneled Cooper pair with respect to equilibrium. In its initial coherently controlled implementation in 1999 [44], a measurable coherence time was achieved of 2 ns. During the years after, optimizations lead to an increase in coherence times of multiple orders of magnitude to $\sim 0.5\ \mu$s [45]. At the time, this was on the low side compared to other superconducting qubits (flux qubits [46, 47], the

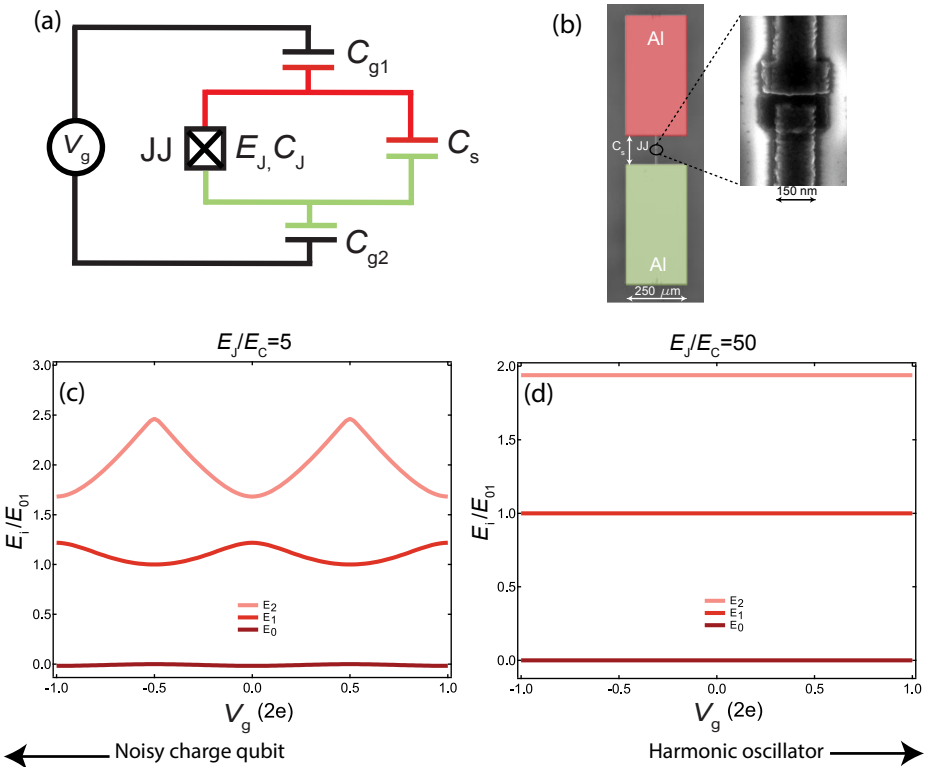


Figure 2.1: Electric circuit model, image and energy levels of a transmon qubit. The transmon qubit **(b)** consists of two aluminum superconducting islands (red and green, false colours), and a Josephson tunnel junction (labeled JJ and magnified in inset, SEM image). The system can be modelled by a non-linear LC-resonator **(a)** with capacitance C_s, which is the capacitance between the islands, and a Josephson junction providing non-linear inductance and additional capacitance C_J. Spurious charged particles on the ground plane cause an offset potential V_g between the islands, capacitively coupled by effective capacitances C_{g1} and C_{g2} . **(c, d)** The first three transmon energy levels as a function of the offset potential for two ratio's of E_J/E_C. This ratio determine both the charge-sensitivity and the anharmonicity. A low ratio makes the qubit more anharmonic, but sensitive to the uncontrolled and noisy charge offset. A higher ratio makes the transmon unsuitable as a qubit, because of the inability to drive qubit transitions selectively. qubit charge-insensitive at the cost of a reduced the anharmonicity. The energies are normalized by E_{01} at $n_g = 0.5$, where we set $E_0 = 0$. Images obtained from [42] with modifications.

quantronium [48]). One of the important limitations of the CPB appeared to be its sensitivity to background charge fluctuations, due to the large contribution of E_C to the overall energy. In 2007, the group of R. Schoelkopf realized [23] that by raising the ratio E_J/E_C, the qubit could be made practically insensitive to charge noise. The arrival at the right design of a transmon qubit can be framed as finding the balance between two of the DiVincenzo's criteria (Chap-

ter 1). Namely, (i) the need for two well-addressable energy levels and (iii) the ability of the qubit to contain information for a long-enough time. As noted, when choosing E_J/E_C too low, criterion (iii) is not met, while choosing E_J/E_C too high, the circuit transforms to a harmonic oscillator (Figure 2.1), violating (i).

Usually, as in this research, a ratio of $40 \lesssim E_J/E_C \lesssim 100$ is chosen with E_C in the range of 200-400 MHz. This is enough to create pulses that drive the E_{01} transition without simultaneously significantly driving E_{12}. Specifically, this allows single-qubit operations to be performed within 20 ns with a probability of leakage to $|2\rangle$ on the order of 10^{-5} [49–51]. The transmon transition frequencies roughly follow the simple relations [23]

$$\begin{aligned} \hbar\omega_{01} &= E_{01} \approx \sqrt{8E_J E_C} - E_C, \\ \hbar\omega_{12} &= E_{12} \approx E_{01} - E_C. \end{aligned} \tag{2.2}$$

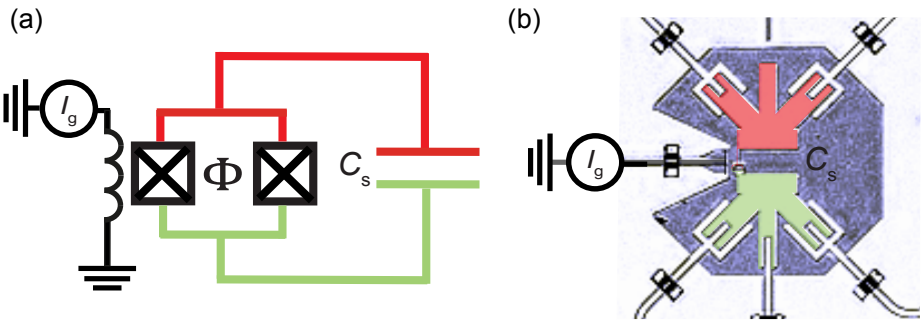


Figure 2.2: Electric circuit model, image and flux-frequency dependence of a flux-tunable transmon qubit. The tunable transmon qubit **(a)** consists of two aluminum superconducting islands (red and green, false colours) coupled by a SQUID loop consisting of two Josephson tunnel junctions. A current source applies a current via the flux-bias line. Through an inductive coupling this current is converted to a flux in the SQUID loop. Image of the transmon obtained from Chapter 5 with modifications.

2.1.3 Flux-tunability using a SQUID loop

In our specific implementation of the transmon, we have in-situ tunability of the transition frequency E_{01}. This tuneability is essential because we perform two-qubit gates by tuning the energy levels of two neighbouring transmons in and out of resonance. Frequency tunability is added to the transmon by replacing the single junction with a pair of junctions (Figure 2.2) with Josephson energies $E_{J,1}$ and $E_{J,2}$. Together, these junctions form a SQUID (superconducting quantum interference device) loop [52]. Effectively in our circuit, this gives rise to a tunable effective Josephson energy as a function of the magnetic flux in the SQUID loop [53].

The Josepshon energy is maximized at zero flux to $E_{J,\max} = E_{J,1} + E_{J,2}$. In this research both junctions are designed to be the same, $E_{J,1} = E_{J,2}$. In this special case, apply-

ing flux to the loop leads to a reduced Josephson energy following $E_J = E_{J,max} |\cos(\pi\Phi/\Phi_0)|$. Consequently, the qubit frequency as a function of flux becomes

$$\hbar\omega_{01}(\Phi) = E_{01}(\Phi) \approx (E_{01,max} + E_C)\sqrt{\cos|\pi\Phi/\Phi_0|} - E_C, \tag{2.3}$$

As for charge, there is a widely observed background noise in magnetic flux in SQUID-based devices [54, 55]. At zero flux, E_J and therefore E_{01} is maximized, leading to a derivative to flux, $\frac{\partial E_{01}}{\partial\Phi} = 0$.

At the maximal frequency, the qubit is protected from this noise, which is therefore referred to as the sweet spot. This sweet spot is generally chosen as the default operation point for the transmon. Usually, one only deviates from this (to first order) flux-insensitive point for flux-pulsed two-qubit gates [25] to enable the transmon to interact with one of its neighbours.

Alternative to the symmetric junction layout used in this thesis, the junctions can be made asymmetric $E_{J,1} \neq E_{J,2}$. This layout has the additional benefit of creating a low-frequency sweet spot (at the cost of limited tuneability). This layout was succesfully used in Refs. [56–58] and its benefits were studied in detail in [59].

2.2 Conventional dispersive readout

2.2.1 From Cavity Quantum Electrodynamics to circuit Quantum Electrodynamics

To perform readout of the transmon and to shield it from the environment, the transmon is coupled to the environment indirectly via a harmonic oscillator; the readout resonator. For qubit readout the matter-like quantum-bit is read out by photons. Studying this type of light-matter interaction at the level of individual photons was first made possible by the introduction of cavity QED, where an atom is placed in a cavity, which is formed by two reflective mirrors. Light, confined in the cavity (the bosonic light-like mode) interacts with the electron energy levels in the atom (fermionic atomic-like modes). If the cavity is on resonance with a two-level atomic transition, the energy in the atom and the cavity begin to swap back and forth following $|1\rangle_a |n-1\rangle_c \leftrightarrow |0\rangle_a |n\rangle_c$, which is between an excited atom and $n-1$ photons in the cavity and a ground-state atom and n photons in the cavity.

Superconducting qubits are generally operated within the circuit QED platform [41], which has a mathematically equivalent description to cavity QED, but operates at a different energy scale. The atom is replaced by the qubit circuit (the artificial atom) and the cavity is replaced by a coplanar waveguide resonator (standing wave in a coaxial transmission line). The first cQED implementation [60] used a CPB coupled to a coplanar waveguide resonator, where a 2-dimensional patterning is used. Alternatively, and in closer analogy to cavity QED, transmons are placed in 3D cavities [24, 31, 61, 62]. For several years, this has been an actively used platform, as achieved qubit coherence times were approximately an order of magnitude higher than standard circuit QED [54], reaching coherence times in excess of 100 μs [63–65]. More recently, coherence in 2D architectures has made a lot of progress. The smaller form factor and ease-of-connectivity to neighbouring qubits have therefore drawn most groups back to the use of coplanar waveguide resonators. Especially, in the form of quantum error

correction which is pursued in this thesis, where qubits have to couple to up to four nearest-neighbouring qubits.

2.2.2 Qubit-cavity coupling: Jaynes-Cummings hamiltonian

For the ideal qubit, with two energy levels only, the system of cavity and atom is described by the Jaynes-Cummings Hamiltonian [66]:

$$\frac{H}{\hbar} = \underbrace{\omega_r \hat{a}^\dagger \hat{a}}_{\text{resonator}} \underbrace{- \frac{\omega_q}{2} \hat{\sigma}_z}_{\text{qubit}} + \underbrace{g\left(\hat{a}^\dagger \hat{\sigma}_- + \hat{a}\hat{\sigma}_+\right)}_{\text{interaction}}. \tag{2.4}$$

The resonator part is described by ω_r the resonator frequency, and the creation and annihilation operator
($\hat{a}^\dagger = \sum_{n=0}^{\infty} \sqrt{n+1}\,|n+1\rangle\langle n|$ and $\hat{a} = \sum_{n=0}^{\infty} \sqrt{n+1}\,|n\rangle\langle n+1|$). The qubit part is described by the qubit frequency ω_q and the qubit's Pauli-z operator $\hat{\sigma}_z = |g\rangle\langle g| - |e\rangle\langle e|$. The interaction part finally uses the qubit-resonator coupling strength g and the lowering and raising operators ($\hat{\sigma}_- = |g\rangle\langle e|$ and $\hat{\sigma}_+ = |e\rangle\langle g|$) of the qubit.

For qubit-resonator detunings $\Delta = \omega_q - \omega_r$ that are large compared to the coupling strength, $|\Delta| \gg g$, the dispersive approximation is valid. This reduces the resonator-qubit interaction to a qubit-state dependent shift of the resonator's frequency and dually, a photon-number dependent shift of the qubit frequency. The Hamiltonian becomes

$$\frac{H_{JC}}{\hbar} = \omega_r \hat{a}^\dagger \hat{a} - \frac{\omega_q}{2}\hat{\sigma}_z + \chi \hat{a}^\dagger \hat{a} \hat{\sigma}_z, \tag{2.5}$$

with $\chi = g^2/\Delta$ the dispersive shift.

This Hamiltonian is of great interest for qubit readout as it allows to determine the qubit state by measuring the resonator's frequency only, i.e without energy exchange between the qubit and its environment. In principle this can lead to an ideal projective non-demolition measurement, yielding equal outcomes when subsequent measurements are performed.

However, this non-demolition character is only featured with low intra-resonator photon numbers, breaking down around a critical photon number $n_{crit} = \Delta^2/4g^2$ [23]. Above this power, resonator and qubit are expected to start exchanging energy, losing the 'projective' character [67, 68]. Also, for transmon qubits, the higher levels (beyond the first two) result in a modification to the dispersive shift $\chi = g^2\alpha/[\Delta(\Delta+\alpha)]$ depending on the anharmonicity $\alpha = \omega_{12} - \omega_{01}$ [23].

2.2.3 Single-shot dispersive readout

The final ingredient required for qubit readout is the ability to infer the resonator's resonance frequency. This is typically done by capacitive coupling it to a feedline [Figure 2.3, (a)]. The coupling strength is usually expressed as the linewidth of the cavity in frequency space κ (ignoring intrinsic loss channels). For a resonator, this width simultaneously sets the characteristic time scale $1/\kappa$ for energy to leak into the cavity (to reach $1 - 1/e \sim 63\%$ of the steady-state population) or out of the cavity (to arrive at $1/e \sim 37\%$ of the initial population).

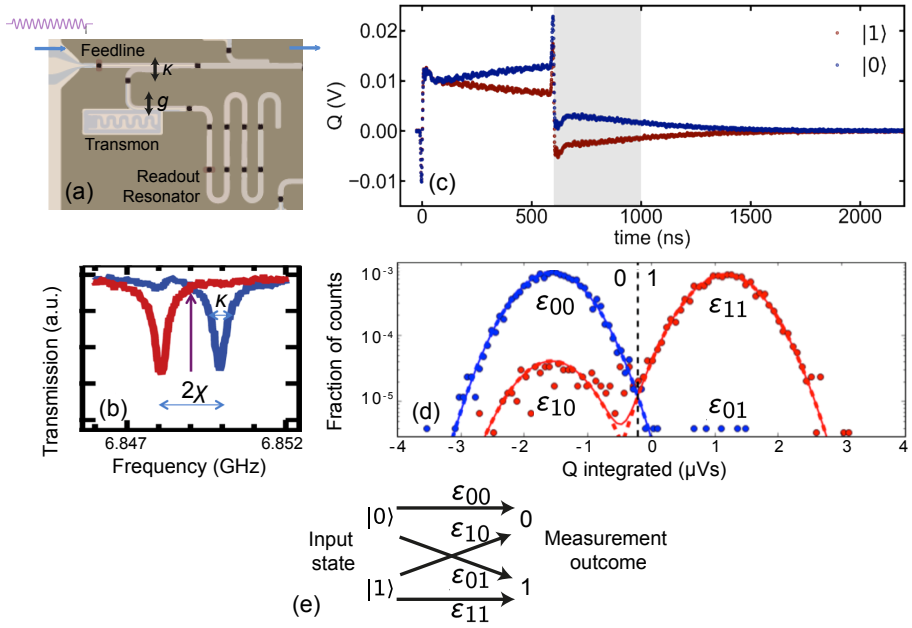


Figure 2.3: Dispersive transmon readout. (a) Device image featuring a transmon qubit coupled to a readout resonator with strength g, which is in turn coupled to a feedline with strength κ. (b) Pulsed feedline transmission near the low-power resonator fundamentals for both qubit states. The dips both have a width κ and are separated by $2\chi = {}^{2g^2\alpha}/_{[\Delta(\Delta+\alpha)]}$. (c) Averaged time-domain transmission measurement for both qubit states, indicating the difference in average signal between the two states (averaged for 2^{15} experimental runs). (d) Single-shot readout histograms for both qubit states. For each individual experimental run, an integrated voltage is recorded (2^{13} in total for each state), which is binned into a histogram. The small overlap between $|0\rangle$ experiments and $|1\rangle$ experiments indicates a high single-shot readout fidelity. The vertical dashed line indicates a threshold voltage for the assigned states 0 and 1. (e) Schematics of the readout error model. The arrows correspond to correct assignment ε_{00}, ε_{11} and incorrect assignment ε_{10}, ε_{01} of the input states $|0\rangle$, $|1\rangle$, respectively. Image of the transmon obtained from Chapter 3 with modifications.

The qubit state is often probed by injecting the feedline with a square readout pulse at a frequency between the resonator resonance dip for qubit in $|0\rangle$ and $|1\rangle$ [Figure 2.3, (a,b)]. This configuration maximizes the phase shift of the reflected pulse, leading to a state-dependent Q-quadrature voltage. This state dependency is most clearly visible in highly averaged transients [Figure 2.3, (c)]. However, in many algorithms (like quantum error correction), the qubit state must be determined in a single shot. The ability to discern the states is revealed by calculating the integrated voltage for each experimental run individually and then binning the values into histograms [Figure 2.3, (d)], grouped for $|0\rangle$ and $|1\rangle$ preparation. For state assignment in further experiments, a threshold voltage can be chosen. Often, this volt-

age is chosen to maximize the average assignment fidelity $\mathcal{F}_a$, which is the probability of assigning the right outcome for both input states on average. $\mathcal{F}_a = 1 - 1/2\,(\varepsilon_{01} + \varepsilon_{10})$, with ε_{01} the probability of assigning 0 for input state $|1\rangle$ and ε_{01}, the probability of assigning 1 to an $|0\rangle$ input.

2.3 Purcell filtering for fast readout

2.3.1 The Purcell effect and the tradeoff between relaxation and measurement speed

Naively, it would seem that the readout process can be sped up by increasing κ_r and g as the first increases the flux of photons that probes the readout resonator and the second increases χ, thereby increasing the state-dependent phase shift in the reflected readout signal. However, there is an important effect on qubit coherence to consider, in which both quantities play a role. Namely, the increased probability of a qubit excitation to leave the transmon via the readout resonator. This effect, known as the Purcell effect, imposes a limit on the qubit relaxation time given by $T_1 < \frac{\Delta^2}{g^2\kappa_r}$ [69]. This means that the product of the measurement photon rate and relaxation time $\kappa_r T_1$ is directly limited by the ratio of Δ and g,

$$\kappa_r T_1 < \left(\frac{\Delta}{g}\right)^2. \tag{2.6}$$

This product can only be increased by increasing Δ or reducing g, which both reduce the dispersive shift χ, leading to less readout visibility [23, 33]. Thus, introducing a tradeoff between relaxation time and measurement speed.

2.3.2 Purcell filtering: how to speed up readout without losing Purcell protection

The readout speed can be increased without running into the Purcell effect by choosing a more advanced readout topology than the traditional scheme of Figure 2.4(a) [70, 71]. This was first demonstrated by adding an additional resonator end (in the form of a stub) at the output side of the feedline. As a whole, this functions as a band-stop filter, with its stop band centered at the qubit frequency [72]. Ideally this filter is transparent to readout photons, yielding a large effective κ, but simultaneously prohibiting the qubit excitation to leak out. Thus, circumventing the limit in Equation (2.6). The same effect was later reached using a band-pass filter resonator [Figure 2.4(b)], with its pass band centered around the readout resonator frequency [33, 35, 73, 74]. This layout is more widely used than the band-stop filter because of two advantages: (1) The fabrication uncertainties of resonator frequencies are typically much lower $\sim$ 30 MHz, than for qubit (sweet spot) frequencies $\sim$ 300 MHz, leading to a higher chance of successful device fabrication. (2) Many processors rely on frequency tuning of qubits for two-qubit gates. The band-stop filter effect is reduced during these operations, while the band-pass filter effect remains intact (even increases for most layouts where $\omega_q < \omega_r$). Assuming the ideal case where readout resonator and Purcell filter are on reso-

nance $\omega_r = \omega_{PF}$, the product $\kappa_r T_1$ is enhanced to [33]

$$\kappa_r T_1 < \left(\frac{\Delta}{g}\right)^2 \left(\frac{2\Delta\omega_r}{\kappa_{PF}\omega_q}\right)^2, \tag{2.7}$$

with κ_{PF}, the linewidth of the Purcell filter. For a typical set of values [$\omega_r/2\pi = 6000$ MHz, $\kappa_{PF}/2\pi = 250$ MHz, $\omega_q/2\pi = 5000$ MHz], the product exceeds the unfiltered limit by a factor $\left(\frac{2\Delta\omega_r}{\kappa_{PF}\omega_q}\right)^2 \sim 100$. This means that for the same readout performance, the Purcell limit is a factor 100 larger.

For multiplexed readout however, this layout comes with a compromise. The condition $\omega_r \sim \omega_{PF}$ is only satisfied if the detuning between ω_r and ω_{PF} is small compared to κ_{PF}. Otherwise, the filter also significantly blocks radiation from the readout resonator towards the output port, leading to the effective κ_{eff} to be lower than κ_r. In practice, this scheme has been used to read out up to nine qubits via one shared PF [35]. In this experiment, a detuning between resonators is chosen of $\sim$ 30 MHz, to still all 'fit' within the pass band of the PF. For applications where qubits are only measured all simultaneously, or where the post-measurement qubit coherence of unmeasured qubits is not to be preserved, this is a suitable solution. However, for the purpose of this thesis, universal QEC, phase information in the data qubits is to be preserved while the ancilla qubits are measured repetitively. This makes this scheme not applicable.

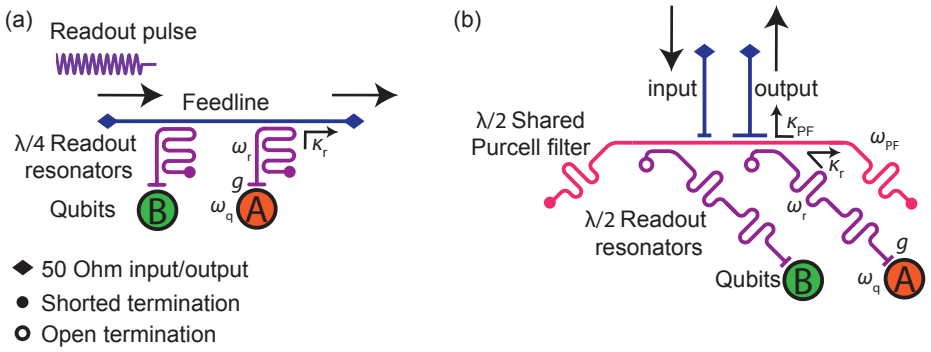


Figure 2.4: Schematic representation of traditional multiplexed two-qubit dispersive readout and multiplexed readout aided by a shared band-pass Purcell filter. (a) Traditional multiplexed readout of two qubits via a feedline and two dedicated readout resonators [70, 71]. Used in Chapters 3 to 5. (b) Multiplexed readout via a band-pass PF. The PF (here) is implemented using a $\lambda/2$ resonator with a voltage antinode in the central region of the filter to allow capacitive coupling to the input, output and readout resonators. Readout resonators also couple capacitively to the qubits. The input port is coupled less strongly to the filter than the output port, making the signal predominantly leave via the output port (minimizing signal loss via the input port). This layout is similar to Ref. [35] where readout resonators are $\lambda/4$ and all couplings to the PF are inductive.

2.3.3 Purcell filtering for QEC: fast and selective readout

During universal QEC, it is essential that phase information in the data qubits is preserved while the ancilla qubits are measured repetitively. The small detunings between resonators, especially combined with fast (large-κ_r) readout resonators, lead to overlap in the filter functions between the readout resonators. This makes it impossible to read out the ancilla qubits, without partially measuring its neighbours, i.e. cross-dephasing. The scheme in Figure 2.5 overcomes this by connecting each readout resonator to an individual Purcell filter [75, 76]. This scheme reduces cross-dephasing for two reasons, (1) the detuning between a set of resonators is not limited by κ_{PF} (as it is for the shared PF scheme). For instance, in a Surface-17 device, we choose to reserve a band of ~ 1 GHz for a set of 9 readout resonators, allowing > 100 MHz spacing. (2) In the limit of large drive detunings $\Delta_{drive} = \omega_{PF} - \omega_{drive}$, the photon number in the readout resonator scales as $\propto 1/\Delta^4_{drive}$ [76], compared to $\propto 1/\Delta^2_{drive}$ for traditional readout or for the shared band-pass filter configuration. This results in significantly less overlap between the filter functions and thereby to a negligible amount of unwanted dephasing. This for the first time allows multiplexed readout to be both fast and selective. In this thesis, this readout scheme is used in Chapter 6, and its implementation is detailed in Ref. [77].

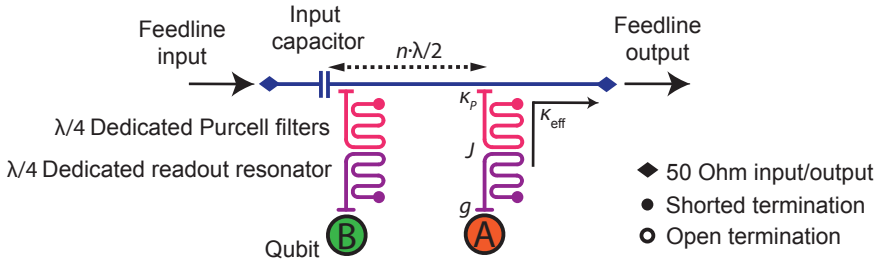
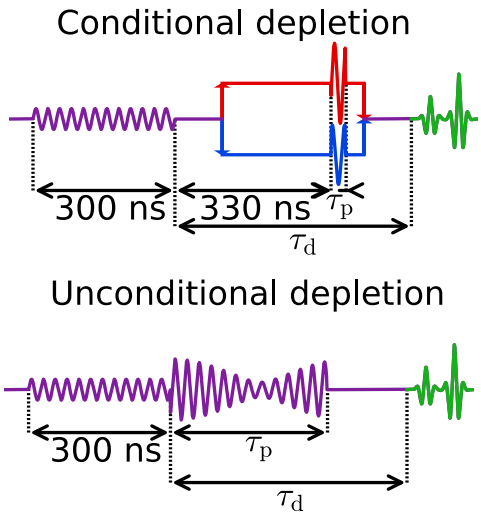


Figure 2.5: (b) Schematic representation of multiplexed readout via dedicated PF-readout resonator pairs, similar to Ref. [76, 78] and as used in Chapter 6. The PFs and readout resonators are both implemented as $\lambda/4$ resonators. All couplings are capacitively: feedline-PF, PF-readout resonator and readout resonator-qubit. An input capacitor is added to provide directionality to the readout signal, predominantly leaving via the output port and thereby minimizing signal loss via the input port. PFs are spaced w.r.t. the input capacitor at $n\lambda/2$ to maximize the coupling.

3

ACTIVE RESONATOR RESET IN THE NONLINEAR DISPERSIVE REGIME OF CIRCUIT QED



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 $\sim$ 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 μs cycle time.

This chapter has been published in *PRApplied* **6**, 034008 (2016).

3.1 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 [29, 34–36, 79–81] 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) [41]. Initially implemented with superconducting qubits, this scheme has since grown to include both semiconducting [82] and hybrid qubit platforms [83, 84]. 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 [41]), 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 [71], the use of parametric amplifiers to improve speed and readout fidelity [31, 32] and null back-action on untargeted qubits [85]), 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 [33, 35, 72] to avoid enhanced qubit relaxation via the Purcell effect [86]. However, increasing κ also enhances qubit dephasing (for a fixed ratio of the dispersive shift χ and κ as desired for high-fidelity readout [87, 88]) by stray photons [89, 90], introducing a compromise. The second approach actively depletes photons using a counter pulse, as recently demonstrated by McClure *et al.* [91]. 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 [41]. Missing analytic forms for this regime, we rely on numerical optimizations by Powell's method [92] to tune up pulses, defined by two or four parameters. Both depletion methods speed up depletion by at least ~ 1250 ns $\sim 5/\kappa$ compared to waiting. To illustrate the benefits for QEC we emulate an ancilla qubit performing parity checks [85, 93] by subjecting our qubit to repeated rounds of coherent operations and measurement. We quantify performance by extracting the mean number of rounds to an un-

expected measurement outcome (i.e. a detection event). With active depletion, we observe an increase in this mean rounds to event, $\overline{\mathrm{RTE}}$, from 15 to 39 and reduce the cycle time to $1\ \mu\mathrm{s} \sim 4/\kappa$. By further fixing the ancilla to remain in the ground state, $\overline{\mathrm{RTE}}$ increases to 75. Simulations [38] indicate that, when including the same intrinsic coherence for surrounding data qubits, a 5-qubit repetition code (studied in [35]) would have a logical error rate below its pseudothreshold [37].

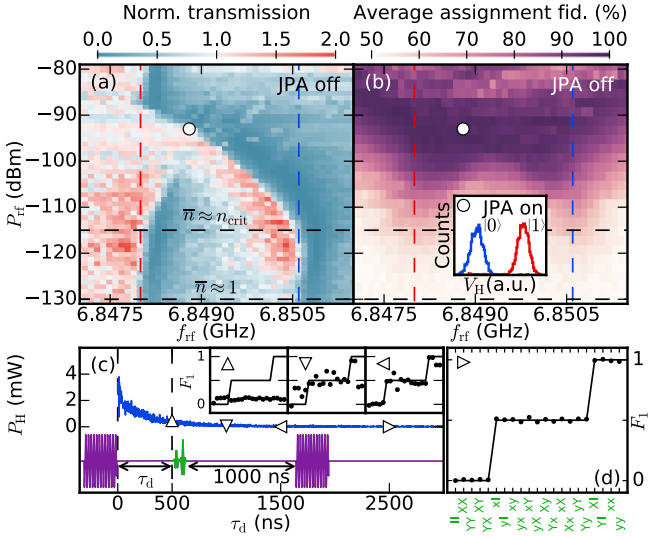


Figure 3.1: (Color online) (a) CW feedline transmission spectroscopy as a function of incident power and frequency near the low- and high-power fundamentals of the resonator. The qubit is simultaneously driven with a weakly saturating tone. The right (left) vertical line indicates the fundamental $f_{\mathrm{r},|0\rangle}$ ($f_{\mathrm{r},|1\rangle}$) in the linear regime. The dot indicates $(P_{\mathrm{rf}}, f_{\mathrm{rf}}) = (-93\ \mathrm{dBm}, 6.8488\ \mathrm{GHz})$ used throughout the experiment. (b) Average assignment fidelity $\mathcal{F}_{\mathrm{a}}$ as a function of P_{rf} and f_{rf} ($\tau_{\mathrm{r}} = 1200$ ns, $\tau_{\mathrm{int}} = 1500$ ns), obtained from histograms with 4000 shots per qubit state. Inset: Turning on the JPA achieves $\mathcal{F}_{\mathrm{a}} = 98.8\%$. (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 $1/\kappa = 250 \pm 2$ ns. Insets and (d): F_1 versus pulse pair for several τ_{d}. The ideal two-step signature is observed at $\tau_{\mathrm{d}} \gtrsim 2500$ ns.

3.2 Experimental results

3.2.1 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 3.4.1). We focus on one qubit-resonator pair for all data presented. This qubit has frequency $f_{\mathrm{q}} = 6.477$ GHz, $T_1 = 25\ \mu$s

and $T_2^{\text{echo}} = 39\ \mu$s. The resonator has a low-power fundamental at $f_{\text{r},|0\rangle} = 6.8506$ GHz ($f_{\text{r},|1\rangle} = 6.8480$ GHz) for qubit in $|0\rangle$ ($|1\rangle$), making the dispersive shift $\chi/\pi = -2.6$ MHz. Note that this shift also corresponds to the qubit detuning per resonator photon. The fundamentals converge to the bare resonator frequency, $f_{\text{r,bare}} = 6.8478$ GHz, at incident power $P_{\text{rf}} \gtrsim -88$ dBm. We calibrate a single-photon power $P_{\text{rf}} = -130$ dBm using photon-number splitting experiments (Figure 3.7) according to [94] and a critical photon number [41] $n_{\text{crit}} = (\Delta^2/4g^2) \approx 33$ ($P_{\text{rf}} \approx -115$ dBm) using $f_{\text{r},|0\rangle} - f_{\text{r,bare}} = g^2/2\pi\Delta$ and $\Delta = 2\pi(f_{\text{q}} - f_{\text{r,bare}})$.

3.2.2 Measurement tune-up and the effect of leftover photons

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

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

where ϵ_{ij} is the probability of incorrectly assigning measurement result j for input state $|i\rangle$. We map $\mathcal{F}_{\text{a}}$ as a function of the power P_{rf} and frequency f_{rf} of a measurement pulse of duration $\tau_{\text{r}} = 1200$ ns [Figure 3.1(b)]. $\mathcal{F}_{\text{a}}$ is maximized at $P_{\text{rf}} = -93$ 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 3.1(a)]. We make two additions to further improve $\mathcal{F}_{\text{a}}$. 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 $\tau_{\text{int}} = 400$ ns) to integrate the homodyne signal before thresholding. This weight function consists of the difference of the averaged transients for $|0\rangle$ and for $|1\rangle$ [95, 96]. These additions achieve $\mathcal{F}_{\text{a}} = 98.8\%$, with $\epsilon_{01} = 0.1\%$ and $\epsilon_{10} = 2.3\%$ [Inset, Figure 3.1(b)], limited by T_1.

The effect of this strong measurement on coherent operations is conveniently illustrated with AllXY measurements [97, 98]. AllXY consists of 21 sequences, two pulses each [Figure 3.1(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 3.1(d)]. Distinct signatures reveal errors in many gate parameters [98]. 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 3.1(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.

3.2.3 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

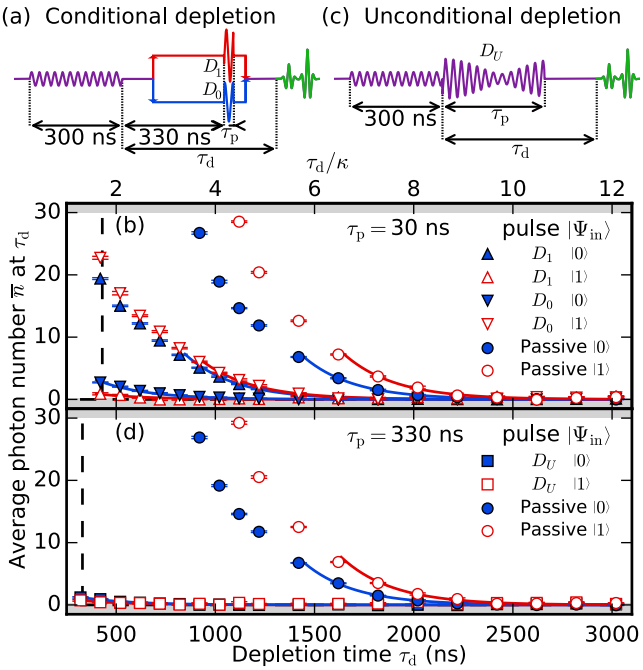


Figure 3.2: (Color online) (a) Pulse scheme for conditional photon depletion. The controller applies a depletion pulse D_0 (at $f_{r,|0\rangle}$) or D_1 (at $f_{r,|1\rangle}$), each with separate amplitude and phase, depending on its declared measurement outcome. (b) Performance of conditional depletion. Average photon number $\overline{n}$ as a function of τ_d for all combinations of input qubit state and depletion pulse. Compared to waiting, conditional depletion saves $\sim$ 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 $\sim$ 1650 (1900) ns for $|0\rangle$ ($|1\rangle$). Exponential best fits (curves) to the data in the linear regime ($\overline{n} \leq 8$) give $1/\kappa = 255 \pm 5$ ns.

described in Section 3.4.2. Measurement noise limits the detector to $\delta\overline{n} \gtrsim 0.3$, providing a dynamic range of two orders of magnitude, suitable for the optimizations that follow.

3.2.4 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 3.2(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 $\overline{n}$ 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 3.4.3). A second optimization at $\tau_d = 500$ ns further optimizes the resulting pulse and converges to $\overline{n} \sim 2.1\ (0.7)$ for $|0\rangle$ ($|1\rangle$), reducing τ_d by at least $5/\kappa$ compared to passive depletion [Figure 3.2(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 [91]), using a universal depletion pulse D_U starting immediately after the measurement pulse [Figure 3.2(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 $\overline{n}$ 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 $\overline{n} \sim 0.8\ (0.4)$ for $|0\rangle$ ($|1\rangle$) and reduces τ_d by $> 6/\kappa$ compared to passive depletion [Figure 3.2(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 [99] may soon help explain these observations and suggest other pulse parameterizations to achieve better depletion.

3.2.5 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 3.3(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, $\overline{\mathrm{RTE}}$. An event is marked by the first qubit measurement outcome deviating from the expected. Imperfections reducing $\overline{\mathrm{RTE}}$ 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 [33]).

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 3.3(b), similar to [22]). 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, $\overline{\mathrm{RTE}}$ is maximized to 14.6 at $\tau_d = 2200$ ns [Figure 3.3(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 $\overline{n} \sim 0.29$ (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 3.4.5). The simple model includes only qubit relaxation and non-photon-induced dephasing (calibrated using standard T_1 and T_2^{echo} mea-

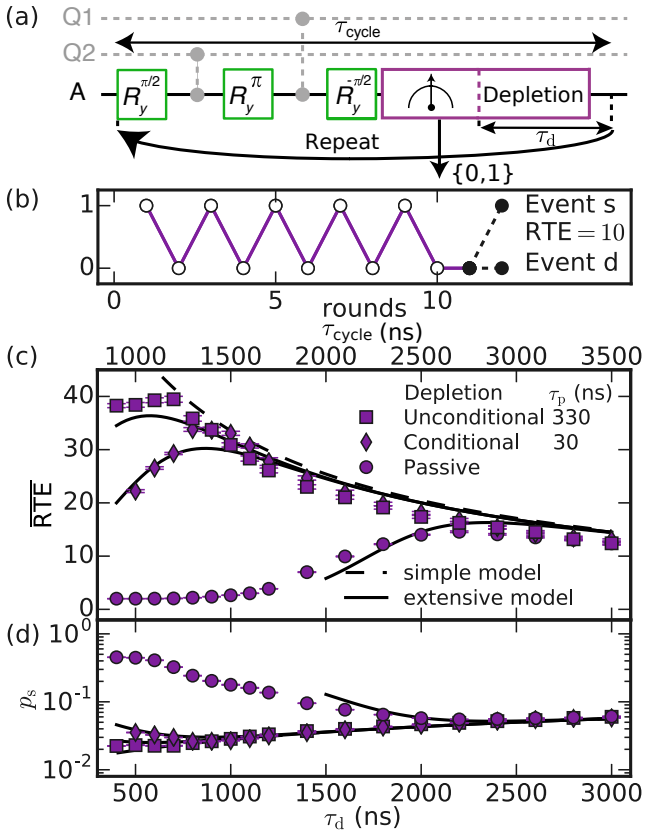


Figure 3.3: (Color online) (a) Block diagram for parity measurements in a repetition code. The ancilla A performs an indirect measurement of the parity of data qubits Q_1 and Q_2 by a coherent interaction step followed by measurement. This emulation replaces conditional-phase gates by idling, reducing 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 3.4.5.

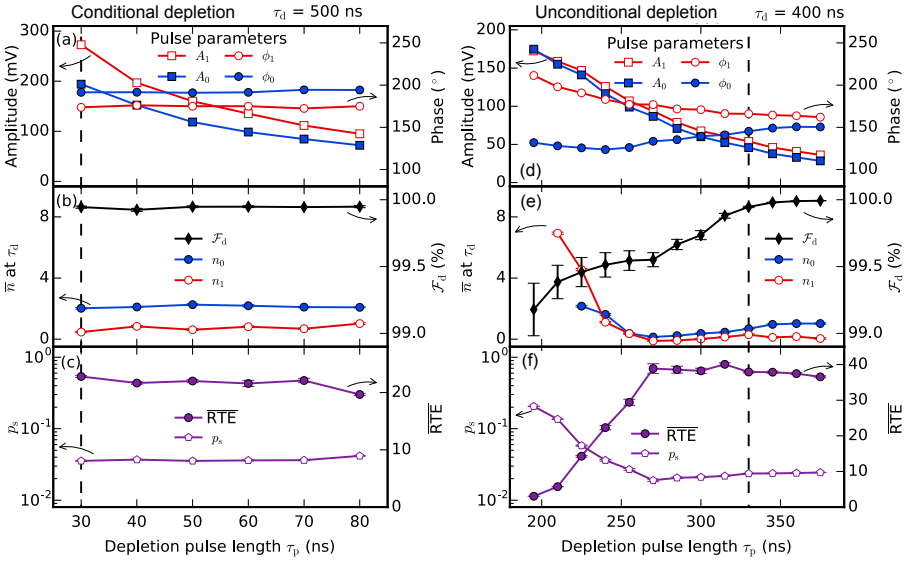


Figure 3.4: Characterization of conditional and unconditional depletion as a function of depletion pulse length τ_p. The dashed lines indicate the pulse lengths for conditional (unconditional) depletion $\tau_p = 30$ ns ($\tau_p = 330$ ns), used in Figures 3.2, 3.3 and 3.5. All data were taken at a fixed $\tau_d = 500$ 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 3.3.

surements). The extensive model also includes photon-induced qubit dephasing and detuning during the coherent step (modeled following [100] with photon dynamics of Figure 3.2), 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 $\overline{n} < 8$. The good agreement between the extensive model and experiment confirms the $\overline{n}$ calibration and demonstrates the nondemolition character of the measurement. The conditions for nondemolition readout in the nonlinear regime have been investigated in Ref. [68].

3.2.6 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 3.4). 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 $\overline{n}$ 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 $\overline{n}$ 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 $\overline{n}$ for $\tau_p = 270$ to 315 ns with the higher $\mathcal{F}_d$ of the longer pulses. However, we did not explore this experimentally.

3.2.7 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 $\Psi_{in} = |0\rangle$ before measurement and depletion (results for $\Psi_{in} = |1\rangle$ are discussed in Section 3.4.4). 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 3.5]. 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 [101].

3.3 Conclusions

The RTE experiments motivate two points for discussion and outlook. First, they highlight the importance of digital feedback [61] in QEC to keep ancillas in $|0\rangle$ as much as possible (as used in a cat code [81]). 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 [102] and the ability to tune without interrupting ongoing error correction [103].

In summary, we have investigated two active methods for fast photon depletion in the nonlinear 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 [68], 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 $\sim$ 500 ns, sufficient to cross the error pseudothreshold in small surface codes at state-of-the-art transmon relaxation times [37].

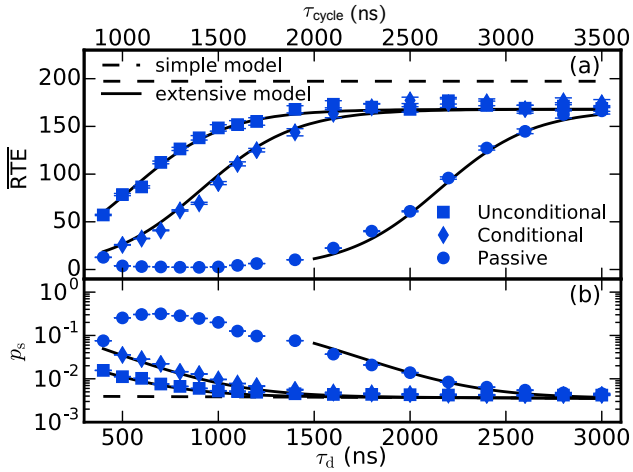


Figure 3.5: (Color online) Emulation of repeating parity measurement for a non-flipping ancilla starting in $|0\rangle$. This variant uses the sequence of Figure 3.3(a) but with opposite polarity on the final $\pi/2$ pulse in order not to flip the ancilla. (a) $\overline{\text{RTE}}$ is no longer sensitive to qubit relaxation during τ_d and reaches a ceiling of $\sim$ 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 3.3.

3.4 Methods

3.4.1 Experimental setup

Figure 3.6 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 [34], 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.

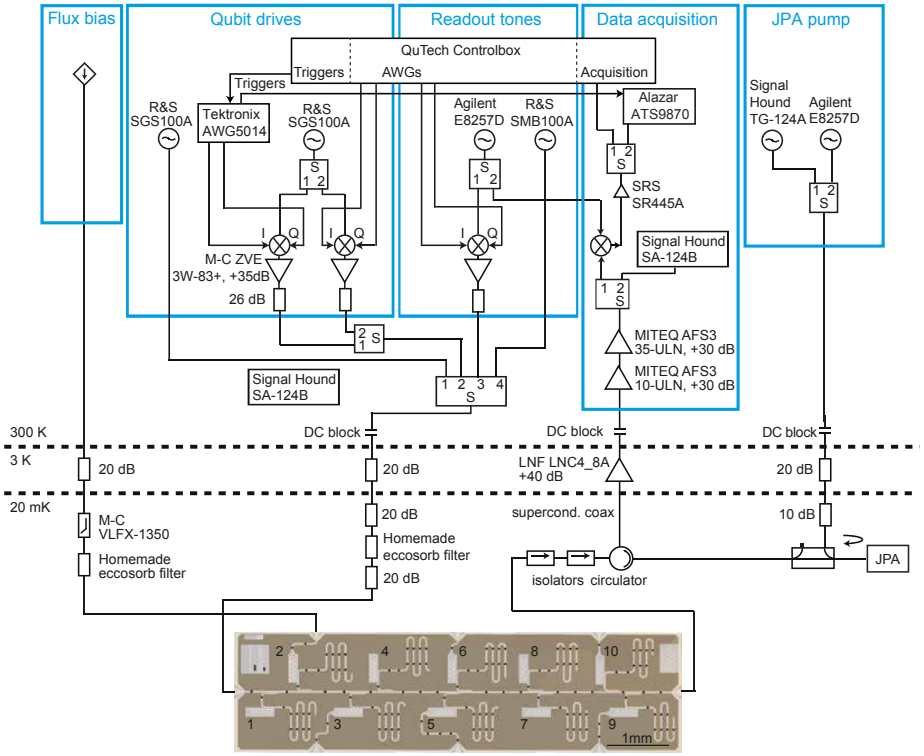


Figure 3.6: Photograph of the cQED chip and complete wiring diagram of electronic components inside and outside the ^{3}He/^{4}He dilution refrigerator (Leiden Cryogenics CF-450). The chip contains ten transmon qubits individually coupled to dedicated readout resonators. All resonators couple capacitively 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.

3.4.2 Photon number calibration

Figure 3.7 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 $\mathcal{E}_{\text{AllXY}} = \alpha\overline{n}+\beta$ 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 $\overline{n}$.

3.4.3 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 [92] in SciPy; *scipy.optimize.fmin_powell* [104].

For conditional depletion, the pulse for $|0\rangle$ ($|1\rangle$) at frequency $f_{r,|0\rangle}$ ($f_{r,|1\rangle}$) is optimized with $\mathcal{E}_{\mathrm{AllXY}}$ 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 $A_{0,\mathrm{init}} = 0.035$ V ($A_{1,\mathrm{init}} = 0.035$ V), equal to half the measurement modulation envelope amplitude, and with an initial phase of $\phi_{0,\mathrm{init}} = 180^\circ$ ($\phi_{1,\mathrm{init}} = 180^\circ$) 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 $\mathcal{E}_{\mathrm{AllXY}}$ 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 $\mathcal{E}_{\mathrm{AllXY}}$ 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 aEach optimizationlgorithm with initial step sizes of 1° and $+1$ mV. Each iteration takes 12 s and each optimization step uses $\sim$ 60 iterations to converge. The total two-step procedure takes $\sim$ 48 minutes in total for the two pulses combined.

For unconditional depletion, the sum of $\mathcal{E}_{\mathrm{AllXY}}$ 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 $A_{0,\mathrm{init}} = A_{1,\mathrm{init}} = 0.035$ V and $\phi_{0,\mathrm{init}} = \phi_{1,\mathrm{init}} = 180^\circ$. 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 $\sim$ 150 iterations to converge and the total two-step procedure takes $\sim$ 2 hours.

3.4.4 Constant excited state QEC emulation

Figure 3.8 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 3.5 but with the qubit initialized in $|1\rangle$. Varying τ_d, we find the optimum tradeoff between

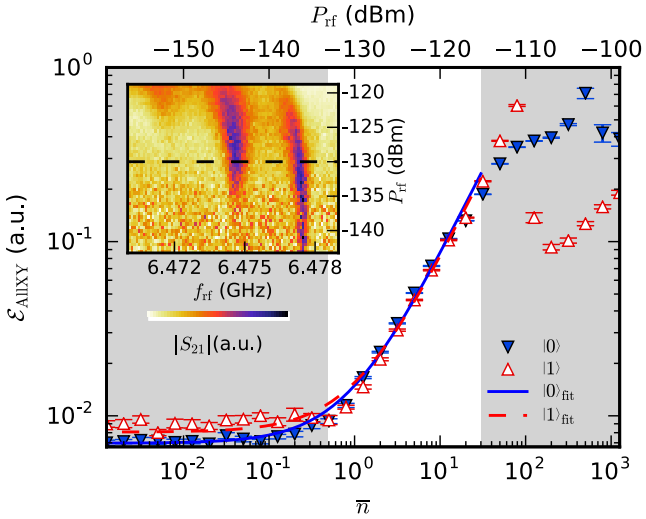


Figure 3.7: Calibration of photon number using AllXY error. $\mathcal{E}_{\rm AllXY}$ measured directly after a readout pulse of 1800 ns duration drives the resonator into a steady-state photon population, $\overline{n}$, for input states $|0\rangle$ and $|1\rangle$. The lines show a bilinear fit to the form $\mathcal{E}_{\rm AllXY} = \alpha\overline{n} + \beta$. Inset: photon-number splitting experiment [94] used to calibrate the single-photon power level, $P_{\rm rf} \sim -130$ dBm.

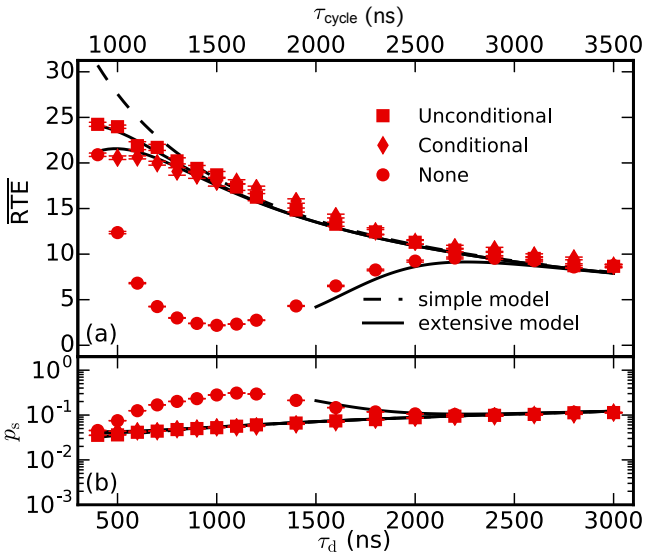


Figure 3.8: Emulated multi-round QEC for a non-flipping ancilla in $|1\rangle$. This variant of the emulation uses the same sequence as Figure 3.5 but with the qubit initialized in $|1\rangle$. (a) Mean rounds to error detection event, $\overline{\rm RTE}$, as a function of $\tau_{\rm d}$. (b) Per-round probability of encountering event of type s as a function of $\tau_{\rm d}$. Added curves correspond to the simple and extensive models described in Section 3.4.5.

errors induced by leftover photons and by relaxation for the three methods. Unconditional depletions performs best, increasing $\overline{\mathrm{RTE}}$ by a factor 2.5 with respect to passive depletion. Note that passive depletion produces a spurious increase in $\overline{\mathrm{RTE}}$ 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.

3.4.5 Theoretical Models

We use two models to compare to data in Figures 3.3, 3.5 and 3.8 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 [105]. Single-qubit gates are approximated as 40 ns decay windows with perfect instantaneous pulses in the middle. This leads to the following scheme: $\tau_\mathrm{d} + 20$ ns of T_1 decay, followed by a $\pi/2$ pulse, then 160 ns of T_2^{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 $\tau_\mathrm{r} = 300$ 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 $1 - \mathcal{F}_\mathrm{d} = 0.1\%$ [Figure 3.4(b)]. The only other possibility is for a single decay event (as we do not allow excitations). To zeroth order in $\tau_\mathrm{r}/T_1 \approx 1/800$, 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. [100]:

$$\frac{d\rho^{\mathrm{qb}}}{dt} = -i\frac{\bar{\omega}_\mathrm{a} + B}{2}[\sigma_\mathrm{z}, \rho^{\mathrm{qb}}] + \gamma_1 \mathcal{D}[\sigma_-]\rho^{\mathrm{qb}} + \frac{\gamma_\phi + \Gamma_\mathrm{d}}{2}\mathcal{D}[\sigma_\mathrm{z}]\rho^{\mathrm{qb}}. \tag{3.1}$$

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_ϕ the pure dephasing rate [$\gamma_\phi = (T_2^{\mathrm{echo}})^{-1} - \frac{1}{2}T_1^{-1} = (177\mu\mathrm{s})^{-1}$]. $\bar{\omega}_\mathrm{a}$ is a constant rotation around the z axis of the Bloch sphere, and so is canceled by the π pulse in the coherent phase. $\Gamma_\mathrm{d} = 2\chi\mathrm{Im}(\alpha_0\alpha_1^*)$ 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

$$\exp\left[-\int \Gamma_\mathrm{d}(t)\right], \tag{3.2}$$

where the integral is taken over the coherent time window. $B = 2\chi\mathrm{Re}(\alpha_0\alpha_1^*)$ 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

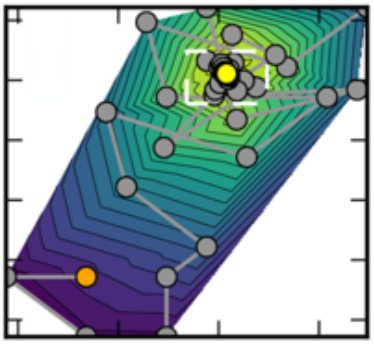
$$\phi_{\text{Stark}} = \int_{t_A} B(t) - \int_{t_B} B(t). \tag{3.3}$$

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 3.2, 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 3.1. As we do not model photon-induced pulse errors, we restrict our modeling to $\overline{n} < 8$, 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 in 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}.

RESTLESS TUNEUP OF HIGH-FIDELITY QUBIT GATES

4



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.

This chapter has been published in *PRApplied* **7**, 041001 (2017).

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 [106, 107] 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 [37, 108] under development by several groups worldwide requires gate errors one order of magnitude below the $\sim 10^{-2}$ threshold of surface code [19, 109].

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 [110], 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 [91, 103, 111–114]. 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 [61]. Passive initialization becomes increasingly inefficient as T_1 steadily increases [115, 116], while feedback-based reset is technically involved [117].

Here, we present a gate tuneup method that dispenses with T_1 initialization and achieves tenfold speedup over the state of the art [112] 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) [118] 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

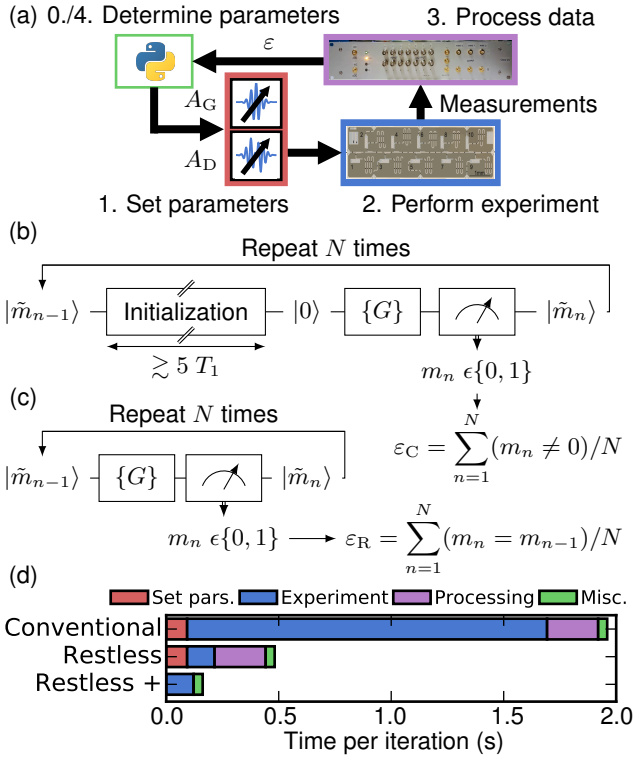


Figure 4.1: (a) A general qubit gate tuneup loop. In conventional tuneup (b), the qubit is initialized before measuring the effect of $\{G\}$. In restless tuneup (c), the qubit is not initialized, and instead m_{n-1} is used to estimate the initial state ($|\tilde{m}_{n-1}\rangle$). (d) Benchmark of 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 rest [1]. For example, when the net ideal gate operation is a bit flip, we can define the error fraction

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

We demonstrate restless tuneup of DRAG pulses [49] on the transmon qubit used in Chapter 3. We choose DRAG pulses (duration $\tau_p = 20$ ns) for their proven ability to reduce gate

[1] except 3.25 μs needed for passive depletion of photons leftover from the 1 μs measurement [113]

error and leakage [51, 97] 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) [50] (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 [119], 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 [120, 121]

$$1 - \varepsilon_\mathrm{C} = A \cdot (p_\mathrm{Cl})^{N_\mathrm{Cl}} + B. \tag{4.2}$$

Here, A and B are constants determined by state preparation and measurement error (SPAM), and $1 - p_\mathrm{Cl}$ is the average depolarizing probability per gate, making $F_\mathrm{Cl} = \frac{1}{2} + \frac{1}{2}p_\mathrm{Cl}$. 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 $\varepsilon_\mathrm{C}(N_\mathrm{Cl})$ decreases monotonically with F_Cl [112].

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

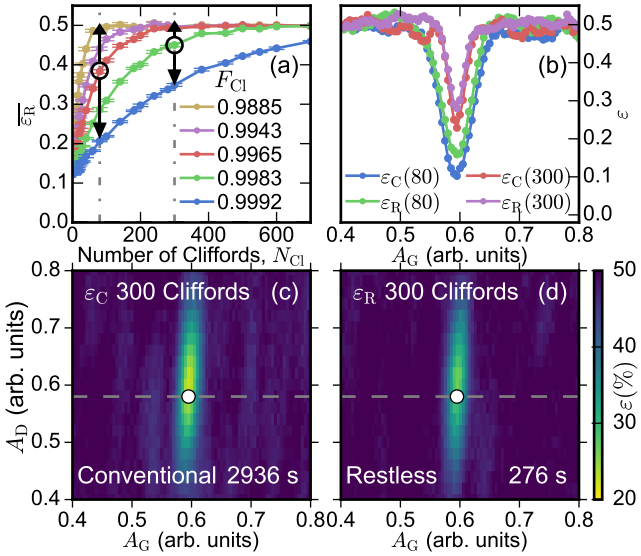


Figure 4.2: (a) Average error fraction of RRB for different F_{Cl} vs N_{Cl}. (b) ε_{C} and ε_{R} as a function of A_{G} for $N_{\mathrm{Cl}} = 80$ and $N_{\mathrm{Cl}} = 300$. The curves are denoted by a dashed line in (c-d). (c-d) ε for $N_{\mathrm{Cl}} = 300$ as a function of A_{G} and A_{D}. White 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 [50, 51]), 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 $\varepsilon_{\mathrm{R}}(N_{\mathrm{Cl}})$ [$\overline{\varepsilon_{\mathrm{R}}}(N_{\mathrm{Cl}})$] 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 $N_{\mathrm{Cl}} = 300$.

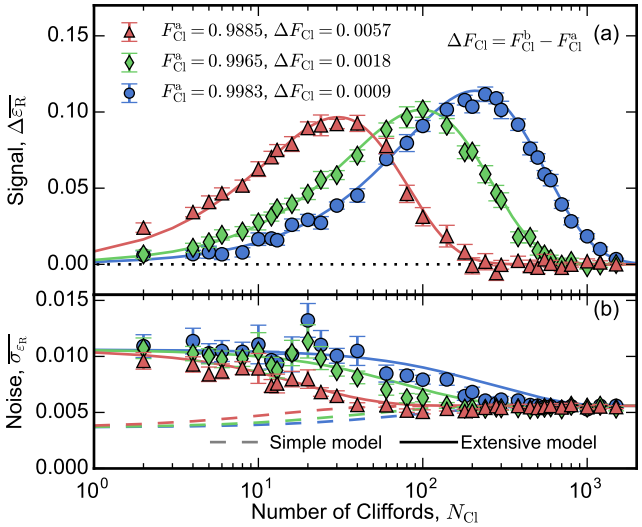


Figure 4.3: (a) Signal $\Delta\overline{\varepsilon_R}$ for a halving of the gate infidelity, plotted as a function N_{Cl} at $F^{a}_{Cl} \sim 0.989$ (red), 0.996 (green) and 0.998 (blue). (b) Noise dependence on N_{Cl} at the same fidelity levels. Added curves are obtained from the two models described 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, $\Delta\overline{\varepsilon_R} = \overline{\varepsilon_R}(F^{b}_{Cl}) - \overline{\varepsilon_R}(F^{a}_{Cl})$, from F^{a}_{Cl} to $F^{b}_{Cl} \approx \frac{1}{2} + \frac{1}{2}F^{a}_{Cl}$ (halving the infidelity). For noise we take $\overline{\sigma_{\varepsilon_R}}$, the average standard deviation of ε_R between F^{a}_{Cl} and F^{b}_{Cl}. We find that the maximal SNR remains ~ 15 for an optimal choice of N_{Cl} that increases with F^{a}_{Cl} (Figure 4.3 and details in Section 4.5.2). This allows tuning in logarithmic time, since reducing error rates $p \to p/2^M$ 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 $\varepsilon_C \to \varepsilon_R$. 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 [122] 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 [113].

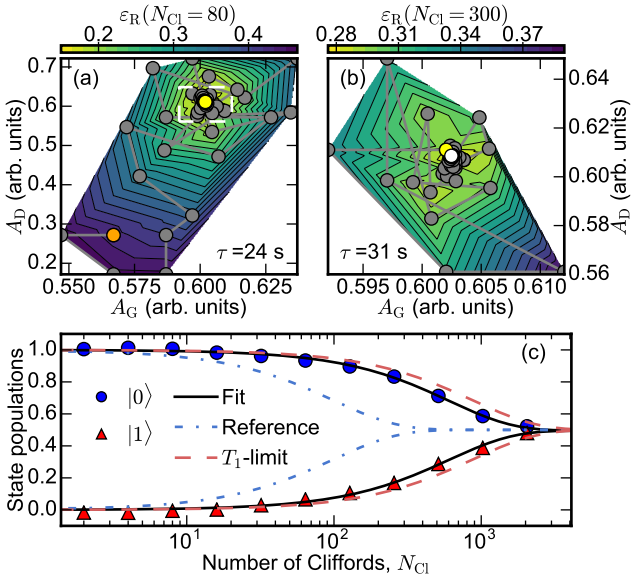


Figure 4.4: Two-parameter restless tuneup using a two-step optimization, first at $N_{Cl} = 80$ (a) and then at $N_{Cl} = 300$ (b). Contour plots show a linear interpolation of ε_R. The starting point, intermediate result and final result are marked by orange, yellow, and white dots respectively. (c) 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 [123] 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 $\Delta A_G \approx -0.03 A_G^{opt}$, $\Delta A_D \approx -0.25 A_D^{opt}$ for the first step, and $\Delta A_G \approx -0.01 A_G^{opt}$, $\Delta A_D \approx -0.08 A_D^{opt}$ 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 μs		19.3 μs	

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 [124]:

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

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 Conlcusions

In summary, we have developed an accurate and robust tuneup method achieving a ten-fold speedup over the state of the art [112]. 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 [125, 126]. Immediate next experiments will extend the restless concept to the tuneup of two-qubit controlled-phase gates [26, 30] exploiting interactions with non-computational states [127], 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 [128], and *QCoDeS* for the framework of instrument drivers [129]. 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 upconverted 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 3) is operated at its coherence sweetspot, with transition frequency 6.47 GHz, -315 MHz anharmonicity, relaxation time $T_1 = 22$ μs and echo time $T_2^{\text{echo}} = 39$ μs. 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 demodu-

lation, integration and discrimination of measurement transients and real-time calculation of ε.

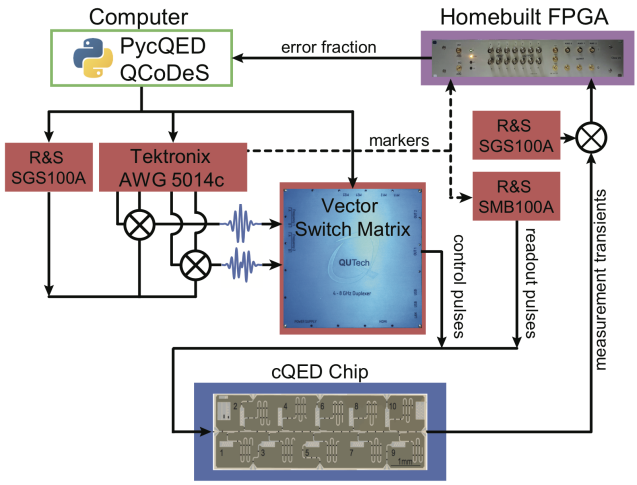


Figure 4.5: Schematic overview of the hardware components used in the numerical tuneup.

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_{G}. 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 $p_{\mathrm{s}}^{(\mathrm{c})}$. Independent measurements of the average and standard deviation of T_1, and extractions of p_{pulse} and $p_{\mathrm{s}}^{(\mathrm{c})}$ 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 $\tau_{\mathrm{RRB}} = \tau_{\mathrm{RO}} + \tau_{\mathrm{Cl}} N_{\mathrm{Cl}}$. The static time $\tau_{\mathrm{RO}} = 4.25\ \mu\mathrm{s}$ is the readout-and-depletion time, whilst the Clifford-dependent

time $\tau_{\mathrm{Cl}} = 37.5$ 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)

$$p_{\mathrm{e}} = p_{\mathrm{s}} +_p N_{\mathrm{Cl}} \times_p p_{\mathrm{c}}.$$

Here, p_{s} is the error contribution due to SPAM, and $p_{\mathrm{c}} = 1 - F_{\mathrm{Cl}}$ 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: $a +_p a +_p \ldots +_p a$ (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:

$$N_{\mathrm{Cl}} \times_p p_{\mathrm{c}} = N_{\mathrm{Cl}} p_{\mathrm{c}} (1-p_{\mathrm{c}})^{N_{\mathrm{Cl}}-1} + \binom{N_{\mathrm{Cl}}}{3} {p_{\mathrm{c}}}^3 (1-p_{\mathrm{c}})^{N_{\mathrm{Cl}}-3} + \ldots.$$

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

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

resulting in a final error rate

$$p_{\mathrm{e}} = p_{\mathrm{s}} + \frac{1}{2}[1 - (1-2p_{\mathrm{c}})^{N_{\mathrm{Cl}}}](1-2p_{\mathrm{s}}). \tag{4.4}$$

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 $\overline{p_{\mathrm{e}}}$. This in turn can be calculated assuming that p_{s} and p_{c} are drawn from a normal distribution, giving

$$\overline{p_{\mathrm{e}}} = \int dp_{\mathrm{s}} dp_{\mathrm{c}}\, p_{\mathrm{e}}(p_{\mathrm{s}}, p_{\mathrm{c}}) \frac{1}{2\pi} \exp\left(\begin{pmatrix} p_{\mathrm{s}} & p_{\mathrm{c}} \end{pmatrix} \Sigma^{-1} \begin{pmatrix} p_{\mathrm{s}} \\ p_{\mathrm{c}} \end{pmatrix} \right) |\Sigma|^{-1/2}.$$

Here, Σ is the covariance matrix;

$$\Sigma = \begin{pmatrix} \mathrm{var}(p_{\mathrm{s}}) & \mathrm{covar}(p_{\mathrm{c}}, p_{\mathrm{s}}) \\ \mathrm{covar}(p_{\mathrm{c}}, p_{\mathrm{s}}) & \mathrm{var}(p_{\mathrm{c}}) \end{pmatrix},$$

with $\overline{p_{\mathrm{c}}}$ ($\mathrm{var}(p_{\mathrm{c}})$) and $\overline{p_{\mathrm{s}}}$ ($\mathrm{var}(p_{\mathrm{s}})$) the means (variances) of p_{c} and p_{s}, respectively, and $\mathrm{covar}(p_{\mathrm{c}}, p_{\mathrm{s}})$ the covariance between p_{c} and p_{s}. The inverse of Σ can be calculated,

$$\Sigma^{-1} = \frac{1}{\mathrm{var}(p_{\mathrm{s}})\mathrm{var}(p_{\mathrm{c}}) - \mathrm{covar}(p_{\mathrm{s}}, p_{\mathrm{c}})^2} \begin{pmatrix} \mathrm{var}(p_{\mathrm{c}}) & -\mathrm{covar}(p_{\mathrm{s}}, p_{\mathrm{c}}) \\ -\mathrm{covar}(p_{\mathrm{s}}, p_{\mathrm{c}}) & \mathrm{var}(p_{\mathrm{s}}) \end{pmatrix}.$$

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

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

From here the integral in p_s can be evaluated:

$$\overline{p_e} = \overline{p_s} + \frac{1}{2}(1-2\overline{p_s})\left(1 - \int dp_e\,(1-2p_c)^{N_{Cl}} \frac{\exp\left(\frac{-(p_c-\overline{p_c})^2}{2\,\mathrm{var}(p_c)}\right)}{\sqrt{2\pi\,\mathrm{var}(p_c)}}\right).$$

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

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

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

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

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

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

and

$$\begin{aligned}\mathrm{var}(p_e) = {} & (1-2\overline{p_c})^{2N_{Cl}}\mathrm{var}(p_s) + {N_{Cl}}^2(1-2\overline{p_s})^2(1-2\overline{p_c})^{2(N_{Cl}-1)}\mathrm{var}(p_c) \\ & + 2N_{Cl}(1-2\overline{p_s})(1-2p_c)^{2N_{Cl}-1}\mathrm{covar}(p_c, p_s).\end{aligned}$$

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 $\overline{p_e}$ and standard deviation σ_{p_e}. Let N_e be the number of erroneous measurements, given as $N_e = N\varepsilon_R$. 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

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

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

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

This leads to the final result:

$$\mathrm{var}(\varepsilon_R) = \frac{1}{N}\overline{p_e}(1-\overline{p_e}) + \frac{N-1}{N}\mathrm{var}(p_e). \tag{4.5}$$

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

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 $\tau_m = 1\ \mu$s, is T_1 rather than noise limited. We can approximate it by perfect state update and measurement at $\tau_b \approx 4\tau_m/7 = 0.57\ \mu$s [38], followed by a rest time $\tau_a = \tau_{RO} - \tau_b = 3.68\ \mu$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

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

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:

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

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_{Cl}, and T_1. In order to calculate the standard deviation, we must then calculate the first derivative, via

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

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 $p_c^{(T_1)} = 1 - F_{Cl}^{(T_1)}$, with $F_{Cl}^{(T_1)}$ as defined in Eq. (3).

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 $\alpha = 8.4 \cdot 10^{-13}$ s^2/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_{Cl}), and below by the acquisition time for 50 such experiments ($f_l = 1/3.7$ s). We find $\overline{T_1} = 21.6$ μs 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 $\Delta\overline{\varepsilon_R}$ and $\overline{\sigma_{\varepsilon_R}}$ shown in Figure 4.3 of the main text, finding good agreement with experiment.

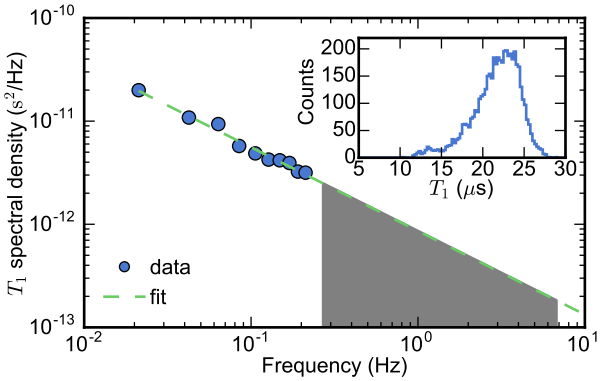


Figure 4.6: Power spectral density of T_1 fluctuations. Main panel: measured single-sided PSD of T_1 fluctuations and best fit (see details in text). The indicated frequency range is that relevant for estimating σ_{T_1} in the RRB experiments of Section 4.5.2. Inset: Histogram of 4914 T_1 measurements.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

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 [130]. These are then used to construct the 24 elements ($G^{\mathrm{GST}}_{\mathrm{Cl}_n}$) of the (single-qubit) Clifford group ($\mathcal{G}_{\mathrm{Cl}}$) according to the decomposition of [119]. 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 $G^{\mathrm{GST}}_{\mathrm{Cl}}|\rho_i\rangle\rangle$ and the target state $|\rho_t\rangle\rangle$:

$$p_n = \sqrt[6]{\prod_{\rho_i} \langle\langle \rho_{\mathrm{t}} | G^{\mathrm{GST}}_{\mathrm{Cl}-n} | \rho_{\mathrm{i}} \rangle\rangle},$$

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

$$|\rho_{\mathrm{t}}\rangle\rangle = G^{\mathrm{Ideal}}_{\mathrm{Cl}-n} |\rho_{\mathrm{i}}\rangle\rangle.$$

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

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
$F_{\mathrm{Cl}}^{\mathrm{GST}}$	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.

4.5.7 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.

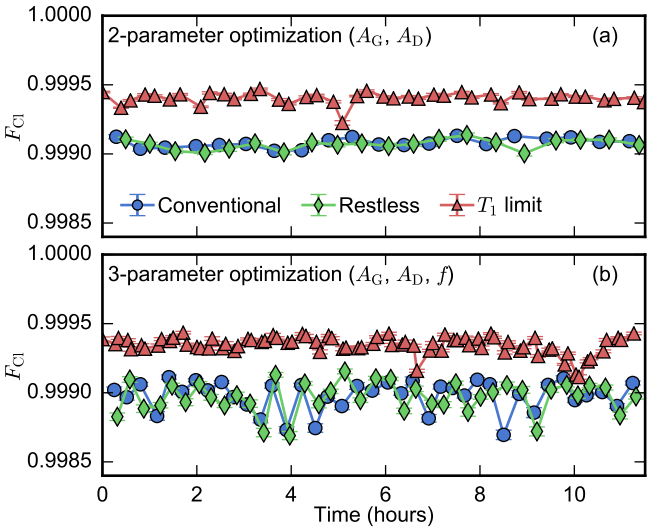
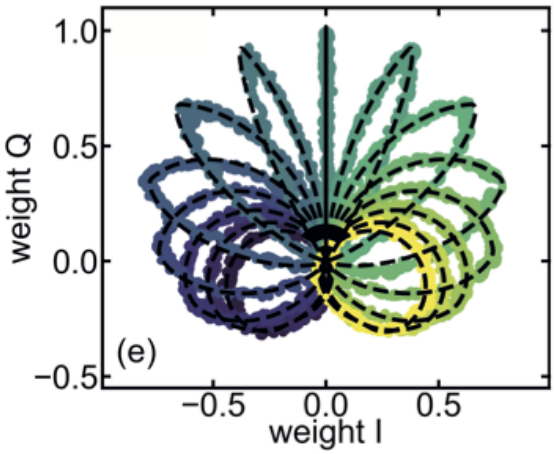


Figure 4.7: Performance comparison of repeated restless and conventional tuneups for two parameters (a) and three parameters (b). Each iteration consists of a conventional tuneup followed by a CRB measurement of F_{Cl}, a restless tuneup followed by a CRB measurement of F_{Cl}, and a T_1 experiment to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

5 GENERAL METHOD FOR EXTRACTING THE QUANTUM EFFICIENCY OF DISPERSIVE QUBIT READOUT IN CIRCUIT QED



We present and demonstrate a general 3-step method for extracting the quantum efficiency of dispersive qubit readout in circuit QED. We use active depletion of post-measurement photons and optimal integration weight functions on two quadratures to maximize the signal-to-noise ratio of non- steady-state homodyne measurement. We derive analytically and demonstrate experimentally that the method robustly extracts the quantum efficiency for arbitrary readout conditions in the linear regime. We use the proven method to optimally bias a Josephon traveling-wave parametric amplifier and to quantify the different noise contributions in the readout amplification chain.

This chapter has been published in *Applied Physics Letters* **9**, 112 (2018).

5.1 Introduction

Many protocols in quantum information processing, like quantum error correction [101, 131], require rapid interleaving of qubit gates and measurements. These measurements are ideally nondemolition, fast, and high fidelity. In circuit QED [23, 41, 60], a leading platform for quantum computing, nondemolition readout is routinely achieved by off-resonantly coupling a qubit to a resonator. The qubit-state-dependent dispersive shift of the resonator frequency is inferred by measuring the resonator response to an interrogating pulse using homodyne detection. A key element setting the speed and fidelity of dispersive readout is the quantum efficiency [132], which quantifies how the signal-to-noise ratio is degraded with respect to the limit imposed by quantum vacuum fluctuations.

In recent years, the use of superconducting parametric amplifiers [133–137] as the front end of the readout amplification chain has boosted the quantum efficiency towards unity, leading to readout infidelity on the order of one percent [31, 32] in individual qubits. Most recently, the development of traveling-wave parametric amplifiers [138, 139] (TWPAs) has extended the amplification bandwidth from tens of MHz to several GHz and with sufficient dynamic range to readout tens of qubits. For characterization and optimization of the amplification chain, the ability to robustly determine the quantum efficiency is an important benchmark.

A common method for quantifying the quantum efficiency η that does not rely on calibrated noise sources compares the information obtained in a weak qubit measurement (expressed by the signal-to-noise-ratio SNR) to the dephasing of the qubit (expressed by the decay of the off-diagonal elements of the qubit density matrix) [140], $\eta = \frac{\mathrm{SNR}^2}{4\gamma_m}$, with $e^{-\gamma_m} = \frac{|\rho_{01}(T)|}{|\rho_{01}(0)|}$, where T is the measurement duration.[1] Previous experimental work [33, 138, 141, 142] has been restricted to fast resonators driven under specific symmetry conditions such that information is contained in only one quadrature of the output field and in steady state. To allow in-situ calibration of η in multi-qubit devices under development [21, 143–146], a method is desirable that does not rely on either of these conditions.

Here, we present and demonstrate a general three-step method for extracting the quantum efficiency of linear dispersive readout in cQED. Our method disposes with previous requirements in both the dynamics and the phase space trajectory of the resonator field, while requiring two easily met conditions: the depletion of resonator photons post measurement [91, 113], and the ability to perform weighted integration of both quadratures of the output field [95, 96]. We experimentally test the method on a qubit-resonator pair with a Josephson TWPA (JTWPA) [138] at the front end of the amplification chain. To prove the generality of the method, we extract a consistent value of η for different readout drive frequencies and drive envelopes. Finally, we use the method to optimally bias the JTWPA and to quantify the different noise contributions in the readout amplification chain.

[1] This definition of quantum effiency applies to phase-preserving amplification where the unavoidable quantum noise from the idler mode of the amplifier is included in the quantum limit. Using this definition, $\eta = 1$ corresponds to an ideal phase preserving amplification.

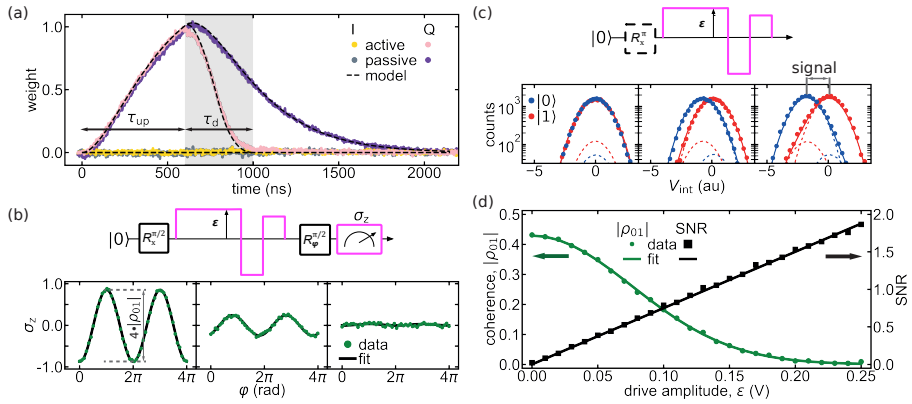


Figure 5.1: The three-step method for extracting the quantum efficiency with active photon depletion. (a) Calibration of the optimal weight functions for the in-phase quadrature I and out-of-phase quadrature Q for active depletion (passive depletion is shown for reference). The measurement pulse consists of a ramp-up of duration $\tau_{up} = 600$ ns and two 200 ns depletion segments ($\tau_d = 400$ ns). The weight functions show the dynamics of the information gain during readout and the effect of the active photon depletion (grey area). Dashed black curves are extracted from a linear model (see Section 5.6). (b) Study of dephasing under variable-strength weak measurement. Observed Ramsey fringes at from left to right $\varepsilon = 0.0, 0.12, 0.25$ V. The measurement pulse, globally scaled with ε, is embedded in a fixed-length ($\tau_{int} = 1100$ ns) Ramsey sequence with final strong fixed-amplitude measurement. The azimuthal angle φ of the final $\pi/2$ rotation is swept from 0 to 4π to discern deterministic phase shifts and dephasing. The coherence $|\rho_{01}|$ is extracted by fitting each fringe with the form $\sigma_z = 2|\rho_{01}|\cos(\varphi + \varphi_0)$. (c) Study of signal-to-noise ratio of variable-strength weak measurement. Histograms of 2^{15} shots at from left to right: $\varepsilon = 0.0, 0.12, 0.25$ V. The qubit is prepared in $|0\rangle$ without (blue) and in $|1\rangle$ with a π pulse (red). Each measurement record is integrated in real time with the weight functions of (a) during $\tau_{int} = 1100$ ns to obtain V_{int}. Each histogram (markers) is fitted with the sum of two Gaussian functions (solid lines), whose individual components are indicated by the dashed lines. From the fits we get the signal, distance between the main Gaussian for $|0\rangle$ and $|1\rangle$, and noise, their average standard deviations. (d) Quantum efficiency extraction. Coherence data is fitted with the form $|\rho_{01}| = be^{-\varepsilon^2/2\sigma^2}$ and signal-to-noise data with the form $\mathrm{SNR} = a\varepsilon$. From the best fits we extract $\eta_e = a^2\sigma^2/2 = 0.165 \pm 0.002$.

5.2 Derivation of the 3-step method

We first derive the method, obtaining experimental boundary conditions. For a measurement in the linear dispersive regime of cQED, the internal field $\alpha(t)$ of the readout resonator, driven

by a pulse with envelope $\varepsilon f(t)$ and detuned by Δ from the resonator center frequency, is described by [100, 140]

$$\frac{\partial \alpha_{|0\rangle/|1\rangle}}{\partial t} = -i\varepsilon f(t) - i(\Delta \pm \chi)\alpha(t) - \frac{\kappa}{2}\alpha(t), \tag{5.1}$$

where κ is the resonator linewidth and 2χ is the dispersive shift. The upper (lower) sign has to be chosen for the qubit in the ground $|0\rangle$ [excited $|1\rangle$] state. We study the evolution of the SNR and the measurement-induced dephasing as a function of the drive amplitude ε, while keeping T constant. We find that the SNR scales linearly, $\mathrm{SNR} = a\varepsilon$, and that coherence elements exhibit a Gaussian dependence, $|\rho_{01}(T,\varepsilon)| = |\rho_{01}(T,0)|\, \mathrm{e}^{-\frac{\varepsilon^2}{2\sigma_\mathrm{m}^2}}$, with a and σ_m dependent on κ, χ, Δ, and $f(t)$. Furthermore, we find (Supplementary material)

$$\eta = \frac{\mathrm{SNR}^2}{4\gamma_\mathrm{m}} = \frac{\sigma_\mathrm{m}^2 a^2}{2} \tag{5.2}$$

provided two conditions are met. The conditions are: i) optimal integration functions [95, 96] are used to optimally extract information from both quadratures, and ii) the intra-resonator field vanishes at the beginning and end; i.e., photons are depleted from the resonator post-measurement.

To meet these conditions, we introduce a three-step experimental method. First, tuneup active photon depletion (or depletion by waiting) and calibration of the optimal integration weights. Second, obtain the measurement-induced dephasing of variable-strength weak measurement by including the pulse within a Ramsey sequence. Third, measure the SNR of variable-strength weak measurement from single-shot readout histograms.

5.3 Experimental setup

We test the method on a cQED test chip containing seven transmon qubits with dedicated readout resonators, each coupled to one of two feedlines (see Section 5.6). We present data for one qubit-resonator pair, but have verified the method with other pairs in this and other devices. The qubit is operated at its flux-insensitive point with a qubit frequency $f_\mathrm{q} = 5.070$ GHz, where the measured energy relaxation and echo dephasing times are $T_1 = 15$ μs and $T_{2,\mathrm{echo}} = 26$ μs, respectively. The resonator has a low-power fundamental at $f_{\mathrm{r},|0\rangle} = 7.852400$ GHz ($f_{\mathrm{r},|1\rangle} = f_{\mathrm{r},|0\rangle} + \chi/\pi = 7.852295$ GHz) for qubit in $|0\rangle$ ($|1\rangle$), with linewidth $\kappa/2\pi = 1.4$ MHz. The readout pulse generation and readout signal integration are performed by single-sideband mixing. Pulse-envelope generation, demodulation and signal processing are performed by a Zurich Instruments UHFLI-QC with 2 AWG channels and 2 ADC channels running at $1.8\ \mathrm{GSample/s}$ with $14-$ and $12-$bit resolution, respectively.

5.4 Experimental results

5.4.1 Extracting the quantum efficiency at the symmetry point

In the first step, we tune up the depletion steps and calibrate the optimal integration weights. We use a measurement ramp-up pulse of duration $\tau_{up} = 600$ ns, followed by a photon-depletion counter pulse [91, 113] consisting of two steps of 200 ns duration each, for a total depletion time $\tau_d = 400$ ns. To successfully deplete without relying on symmetries that are specific to a readout frequency at the midpoint between ground and excited state resonances (i.e., $\Delta = 0$), we vary 4 parameters of the depletion steps (details provided in the supplementary material). From the averaged transients of the finally obtained measurement pulse, we extract the optimal integration weights given by [95, 96] the difference between the averaged transients for $|0\rangle$ and $|1\rangle$ [Figure 5.1(a)]. The success of the active depletion is evidenced by the nulling at the end of τ_d. In this initial example, we connect to previous work by setting $\Delta = 0$, leaving all measurement information in one quadrature.

We next use the tuned readout to study its measurement-induced dephasing and SNR to finally extract η. We measure the dephasing by including the measurement-and-depletion pulse in a Ramsey sequence and varying its amplitude, ε [Figure 5.1(b)]. By varying the azimuthal angle of the second qubit pulse, we allow distinguishing dephasing from deterministic phase shifts and extract $|\rho_{01}|$ from the amplitude of the fitted Ramsey fringes. The SNR at various ε is extracted from single-shot readout experiments preparing the qubit in $|0\rangle$ and $|1\rangle$ [Figure 5.1(c)]. We use double Gaussian fits in both cases, neglecting measurement results in the spurious Gaussians to reduce corruption by imperfect state preparation and residual qubit excitation and relaxation. As expected, as a function of ε, $|\rho_{01}|$ decreases following a Gaussian form and the SNR increases linearly [Figure 5.1(d)]. The best fits to both dependencies give $\eta_e = 0.165 \pm 0.002$. Note that both dephasing and SNR measurements include ramp-up, depletion and an additional $\tau_{buffer} = 100$ ns, making the total measurement window $\tau_{int} = \tau_{up} + \tau_d + \tau_{buffer} = 1100$ ns.

5.4.2 Test the method: extracting the quantum efficiency in generalized conditions

We next demonstrate the generality of the method by extracting η as a function of the readout drive frequency. We repeat the method at fifteen readout drive detunings over a range of 2.8 MHz $\sim \kappa/\pi \sim 14\chi/\pi$ around $\Delta = 0$ [Figure 5.2(a,b)]. Furthermore, we compare the effect of using optimal weight functions versus square weight functions, and the effect of using active versus passive photon depletion. The square weight functions correspond to a single point in phase space during τ_{int}, with unit amplitude and an optimized phase maximizing SNR. We satisfy the zero-photon field condition by depleting the photons actively with $\tau_{int} = 1100$ ns (as in Figure 5.1) or passively by waiting with $\tau_{int} = 2100$ ns. When information is extracted from both quadratures using optimal weight functions, we measure an average $\eta_e = 0.167$ with 0.04 standard deviation. The extracted optimal integration functions in the time domain [Figure 5.2(c,d)] show how the resonator returns to the vacuum for both active and passive depletion. Square weight functions are not able to track the measurement dynamics in the time domain (even at $\Delta = 0$), leading to a reduction in η_e. Figures 5.2(e,f)

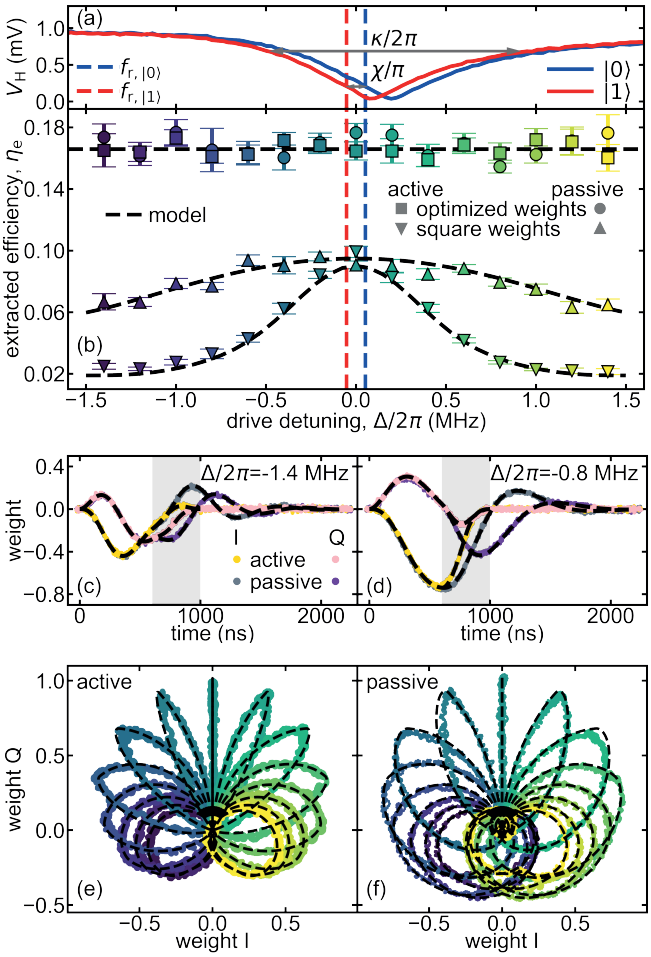


Figure 5.2: (a) Pulsed feedline transmission near the low-power resonator fundamentals. The qubit is prepared in $|0\rangle$ without (blue) and in $|1\rangle$ with a π pulse (red). The data fits $\kappa/2\pi = 1.4$ MHz and $f_{r,|0\rangle} = 7.852400$ GHz ($f_{r,|1\rangle} = 7.852295$ GHz), indicated by the dashed vertical lines. (b) Quantum efficiency extraction as a function of Δ using the pulse timings and three-step method of Figure 5.1. We use both the active depletion ($\tau_{\text{int}} = 1100$ ns) and passive depletion schemes ($\tau_{\text{int}} = 2100$ ns) and assess the benefit of optimal weights to standard square integration weights. (c,d) Optimal weight functions for I and Q at $\Delta/2\pi = -1.4$ MHz, -0.8 MHz [as in Figure 5.1(a)]. (e, f) Parametric plot of the optimal weight functions at all measured Δ [marker colors correspond to (a)]. Dashed black curves (b-f) are extracted from a linear model (see Section 5.6).

show the weight functions in phase space. The opening of the trajectories with detuning illustrates the rotating optimal measurement axis during measurement and leads to a further reduction of increase of η_e when square weight functions are used. The dynamics and the η_e

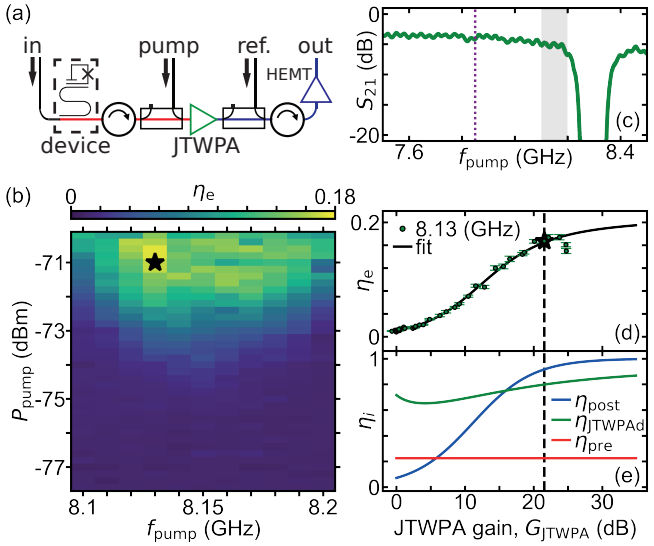


Figure 5.3: JTWPA pump tuneup to maximize the quantum efficiency and amplification chain modeling. (a) Simplified setup diagram, showing the input paths for the readout signal carrying the information on the qubit state and the added pump tone biasing the JTWPA. Both microwave tones are combined in the JTWPA amplifying the small readout signal. (b) η_e as a function of pump power and frequency. (c) CW low-power transmission of the JTWPA showing the dip in transmission due to the dispersion feature near 8.3 GHz and low-power insertion loss of $\sim$ 4.0 dB near $f_{r,|0\rangle}$ (dashed vertical line). The grey area indicates the frequency range of (b). S_{21} is obtained by measuring and comparing the output power when selecting the pump input or the reference input (input lines are duplicates and calibrated up to the directional couplers at room temperature). (d) Parametric plot of η_e at $f_{pump} = 8.13$ GHz and independently measured JTWPA gain. The fit (line) uses a three-stage model with $\eta(G_{JTWPA}) = \eta_{pre} \times \eta_{JTWPAd}(G_{JTWPA}) \times \eta_{post}(G_{JTWPA})$ [model details in the main text]. (e) Plots of the best-fit η_{pre}, $\eta_{JTWPAd}(G_{JTWPA})$ and $\eta_{post}(G_{JTWPA})$. The stars (b, d) and vertical dashed lines (d, e) indicate ($P_{pump} = -71.0$ dBm, $f_{pump} = 8.13$ GHz, $\eta = 0.1670$, $G_{JTWPA} = 21.6$ dB) used throughout the experiment.

dependence on Δ are excellently described by the linear model, which uses Equation (5.1), the separately calibrated κ and χ [Figure 5.2(a)] and $\eta = 0.1670$ (details in the supplementary material). Furthermore, the matching of the dynamics and depletion pulse parameters (see Section 5.6) when using active photon depletion confirm the numerical optimization techniques.

To further test the robustness of the method to arbitrary pulse envelopes, we have used a measurement-and-depletion pulse envelope $f(t)$ resembling a typical Dutch skyline. The pulse envelope outlines five canal houses, of which the first three ramp up the resonator and the latter two are used as the tunable depletion steps. Completing the three steps, we extract (see Section 5.6) $\eta_e = 0.167 \pm 0.005$, matching our previous value to within error.

5.4.3 Use the method: optimize TWPA biasing

We use the proven method to optimally bias the JTWPA and to quantify the different noise contributions in the readout chain. To this end, we map η_e as a function of pump power and frequency, just below the dispersive feature of the JTWPA, finding the maximum $\eta_e = 0.1670$ at ($P_{pump} = -71.0$ dBm, $f_{pump} = 8.13$ GHz) [Figures 5.3(a-c)]. We next compare the obtained η_e at the optimal bias frequency to independent low-power measurements of the JTWPA gain G_{JTWPA} we find $G_{JTWPA} = 21.6$ dB at the optimal bias point. We fit this parametric plot with a three-stage model, with noise contributions before, in and after the JTWPA, $\eta(G_{JTWPA}) = \eta_{pre} \times \eta_{JTWPAd}(G_{JTWPA}) \times \eta_{post}(G_{JTWPA})$. The parameter η_{pre} captures losses in the device and the microwave network between the device and the JTWPA and is therefore independent of G_{JTWPA}. The JTWPA has a distributed loss along the amplifying transmission line, which is modeled as an array of interleaved sections with quantum-limited amplification and sections with attenuation adding up to the total insertion loss of the JTWPA (as in [138]). Finally, the post-JTWPA amplification chain is modeled with a fixed noise temperature, whose relative noise contribution diminishes as G_{JTWPA} is increased. The best fit [Figures 5.3(d,e)] gives $\eta_{pre} = 0.22$, consistent with 50% photon loss due to symmetric coupling of the resonator to the feedline input and output, an attenuation of the microwave network between device and JTWPA of 2 dB and residual loss in the JTWPA of 27%. We fit a distributed insertion loss of the JTWPA of 4.6 dB, closely matching the separate calibration of 4.2 dB [Figure 5.3(c)]. Finally, we fit a noise temperature of 2.6 K, close to the HEMT amplifier's factory specification of 2.2 K.

We identify room for improving η_e to ~ 0.5 by implementing Purcell filters with asymmetric coupling [33, 75] (primarily to the output line) and decreasing the insertion loss in the microwave network, by optimizing the setup for shorter and superconducting cabling between device and JTWPA.

5.5 Conclusions

In conclusion, we have presented and demonstrated a general three-step method for extracting the quantum efficiency of linear dispersive qubit readout in cQED. We have derived analytically and demonstrated experimentally that the method robustly extracts the quantum efficiency for arbitrary readout conditions in the linear regime. This method will be used as a tool for readout performance characterization and optimization.

5.6 Modeling and experimental methods

This section provides additional sections and figures. In Section 5.6.1, we present details of the linear model we use to describe the resonator and qubit dynamics during linear dispersive readout. In Section 5.6.2, we describe how we evaluated these expressions to obtain the dashed lines in Figure 5.2, to which experimental results are compared. In Section 5.6.3, we show that Equation (5.2) follows from the linear model. Section 5.6.4 provides the cost function used for the optimization of depletion pulses. Figure 5.4 supplies the optimized depletion pulse parameters as a function of Δ and the SNR and coherence as a function of the drive amplitude and Δ. Figure 5.5 shows the extraction of η_e for an alternative pulse shape. Finally, Figure 6.6 provides a full wiring diagram and a photograph of the device.

5.6.1 Modeling of resonator dynamics and measurement signal

In this section, we give the expressions that model the resonator dynamics and measured signal in the linear dispersive regime.

In general, the measured homodyne signal consists of in-phase (I) and in-quadrature (Q) components, given by [100]

$$\begin{aligned} V_{\mathrm{I},|i\rangle}(t) &= V_0 \left(\sqrt{2\kappa\eta}\mathrm{Re}(\alpha_{|i\rangle}(t)) + n_{\mathrm{I}}(t) \right), \\ V_{\mathrm{Q},|i\rangle}(t) &= V_0 \left(\sqrt{2\kappa\eta}\mathrm{Im}(\alpha_{|i\rangle}(t)) + n_{\mathrm{Q}}(t) \right). \end{aligned} \tag{5.3}$$

Here, V_0 is an irrelevant gain factor and n_{I}, n_{Q} are continuous, independent Gaussian white noise terms with unit variance, $\langle n_j(t) n_k(t') \rangle = \delta_{jk}\delta(t-t')$, while the internal resonator field $\alpha_{|i\rangle}$ follows Equation (5.1) for $i \in \{0, 1\}$. In the shunt resonator arrangement used on the device for this work, the measured signal also includes an additional term describing the directly transmitted part of the measurement pulse. We omitted this term here, as it is independent of the qubit state, and thus is irrelevant for the following, as we will exclusively encounter the signal difference $V_{\mathrm{int},|1\rangle} - V_{\mathrm{int},|0\rangle}$.

For state discrimination, the homodyne signals are each multiplied with weight functions, given by the difference of the averaged signals, then summed and integrated over the measurement window of duration T:

$$V_{\mathrm{int},|i\rangle} = \int_0^T w_{\mathrm{I}} V_{\mathrm{I},|i\rangle} + w_{\mathrm{Q}} V_{\mathrm{Q},|i\rangle}\, dt. \tag{5.4}$$

The optimal weight functions [95, 96] are given by the difference of the average signal

$$w_{\mathrm{I/Q}} = \langle V_{\mathrm{I/Q},|1\rangle} - V_{\mathrm{I/Q},|0\rangle} \rangle. \tag{5.5}$$

As an alternative to optimal weight functions, often constant weight functions are used

$$w_{\mathrm{I}} = \cos\phi_w, \quad w_{\mathrm{Q}} = \sin\phi_w, \tag{5.6}$$

where the demodulation phase ϕ_w is usually chosen as to maximize the SNR (see below).

We define the signal S as the absolute separation between the average V_{int} for $|1\rangle$ and $|0\rangle$. In turn, we define the noise N as the standard deviation of $V_{\mathrm{int},|i\rangle}$, which is independent of $|i\rangle$. Thus,

$$S = \left| \langle V_{\mathrm{int},|1\rangle} - V_{\mathrm{int},|0\rangle} \rangle \right|,$$
$$N^2 = \langle V_{\mathrm{int}}^2 \rangle - \langle V_{\mathrm{int}} \rangle^2.$$

The signal-to-noise ratio SNR is then given as

$$\mathrm{SNR} = \frac{S}{N}. \tag{5.7}$$

The measurement pulse leads to measurement-induced dephasing. Experimentally, the dephasing can be quantified by including the measurement pulse in a Ramsey sequence. The coherence elements of the qubit density matrix are reduced due to the pulse as [100]

$$|\rho_{01}(\varepsilon)| = \mathrm{e}^{-\gamma_{\mathrm{m}}} |\rho_{01}(\varepsilon = 0)|,$$

where

$$\gamma_{\mathrm{m}} = 2\chi \int_0^T \mathrm{Im}(\alpha_{|0\rangle} \alpha^*_{|1\rangle}) dt. \tag{5.8}$$

Thus, γ_{m} scales with ε^2, and the coherence elements decay as a Gaussian in ε.

5.6.2 Comparison of experiment and model

We here describe how we compared the theoretical model given by the previous section and Equation (5.1) to the experimental data as presented in Figure 5.2.

In panels (c)-(f) of Figure 5.2, we compare the measured weight functions to a numerical evaluation of Equation (5.1). The dashed lines in those panels are obtained by numerically integrating Equation (5.1), using the ϵ and Δ applied in experiment, and with the resonator parameters κ and χ that are obtained from resonator spectroscopy [presented in panel (a)]. From the resulting $\alpha_{|i\rangle}$ we then evaluate Equations (5.3) and (5.5) to obtain $w_{\mathrm{I/Q}}$, presented in panels (c)-(f). The scale factor V_0 was chosen to best represent the experimental data.

In order to model the data presented in panel (b), we further inserted the $\alpha_{|i\rangle}$ into Equations (5.7) and (5.8), and finally into Equation (5.2) to obtain η_{e}. This step is performed for both optimal weights and constant weights, Equations (5.5) and (5.6). As shown in Figure 5.2, the result depends on pulse shape and Δ when using square weights, but does not when using optimal weights. The value for η in Equation (5.3) is chosen as the average of η_{e} for optimal weight functions, $\eta_{\mathrm{e}} = 0.167$.

5.6.3 Derivation of Equation (5.2)

With the definitions of the previous sections, we now show that (5.2) holds for arbitrary pulses and resonator parameters if optimal weight functions are used, so that η_{e} in Figure 5.2 indeed coincides with η in Equation (5.3).

Using optimal weight functions, we can evaluate Equation (5.7) in terms of $\alpha_{|i\rangle}$ by inserting Equations (5.5) and (5.3), obtaining for the signal S:

$$S_{\text{opt}} = 2\kappa\eta V_0^2 \int_0^T \left|\alpha_{|1\rangle} - \alpha_{|0\rangle}\right|^2 dt.$$

For the noise N, we obtain

$$\begin{aligned} N_{\text{opt}}^2 &= V_0^2 \left\langle \int_0^T \left(w_{\text{I}} n_{\text{I}} + w_{\text{Q}} n_{\text{Q}}\right)^2 dt \right\rangle \\ &= 2\kappa\eta V_0^4 \int_0^T \left|\alpha_{|1\rangle} - \alpha_{|0\rangle}\right|^2 dt, \end{aligned}$$

where we used the white noise property of $n_{\text{I,Q}}(t)$.

The SNR is then given by

$$\text{SNR}_{\text{opt}} = \frac{S_{\text{opt}}}{N_{\text{opt}}} = \sqrt{2\kappa\eta \int_0^T \left|\alpha_{|1\rangle} - \alpha_{|0\rangle}\right|^2 dt}. \tag{5.9}$$

Note that the $\alpha_{|i\rangle}$ scale linearly with the amplitude ε due to the linearity of Equation (5.1), so that the SNR scales linearly with ε as well.

We now show that the γ_{m} and SNR are related by Equation (5.2), independent of resonator and pulse parameters. For that, we need to make use of constraint (ii), namely that the resonator fields $\alpha_{|i\rangle}$ vanish at the beginning and end of the integration window. We then can write

$$\begin{aligned} 0 &= \left[\left|\alpha_{|0\rangle} - \alpha_{|1\rangle}\right|^2\right]_0^T \\ &= \int_0^T \partial_t \left|\alpha_{|0\rangle} - \alpha_{|1\rangle}\right|^2 dt \\ &= 2\int_0^T \text{Re}\left((\alpha_{|1\rangle}^* - \alpha_{|0\rangle}^*)\partial_t(\alpha_{|1\rangle} - \alpha_{|0\rangle})\right) dt, \end{aligned}$$

where the first equality is ensured by requirement (ii), and the second equality follows from rewriting as the integral of a differential.

We insert the differential equation (5.2) into this expression, obtaining

$$\begin{aligned} &\text{Re}\int_0^T \left(\alpha_{|1\rangle}^* - \alpha_{|0\rangle}^*\right) \times \\ &\left(\left(-i\Delta - \frac{\kappa}{2}\right)\left(\alpha_{|1\rangle} - \alpha_{|0\rangle}\right) - i\chi\left(\alpha_{|1\rangle} + \alpha_{|0\rangle}\right)\right) dt = 0. \end{aligned}$$

Isolating the κ term and dropping purely imaginary Δ and χ terms, we obtain

$$\begin{aligned}
&\frac{\kappa}{2}\int_0^T \left|\alpha_{|1\rangle}-\alpha_{|0\rangle}\right|^2 dt \\
=&-\mathrm{Re}\left(i\chi\int_0^T \left(\alpha_{|1\rangle}+\alpha_{|0\rangle}\right)(\alpha^*_{|1\rangle}-\alpha^*_{|0\rangle})dt\right) \\
=&-\mathrm{Re}\left(i\chi\int_0^T \left(|\alpha_{|1\rangle}|^2-|\alpha_{|0\rangle}|^2+2i\mathrm{Im}(\alpha_{|0\rangle}\alpha^*_{|1\rangle})\right)dt\right) \\
=&2\chi\int_0^T \mathrm{Im}(\alpha_{|0\rangle}\alpha^*_{|1\rangle})dt.
\end{aligned}$$

Comparing the first and last line with Equations (5.9) and (5.8), respectively, this equality shows indeed that the SNR, when defined with optimal integration weights, and the measurement-induced dephasing γ_m are related by Equation (5.2), independent of the resonator parameters κ, χ, and the functional form $\varepsilon f(t)$ of the drive.

5.6.4 Depletion tuneup

Here, we provide details on the depletion tuneup. The depletion is tuned by optimizing the amplitude and phase of both depletion steps (Figure 5.4) using the Nelder-Mead algorithm with a cost function that penalizes non-zero averaged transients for both $|0\rangle$ and $|1\rangle$ during a $\tau_\mathrm{c} = 200$ ns time window after the depletion. The transients are obtained by preparing the qubit in $|0\rangle$ ($|1\rangle$) and averaging the time-domain homodyne voltages $V_{\mathrm{I},|0\rangle}$ and $V_{\mathrm{Q},|0\rangle}$ ($V_{\mathrm{I},|1\rangle}$ and $V_{\mathrm{Q},|1\rangle}$) of the transmitted measurement pulse for 2^{15} repetitions. The cost function consists of four different terms. The first two null the transients in both quadratures post-depletion. The last two additionally penalize the difference between the transients for $|0\rangle$ and $|1\rangle$ with a tunable factor d. In the experiment, we found reliable convergence of the depletion tuneup for $d = 10$.

$$\begin{aligned}
cost = &\sqrt{\int_{\tau_\mathrm{up}+\tau_\mathrm{d}}^{\tau_\mathrm{up}+\tau_\mathrm{d}+\tau_\mathrm{c}} \langle V_{\mathrm{I},|0\rangle}(t)\rangle^2+\langle V_{\mathrm{Q},|0\rangle}(t)\rangle^2 dt} \\
&+\sqrt{\int_{\tau_\mathrm{up}+\tau_\mathrm{d}}^{\tau_\mathrm{up}+\tau_\mathrm{d}+\tau_\mathrm{c}} \langle V_{\mathrm{I},|1\rangle}(t)\rangle^2+\langle V_{\mathrm{Q},|1\rangle}(t)\rangle^2 dt} \\
&+d\sqrt{\int_{\tau_\mathrm{up}+\tau_\mathrm{d}}^{\tau_\mathrm{up}+\tau_\mathrm{d}+\tau_\mathrm{c}} \langle V_{\mathrm{I},|1\rangle}(t)-V_{\mathrm{I},|0\rangle}(t)\rangle^2 dt} \\
&+d\sqrt{\int_{\tau_\mathrm{up}+\tau_\mathrm{d}}^{\tau_\mathrm{up}+\tau_\mathrm{d}+\tau_\mathrm{c}} \langle V_{\mathrm{Q},|1\rangle}(t)-V_{\mathrm{Q},|0\rangle}(t)\rangle^2 dt}.
\end{aligned}$$

In Figures 5.4(b,c), we show the obtained depletion pulse parameters for different values of Δ. As a comparison, we show the parameters that are predicted by numerically integrating Equation (5.1), with resonator parameters extracted from Figure 5.2(a), and numerically finding the depletion pulse parameters that lead to $\alpha_{|0,1\rangle}(T) = 0$.

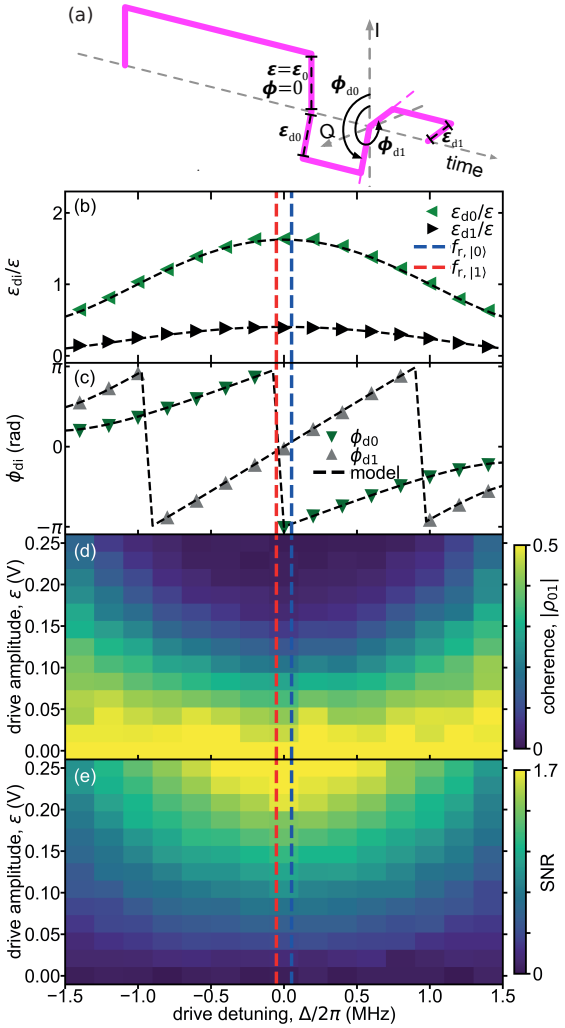


Figure 5.4: Depletion pulse parameters, coherence and SNR as a function of detuning. (a) The measurement pulse consists of a ramp-up of duration $\tau_{\text{up}} = 600$ ns, fixed phase $\phi = 0$ and amplitude ε (fixed during tuneup to $\varepsilon = \varepsilon_0 = 0.25$ V) and two 200 ns depletion segments ($\tau_{\text{d}} = 400$ ns) with each a tunable phase (φ_{d0}, φ_{d1}) and amplitude (ε_{d0}, ε_{d1}). (b,c) Depletion pulse parameters from the depletion optimizations used in Figure 5.2. Dashed vertical lines indicate $f_{\text{r},|0\rangle}$ (blue) and $f_{\text{r},|1\rangle}$ (red). Dashed black curves are extracted from the linear model (see Section 5.6.4). Coherence (d) and SNR (e) as a function of drive amplitude and detuning. At non-zero ε, SNR is maximal (coherence is minimal) at the midpoint frequency $\Delta = 0$ and decreases (increases) with detuning.

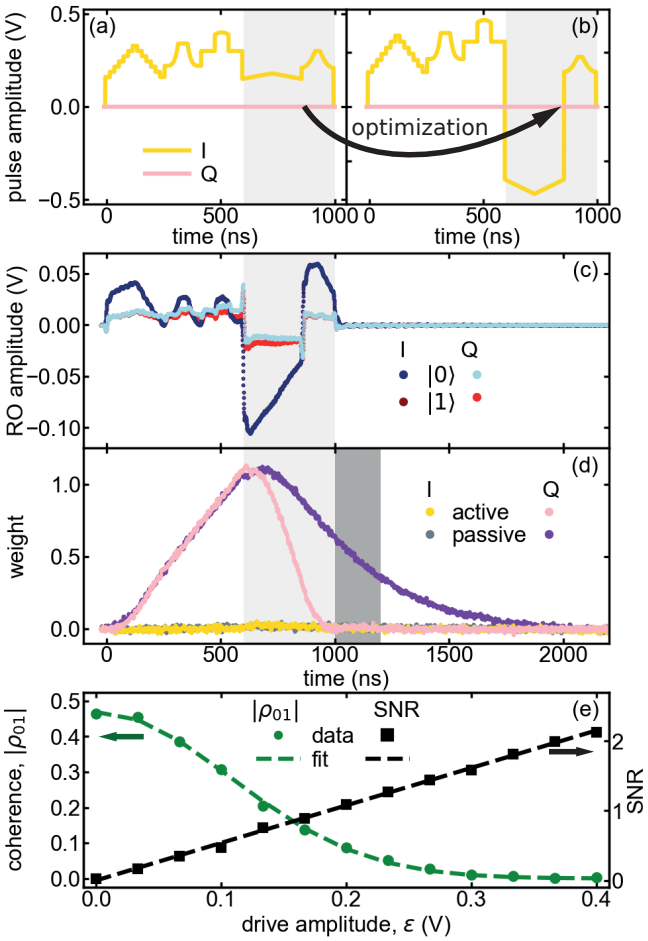


Figure 5.5: The three-step method for quantum efficiency extraction with a pulse envelope consisting of seventeenth-century Dutch canal house façade outlines. (a) Pulse envelope with five façades, of which the first three ramp up the resonator with duration $\tau_{\mathrm{up}} = 600$ ns, fixed phase $\varphi = 0$ and amplitude ε (fixed during tuneup to $\varepsilon = \varepsilon_0 = 0.4$ V) and the last two are 240 ns and 160 ns depletion segments ($\tau_{\mathrm{d}} = 400$ ns) with each a tunable phase and amplitude. (b) Optimized depletion pulse with $\varepsilon_{\mathrm{d0}} = 1.68\varepsilon$, $\varepsilon_{\mathrm{d1}} = 0.58\varepsilon$, $\varphi_{\mathrm{d0}} = 1.005\pi$ rad, $\varphi_{\mathrm{d1}} = 0.007\pi$ rad. (c) Averaged feedline transmission of the optimized depletion pulse. The qubit is prepared in $|0\rangle$ (blue) and in $|1\rangle$ (red). (d) Optimal weight functions extracted for the depletion pulse (purple) and as a reference, weight functions are shown for passive depletion ($\varepsilon_{\mathrm{d0}} = \varepsilon_{\mathrm{d1}} = 0$ V). (d) Quantum efficiency extraction using 13 values of ε. The best-fit values give $\eta_{\mathrm{e}} = 0.167 \pm 0.005$.

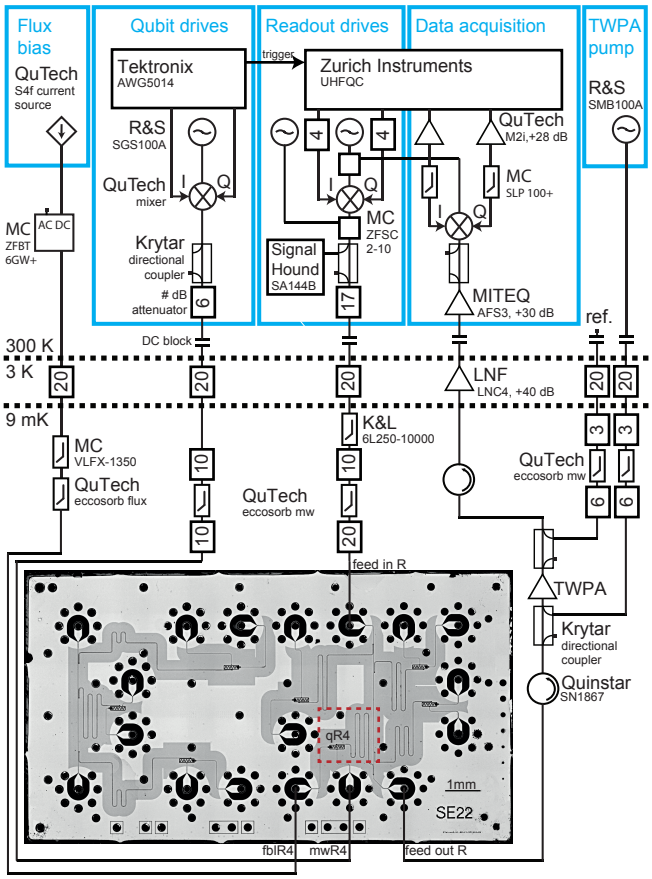
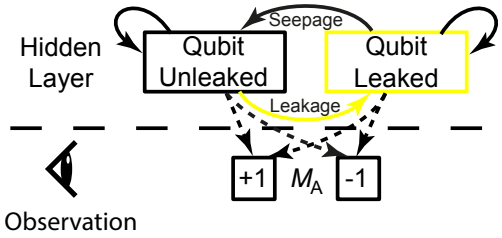


Figure 5.6: Photograph of cQED chip (identical design as the one used) and complete wiring diagram of electronic components inside and outside the ^{3}He/^{4}He dilution refrigerator (Leiden Cryogenics CF-CS81). The test chip contains seven transmon qubits individually coupled to dedicated microwave drive lines, flux bias lines and readout resonators. The three (four) resonators on the left (right) side couple capacitively to the left (right) feedline traversing the chip from top to bottom. All 18 connections are made from the back side of the chip and reach the front through vertical coax lines [21]. Each vertical coax line consists of a central through-silicon via (TSV) that carries the signal and seven surrounding TSVs acting as shield connecting the front and back side ground planes. Other, individual TSVs interconnect front side and back side ground planes to eliminate chip modes.

6

PROTECTING QUANTUM ENTANGLEMENT FROM LEAKAGE AND QUBIT ERRORS VIA REPETITIVE PARITY MEASUREMENTS



Protecting quantum information from errors is essential for large-scale quantum computation. Quantum error correction (QEC) encodes information in entangled states of many qubits, and performs parity measurements to identify errors without destroying the encoded information. However, traditional QEC cannot handle leakage from the qubit computational space. Leakage affects leading experimental platforms, based on trapped ions and superconducting circuits, which use effective qubits within many-level physical systems. We investigate how two-transmon entangled states evolve under repeated parity measurements, and demonstrate the use of hidden Markov models to detect leakage using only the record of parity measurement outcomes required for QEC. We show the stabilization of Bell states over up to 26 parity measurements by mitigating leakage using postselection, and correcting qubit errors using Pauli-frame transformations. Our leakage identification method is computationally efficient and thus compatible with real-time leakage tracking and correction in larger quantum processors.

This chapter has been published in *Science Advances* **6**, 12 (2020).

6.1 Introduction

Large-scale quantum information processing hinges on overcoming errors from environmental noise and imperfect quantum operations. Fortunately, the theory of QEC predicts that the coherence of single degrees of freedom (logical qubits) can be better preserved by encoding them in ever-larger quantum systems (Hilbert spaces), provided the error rate of the constituent elements lies below a fault-tolerance threshold [101]. Experimental platforms based on trapped ions and superconducting circuits have achieved error rates in single-qubit gates [30, 147, 148], two-qubit gates [30, 148, 149], and qubit measurements [33, 76, 113, 147] at or below the threshold for popular QEC schemes such as surface [18, 19] and color codes [150]. They therefore seem well poised for the experimental pursuit of quantum fault tolerance. However, a central assumption of textbook QEC, that error processes can be discretized into bit flips (X), phase flips (Z) or their combination ($Y = iXZ$) only, is difficult to satisfy experimentally. This is due to the prevalent use of many-level systems as effective qubits, such as hyperfine levels in ions and weakly anharmonic transmons in superconducting circuits, making leakage from the two-dimensional computational space of effective qubits a threatening error source. In quantum dots and trapped ions, leakage events can be as frequent as qubit errors [151, 152]. However, even when leakage is less frequent than qubit errors as in superconducting circuits [30, 149], if ignored, leakage can produce the dominant damage to encoded logical information. To address this, theoretical studies propose techniques to reduce the effect of leakage by periodically moving logical information, and removing leakage when qubits are free of logical information [153–156]. Alternatively, more hardware-specific solutions have been proposed for trapped ions [157] and quantum dots [158]. In superconducting circuits, recent experiments have demonstrated single- and multi-round parity measurements to correct qubit errors with up to 9 physical qubits [34–36, 62, 78, 145, 159–161]. Parallel approaches encoding information in the Hilbert space of single resonators using cat [81] and binomial codes [162] used transmon-based photon-parity checks to approach the break-even point for a quantum memory. However, no experiment has demonstrated the ability to detect and mitigate leakage in a QEC context.

In this report, we experimentally investigate leakage detection and mitigation in a minimal QEC system. Specifically, we protect an entangled state of two transmon data qubits (Q_{DH} and Q_{DL}) from qubit errors and leakage during up to 26 rounds of parity measurements via an ancilla transmon (Q_{A}). Performing these parity checks in the Z basis protects the state from X errors, while interleaving checks in the Z and X bases protects it from general qubit errors (X, Y and Z). Leakage manifests itself as a round-dependent degradation of data-qubit correlations ideally stabilized by the parity checks: $\langle Z \otimes Z \rangle$ in the first case and $\langle X \otimes X \rangle$, $\langle Y \otimes Y \rangle$, and $\langle Z \otimes Z \rangle$ in the second. We introduce hidden Markov models (HMMs) to efficiently detect data-qubit and ancilla leakage, using only the string of parity outcomes, demonstrating restoration of the relevant correlations. Although we use postselection here, the low technical overhead of HMMs makes them ideal for real-time leakage correction in larger QEC codes.

6.2 Results

6.2.1 A mimimal QEC setup

Repetitive parity checks can produce and stabilize two-qubit entanglement. For example, performing a $Z\otimes Z$ parity measurement (henceforth a ZZ check) on two data qubits prepared in the unentangled state $|{+}{+}\rangle = (|0\rangle + |1\rangle)\otimes(|0\rangle + |1\rangle)/2$ will ideally project them to either of the two (entangled) Bell states $\left|\Phi^+\right\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ or $\left|\Psi^+\right\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$, as signaled by the ancilla measurement outcome M_A. Further ZZ checks will ideally leave the entangled state unchanged. However, qubit errors will alter the state in ways that may or may not be detectable and/or correctable. For instance, a bit-flip (X) error on either data qubit, which transforms $\left|\Phi^+\right\rangle$ into $\left|\Psi^+\right\rangle$, will be detected because X anti-commutes with a ZZ check. The corruption can be corrected by applying a bit flip on either data qubit because this cancels the original error ($X^2 = I$) or completes the operation $X\otimes X$, of which $\left|\Phi^+\right\rangle$ and $\left|\Psi^+\right\rangle$ are both eigenstates. The correction can be applied in real time using feedback [62, 78, 159, 163] or kept track of using Pauli frame updating (PFU) [35, 164]. We choose the latter, with PFU strategy "X on Q_DH". Phase-flip errors are not detectable since Z on either data qubit commutes with a ZZ check. Such errors transform $\left|\Phi^+\right\rangle$ into $\left|\Phi^-\right\rangle = (|00\rangle - |11\rangle)/\sqrt{2}$ and $\left|\Psi^+\right\rangle$ into $\left|\Psi^-\right\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$. Finally, Y errors produce the same signature as X errors. Our PFU strategy above converts them into Z errors. Crucially, by interleaving checks of type ZZ and XX (measuring $X\otimes X$), arbitrary qubit errors can be detected and corrected. The ZZ check will signal either X or Y error, and the XX check will signal Z or Y, providing a unique signature in combination.

6.2.2 Generating entanglement by measurement

Our parity check is an indirect quantum measurement involving coherent interactions of the data qubits with Q_A and subsequent Q_A measurement [85] (Figure 6.1A). The coherent step maps the data-qubit parity onto Q_A in 120 ns using single-qubit (SQ) and two-qubit controlled-phase (CZ) gates [149]. Gate characterizations Table 6.1 indicate state-of-the-art gate errors $e_\mathrm{SQ} = \{0.08 \pm 0.02, 0.14 \pm 0.016, 0.21 \pm 0.06\}\%$ and $e_\mathrm{CZ} = \{1.4 \pm 0.6, 0.9 \pm 0.16\}\%$ with leakage per CZ $L_1 = \{0.27 \pm 0.12, 0.15 \pm 0.07\}\%$. We measure Q_A with a 620-ns pulse including photon depletion [91, 113], achieving an assignment error $e_\mathrm{a} = 1.0 \pm 0.1\%$. We avoid data-qubit dephasing during the Q_A measurement by coupling each qubit to a dedicated readout resonator and a dedicated Purcell filter [76] (Figure 6.5). The parity check has a cycle time of 740 ns, corresponding to only $2.5 \pm 0.2\%$ and $5.0 \pm 0.3\%$ of the data-qubit echo dephasing times Table 6.1.

The parity measurement performance can be quantified by correlating its outcome with input and output states. We first quantify the ability to distinguish even- ($|00\rangle$, $|11\rangle$) from odd-parity ($|01\rangle$, $|10\rangle$) data-qubit input states, finding an average parity assignment error $e_\mathrm{a,ZZ} = 5.1 \pm 0.2\%$. Second, we assess the ability to project onto the Bell states by performing a ZZ check on $|{+}{+}\rangle$ and reconstructing the most-likely physical data-qubit output density matrix ρ, conditioning on $M_\mathrm{A} = \pm 1$. When tomographic measurements are performed

simultaneously with the Q_A measurement, we find Bell-state fidelities $F_{|\Phi^+\rangle|M_A=+1} = \langle\Phi^+|\rho_{M_A=+1}|\Phi^+\rangle = 94.7 \pm 1.9\%$ and $F_{|\Psi^+\rangle|M_A=-1} = 94.5 \pm 2.5\%$ (Figure 6.1, C and D). We connect $|\Psi^+\rangle$ to $|\Phi^+\rangle$ by incorporating the PFU into the tomographic analysis, obtaining $F_{|\Phi^+\rangle} = 94.6 \pm 0.9\%$ without any postselection (Figure 6.1E). The nondemolition character of the ZZ check is then validated by performing tomography only once the Q_A measurement completes. We include an echo pulse on both data qubits during the Q_A measurement to reduce intrinsic decoherence and negate residual coupling between data qubits and Q_A (Figure 6.7). The degradation to $F_{|\Phi^+\rangle} = 91.8 \pm 0.5\%$ is consistent with intrinsic data-qubit decoherence under echo and confirms that measurement-induced errors are minimal.

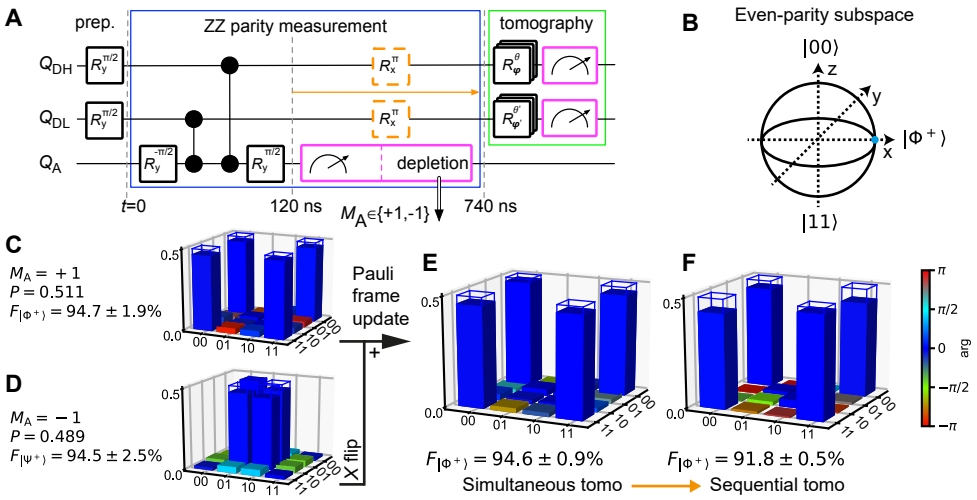


Figure 6.1: Entanglement genesis by ZZ parity measurement and Pauli frame update. (**A**) Quantum circuit for a parity measurement of the data qubits via coherent operations with ancilla Q_A and Q_A measurement. Tomography reconstructs the data-qubit output density matrix (ρ). Echo pulses (orange) are applied halfway the Q_A measurement when performing tomography sequential to the Q_A measurement. (**B**) Bloch-sphere representation of the even-parity subspace with a marker on $|\Phi^+\rangle$. (**C** to **F**) Plots of ρ with fidelity to the Bell states (indicated by frames) for tomography simultaneous with Q_A measurement (C to E) and sequential to Q_A measurement (F). (C)[(D)] Conditioning on $M_A = +1[-1]$ ideally generates $|\Phi^+\rangle$ [$|\Psi^+\rangle$] with equal probability P. (E)[(F)] PFU applies bit-flip correction (X on Q_{DH}) for $M_A = -1$ and reconstructs ρ using all data for simultaneous [sequential] tomography.

6.2.3 Protecting entanglement from bit flips and the observation of leakage

QEC stipulates repeated parity measurements on entangled states. We therefore study the evolution of $F_{|\Phi^+\rangle} = (1 + \langle X\otimes X\rangle - \langle Y\otimes Y\rangle + \langle Z\otimes Z\rangle)/4$ and its constituent correlations as a function of the number M of checks (Figure 6.2A). When performing PFU using the

first ZZ outcome only (ignoring subsequent outcomes), we observe that $F_{|\Phi^+\rangle}$ witnesses entanglement (> 0.5) during 10 rounds and approaches randomization (0.25) by $M = 25$ (Figure 6.2B). The constituent correlations also decay with simple exponential forms. A best fit of the form $\langle Z\otimes Z\rangle[M] = a \cdot e^{-M/\upsilon_{ZZ}} + b$ gives a decay time $\upsilon_{ZZ} = 9.0 \pm 0.9$ rounds; similarly, we extract $\upsilon_{XX} = 11.7 \pm 1.0$ rounds (Figure 6.2, C and D). By comparison, we observe that Bell states evolving under dynamical decoupling only (no ZZ checks, see Figure 6.8) decay similarly ($\upsilon_{ZZ} = 8.6 \pm 0.3$, $\upsilon_{XX} = 12.8 \pm 0.4$ rounds). These similarities indicate that intrinsic data-qubit decoherence is also the dominant error source in this multi-round protocol.

To demonstrate the ability to detect X and Y but not Z errors, we condition the tomography on signaling no errors during M rounds. This boosts $\langle Z\otimes Z\rangle$ to a constant, while the undetectability of Z errors only allows slowing the decay of $\langle X\otimes X\rangle$ to $\upsilon_{XX} = 33.2 \pm 1.7$ rounds (and of $\langle Y\otimes Y\rangle$ to $\upsilon_{YY} = 31.3 \pm 1.9$ rounds). Naturally, this conditioning comes at the cost of the postselected fraction f_{post} reducing with M (Figure 6.9).

Moving from error detection to correction, we consider the protection of $|\Phi^+\rangle$ by tracking X errors and applying corrections in post-processing. The correction relies on the final two M_A only, concluding even parity for equal measurement outcomes and odd parity for unequal. For this small-scale experiment, this strategy is equivalent to a decoder based on minimum-weight perfect matching (MWPM) [19, 38], justifying its use. Because our PFU strategy converts Y errors into Z errors, one expects a faster decay of $\langle X\otimes X\rangle$ compared to the no-error conditioning; indeed, we observe $\upsilon_{XX} = 11.8 \pm 1.0$ rounds. Most importantly, correction should lead to a constant $\langle Z\otimes Z\rangle$. While $\langle Z\otimes Z\rangle$ is clearly boosted, a weak decay to a steady state $\langle Z\otimes Z\rangle = 0.73 \pm 0.03$ is also evident (Figure 6.2D). As previously observed in Ref. [163], this degradation is the hallmark of leakage [see also [35, 159]]. We additionally compare the experimental results to simulations using a model that assumes ideal two-level systems [38] (no leakage) based on independently calibrated parameters of Table 6.1 (Figure 6.12 A to D). At $M = 1$ model and experiment coincide for all correction strategies. At larger M 'first' and 'final' correction strategies deviate significantly, consistent with a gradual build-up of leakage, which we now turn our focus to.

6.2.4 Leakage detection using hidden Markov models

Both ancilla and data-qubit leakage in our experiment can be inferred from a string $\vec{M}_A = (M_A[m = 0], \ldots, M_A[m = M])$ of measurement outcomes. Leakage of Q_A to the second excited transmon state $|2\rangle$ produces $M_A = -1$ because measurement does not discern it from $|1\rangle$. This leads to the pattern $\vec{M}_A = (\ldots -1, -1\ldots)$ until Q_A seeps back to $|1\rangle$ (coherently or by relaxation), as it is unaffected by subsequent $\pi/2$ rotations (Figure 6.3C). Leakage of a data qubit (Figure 6.3B) leads to apparent repeated errors [signaled by $\vec{M}_A = (\ldots +1, +1, -1, -1\ldots)$], as the echo pulses only act on the unleaked qubit. This is equivalent to a pattern of repeated error signals in the data-qubit syndrome $s_D[m] := M_A[m] \cdot M_A[m-2]$ — $s_D = (\ldots, -1, -1, -1, \ldots)$. (We call $s_D[m] = -1$ an

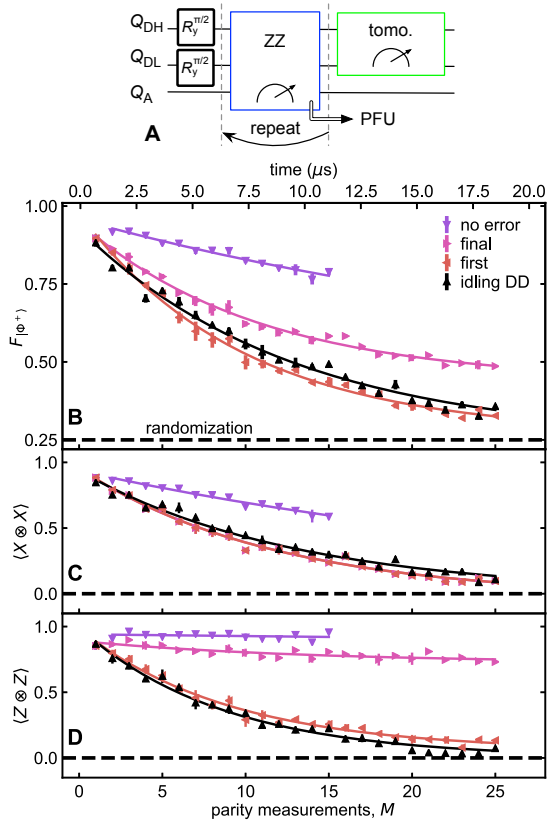


Figure 6.2: Protecting entanglement from bit flips with repeated ZZ checks. (**A**) The quantum circuit of Figure 6.1A extended with M rounds of repeated ZZ checks. (**B**) Fidelity to $|\Phi^+\rangle$ as a function of M. 'No error' postselects the runs in which no bit flip is detected. 'Final' applies PFU based on the last two outcomes (equivalent to minimum-weight perfect matching). 'First' uses the first parity outcome only. 'Idling DD' are Bell states evolving under dynamical decoupling only (quantum circuit in Figure 6.8). (**C**) Corresponding $\langle X \otimes X \rangle$. 'final' coincides with 'first'. (**D**) Corresponding $\langle Z \otimes Z \rangle$. The weak degradation observed for 'final' is the hallmark of leakage. Curves in (B to D) are best fits of a simple exponential decay.

error signal as in the absence of noise $s_D[m] = +1$, while the measurements $M_A[m]$ will still depend on the ZZ parity.)

Neither of the above patterns is entirely unique to leakage; each may also be produced by some combination of qubit errors. Therefore, we cannot unambiguously diagnose an individual experimental run of corruption by leakage. However, given a set of ancilla measurements $M_A[0], \ldots, M_A[m]$, the likelihood $L_{\text{comp,Q}}(\vec{M}_A)$ that qubit Q is in the computational subspace during the final parity checks is well-defined. In this work, we infer $L_{\text{comp,Q}}(\vec{M}_A)$ by using a hidden Markov model (HMM) [165], which treats the system as leaking out of and seeping back to the computational subspace in a stochastic fashion between each mea-

surement round (a leakage HMM in its simplest form is shown in Figure 6.3A, and further described in Sections 6.5.1 to 6.5.3). This may be extended to scalable leakage detection (for the purposes of leakage mitigation) in a larger QEC code, by using a separate HMM for each data qubit and ancilla. To improve the validity of the HMMs, we extend their internal states to allow the modeling of additional noise processes in the experiments (detailed in Sections 6.5.4 and 6.5.5).

Before assessing the ability of our HMMs to improve fidelity in a leakage mitigation scheme, we first validate and benchmark them internally. A common method to validate the HMM's ability to model the experiment is to compare statistics of the experimentally-generated data to a simulated data set generated by the model itself. As we are concerned only with the ability of the HMM to discriminate leakage, $L_{\mathrm{comp,Q}}(\vec{M}_A)$ provides a natural metric for comparison. In Figure 6.3, D and E, we overlay histograms of 10^5 experimental and simulated experiments, binned according to $L_{\mathrm{comp,Q}}(\vec{M}_A)$, and observe excellent agreement. To further validate our model, we calculate the Akaike information criterion [166]:

$$A(H) = 2n_{\mathrm{p},H} - 2\log\left[\max_{p_i} L(\{\vec{o}\}|H\{p_i\})\right], \tag{6.1}$$

where $L(\vec{o}|M)$ is the likelihood of making the set of observations $\{\vec{o}\}$ given model H (maximized over all parameters p_i in the model, as listed in Table 6.2.), and $n_{\mathrm{p},M}$ is the number of parameters p_i. The number $A(H)$ is rather meaningless by itself; we require a comparison model $H^{(\mathrm{comp})}$ for reference. Our model is preferred over the comparison model whenever $A(H) > A(H^{(\mathrm{comp})})$. For comparison, we take the target HMM H, remove all parameters describing leakage, and re-optimize. We find the difference $A(H)-A(H^{(\mathrm{comp})}) = 1.1\times10^5$ for the data-qubit HMM, and 2.1×10^4 for the ancilla HMM, giving significant preference for the inclusion of leakage in both cases. [The added internal states beyond the simple two-state HMMs clearly improves the overlap in histograms, Figure 6.14, A and B. The added complexity is further justified by the Akaike information criterion , see Section 6.5.6].

The above validation suggests that we may assume that the ratio of actual leakage events at a given $L_{\mathrm{comp,Q}}$ is well approximated by $L_{\mathrm{comp,Q}}$ itself (which is true for the simulated data). Under this assumption, we expose the HMMs discrimination ability by plotting its receiver operating characteristic [167] (ROC). The ROC (Figure 6.3F) is a parametric plot (sweeping a threshold $L^{\mathrm{th}}_{\mathrm{comp,Q}}$) of the true positive rate TPR (the fraction of leaked runs correctly identified) versus the false positive rate FPR (the fraction of unleaked runs wrongly identified). Random rejection follows the line $y = x$; the better the detection the greater upward shift. Both ROCs indicate that most of the leakage ($\mathrm{TPR} = 0.7$) can be efficiently removed with $\mathrm{FPR} \sim 0.1$. Individual mappings of TPR and FPR as a function of $L^{\mathrm{th}}_{\mathrm{comp,Q}}$ can be found in Figure 6.13, A and B. Further rejection is more costly, which we attribute to these leakage events being shorter-lived. This is because the shorter a leakage event, the more likely its signature is due to (a combination of) qubit errors. Fortunately, shorter leakage events are also less damaging. For instance, a leaked data qubit that seeps back within the same round may be indistinguishable from a relaxation event, but also has the same effect on encoded logical information [154].

We now verify and externally benchmark our HMMs by their ability to improve $\langle Z \otimes Z \rangle$ by rejecting data with a high probability of leakage. To do this, we set a threshold $L^{\text{th}}_{\text{comp,Q}}$, and reject experimental runs whenever $L_{\text{comp,Q}}(\vec{M}_A) < L^{\text{th}}_{\text{comp,Q}}$. For both HMMs we choose $L^{\text{th}}_{\text{comp,Q}}$ to achieve $\text{TPR} = 0.7$. With this choice, we observe a restoration of $\langle Z \otimes Z \rangle$ to its first-round value across the entire curve (Figure 6.3G), mildly reducing f_{post} to 0.82 (averaged over M). This restoration from leakage is confirmed by the 'final + HMM' data matching the no-leakage model results in Figure 6.12, A to D. As low $L_{\text{comp,Q}}(\vec{M}_A)$ is also weakly correlated with qubit errors, the gain in $\langle Z \otimes Z \rangle$ is partly due to false positives. Of the ~ 0.13 increase at $M = 25$, we attribute 0.07 to actual leakage (estimated from the ROCs). By comparison, the simple two-state HMM, leads to a lower improvement, whilst rejecting a larger part of the data (Figure 6.14G), ultimately justifying the increased HMM complexity in this particular experiment.

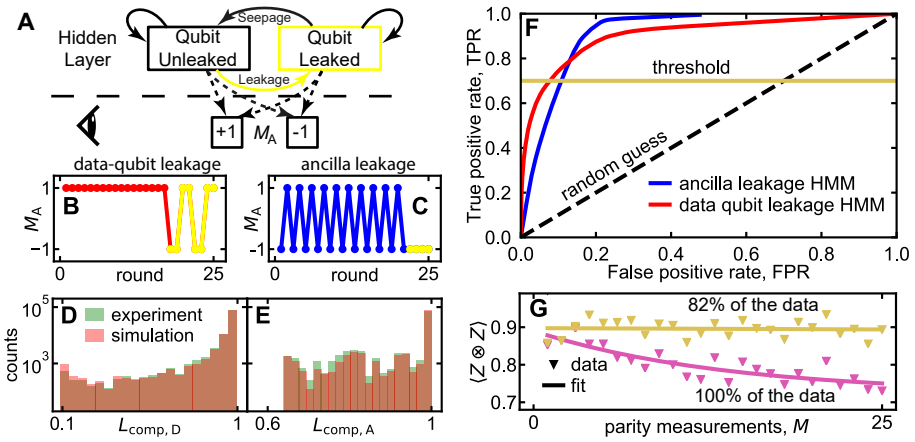


Figure 6.3: Leakage detection and mitigation during repeated ZZ checks using hidden Markov models (HMMs). (**A**) Simplified HMM. In each round, a hidden state (leaked or unleaked) (top) is updated probabilistically (full arrows), and produces an observable M_{A} (bottom) with state-dependent probabilities (dashed arrows). After training, the HMM can be used to assess the likelihood of states given a produced string $\vec{M}_A$ of M_{A}. (**B**) Example $\vec{M}_A$ for a data-qubit leakage event (yellow markers), showing the characteristic pattern of repeated errors. (**C**) Example $\vec{M}_A$ for Q_{A} leakage signalled by constant $M_{\text{A}} = -1$. (**D**) Histograms of 10^5 $\vec{M}_A$ with $M = 25$, obtained both experimentally, and simulated by the HMM optimized to detect data-qubit leakage, binned according to the likelihood (Equation (6.2) and Section 6.5.2) of the data qubits being unleaked (as assessed from the trained HMM). HMM training suggests 5.6% total data-qubit leakage at $M = 25$ [calculated from Table 6.2 as the steady-state fraction $p_{\text{leak}}/(p_{\text{leak}} + p_{\text{seep}})$]. (**E**) Corresponding histograms using the HMM optimized for Q_{A} leakage. This HMM suggests 3.8% total Q_{A} leakage. (**F**) Receiver operating characteristics for the trained HMMs. (**G**) $\langle Z \otimes Z \rangle$ after M ZZ checks and correction based on the 'final' outcomes, without (same data as in Figure 6.2D) and with leakage mitigation by postselection ($\text{TPR} = 0.7$).

6.2.5 Protecting entanglement from general qubit errors and mitigation of leakage

We finally demonstrate leakage mitigation in the more interesting scenario where $\left|\Phi^{+}\right\rangle$ is protected from general qubit error by interleaving ZZ and XX checks [78, 163]. ZZ may be converted to XX by adding $\pi/2$ y rotations on the data qubits simultaneous with those on Q_{A}. This requires that we change the definition of the syndrome to $s_{\mathrm{D}}[m] = M_{\mathrm{A}}[m] \cdot M_{\mathrm{A}}[m-1] \cdot M_{\mathrm{A}}[m-2] \cdot M_{\mathrm{A}}[m-3]$, as we need to 'undo' the interleaving of the ZZ and XX checks to detect errors. For an input state $|{+}0\rangle = (|0\rangle + |1\rangle)/\sqrt{2} \otimes |0\rangle$, a first pair of checks ideally projects the data qubits to one of the four Bell states with equal probability. Expanding the PFU to X and/or Z on Q_{DH} we find $F_{\left|\Phi^{+}\right\rangle} = 83.8 \pm 0.8\%$ (Figure 6.10). For subsequent rounds, the 'final' strategy now relies on the final three M_{A}. We observe a decay towards a steady state $F_{\left|\Phi^{+}\right\rangle} = 73.7 \pm 0.9\%$ (Figure 6.4), consistent with previously observed leakage. We battle this decay by adapting the HMMs (detailed in Sections 6.5.4 and 6.5.5). We find an improved ROC for Q_{A} leakage (Figure 6.11). For data-qubit leakage however, the ROC is degraded. This is to be expected — when one data qubit is leaked in this experiment, the ancilla effectively performs interleaved Z and X measurements on the unleaked qubit. This leads to a signal of random noise $P(s_{\mathrm{D}}[m] = -1) = 0.5$, which is less distinguishable from unleaked experiments $P(s_{\mathrm{D}}[m] = -1) \sim 0$ than the signal of a leaked data-qubit during the $\langle Z \otimes Z \rangle$-only experiment $P(s_{\mathrm{D}}[m] = -1) \sim 1$. Most importantly, thresholding to $\mathrm{TPR} = 0.7$ restores $\langle X \otimes X \rangle$ and $\langle Z \otimes Z \rangle$, leading to an almost constant $F_{\left|\Phi^{+}\right\rangle} = 82.8 \pm 0.2\%$ with $f_{\mathrm{post}} = 0.81$ (averaged over M), as expected from the no-leakage model results in Figure 6.12, E to H. In this experiment, the simple two-state HMMs performs almost identically compared to the complex HMM, achieving Bell-state fidelities within 2% whilst retaining the same amount of data (Figure 6.14N).

6.3 Discussion

This HMM demonstration provides exciting prospects for leakage detection and correction. In larger systems, independent HMMs can be dedicated to each qubit because leakage produces local error signals [155]. An HMM for an ancilla only needs its measurement outcomes while a data-qubit HMM only needs the outcomes of the nearest-neighbor ancillas [details in Section 6.5.7]. Therefore, the computational power grows linearly with the number of qubits, making the HMMs a small overhead when running parallel to MWPM. HMM outputs could be used as inputs to MWPM, allowing MWPM to dynamically adjust its weights. The outputs could also be used to trigger leakage reduction units [153–156] or qubit resets [168].

In summary, we have performed the first experimental investigation of leakage detection during repetitive parity checking, successfully protecting an entangled state from qubit errors and leakage in a circuit quantum electrodynamics processor. Future work will extend this protection to logical qubits, e.g., the 17-qubit surface code [37, 38]. The low technical overhead and scalability of HMMs is attractive for performing leakage detection and correction in real time using the same parity outcomes as traditionally used to correct qubit errors only.

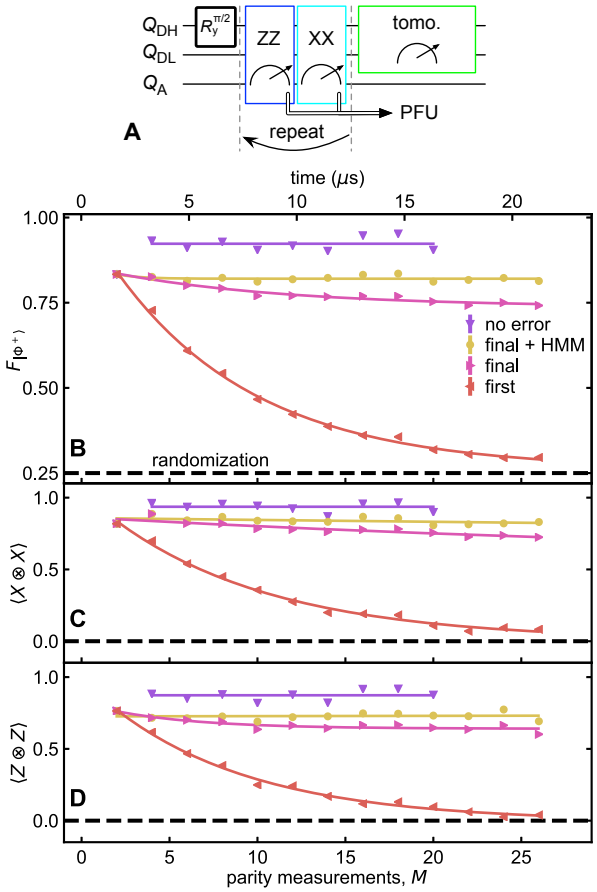


Figure 6.4: Protecting entanglement from general qubit error and leakage. (**A**) Simplified quantum circuit with preparation, repeated pairs of ZZ and XX checks, and data-qubit tomography. (**B**) Fidelity to $|\Phi^+\rangle$ as a function of M, extracted from the data-qubit tomography. 'No error' postselects the runs in which no error is detected (postselected fraction in Figure 6.9). 'Final' applies PFU based on the last three outcomes (equivalent to mimimum-weight perfect matching). 'Final + HMM' includes mitigation of leakage. 'First' uses only the first pair of parity outcomes. (**C** and **D**) Corresponding $\langle X \otimes X \rangle$ and $\langle Z \otimes Z \rangle$. Curves in (B to D) are best fits of a simple exponential decay.

6.4 Materials and methods

6.4.1 Device

Our quantum processor (Figure 6.5) follows a three-qubit-frequency extensible layout with nearest-neighbor interactions that is designed for the surface code [21]. Our chip contains low- and high-frequency data qubits (Q_{DL} and Q_{DH}), and an intermediate-frequency ancilla (Q_A). Single-qubit gates around axes in the equatorial plane of the Bloch sphere are

performed via a dedicated microwave drive line for each qubit. Two-qubit interactions between nearest neighbors are mediated by a dedicated bus resonator (extensible to four per qubit) and controlled by individual tuning of qubit transition frequencies via dedicated flux-bias lines [25]. For measurement, each qubit is dispersively coupled to a dedicated readout resonator (RR) which is itself connected to a common feedline via a dedicated Purcell resonator (PR). The RR-PR pairs allow frequency-multiplexed readout of selected qubits with negligible backaction on untargeted qubits [76].

6.4.2 Setup

A full wiring diagram of the setup is provided in (Figure 6.6). All operations are controlled by a fully digital device, the central controller (CC7), which takes as input a binary in an executable quantum instruction set architecture [eQASM [169]], and outputs digital codeword triggers based on the execution result of these instructions. These digital codeword triggers are issued every 20 ns to arbitrary waveform generators (AWGs) for single-qubit gates and two-qubit gates, a vector switch matrix (VSM) for single-qubit gate routing and a readout module (AWG and acquisition) for frequency-multiplexed readout. Single-qubit gate generation, readout pulse generation and readout signal integration are performed by single-sideband mixing. The measurement signal is amplified with a JTWPA [138] at the front end of the amplification chain. Following Ref. [170], we extract an overall measurement efficiency $\eta = 48 \pm 1.0\%$ by comparing the integrated signal-to-noise ratio of single-shot readout to the integrated measurement-induced dephasing.

6.4.3 Cross-measurement-induced dephasing of data qubits

During ancilla measurement, data-qubit coherence is susceptible to intrinsic decoherence, phase shifts via residual ZZ interactions and cross-measurement-induced dephasing [76, 85]. For the single-data-qubit subspace we investigate the different contributions experimentally and assess the benefit of an echo pulse on the data qubits halfway through the ancilla measurement. We study this by including the ancilla measurement (with amplitude ε) in a Ramsey-type sequence (Figure 6.7A). By varying the azimuthal phase of the second $\pi/2$ pulse, we obtain Ramsey fringes from which we extract the coherence $|\rho_{01}|$ and phase $arg\left(\rho_{01}\right)$. Several features of these curves explain the need for the echo pulse on the data qubits. Firstly, at $\varepsilon = 0$, the echo pulse improves data-qubit coherence (for both ancilla states) by reducing the effect of low-frequency noise (Figure 6.7, B and C). This is confirmed by individual Ramsey and echo experiments. Secondly, the echo pulse almost perfectly cancels ancilla-state dependent phase shifts due to residual ZZ interactions (Figure 6.7, D and E). When gradually turning on the ancilla measurement towards the nominal value $\varepsilon = 1$, we furthermore observe that: thirdly, the echo pulse almost perfectly cancels the measurement-induced Stark shift (Figure 6.7, D and E). When increasing the measurement amplitude beyond the operation amplitude (indicated by the vertical dashed lines), we see rapid non-Gaussian decay of data-qubit coherence. We attribute this to measurement-induced relaxation of the ancilla: via

the ZZ interaction, this can lead to probabilistic phase shifts on the data qubit. This effect is stronger for Q_{DL} than for Q_{DH} due to its higher residual interaction with Q_A (Table 6.1).

Gate and Coherence Parameters	Q_{DL}	Q_A	Q_{DH}
operating qubit frequency, $\omega_{op}/2\pi$ (GHz)	5.02	5.79	6.88†
max. qubit frequency, $\omega_{max}/2\pi$ (GHz)	5.02	5.79	6.91
anharmonicity, $\alpha/2\pi$ (MHz)	-306	-308	-331
coherence time (at $\omega_{op}/2\pi$), T_2^{echo} (μs)	29.6$\pm$2.7	21.7$\pm$1.4	14.7$\pm$0.9
relaxation time (at $\omega_{op}/2\pi$) T_1 (μs)	25.3$\pm$1.2	17.0$\pm$0.6	25.6$\pm$1.2
Ramsey deph. time (at $\omega_{op}/2\pi$), T_2^* (μs)	24.5$\pm$2.0	14.6$\pm$1.2	5.9$\pm$0.7
mean error / single-qubit gate††††, e_{SQ} (%)	0.08$\pm$0.02	0.14$\pm$0.016	0.21$\pm$0.06
resonance exchange coupling, $J_1/2\pi$ (MHz)	17.2	14.3	
bus resonator frequency, $\sim \omega_{bus}/2\pi$ (GHz)	8.5	8.5	
error per CZ†††††, e_{CZ} (%)	1.4$\pm$0.6	0.9$\pm$0.16	
leakage per CZ†††††, L_1 (%)	0.27$\pm$0.12	0.15$\pm$0.07	
ZZ coupling (at $\omega_{op}/2\pi$), $\zeta_{ZZ}/2\pi$ (MHz)	0.95	0.33	
Measurement Parameters	Q_{DL}	Q_A	Q_{DH}
readout pulse frequency, $\omega_{ro}/2\pi$ (GHz)	7.225	7.420	7.838
readout resonator frequency, $\omega_{ro}/2\pi$ (GHz)	7.275	7.385	7.867
Purcell resonator frequency, $\omega_{ro}/2\pi$ (GHz)	7.260	7.405	7.872
qubit-RR coupling, $g_{01,RR}/2\pi$ (MHz)	202	188	135
PF-RR coupling, $J_{RR,PF}/2\pi$ (MHz)	48	30	38
dispersive shift qubit-RR, χ_{RR}/π (MHz)	-2.5	-5.3	$-2.8^{\dagger\dagger}$
dispersive shift qubit-PF, χ_{PF}/π (MHz)	-1.5	-4.7	$-2.8^{\dagger\dagger}$
critical photon number, n_{crit}	2.3	2.7	2.4
intra-resonator photon number RR, n_{RR}		1.2	
quantum efficiency, η (%)		48$\pm$1.0	
Average assignment error, e_a (%)	9.0†††	1.0$\pm$0.1	16†††
Measurement integration time, τ_{int} (ns)	600	600	600

Table 6.1: Measured parameters of the three-transmon device. † Q_{DH} is operated 30 MHz below its maximum frequency to avoid spurious interaction with a spurious two-level system. †† The Purcell mode and readout resonator mode of Q_{DH} have near-perfect hybridization (with qubit at $\omega_{op}/2\pi$), making them indistinguishable. ††† Single-shot readout on the data qubits was not optimized. †††† Single-qubit gates are characterized using Clifford randomized benchmarking [171] ††††† Two-qubit gates are characterized using interleaved RB [30, 171] with a leakage-extraction modification [149].

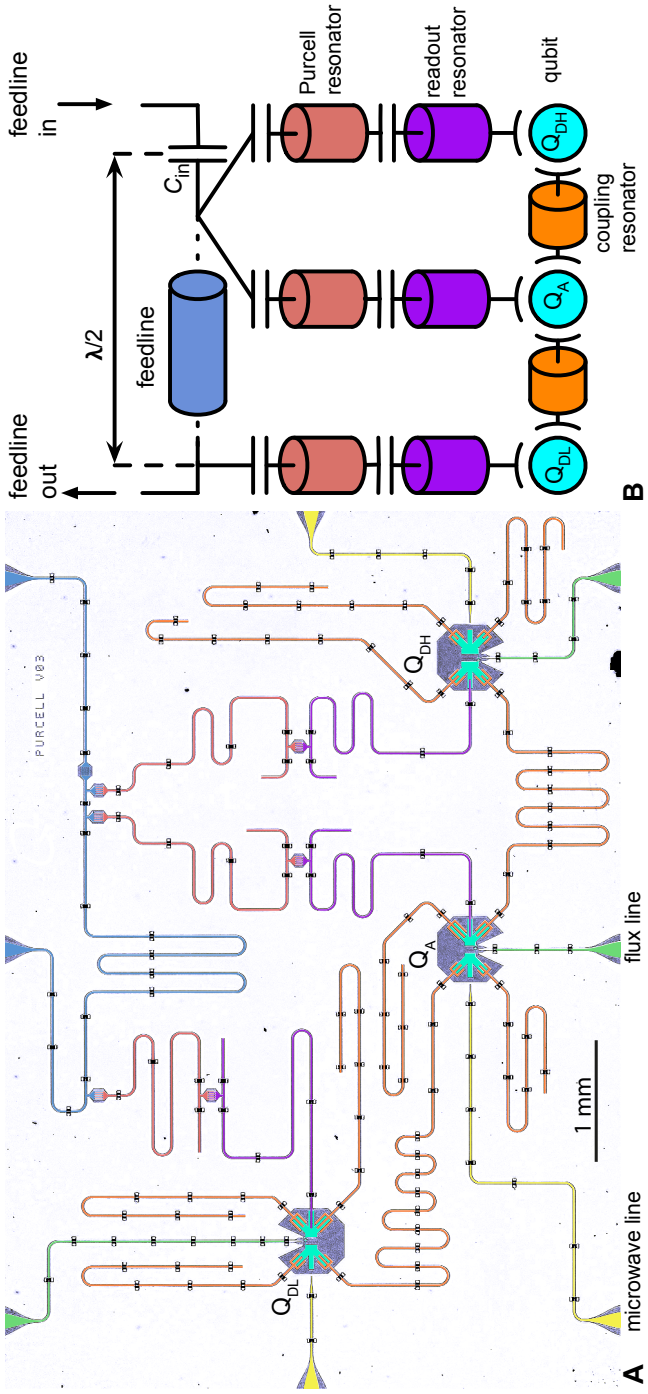


Figure 6.5: False-colored photograph (a) and simplified circuit diagram (b) of the quantum processor with corresponding colors.

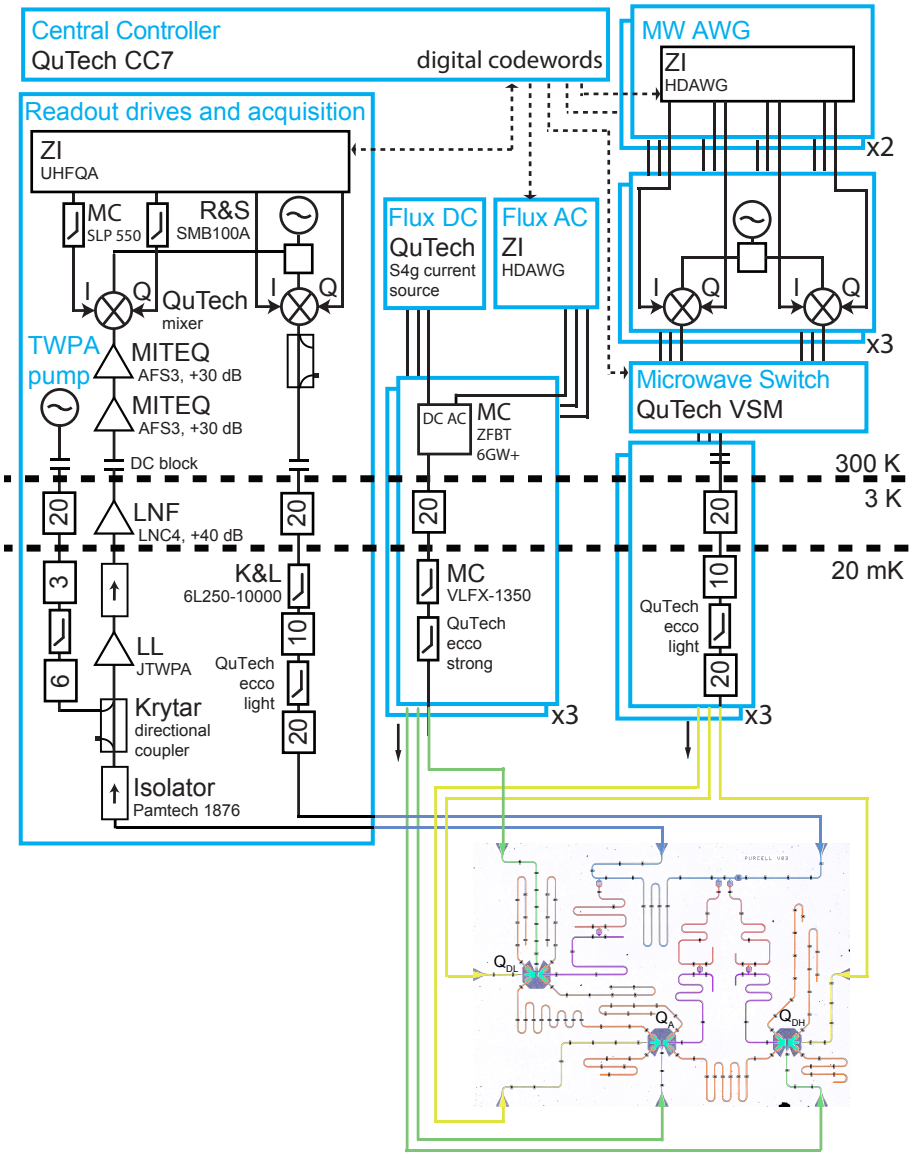


Figure 6.6: Complete wiring diagram of electronic components inside and outside the ^{3}He/^{4}He dilution refrigerator (Leiden Cryogenics CF-CS81).

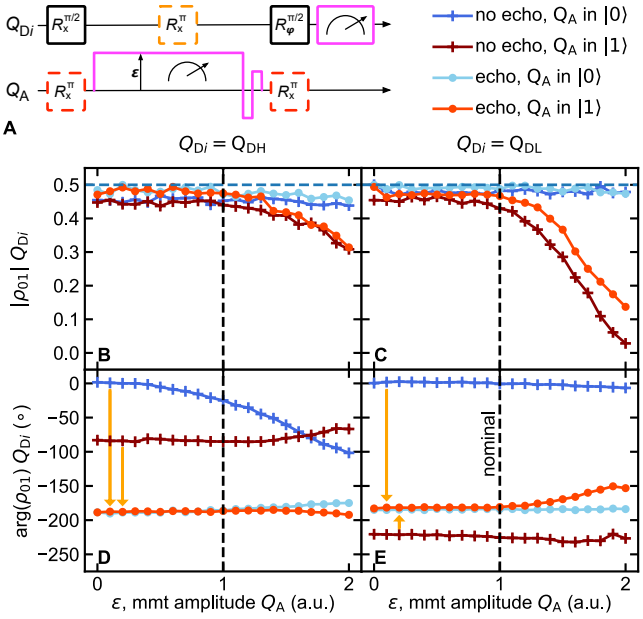


Figure 6.7: Study of data-qubit coherence and phase accrual during ancilla measurement. (**A**) Quantum circuit to extract data-qubit coherence and phase with or without echo pulse (orange) and with or without excitation in the ancilla. (**B** and **C**) Data-qubit coherence as a function of ancilla measurement amplitude. (**D** and **E**) Data-qubit phase as a function of ancilla measurement amplitude.

6.4.4 Uncertainty calculations

All quoted uncertainties are an estimation of standard error of the mean (SEM). SEMs for the independent device characterizations (Section 6.2.2, Table 6.2) are either obtained from at least three individually fitted repeated experiments (T_2^{echo}, T_1, T_2^*, η, e_{a}, $e_{\text{a,ZZ}}$) or in the case that the quantitiy is only measured once (e_{SQ}, e_{CZ}, L_1), the SEM is estimated from least-squares fitting by the LmFit fitting module using the covariance matrix [172].

SEMs in the first-round Bell-state fidelities (Figures 6.1 and 6.10, Sections 6.2.2 and 6.2.5) are obtained through bootstrapping. For bootstrapping, a data-set (in total 4096 runs with each 36 tomographic elements and 28 calibration points) is subdivided into four subsets and tomography is performed on each of these subsets individually. As verification, subdivision was performed with eight subsets leading to similar SEMs.

SEMs in the multi-round experiment parameters (steady-state fidelities, decay constants) are also estimated from least-squares fitting by the LmFit fitting module using the covariance matrix [172] (Sections 6.2.3 to 6.2.5).

6.5 Models

6.5.1 Hidden Markov models

HMMs provide an efficient tool for indirect inference of the state of a system given a set of output data [165]. A hidden Markov model describes a time-dependent system as evolving between a set of N_h hidden states $\{h\}$ and returning one of N_o outputs $\{o\}$ at each timestep m. The evolution is stochastic: the system state $H[m]$ of the system at timestep m depends probabilistically on the state $H[m-1]$ at the previous timestep, with probabilities determined by a $N_h \times N_h$ transition matrix A

$$A_{h,h} = P(H[m] = h \mid H[m-1] = h').$$

The user cannot directly observe the system state, and must infer it from the outputs $O[m] \in \{o\}$ at each timestep m. This output is also stochastic: $O[m]$ depends on $H[m]$ as determined by a $N_o \times N_h$ output matrix B

$$B_{o,h} = P(O[m] = o \mid H[m] = h).$$

If the A and B matrices are known, along with the expected distribution $\vec{\pi}^{(\text{prior})}[0]$ of the system state over the N_h possibilities,

$$\pi_h^{(\text{prior})}[1] = P(H[1] = h),$$

one may simulate the experiment by generating data according to the above rules. Moreover, given a vector $\vec{o}$ of observations, we may calculate the distribution $\vec{\pi}[m]$ over the possible states at a later time m,

$$\pi_h^{(\text{post})}[m] = P(H[m] = h | O[1] = o_1, \ldots, O[m] = o_m),$$

by interleaving rounds of Markovian evolution,

$$\begin{aligned}
&\pi_h^{(\text{prior})}[m] \\
&:= P(H[m] = n | O[1] = o_1, \ldots, O[m-1] = o_{m-1}) \\
&= \sum_{h'} A_{h,h'} \pi_{h'}^{(\text{post})}[m-1],
\end{aligned}$$

and Bayesian update,

$$\pi_h^{(\text{post})}[m] = \frac{B_{o_m,h}\, \pi_h^{(\text{prior})}[m]}{\sum_{h'} B_{o_m,h'} \pi_{h'}^{(\text{prior})}[m]}.$$

6.5.2 Hidden Markov models for QEC experiments

To maximize the discrimination ability of HMMs in the various settings studied in this work, we choose different quantities to use for our output vectors $\vec{o}$. In all experiments in this work, the signature of a leaked ancilla is repeated $M_A[m] = -1$, and so we choose $\vec{o} = \vec{M}_A$.

By contrast, the signature of leaked data qubits in both experiments may be seen as an increased error rate in their corresponding syndromes $\vec{s_{\mathrm{D}}}$, and we choose $\vec{o} = \vec{s_{\mathrm{D}}}$ for the corresponding HMMs.

One may predict the computational likelihood for data-qubit (D) leakage at timestep M in the ZZ-check experiment given $\vec{\pi}(M)$. In particular, once we have declared which states h correspond to leakage, we may write

$$L_{\mathrm{comp,D}} = \sum_{h \text{ unleaked}} \pi_h^{(\mathrm{post})}[M]. \tag{6.2}$$

However, in the repeated ZZ-check experiment, the ancilla (A) needs to be within the computational subspace for two rounds to perform a correct parity measurement. Therefore, the computational likelihood is slightly more complicated to calculate,

$$L_{\mathrm{comp,A}}[M] = \frac{\sum_{h,h' \text{ unleaked}} B_{o_m,h}\, A_{h,h'} \pi_{h'}^{(\mathrm{post})}[M-1]}{\sum_{h,h'} B_{o_m,h} A_{h,h'} \pi_{h'}^{(\mathrm{post})}[M-1]}.$$

In the interleaved ZZ- and XX-check experiment, the situation is more complicated as we require data from the final two parity checks to fully characterize the quantum state. This implies that we need unleaked data qubits for the last two rounds and unleaked ancillas for the last three. The likelihood of the latter may be calculated by similar means to the above.

6.5.3 Simplest models for leakage discrimination

One need not capture the full dynamics of the quantum system in a HMM to infer whether a qubit is leaked. This is of critical importance if we wish to extend this method for the purposes of leakage mitigation in a large QEC code [as we discuss in Section 6.5.7]. The simplest possible HMM (Figure 6.3A) has two hidden states: $H[m] = 1$ if the qubit(s) in question are within the computational subspace, and $H[m] = 2$ if Q_{A} (or either data qubit) is leaked. (The labels 1 and 2 are arbitrary here, and explicitly have no correlation with the qubit states $|1\rangle$ and $|2\rangle$.) Then, the 2×2 transition matrix simply captures the leakage and seepage rates of the system in question:

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + p_{\mathrm{leak}} \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} + p_{\mathrm{seep}} \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}.$$

The 2×2 output matrices then capture the different probabilities of seeing output $O[m] = 0$ or $O[m] = 1$ when the qubit(s) are leaked or unleaked:

$$B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + p_{0,1} \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} \tag{6.3}$$

$$+ p_{1,0} \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}. \tag{6.4}$$

When studying data-qubit leakage, $p_{0,1}$ simply captures the rate of errors within the computational subspace. Then, in the repeared ZZ-check experiment, $p_{1,0}$ captures events such as ancilla or measurement errors that cancel the error signal of a leakage event. However, in the interleaved ZZ and XX experiment, a leaked qubit causes the syndrome to be random, so we expect $p_{1,0} \sim 0.5$. When studying ancilla leakage, $p_{1,0}$ is simply the probability of $|2\rangle$ state being read out as $|0\rangle$, and is also expected to be close to 0. However, $p_{0,1} \sim 0.5$, as we do not reset Q_A or the logical state between rounds of measurement, and thus any measurement in isolation is roughly equally-likely to be 0 or 1. In all situations, we assume that the system begins in the computational subspace — $\pi_n(0) = \delta_{n,0}$. With this fixed, we may choose the parameters p_leak, p_seep, $p_{0,1}$ and $p_{1,0}$ to maximize the likelihood $L(\{\vec{o}\})$ of observing the recorded experimental data $\{o\}$. (Note that $L(\{\vec{o}\})$ is not the computational likelihood $L_\mathrm{comp,Q}$.)

6.5.4 Modeling additional noise

The simple model described above does not completely capture all of the details of the stabilizer measurements $\vec{M}_A$. For example, the data-qubit HMM will overestimate the leakage likelihood when an ancilla error occurs, as this gives a signal with a time correlation that is unaccounted for. As the signature of a leakage event in a fully fault-tolerant code will be large Section 6.5.7, we expect these details to not significantly hinder the simple HMM in a large-scale QEC simulation. However, this lack of accuracy makes evaluating HMM performance somewhat difficult, as internal metrics may not be so trustworthy. We also risk overestimating the HMM performance in our experiment, as our only external metrics for success (e.g., fidelity) do just as poorly when errors occur near the end of the experiment as they do when leakage occurs. Therefore, we extend the set of hidden states in the HMMs to account for ancilla and measurement errors, and to allow the ancilla HMM to keep track of the stabilizer state. To attach physical relevance to the states in our Markovian model, and to limit ourselves to the noise processes that we expect to be present in the system, we generalize Equations (6.3) and (6.4) to a linearly-parametrized model,

$$A = A_0 + \sum_\mathrm{err} p_\mathrm{err} D^{(\mathrm{A})}_\mathrm{err}, \qquad B = B_0 + \sum_\mathrm{err} p_\mathrm{err} D^{(\mathrm{B})}_\mathrm{err}.$$

Here, we choose the matrices $D^{(\mathrm{A})}_i$ and $D^{(\mathrm{B})}_i$ such that the error rates $p^{(\mathrm{A})}_i, p^{(\mathrm{B})}_i$ correspond to known physical processes. (We add the superscripts (A) and (B) here to the D matrices to emphasize that each error channel only appears in one of the two above equations.)

The error generators $D^{(\mathrm{A})}$, $D^{(\mathrm{B})}$ may be identified as derivatives of A with respect to these error rates:

$$D^{(\mathrm{A})}_i = \frac{\partial A}{\partial p^{(\mathrm{A})}_i}, \qquad D^{(\mathrm{B})}_i = \frac{\partial B}{\partial p^{(\mathrm{B})}_i}.$$

This may be extended to calculate derivatives of the likelihood $L(\{\vec{o}\})$ (or more practically, the log-likelihood) with respect to the various parameters p_i. This allows us to obtain the maximum likelihood model within our parametrization via gradient descent methods (in particular

the Newton-CG method), instead of resorting to more complicated optimization algorithms such as the Baum-Welch algorithm [165]. All models were averaged over between 10 and 20 optimizations using the Newton-CG method in scipy [104], calculating likelihoods, gradients and Hessians over 10,000-20,000 experiments per iteration, and rejecting any failed optimizations. As the signal of ancilla leakage is identical to the signal for even ZZ and XX parities with ancilla in $|1\rangle$ and no errors, we find that the optimization is unable to accurately estimate the ancilla leakage rate, and so we fix this in accordance with independent calibration to 0.0040/round using averaged homodyne detection of $|2\rangle$ (making use of a slightly different homodyne voltage for $|1\rangle$ and $|2\rangle$).

6.5.5 Hidden Markov models used in Figures 6.2 and 6.4

Different Markov models (with independently optimized parameters) were used to optimize ancilla and data-qubit leakage estimation for both the ZZ experiment and the experiment interleaving ZZ and XX checks. This lead to a total of four HMMs, which we label H_{ZZ}-D, H_{ZZ}-A, $H_{ZZ,XX}$-D and $H_{ZZ,XX}$-A. A complete list of parameter values used in each HMM is given in Table 6.2. We now describe the features captured by each HMM. As we show in Section 6.5.6, these additional features are not needed to increase the error mitigation performance of the HMMs, but rather to ensure their closeness to the experiment and increase trust in their internal metrics.

To go beyond the simple HMM in the ZZ-check experiment when modeling data-qubit leakage (H_{ZZ}-D), we need to include additional states to account for the correlated signals of ancilla and readout error. If we assume data-qubit errors (that remain within the logical subspace) are uncorrelated in time, they are already well-captured in the simple model. This is because any single error on a data qubit may be decomposed into a combination of Z errors (which commute with the measurement and thus are not detected) and X errors (which anti-commute with the measurement and thus produce a single error signal $s_D[m] = 1$), and is thus captured by the $p_{0,1}$ parameter. When one of the data qubits is leaked, uncorrelated X errors on the other data qubit cancel the constant $s_D[m] = -1$ signal for a single round, and are thus captured by the $p_{1,0}$ parameter. However, errors on the ancilla, and readout errors, give error signals that are correlated in time (separated by 1 or 2 timesteps, respectively). This may be accounted for by including extra 'ancilla error states'. These may be most easily labeled by making the h labels a tuple $h = (h_0, h_1)$, where h_0 keeps track of whether or not the qubit is leaked, and $h_1 = 1, 2, 3$ keeps track of whether or not a correlated error has occurred. In particular, we encode the future syndrome for 2 cycles in the absence of error on h_1, allowing us to account for any correlations up to 2 rounds in the future. This extends the model to a total of $4 \times 2 = 8$ states. The transition and output matrices in the absence of error for the unleaked $h_0 = 0$ states may then be written in a compact form (noting that leakage errors cancel out with correlated ancilla and readout errors to give $s_D[m] = +1$),

$$[A_0]_{(h_0,h_1//2),(h_0,h_1)} = 1, \qquad [B_0]_{-1^{h_0+h_1},(h_0,h_1)} = 1, \tag{6.5}$$

where the double slash $//$ refers to integer division.

Let us briefly demonstrate how the above works for ancilla error in the system. Suppose the system was in the state $h = (0,3)$ at time m. It would output $M_A[m] = -1$, and then evolve to $h = (0,3//2) = (0,1)$ at time $m+1$ (in the absence of additional error). Then, it would output a second error signal $[M_A[m+1] = -1]$ and finally decay back to the $h = (0,1//2) = (0,0)$ state. This gives the HMM the ability to model ancilla error as an evolution from $h = (0,0)$ to $h(0,3)$. Formally, we assign the matrix $D^{(A)}_{\text{ancilla}}$ to this error process, and following this argument we have

$$[D^{(A)}_{\text{ancilla}}]_{(0,0),(0,0)} = -1, \quad [D^{(A)}_{\text{ancilla}}]_{(0,3),(0,0)} = 1.$$

The corresponding error rate p_{ancilla} is then an additional free parameter to be optimized to maximize the likelihood. To finish the characterization of this error channel, we need to consider the effect of ancilla error in states other than $h = (0,0)$. Two ancilla errors in the same timestep cancel, but two ancilla errors in subsequent timesteps will cause the signature $s_D = \ldots, -1, +1, -1, \ldots$. This may be captured by an evolution from $h = (0,2)$ to $h = (0,3)$ [instead of $h = (0,1)$], which implies we should set

$$[D^{(A)}_{\text{ancilla}}]_{(0,1),(0,3)} = -1, \quad [D^{(A)}_{\text{ancilla}}]_{(0,2),(0,3)} = 1.$$

(Note that A_0 already captures a decay from $h = (0,2) \to (0,1) \to (0,0)$, which will give the desired signal.) We note that this also matches the signature of readout error, which can then be captured by a separate error channel $D^{(A)}_{\text{readout}}$ which increases this correlation

$$[D^{(A)}_{\text{readout}}]_{(0,1),(0,2)} = -1, \quad [D^{(A)}_{\text{readout}}]_{(0,3),(0,2)} = 1.$$

One can then check that ancilla errors in $h = (0,3)$ should cause the system to remain in $h = (0,3)$, and that ancilla or readout errors in $h = (0,1)$ should evolve the system to $h = (0,2)$. We note that this model cannot account for the $s_D = \ldots -1, +1, +1, +1, -1$ signature of readout error at time m and $m+2$, but adjusting the model to include this has negligible effect.

Ancilla error in the ZZ-check experiment when the data qubits are leaked has the same correlated behavior as when they are not, but may occur at a different rate. This requires that we define a new matrix $D^{(A)}_{\text{ancilla,leaked}}$ by

$$[D^{(A)}_{\text{ancilla,leaked}}]_{(1,j),(1,k)} = [D^{(A)}_{\text{ancilla}}]_{(0,j),(0,k)},$$

with a separate error rate $p_{\text{ancilla,leaked}}$. As we do not expect the readout of the ancilla to be significantly affected by whether the data qubit is leaked, we do not add an extra parameter to account for this behavior, and instead simply set

$$[D^{(A)}_{\text{readout}}]_{(1,j),(1,k)} = [D^{(A)}_{\text{readout}}]_{(0,j),(0,k)}.$$

We also assume that leakage p_{leak} and seepage p_{seep} rates are independent of these correlated errors (i.e., $[D^{(A)}_{\text{leak}}]_{(0,j),(0,k)}, [D^{(A)}_{\text{seep}}]_{(1,j),(1,k)}] \in \{0,-1\}$). We then assume that the first measurement made following a leakage/seepage event is just as likely to have an

additional error (corresponding to an evolution to $(h_0, 1)$) or not (corresponding to an evolution to $(h_0, 0)$). We finally account for data-qubit error in the output matrices in the same way as in the simple model, but with different error rates $p_{\text{data,leaked}}$ for the leaked states $(1, h_1)$ and p_{data} for the unleaked states $(0, h_1)$.

There are a few key differences between the interleaved ZZ—XX and ZZ experiments that need to be captured in the data-qubit HMM $H_{ZZ,XX} - D$. Firstly, as the syndrome is now given by $s_D[m] = M_A[m] \cdot M_A[m-1] \cdot M_A[m-2] \cdot M_A[m-3]$, ancilla and classical readout error can then generate a signal stretching up to 4 steps in time. This implies that we require 2^4 possibilities for h_1 to keep track of all correlations. However, as a leaked data qubit makes ancilla output random in principle, we no longer need to keep track of the ancilla output upon leakage. This implies that we can accurately model the experiment with $16 + 1 = 17$ states, which we can label by $h \in \{2, (1, h_1)\}$. The A_0 and B_0 matrices in the unleaked states $(1, h_1)$ follow Equation (6.5), and we fix $[A_0]_{2,2} = 1$ (as in the absence of p_{seep} a leaked state stays leaked). However, we allow for some bias in the leaked state error rate - $B_{-1,2} = p_{\text{data,leaked}}$ is not fixed to 0.5. (For example, this accounts for a measurement bias towards a single state, which will reduce the error rate below 0.5.) The non-zero elements in the matrices $D^{(A)}_{\text{ancilla}}$ and $D^{(A)}_{\text{readout}}$ may be written:

$$\begin{aligned}
[D^{(A)}_{\text{ancilla}}]_{(1,h_1//2),(1,h_1)} &= -1, \\
[D^{(A)}_{\text{ancilla}}]_{(1,h_1//2\oplus 5)} &= 1, \\
[D^{(A)}_{\text{readout}}]_{(1,h_1//2),(1,h_1)} &= -1, \\
[D^{(A)}_{\text{readout}}]_{(1,h_1//2\oplus 15)} &= 1.
\end{aligned}$$

Here, $a \oplus b$ refers to addition of each binary digit of a and b modulo 2. We may use this formalism to additionally keep track of Y data-qubit errors, which show up as correlated errors on subsequent XX and ZZ stabilizer checks, by introducing a new error channel

$$[D^{(A)}_{\text{data,Y}}]_{(1,h_1//2),(1,h_1)} = -1, \quad [D^{(A)}_{\text{data,Y}}]_{(1,h_1//2\oplus 3)} = 1,$$

with a corresponding error rate $p_{\text{data,Y}}$. As before, we assume that leakage occurs at a rate p_{leak} independently of h_1, and that seepage takes the system either to the state with either no error signal $h = (1, 0)$ or one error signal $h = (0, 1)$ with a rate p_{seep}.

As the output used for the $H_{ZZ} - A$ HMM is the pure measurement outcomes M_A, the dominant signal that must be accounted for is that of the stabilizer ZZ itself. This either causes a constant signal $M_A[m] = M_A[m-1]$ or a constant flipping signal $M_A[m] = -M_A[m-1]$. This cannot be accounted for in the simple HMM, as it cannot contain any history in a single unleaked state. To deal with this, we extend the set of states in the $H_{ZZ} - A$ HMM to include both an estimate of the ancilla state $a \in \{0, 1, 2\}$ at the point of measurement, and the stabilizer state $s \in \{0, 1\}$, and label the states by the tuple (a, s). The ancilla state then immediately defines the device output in the absence of any error:

$$[B_0]_{1,(0,s)} = [B_0]_{-1,(1,s)} = [B_0]_{-1,(2,s)} = 1,$$

while the stabilizer state defines the transitions in the absence of any error or leakage:

$$[A_0]_{(a+s \bmod 2,s)(a,s)} = 1 \text{ if } a < 2, \qquad [A_0]_{(2,s),(2,s)} = 1.$$

The only thing that affects the output matrices is readout error:

$$\begin{aligned}
&[D^{(B)}_{\text{readout}}]_{1,(0,s)} \\
&= [D^{(B)}_{\text{readout}}]_{-1,(1,s)} = [D^{(B)}_{\text{readout}}]_{-1,(2,s)} = -1, \\
&[D^{(B)}_{\text{readout}}]_{-1,(0,s)} \\
&= [D^{(B)}_{\text{readout}}]_{1,(1,s)} = [D^{(B)}_{\text{readout}}]_{1,(2,s)} = 1.
\end{aligned}$$

Data-qubit errors flip the stabilizer with probability p_{data}:

$$\begin{aligned}
[D^{(A)}_{\text{data}}]_{(a,s),(a',s)} &= -[A_0]_{(a,s),(a',s)}, \\
[D^{(A)}_{\text{data}}]_{(a,s),(a',1-s)} &= [A_0]_{(a,s),(a',s)}.
\end{aligned}$$

Ancilla errors flip the ancilla with probability p_{ancilla}, but these are dominated by T_1 decay, and so are highly asymmetric. To account for this, we used different error rates $p_{\text{anc},a,a'}$ for the four possible combinations of ancilla measurement at time $m-1$ and expected ancilla measurement at time m:

$$\begin{aligned}
[D^{(A)}_{\text{anc},a,a'}]_{(a,s)(a',s)} &= -[A_0]_{(a,s),(a',s)}, \\
[D^{(A)}_{\text{anc},a,a'}]_{(a+1 \bmod 2,s)(a',s)} &= -[A_0]_{(a,s),(a',s)}.
\end{aligned}$$

(Note that this asymmetry could not be accounted for in the data-qubit HMM as the state of the ancilla was not contained within the output vector.) As with the data-qubit HMMs, we assume that ancilla leakage is HMM-state independent, as it is dominated by CZ gates during the time that the ancilla is either in $|+\rangle$ or $|-\rangle$. We also assume that leakage (with rate p_{leak}) and seepage (with rate p_{seep}) have equal chances to flip the stabilizer state, as ancilla leakage has a good chance to cause additional error on the data qubits.

The ancilla-qubit HMMs need little adjustment between the ZZ-check experiment and the experiment interleaving ZZ and XX checks. The $H_{\text{ZZ,XX}}$-A HMM behaves almost identically to the H_{ZZ}-A HMM, but we include in the state information on the XX stabilizer as well as the ZZ stabilizer. This leaves the states indexed as (a, s_1, s_2). The HMM needs to also keep track of which stabilizer is being measured. This may be achieved by shuffling the stabilizer labels at each timestep: for $a = 0, 1$, we set

$$[A_0]_{(a+s_1 \bmod 2,s_2,s_1)(a,s_1,s_2)} = 1.$$

Other than this, the HMM follows the same equations as above (with the additional index added as expected.)

Error type	H_{ZZ}A	H_{ZZ}D	$H_{ZZ,XX}$A	$H_{ZZ,XX}$D
leakage [p_{leak}]	0.0040*	0.0064	0.0040*	0.0064
seepage [p_{seep}]	0.101	0.108	0.101	0.103
data-qubit error [p_{data}]	0.042	0.050	0.045	0.030
during leakage [$p_{\mathsf{data,leaked}}$]	-	0.155	-	0.489
Y error (additional) [$p_{\mathsf{data,Y}}$]	-	-	-	0.014
readout error [p_{readout}]	0.011	0.004	0.027	0.014
ancilla error [p_{ancilla}]	0.028	0.030	-	0.029
($M_A[m-1]=1$, $M_A[m]=1$) [$p_{\mathsf{anc},0,0}$]	-	-	0.001	-
($M_A[m-1]=1$, $M_A[m]=-1$) [$p_{\mathsf{anc},0,1}$]	-	-	0.021	-
($M_A[m-1]=-1$, $M_A[m]=1$) [$p_{\mathsf{anc},1,0}$]	-	-	0.044	-
($M_A[m-1]=-1$, $M_A[m]=-1$) [$p_{\mathsf{anc},1,1}$]	-	-	0.058	-
during leakage [$p_{\mathsf{ancilla,leaked}}$]	-	0.113	-	-

Table 6.2: Values of error rates used in the various HMMs in this work. All values are obtained by optimizing the likelihood of observing the given syndrome data except for the ancilla leakage rate (denoted *) which is directly obtained from the experiments (as noted in the text).

6.5.6 Performance of the simple hidden Markov model

In this section we detail the performance of the simple HMMs, as described in Figure 6.3A and Section 6.5.3 of the main text. In Figure 6.14, A and B, we plot a histogram of the computational likelihoods $L_{\mathsf{comp,Q}}$ of 10^5 simulated and actual ZZ experiments as calculated with the simple HMMs $H_{ZZ} - D^{(\mathsf{simple})}$ and $H_{ZZ} - A^{(\mathsf{simple})}$. This can be compared with Figure 6.3, B and C, of the main text. We plot similar histograms for the interleaved ZZ—XX experiment in Figure 6.14, H and I. We see reasonable agreement, but noticeably worse agreement than that in the detailed model. This is underscored by the Akaike information criterion (Equation (6.1) of the main text), which is significantly reduced compared to the more detailed HMMs:

$$\begin{aligned} A(H_{ZZ} - D) - A(H_{ZZ} - D^{(\mathsf{simple})}) \\ = 4.5 \times 10^5 \\ A(H_{ZZ} - A) - A(H_{ZZ} - A^{(\mathsf{simple})}) \\ = 5.9 \times 10^6 \\ A(H_{ZZ,XX} - D) - A(H_{ZZ,XX} - D^{(\mathsf{simple})}) \\ = 1.5 \times 10^5 \\ A(H_{ZZ,XX} - A) - A(H_{ZZ,XX} - A^{(\mathsf{simple})}) \\ = 1.6 \times 10^6. \end{aligned}$$

Indeed, in all cases the Akaike information criterion for the simple HMM is lower than that for the detailed HMM without leakage. This makes complete sense, as even though the sim-

ple HMMs might capture leakage fairly well, the additional effects captured in the detailed HMMs are far more dominant in the measurement signals than that of leakage. As such, the internal metrics, such as the ROC curves (Figure 6.15) for the simplified model are significantly less trustworthy than those of the detailed model. This exemplifies the need for external HMM verification, as achieved in the main text by testing the HMM in a leakage mitigation scheme. We now repeat this verification procedure for the simple model. We see that in the ZZ experiment the performance is significantly degraded; although the flat line in the $\langle Z \otimes Z \rangle$ curve is restored after about 8 parity checks, it requires rejecting 47% of the data, and is restored to a point $\sim$ 8% below the performance of the detailed HMM. By contrast, the simple HMM performs almost identically to the complex HMM in the interleaved ZZ—XX experiment, achieving Bell-state fidelities within 2% whilst retaining the same amount of data. As the signal from a large-scale QEC code is more similar to the latter experiment than the former (See Section 6.5.7), this strongly suggests that the detailed modeling used in this text will not be needed in such experiments.

6.5.7 Hidden Markov models for large-scale QEC

The hidden Markov models used in this text provide an exciting prospect for the indirect detection of leakage on both data qubits and ancillas in a QEC code. This is essential for accurate decoding of stabilizer measurements made during QEC. Furthermore, this idea can be combined with proposals for leakage reduction [153–156] to target such efforts, reducing unnecessary overhead. As leakage does not spread in superconducting qubits (to lowest order), and gives only local error signals [155], such a scheme would require a single HMM per (data and ancilla) qubit. Each individual HMM needs only to process the local error syndrome, and as demonstrated in this work, completely independent HMMs may be used for the detection of nearby data-qubit and ancilla leakage. This implies that the computational overhead of leakage detection via HMMs in a larger QEC code will grow only linearly with the system size. Previous leakage reduction units are designed to act as the identity on the computational subspace (up to additional noise), so we do not require perfect discrimination between leaked and computational states. However, optimizing this discrimination (and investigating threshold levels for the application of targeted leakage reduction) will boost the code performance. Also, near-perfect discrimination could allow for the direct resetting of leaked data qubits [168], which would completely destroy an error correcting code if not targeted.

On the other hand, for implementation on classical hardware within the sub-1 μs QEC cycle time on superconducting qubits [38], one may wish to strip back some of the optimization used in this work. The minimal HMM that could be used in QEC for detection has only two states, leaked and unleaked (Figure 6.3A), and 2^{n_A} outputs, where n_A is the number of neighboring ancilla on which a signature of leakage is detected. (For the surface code, $n_A \leq 4$ in all situations.) Such a simple model cannot perfectly deal with correlated errors, such as ancilla errors (which give multiple error signals separated in time). However, this should only cause a slight reduction in the discrimination capability whenever such correlations remain

local. If the loss in accuracy is acceptable, one may store only $\pi_0^{(\text{post})}$, and update it following a measurement $M_\text{A}[m]$ as

$$\begin{aligned}
&\pi_0^{(\text{prior})}[m] \\
&= (A_{0,0} - A_{0,1})\pi_0^{(\text{post})}[m-1] + A_{0,1}, \\
&\pi_0^{(\text{post})}[m] \\
&= \frac{\pi_0^{(\text{prior})}[m] B_{M_\text{A}[m],0}}{B_{M_\text{A}[m],1} + \pi_0^{(\text{prior})}(B_{M_\text{A}[m],0} - B_{M_\text{A}[m],1})},
\end{aligned}$$

which is trivial compared to the overhead for most QEC decoders.

A key question about the use of HMMs for leakage detection in future QEC experiments is whether leakage in larger codes is reliably detectable. In previous theoretical work [173], data-qubit leakage in repetition codes has been sometimes hidden, a phenomenon known as 'leakage paralysis' or 'silent stabilizer' [174]. This effect occurs when the relative phase φ accumulated between the $|20\rangle$ and $|21\rangle$ states during a CZ gate is a multiple of π. In the absence of additional error, an indirect measurement of the data qubit via an ancilla would return a result $\frac{\varphi}{\pi}$ mod 2. (By comparison, if $\varphi = \pi/2$, the ancilla would return measurements of 0 or 1 at random.) This is then identical to the measurement of a data qubit in the $|\frac{\varphi}{\pi} \bmod 2\rangle$ state, and no discrimination between the two may be achieved. However, in an N-qubit parity check S, the ancilla continues to accumulate phase from the other qubits, reducing this to an $N-1$-qubit effective parity check S' (plus a well-defined, constant phase). Such a parity check may no longer commute with other effective parity checks R' that share the leaked qubit, even though we would require $[S, R] = 0$ in stabilizer QEC. This is demonstrated in our second experiment measuring both ZZ and XX parity checks; though these commute when no data qubit is leaked, leakage reduces the checks to non-commuting Z and X measurements (of the unleaked data qubit). (In the ZZ experiment, the leakage paralysis was broken by the echo pulse on the data qubits, which flips the effective stabilizer of a leaked qubit at each round.) The repeated measurement of these non-commuting operators generates random results, similar to the case when $\varphi = \pi/2$. To the best of our knowledge, in all fully fault-tolerant stabilizer QEC codes, the removal of a single data qubit breaks the commutativity of at least two neighboring stabilizers. As such, data-qubit leakage will always be detectable in QEC experiments with superconducting circuits.

Beyond the proof-of-principle argument above, one might question whether the signal of leakage is improved or reduced when going from our prototype experiment to a larger QEC code, and when the underlying physical-qubit error rate is reduced. Fortunately, we can expect an improvement in the HMM discrimination capability in both situations. To see this, consider the example of a data qubit which is either leaked at round 1 with probability p_leak or never leaks. Let us further assume that in the absence of leakage, a number of neighboring ancillas n_A incur errors (where the parity check reports a flip) at a rate p, whereas in the presence of leakage these ancillas incur errors at a rate 0.5. (For example, in the bulk of the

surface code, $n_A = 4$.) The computational likelihood at round $m > 0$ after seeing e errors may be calculated as

$$L_{\text{comp,Q}}[m] = \frac{(1-p_{\text{leak}})p^e(1-p)^{mn_A-e}}{(1-p_{\text{leak}})p^e(1-p)^{mn_A-e} + p_{\text{leak}}(0.5)^{mn_A}}.$$

If the data qubit was leaked, $e \sim mn_A/2$, and the computational likelihood on average is approximately

$$L_{\text{comp}}[m] \sim \frac{1-p_{\text{leak}}}{p_{\text{leak}}}\left(\frac{p^{n_A/2}(1-p)^{n_A/2}}{0.5^{n_A}}\right)^m,$$

which is of the form

$$L_{\text{comp}}[m] = Ae^{-\lambda m}, \qquad A = \frac{1-p_{\text{leak}}}{p_{\text{leak}}},$$
$$\lambda = \log\left(2^{n_A}p^{-\frac{n_A}{2}}(1-p)^{-\frac{n_A}{2}}\right).$$

We see that the signal of leakage ($L_{\text{comp}}[m] \to 0$) switches on exponentially in time, with a rate proportional to $\log(p^{-n_A/2})$. Any decrease in p (from better qubits) or increases in n_A (from additional ancillas surrounding the leaked qubit in a QEC code) will serve to increase, and not decrease this rate. The exponential decay constant is inversely proportional to the leakage rate (as this corresponds to an initial HMM skepticism towards unlikely leakage events). However, as the likelihood 'switch' is exponential, a decrease in p_{leak} by even an order of magnitude should only increase the time before definite detection by a single step or so. The above analysis is complicated in a real scenario, as single physical errors give correlated detection signals, and as leakage may occur at any time, and as leaked qubits may seep. Correlations in the detection signals will serve to renormalize the switching time λ (but not remove the generic feature of exponential onset). Seepage causes individual leakage events to be finite (with some average lifetime T_{seep}); an individual leakage event of length $\ll \lambda^{-1}$ will not be detectable by the HMM. However, when the system returns to the computational subspace in such a short period of time, the leakage event may be treated as a 'regular' error, and does not need complicated leakage-detection hardware for fault tolerance. For example, a leakage event followed by immediate decay to $|1\rangle$ is indistinguishable from a direct transition to $|1\rangle$ for all practical purposes in QEC.

6.6 Additional Figures

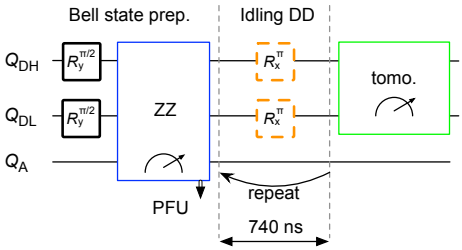


Figure 6.8: Quantum circuit for Bell-state idling experiments under dynamical decoupling.

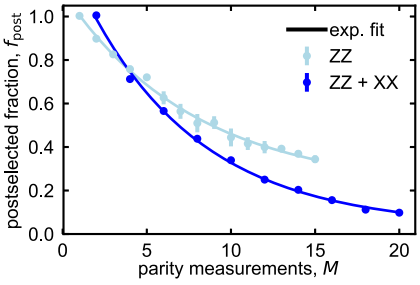


Figure 6.9: Postselected fractions for the 'no error' conditioning in Figures 6.2 and 6.4.

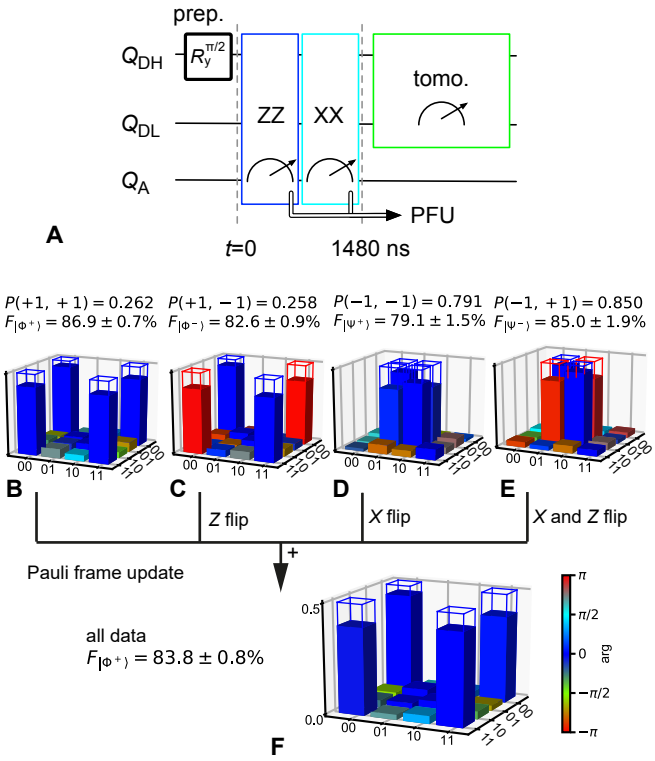


Figure 6.10: Generating entanglement by sequential ZZ and XX parity measurements and PFU. (**A**) Simplified quantum circuit for preparation, ZZ and XX measurements, sequential data-qubit state tomography and PFU. (**B** to **E**) Manhattan-style plots of the reconstructed data-qubit density matrix conditioned on the ancilla measurement outcomes with occurrence and fidelity to the four expected Bell states. (**F**) We use the two-bit outcome of the parity checks to apply a PFU that transforms all runs ideally to $|\Phi^+\rangle$. Frames on the tomograms indicate the Bell states ideally produced.

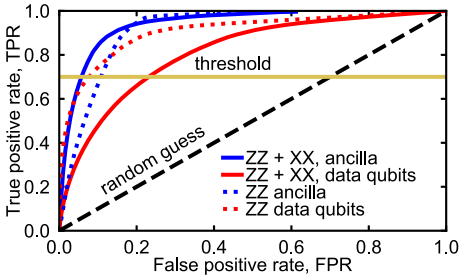


Figure 6.11: Receiver operating characteristics (ROCs) for mitigation of data-qubit and ancilla leakage during interleaved ZZ and XX checks. Data-qubit and ancilla leakage are each discerned via a dedicated HMM (full curves). For comparison, the ROCs for the HMMs for repeated ZZ checks only are also shown (dotted curves, same data as in Figure 6.3F).

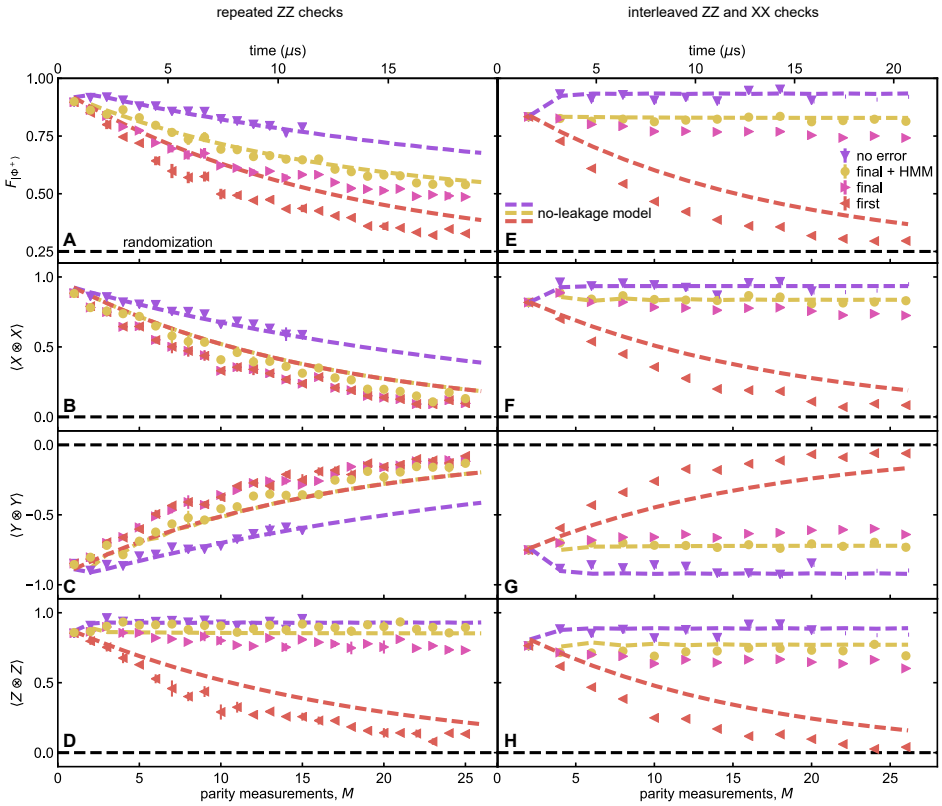


Figure 6.12: Comparison of experimental data and no-leakage modeling of the repeated parity check experiments of Figures 6.2 and 6.4. Simulations use the independently measured T_2^{echo}, T_1, e_a, e_{SQ}, e_{CZ}) of Table 6.1. This modeling uses two-level systems (no leakage) following Ref. [38], which uses quantumsim [175]. As expected, the modeling is outperforming the experiment for 'first' and 'final' correction strategies as the modeling does not include leakage. It however shows an excellent matching for the 'no error' conditioning (which rejects both qubit errors and leakage). The 'final + HMM' is excellently matching the 'final' modeling curve, confirming the leakage detection capability of the HMMs.

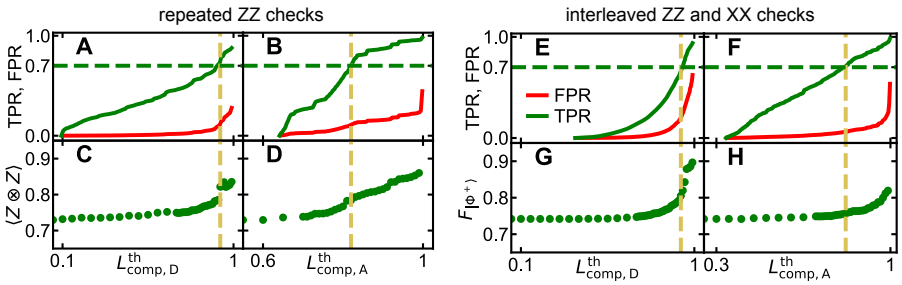


Figure 6.13: Leakage mitigation for the repeated parity check experiments as a function of the chosen threshold. (**A**) [(**B**)] TPR, FPR as a function of the chosen computational-space likelihood threshold for the repeated parity check experiments of Figures 6.2 and 6.3 for data-qubit leakage [ancilla leakage] at $M = 25$. (**C**) [(**D**)] The improvement in repeated ZZ checks is expressed as the increase in $\langle Z \otimes Z \rangle$ for data-qubit leakage [ancilla leakage]. Horizontal dashed lines indicate the chosen threshold $\mathrm{TPR} = 0.7$ (Figure 6.3, F and G) and vertical dashed lines indicate the accompanying computational-space likelihoods. (**E**-**H**) Similar plots for leakage rejection for interleaved ZZ and XX checks (Figure 6.4) at $M = 26$. The protocol improvement is here expressed as an increase of $F_{|\Phi^+\rangle}$.

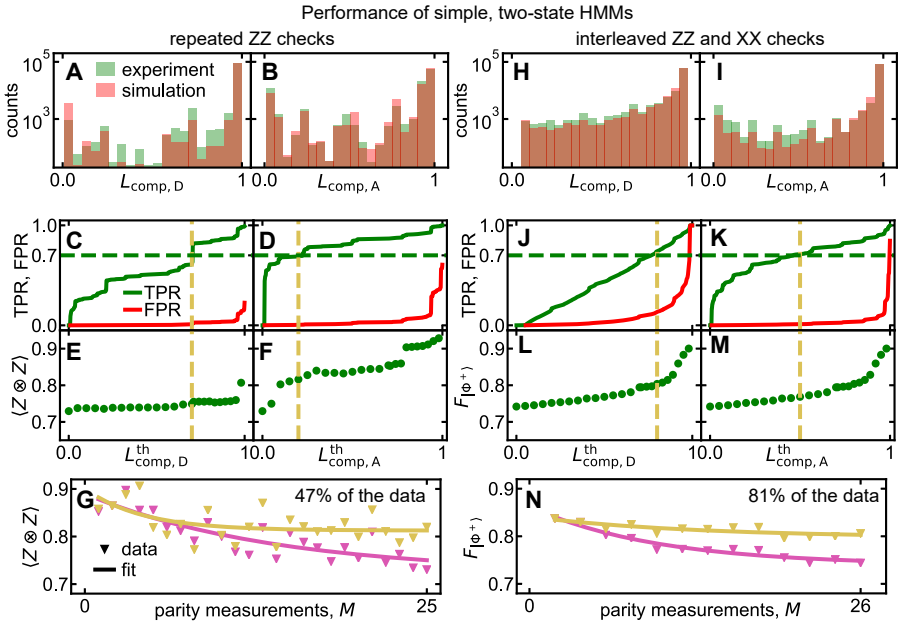


Figure 6.14: Leakage mitigation for the simple, two-state HMMs for repeated parity check experiments as a function of the chosen threshold. (**A**) [(**B**)] Histograms of 10^5 $\vec{M}_A$ with $M = 25$ for repeated ZZ checks (as in Figure 6.3D [Figure 6.3E]). HMM training suggests 3.6% [20%] total data-qubit [ancilla] leakage at $M = 25$. (**C**) [(**D**)] TPR, FPR as a function of the chosen computational-space likelihood threshold for the repeated parity check experiments of Figure 6.2 for data-qubit leakage [ancilla leakage] at $M = 25$. (**E**) [(**F**)] The improvement in repeated ZZ checks is expressed as the increase in $\langle Z \otimes Z \rangle$ for data-qubit leakage [ancilla leakage]. Horizontal dashed lines indicate the chosen threshold TPR $= 0.7$ and vertical dashed lines indicate the accompanying computational-space likelihoods (as in Figure 6.13). (**G**) $\langle Z \otimes Z \rangle$ after M ZZ checks and correction based on the 'final' outcomes, without (same data as in Figure 6.2D) and with leakage mitigation by postselection (TPR $= 0.7$). (**H**-**M**) Similar plots for simple-HMM leakage rejection for interleaved ZZ and XX checks (Figure 6.4) at $M = 26$. (**N**) $F_{|\Phi^+\rangle}$ after M interleaved checks and correction based on the 'final' outcomes, without (same data as in Figure 6.4) and with leakage mitigation by postselection (TPR $= 0.7$). The protocol improvement (L, M and N) is here expressed as an increase of $F_{|\Phi^+\rangle}$.

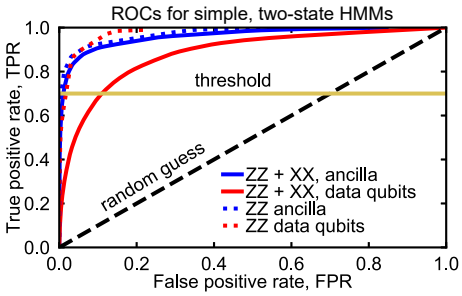


Figure 6.15: Receiver operating characteristics (ROCs) for leakage mitigation as in Figure 6.11, but using simple two-state HMMs.

7 CONCLUSIONS AND OUTLOOK

Transmon qubits are promising building blocks for fault-tolerant quantum computers. The work presented in this thesis advances these qubits towards this goal, mostly by focusing on the ability to read out qubits in a fast, efficient and scalable manner. Ultimately, we implemented an error correction scheme that not only detects arbitrary qubit errors but also detects leakage from the qubit computational space. In this chapter, we summarize the results. Then, the implications of these results for fault-tolerant quantum computing are investigated by simulating a 17-qubit quantum processor with the performance of the three-qubit device in Chapter 6.

7.1 Summary

The work in this thesis realizes several contributions to fault-tolerant quantum computing with superconducting circuits.

Important aspects of readout have been addressed by:

- speeding up readout by actively removing measurement photons after ancilla readout to ready the ancilla for the following quantum error correction cycle,
- speeding up readout by using the qubit for the characterization and optimization of the readout amplification chain,
- speeding up readout by the use of dedicated Purcell filters, while avoiding spurious dephasing of untargeted qubits during measurement,
- using repetitive protocols for gate tuning, measurement benchmarking and parity-check benchmarking,

This has allowed us to reach important milestones towards protecting logical information in a fully fault-tolerant quantum computer:

- performing repeated parity measurements with a high fidelity and a short cycle time to protect a state from arbitrary qubit errors (X, Y and Z),
- proposing and demonstrating the first protocol for the mitigation of leakage to non-computational states during repetitive quantum error correction.

The above results show that an architecture has been put together and tested with all necessary components for the preservation of logical information with larger numbers of qubits. An important next milestone is to preserve a logical qubit in a distance-three layout (counting the minimum number of single-qubit gates required for a logical operation), where logical information is protected from a single arbitrary error in the lattice. A widely pursued implementation of this goal is the Surface-17 layout, with nine data qubits and eight ancillas [37].

7.2 The projected performance of Surface-17 [176, 177]

Over the course of this thesis, several Surface-17 devices have been characterized. However, the performance of these devices was not yet good enough to perform quantum error correction. This was mainly caused by reduced coherence times and inaccuracy in frequency targeting in these larger devices. At the time of writing, fortunately, these issues have been largely resolved, with reported coherence times (T_1, T_2^{echo}) in the range $50 - 100$ μs for seven-qubit devices. Performing a quantum error correction experiment on these devices however remains an outstanding challenge.

Towards the realization of Surface-17, detailed simulations have been performed to predict its performance. As input for these simulations, experimentally obtained parameters are used such as single-qubit gate errors, two-qubit gate errors, measurement errors and qubit

coherence times. A high level of detail is reached by modeling the processor using full-density-matrix simulations [38]. These simulations indicate that with currently obtained errors and coherence times of $\sim$ 30 μs, the logical qubit outperforms the constituent physical qubits, i.e. its performance is beyond the memory break-even point. However, an important omission in these simulations is the assumption of having perfect two-level systems, i.e. no leakage to higher levels. The investigations in Chapter 6 have shown that, even if leakage errors are small compared to other error sources, the leakage build-up over multiple QEC rounds, has a significant effect on the overall performance. This strongly motivates the inclusion of leakage into the error models.

Leakage has been included in the theoretical studies of Ref. [176]. As a first step, to test against an experiment, the 3-qubit experimental results of Chapter 6 were reproduced. Next, we assess the performance of a Surface-17 logical qubit (Figure 7.1) given the achieved state-of-the-art parameters and compare the decay of logical information to the physical-qubit $T_1 = T_\phi/2 = 21$ μs. With these parameters, the best possible decoder or 'upper bound (UB)' decoder, would perform close to the memory break-even point. A realistic and often-used minimum-weight-perfect-matching (MWPM) [22] decoder would however fall significantly below this threshold. Evidently, improved qubit coherence and/or the mitigation of leakage are required to experimentally demonstrate the memory break-even point.

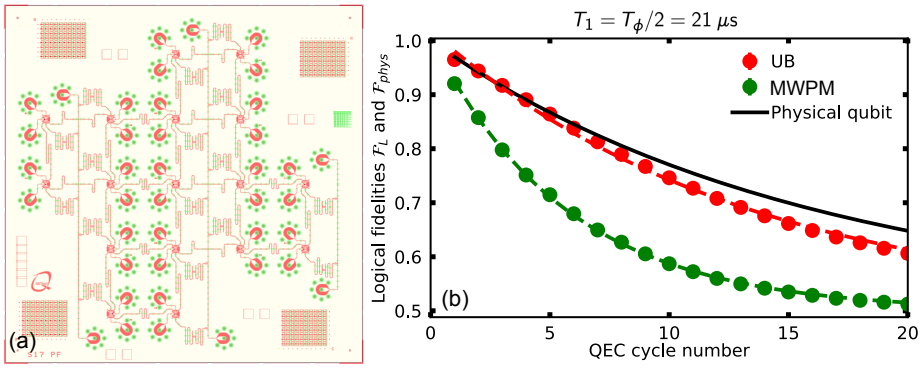


Figure 7.1: Layout and simulated logical qubit performance in Surface-17, based on achieved experimental parameters in Chapter 6. (a) Layout of a Surface-17 device with dedicated Purcell filter resonators for each qubit. Image obtained from [77, 178]. (b) Simulation of the performance of Surface-17, assuming the error model parameters fitting the Surface-3 experiment, when decoding with a MWPM decoder (green) and the decoder upper bound (red). The device itself performs just below the break-even point, where it would outperform the constituent physical qubits with an average decay rate $T_1 = T_\phi/2 = 21$ μs.

7.3 Leakage mitigation in Surface-17 [176, 177]

To improve the simulated performance, the leakage mitigation strategy with hidden Markov models (HMMs) (introduced in Chapter 6), was extended for this larger experiment. There still

is one HMM per qubit, but the data-qubit HMMs now have measurement result inputs from up to four neighboring ancillas. These additional inputs significantly improve the detectability of data-qubit leakage, compared to the single-ancilla results.

Most importantly, we analyze how HMM leakage mitigation improves the logical qubit performance by sweeping the leakage-likelihood threshold (Figure 7.2). Without postselection, the logical qubit for the MWPM decoder outperforms the physical qubit (with $T_1 = T_\phi/2 = 30\ \mu$s). When sweeping the threshold, MWPM reaches the break-even point at a post-selection fraction of $\sim 45\%$. This indicates that indeed, a logical qubit can be made with currently achieved experimental performance, even when it suffers from leakage.

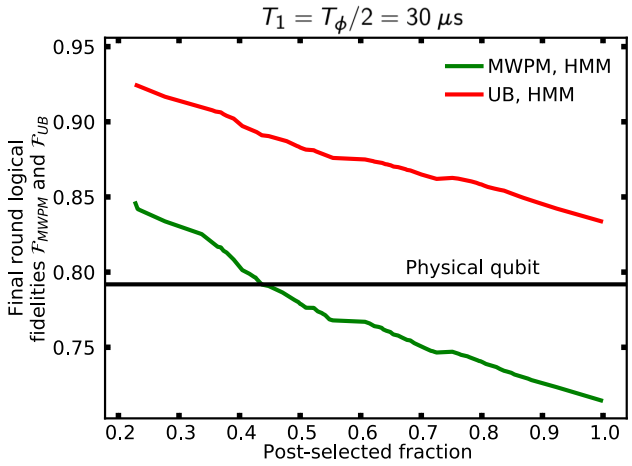


Figure 7.2: Simulated logical qubit performance in Surface-17 with leakage mitigation after 20 QEC rounds. Simulation of the performance of the UB decoder (red) and the MWPM decoder (green) as indicated by the achieved logical fidelity in the final round as a function of postselection of data-qubit leakage over a range of thresholds both utilizing the predictions given by the HMMs. Physical qubits were simulated with $T_1 = T_\phi/2 = 30\ \mu$s. Other parameters are obtained from Chapter 6.

Leakage detection presents an important advance for quantum error correction. Leakage mitigation by postselection presents the most straightforward way of handling the degrading effects that leakage has on the performance of the logical qubit. However, postselection is only viable for near-term experiments as the postselected fraction reduces linearly with the number of qubits. As such, the integration of the HMM leakage detection with MWPM and real-time correction schemes is an important task. This is difficult as both the decoder and the HMM have to keep up with the experimental cycle time of ~ 500 ns. Fortunately for the HMM, it has little complexity compared to the MWPM decoder itself, making it a small overhead.

ACKNOWLEDGEMENTS

The past years have formed me in many ways, both professionally and privately. It goes without saying that there are many people that played an important role in this. The list is long, so please forgive me if I forget some of you. First, I'd like to express my deepest gratitude to my parents **Saskia** and **Kees**. You have unconditionally supported me throughout my studies and career, and have actively stimulated me to go 'back to school' for this degree.

My fascination for quantum computing started in 2012 when I was scouting for a final Master thesis project. Before, my interests were scattered from general engineering, to business consulting to even a flirt with finance. I was invited by **Julia** to take a look in the quantum lab where she just started her project. I met **Leo** and **Diego** and was immediately excited by the drive and enthusiasm that the lab was breathing. Two days later, I started the project. During this initial 9 months with the group, I became more and more drawn by the art of experimental physics and was gradually intrigued by controlling nature at such a fundamental level. I had never imagined that science could be so alluring and addictive.

Leo, I want to thank you for offering me the position and being my supervisor, mentor and friend. I admire your drive for perfection, your highest standards in scientific integrity and never-ending enthusiasm as a teacher. The countless sessions preparing a cooldown, discussing the course of an experiment and fine tuning a paper have each brought me one step closer to the finish line. I'm grateful for the many opportunities you gave me to take part in major scientific events around the world. They almost make me forget about the blood, sweat and tears that went into our monthly love letters to IARPA. I also want to thank you for supporting Jules and me in starting a company in a very early stage. Without your openness to this idea we would not have taken off.

I want to thank all comity and ceremony members for their critical review. **Prof. VanderSypen**, thanks for being my second promotor and for being a continuous observer of this PhD project. Moreover, I want to thank the independent comity members for their effort to take part in the opposition of this thesis, **prof. M. Möttönen**, **prof. B. Huard**, **prof. D.P. DiVincenzo**, **prof. S.D.C. Wehner** and **prof. G.A. Steele**. On the defending side, I want to thank **Joris**. You have been one of my dearest friends since high school. The countless Mario party sessions, sailing trips and questionable party locations form an indestructible fundament for our friendship. As hard as it sometimes is to see each other, so easy it is to pick up where we left when we do. I was proud to support the defense of your dermatology thesis and happy you are my paranymph now. **Jules**, our friendship has always involved exploration. Often, this meant finding a team of like-minded to travel to places around the world, from the forests of Sumatra, to the high plateaus of Kyrgyzstan, to the savanne of Uganda. During these trips, we explored ideas to one day start a company. Halfway this PhD, we decided Qblox was it. I

thank you for taking the first step into this adventure by joining the lab. It is a great pleasure to explore the world with you in this new form.

Life was not always easy in the lab as the standards were high and the heat was always on. But during our scientific endeavors, there was always a sense of companionship, trust and mutual support. I want to thank you all for the time I have had in the DiCarlo lab. **Adriaan**, I was lucky to have you as my scientific twin brother. Working together was however sometimes conflicting. You setting up structure to build a more sustainable experiment, and me, recklessly cutting corners towards our end goal. Zooming out, this was a great combination. I'm happy that you will stay in quantum and hope that we can keep bundeling our forces. **Xiang**, you were the bridge between computer science and physics. You have taught me a lot about computing and have changed my vision on how to build one. All the best with running your own group. I will never forget stepping into your 'spicy nightmare' dinner party! **Tom**, thinking about working together on the depletion project back in the day makes me smile. Especially, the mismatch between experiment and simulation that was hunting us for so long, which appeared to be rooted in a trivial miscommunication. Your rapid transition towards an assistant professor still amazes me. **Rene**. I'm thankful that you joined for a Master's project. Your independence in figuring out the Purcell filter modeling and design was remarkable. You were the right man at the right time. I hope to be working with you again at some point. **Brian**, experimentalists and theorists are usually comfortable in their own domain. You baffled everyone and crushed my ego, showing how easy it was for you to switch sides. I'll never forget our bike ride and beers around the SF bay area. **Jacob**, Thanks for being a central figure in the team. You were not only oil to the FPGA machinery but also to the team and you often aligned the noses to reach our monthly goals. **Miguel**, splendid how you have picked up the work on the CC. You naturally picked up the bridge function between physics and computer science and even master the art of making a perfect Bacalhau. **Raymond, Raymond and Marijn** Thanks for your support and wisdom. It is almost never that you help someone in the way they expect. It is mostly your thorough questioning that brings up the real problem that helps research to move forward. **Wouter, Martin, Nadia** and **Duije**, your immense knowledge of computer science and microwave engineering have pushed the group and my personal interest in new directions. I want to thank you for teaching me a lot and hope to continue working with you. **Florian**, thanks for all the fun and sarcasm in the lab. Your no-nonsense attitude and iron motivation to keep fabricating those quirky qubits: great. Licking the back of peoples head after too many drinks: not so great. I'm looking forward to making more hikes in Oregon and bike rides in 'het Westland'. **Thijs**, great to see how you have inherited Flo's love for the quirky nanowire qubits. If someone can make them work besides Flo, it's you. **Alessandro**, thanks for all your pirating in the clean room. Without your creativity, there would not have been a single experiment in this thesis. **Nandini**, thanks for your persistence. You have played a crucial role in the Purcell-chip fabrication. I wish you all the best in turning your hard work into a PhD soon. **Marc**, thanks for being the cowboy in chip design software. Your new angle has greatly contributed to the Purcell filter chip. **Ramiro**, you have reached fantastic results doing almost all aspects of experimental quantum computing, from chip design to algorithm. I wish you the best in rounding up the thesis and with

your continuation in Barcelona. **Xavier**, thanks for your lightness and the fun times in the lab. **Slava and Boris**, thanks for deeply diving into the leakage detection simulations. You have shown how well our method scales. **Stefano**, you were the light spirit in the lab. I hope your life and career is taking good shape in the SF bay area. **Chris**, thanks for being the creative spirit in the lab. In journal clubs you always enlightened us with your deep understanding of what the work was actually about and what it could have been. At the same time, you keep entertaining us with your utterly dark humor. Your live performance of rage against the machine on your graduation will forever echo my brain. Epic. **Nathan**, thanks for being a patient teacher with me and Adriaan. I vividly remember how you took the time and effort to help us understand the processes in La Maserati at every stage of the cooldown. I wish you all the best in running your lab with J.P. in Australia. **Diego**, although we did not work together during the PhD, I still want to thank you for initiating me during my Master's. I often think about the time we sat next to La Ferrari in F033 during long days. Your defense in 2014 and your staggering achievements, triggered me to start this PhD project. **Marios**, sad you have left the group but fortunately, you found a better fit for your PhD. I'm happy this lead to such fantastic results. **Hany, Jorge, Matt**, good to see you guys are picking up the torch so nicely. I'm impressed by your latest presentations. I'm looking forward to see how easy it actually is to run the Surface-17 experiment.

Dave, thanks for combining work with an awesome amount of fun. The time with you in NL was really nice, and of course together representing Rotterdam. Good times! **Jeanette, Roman, Lester** it was fantastic to have you come over many times. Also, many thanks for welcoming to Anne-Marije and me during our visits in Portland. I wish you and the whole Intel team all the best!

I want to express my gratitude to **the leadership and support staff of QuTech** and congratulate them with a successful transformation. Within the past few years, the Quantum Transport research group has grown to become a professional institute with several hundreds of members. At the same time, from an almost pure scientific focus, QuTech has embraced a diversification towards engineering and now also business. Despite the natural distancing of people that the growth brings along, the institute has managed to hold onto its family feeling and traditions. I aspire to remain part of this growing ecosystem of science, engineering and business for many years to come. **Jérémy, Garrelt, Thorsten, Nora**, thanks for seeing something in Qblox from early on. Your trust means the world for our first stage. **Kees, Freeke, Marieke, Anne, Renate**, you have pushed forward the spin-out of Qblox in many ways. Thanks to your hard work Qblox is now a separate entity but with strong and durable connection to QuTech. I'm happy to be part of the now quickly growing Quantum Delft ecosystem.

Julia, Suus, Bas, Max, It was good to have 'old friends' walking around at QuTech. Your choices for a PhD have inspired me (although with some latency).

Apart from colleagues, the support of family and friends has kept me motivated throughout the years. **Roos**, thanks for being my caring older sister and leading me the way to Delft (and later Rotterdam). It is nice to have you, **Bas**, **Noor** and **Jolijn** so close by. **Abel, Willeke, Floor, Karel, Joep, Irene, Laura, Marijke, Joke**, thanks for being a warm and loving family.

Joke, although you are strictly not my mother in law (yet), I'm happy you have become part of my family over the past years. The family surf trips to Bonaire were the best possible contrast to the PhD work. Actually, not always contrast as Chapter 6 was partially written on the island. **Maria & Cees**, thanks for your routinely expressed interest in and support for my work in Delft. You are always very welcome.

I want to thank the many friends for bringing a lot of support but also for the necessary distraction. **Jan Willem**, together we started our physics adventures and first marathon. Although we now live on either side of the country, we should keep up with our long evenings of many beers and deep conversations. BOZ, Roffa Fawaka, Rode Pot, Denktank Tijgers, you make my world: **Joris, Marlot, Yvette, Dorien, Ellen, Luuk, Katrien, Max, Iris, Julia, Lex, Andries, Nadia, Jules, Perijne, Bush, Alessandra, Jean-Luc, Jotte, Kees, Ida, Eva, Bernold, Bram, Sophie, Nienke, Maaike, Aster, Ivanka, Josje**. **Buren van de Claes de Vrieselaan**, thanks for being the most awesome neighbors. From spontaneous Friday night drinks to the battle against speeding cars in the street. **Wouter**, thanks for the many times we worked together on the houses. Your design work is putting an important stamp on Qblox and on this thesis. Keep on swinging your mighty brand 'lasso'. **The Qblox team; Jules, Wouter, Jordy, Marijn, Callum, Maxime, Jay, Victor**. Happy to be making the greatest technology with such a talented, loyal and enthusiastic team. Full steam ahead!

Last but certainly not least, **Anne-Marije**, thank you for transforming me into the happiest man alive. I love driving home with you every day (although often a bit passed the planned time). You have supported me and this thesis in many ways. It is great to have you as my sparring partner. I'm happy you have started your PhD and hope I am bringing you the same support. I look forward to our future together.

REFERENCES

[1] R. P. FEYNMAN. *Simulating physics with computers*. International Journal of Theoretical Physics, **21** (6-7), 467–488, 1982. (see page: 1)

[2] 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: 1)

[3] L. K. GROVER. *Quantum mechanics helps in searching for a needle in a haystack*. Phys. Rev. Lett., **79**, 325, 1997. (see page: 1)

[4] A. W. HARROW, A. HASSIDIM, and S. LLOYD. *Quantum algorithm for linear systems of equations*. Phys. Rev. Lett., **103**, 150 502, 2009. (see page: 1)

[5] F. BLOCH. *Nuclear induction*. Phys. Rev., **70**, 460–474, 1946. (see page: 3)

[6] M. BORN. *Zur quantenmechanik der stoßvorgänge*. Zeitschrift für Physik, **37** (12), 863–867, 1926. (see page: 3)

[7] B. HENSEN, H. BERNIEN, A. E. DRÉAU, A. REISERER, N. KALB, M. S. BLOK, J. RUITENBERG, R. F. VERMEULEN, R. N. SCHOUTEN, C. ABELLÁN, *et al.* *Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres*. Nature, **526** (7575), 682–686, 2015. (see page: 3)

[8] 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)

[9] T. D. LADD, F. JELEZKO, R. LAFLAMME, Y. NAKAMURA, C. MONROE, and J. L. O'BRIEN. *Quantum computers.* Nature, **464** (7285), 45–53, 2010. (see page: 3)

[10] D. P. DIVINCENZO. *The physical implementation of quantum computation*. Fortschritte der Physik, **48**, 771, 2000. (see page: 3)

[11] 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. Nature Physics, **14** (6), 595–600, 2018. (see page: 4)

[12] 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 page: 4)

[13] J. PRESKILL. *Quantum computing in the NISQ era and beyond*. Quantum, **2**, 79, 2018. (see page: 4)

[14] P. W. SHOR. *Scheme for reducing decoherence in quantum computer memory*. Phys. Rev. A, **52**, R2493–R2496, 1995. (see page: 4)

[15] A. R. CALDERBANK and P. W. SHOR. *Good quantum error-correcting codes exist*. Phys. Rev. A, **54**, 1098, 1996.

[16] A. M. STEANE. *Error correcting codes in quantum theory*. Phys. Rev. Lett., **77**, 793, 1996.

[17] S. B. BRAVYI and A. Y. KITAEV. *Quantum codes on a lattice with boundary*. arXiv:quant-ph/9811052, 1998. (see page: 5)

[18] R. RAUSSENDORF and J. HARRINGTON. *Fault-tolerant quantum computation with high threshold in two dimensions.* Phys. Rev. Lett., **98**, 190 504, 2007. (see pages: 4, 70)

[19] A. G. FOWLER, M. MARIANTONI, J. M. MARTINIS, and A. N. CLELAND. *Surface codes: Towards practical large-scale quantum computation.* Phys. Rev. A, **86**, 032 324, 2012. (see pages: 4, 5, 38, 70, and 73)

[20] C. GIDNEY and M. EKER. *How to factor 2048 bit rsa integers in 8 hours using 20 million noisy qubits.* arXiv:1905.09749, 2019. (see page: 4)

[21] R. VERSLUIS, S. POLETTO, N. KHAMMASSI, B. 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**, 034 021, 2017. (see pages: 5, 54, 67, and 78)

[22] 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 pages: 5, 26, and 103)

[23] 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**, 042 319, 2007. (see pages: 5, 13, 14, 16, 18, and 54)

[24] H. PAIK, D. I. SCHUSTER, L. S. BISHOP, G. KIRCHMAIR, G. CATELANI, A. P. SEARS, B. R. JOHNSON, M. J. REAGOR, L. FRUNZIO, L. I. GLAZMAN, S. M. GIRVIN, M. H. DEVORET, and R. J. SCHOELKOPF. *Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture.* Phys. Rev. Lett., **107**, 240 501, 2011. (see pages: 5, 15)

[25] 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, 15, and 79)

[26] 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: 45)

[27] A. DEWES, R. LAURO, F. R. ONG, V. SCHMITT, P. MILMAN, P. BERTET, D. VION, and D. ESTEVE. *Quantum speeding-up of computation demonstrated in a superconducting two-qubit processor*. Phys. Rev. B, **85**, 140 503, 2012.

[28] E. LUCERO, R. BARENDS, Y. CHEN, J. KELLY, M. MARIANTONI, A. MEGRANT, P. O'MALLEY, D. SANK, A. VAINSENCHER, J. WENNER, T. WHITE, Y. YIN, A. N. CLELAND, *et al. Computing prime factors with a Josephson phase qubit quantum processor*. Nat. Phys., **8**, 719, 2012.

[29] 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 pages: 6, 22)

[30] 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: 6, 45, 70, and 80)

[31] 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**, 050 507, 2012. (see pages: 6, 15, 22, and 54)

[32] 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**, 050 506, 2012. (see pages: 22, 54)

[33] 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**, 190 504, 2014. (see pages: 6, 18, 19, 22, 26, 54, 60, and 70)

[34] D. RISTÈ, S. POLETTO, 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. Comm., **6**, 6983, 2015. (see pages: 6, 22, 30, and 70)

[35] J. KELLY, R. BARENDS, A. FOWLER, A. MEGRANT, E. JEFFREY, T. WHITE, D. SANK, J. MUTUS, B. CAMPBELL, Y. CHEN, *et al. State preservation by repetitive error detection in a superconducting quantum circuit*. Nature, **519** (7541), 66–69, 2015. (see pages: 6, 18, 19, 22, 23, 71, and 73)

[36] A. D. CÓRCOLES, 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. Comm., **6**, 6979, 2015. (see pages: 6, 22, and 70)

[37] Y. TOMITA and K. M. SVORE. *Low-distance surface codes under realistic quantum noise*. Phys. Rev. A, **90**, 062 320, 2014. (see pages: 6, 23, 30, 38, 77, and 102)

[38] T. O'BRIEN, B. TARASINSKI, and L. DICARLO. *Density-matrix simulation of small surface codes under current and projected experimental noise*. npj Quantum Information, **3** (1), 39, 2017. (see pages: 6, 23, 49, 73, 77, 92, 97, and 103)

[39] M. A. ROL. *Control for programmable superconducting quantum circuits*. Ph.D. thesis, TU Delft, 2020. (see page: 7)

[40] B. D. JOSEPHSON. *Possible new effects in superconductive tunnelling*. Phys. Lett., **1** (7), 251–253, 1962. (see page: 12)

[41] 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**, 062 320, 2004. (see pages: 12, 15, 22, 24, and 54)

[42] C. C. BULTINK. *First applications of feedback control of transmon qubits: fast reset and real-time detection of quasiparticle tunneling*. Master's thesis, TU Delft, 2012. (see page: 13)

[43] 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: 12)

[44] 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: 12)

[45] A. WALLRAFF, D. I. SCHUSTER, A. BLAIS, L. FRUNZIO, J. MAJER, M. H. DEVORET, S. M. GIRVIN, and R. J. SCHOELKOPF. *Approaching unit visibility for control of a superconducting qubit with dispersive readout*. Phys. Rev. Lett., **95**, 060 501, 2005. (see page: 12)

[46] P. BERTET, I. CHIORESCU, G. BURKARD, K. SEMBA, C. J. P. M. HARMANS, D. P. DIVINCENZO, and J. E. MOOIJ. *Dephasing of a superconducting qubit induced by photon noise*. Phys. Rev. Lett., **95**, 257 002, 2005. (see page: 12)

[47] F. YOSHIHARA, K. HARRABI, A. O. NISKANEN, Y. NAKAMURA, and J. S. TSAI. *Decoherence of flux qubits due to* $1/f$ *flux noise*. Phys. Rev. Lett., **97**, 167 001, 2006. (see page: 12)

[48] D. VION, A. AASSIME, A. COTTET, P. JOYEZ, H. POTHIER, C. URBINA, D. ESTEVE, and M. H. DEVORET. *Manipulating the quantum state of an electrical circuit*. Science, **296**, 886, 2002. (see page: 13)

[49] 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**, 110 501, 2009. (see pages: 14, 39)

[50] S. ASAAD, C. DICKEL, S. POLETTO, 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**, 16 029, 2016. (see pages: 40, 41)

[51] Z. CHEN, J. KELLY, C. QUINTANA, R. BARENDS, 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**, 020 501, 2016. (see pages: 14, 40, and 41)

[52] R. C. JAKLEVIC, J. LAMBE, A. H. SILVER, and J. E. MERCEREAU. *Quantum interference effects in Josephson tunneling*. Phys. Rev. Lett., **12**, 159–160, 1964. (see page: 14)

[53] 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**, 180 502, 2008. (see page: 14)

[54] W. D. OLIVER and P. B. WELANDER. *Materials in superconducting quantum bits*. MRS bulletin, **38** (10), 816–825, 2013. (see page: 15)

[55] F. LUTHI, T. STAVENGA, O. W. ENZING, A. BRUNO, C. DICKEL, N. K. LANGFORD, M. A. ROL, T. S. JESPERSEN, J. NYGÅRD, P. KROGSTRUP, and L. DICARLO. *Evolution of nanowire transmon qubits and their coherence in a magnetic field*. Phys. Rev. Lett., **120**, 100 502, 2018. (see page: 15)

[56] 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. Comm., **8**, 1715, 2017. (see page: 15)

[57] C. DICKEL, J. J. WESDORP, N. K. LANGFORD, S. PEITER, R. SAGASTIZABAL, A. BRUNO, B. CRIGER, F. MOTZOI, and L. DICARLO. *Chip-to-chip entanglement of transmon qubits using engineered measurement fields*. Phys. Rev. B, **97**, 064 508, 2018.

[58] 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), 2018. (see page: 15)

[59] 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), 044 003, 2017. (see page: 15)

[60] A. WALLRAFF, D. I. SCHUSTER, A. BLAIS, L. FRUNZIO, R.-S. HUANG, J. MAJER, S. KUMAR, S. M. GIRVIN, and R. J. SCHOELKOPF. *Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics*. Nature, **431**, 162–167, 2004. (see pages: 15, 54)

[61] 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**, 240 502, 2012. (see pages: 15, 29, and 38)

[62] D. RISTÈ, M. DUKALSKI, C. A. WATSON, G. DE LANGE, M. J. TIGGELMAN, Y. M. BLANTER, K. W. LEHNERT, R. N. SCHOUTEN, and L. DICARLO. *Deterministic entanglement of superconducting qubits by parity measurement and feedback*. Nature, **502** (7471), 350–354, 2013. (see pages: 15, 70, and 71)

[63] C. RIGETTI, J. M. GAMBETTA, S. POLETTO, B. L. T. PLOURDE, J. M. CHOW, A. D. CÓRCOLES, J. A. SMOLIN, S. T. MERKEL, J. R. ROZEN, G. A. KEEFE, M. B. ROTHWELL, M. B. KETCHEN, and M. STEFFEN. *Superconducting qubit in a waveguide cavity with a coherence time approaching 0.1 ms*. Phys. Rev. B, **86**, 100 506, 2012. (see page: 15)

[64] 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. Comm., **4**, 1913, 2013.

[65] K. SERNIAK, S. DIAMOND, M. HAYS, V. FATEMI, S. SHANKAR, L. FRUNZIO, R. SCHOELKOPF, and M. DEVORET. *Direct dispersive monitoring of charge parity in offset-charge-sensitive transmons*. Phys. Rev. Applied, **12**, 014 052, 2019. (see page: 15)

[66] E. T. JAYNES and F. W. CUMMINGS. *Comparison of quantum and semiclassical radiation theories with application to the beam maser*. Proceedings of the IEEE, **51** (1), 89–109, 1963. (see page: 16)

[67] M. D. REED, L. DICARLO, B. R. JOHNSON, L. SUN, D. I. SCHUSTER, L. FRUNZIO, and R. J. SCHOELKOPF. *High-fidelity readout in circuit quantum electrodynamics using the Jaynes-Cummings nonlinearity*. Phys. Rev. Lett., **105**, 173 601, 2010. (see page: 16)

[68] D. SANK, Z. CHEN, M. KHEZRI, J. KELLY, R. BARENDS, B. CAMPBELL, Y. CHEN, B. CHIARO, A. DUNSWORTH, A. FOWLER, *et al. Measurement-induced state transitions in a superconducting qubit: Beyond the rotating wave approximation*. Physical review letters, **117** (19), 190 503, 2016. (see pages: 16, 28, and 30)

[69] E. M. PURCELL, H. C. TORREY, and R. V. POUND. *Resonance absorption by nuclear magnetic moments in a solid*. Phys. Rev., **69**, 37–38, 1946. (see page: 18)

[70] Y. CHEN, D. SANK, P. O'MALLEY, T. WHITE, R. BARENDS, B. CHIARO, J. KELLY, E. LUCERO, M. MARIANTONI, A. MEGRANT, C. NEILL, A. VAINSENCHER, J. WENNER, *et al. Multiplexed dispersive readout of superconducting phase qubits*. Appl. Phys. Lett., **101** (18), 182601, 2012. (see pages: 18, 19)

[71] J. P. GROEN, D. RISTÈ, L. TORNBERG, 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**, 090 506, 2013. (see pages: 18, 19, and 22)

[72] 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), 203 110, 2010. (see pages: 18, 22)

[73] E. A. SETE, J. M. MARTINIS, and A. N. KOROTKOV. *Quantum theory of a bandpass purcell filter for qubit readout*. Phys. Rev. A, **92**, 012 325, 2015. (see page: 18)

[74] N. T. BRONN, Y. LIU, J. B. HERTZBERG, A. D. CÓRCOLES, A. A. HOUCK, J. M. GAMBETTA, and J. M. CHOW. *Broadband filters for abatement of spontaneous emission in circuit quantum electrodynamics*. Applied Physics Letters, **107** (17), 172 601, 2015. https://doi.org/10.1063/1.4934867. (see page: 18)

[75] T. WALTER, P. KURPIERS, S. GASPARINETTI, P. MAGNARD, A. POTOČNIK, Y. SALATHÉ, M. PECHAL, M. MONDAL, M. OPPLIGER, C. EICHLER, and A. WALLRAFF. *Rapid High-Fidelity Single-Shot Dispersive Readout of Superconducting Qubits*. Phys. Rev. Appl., **7** (5), 054 020, 2017. (see pages: 20, 60)

[76] 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**, 034 040, 2018. (see pages: 20, 70, 71, and 79)

[77] R. VOLLMER. *Fast and scalable readout for fault-tolerant quantum computing with superconducting qubits*. Master's thesis, TU Delft, 2018. (see pages: 20, 103)

[78] C. K. ANDERSEN, A. REMM, S. LAZAR, S. KRINNER, J. HEINSOO, J.-C. BESSE, M. GABUREAC, A. WALLRAFF, and C. EICHLER. *Entanglement stabilization using ancilla-based parity detection and real-time feedback in superconducting circuits*. npj Quantum Information, **5** (1), 69, 2019. (see pages: 20, 70, 71, and 77)

[79] 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. Comm., **5**, 11 526, 2016. (see page: 22)

[80] D. NIGG, M. MÜLLER, E. A. MARTINEZ, P. SCHINDLER, M. HENNRICH, T. MONZ, M. A. MARTIN-DELGADO, and R. BLATT. *Quantum computations on a topologically encoded qubit.* Science, **345** (6194), 302–305, 2014.

[81] 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: 22, 29, and 70)

[82] K. D. PETERSSON, L. W. MCFAUL, 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: 22)

[83] T. W. LARSEN, K. D. PETERSSON, F. KUEMMETH, T. S. JESPERSEN, P. KROGSTRUP, J. NYGÅRD, and C. M. MARCUS. *Semiconductor-nanowire-based superconducting qubit*. Phys. Rev. Lett., **115**, 127 001, 2015. (see page: 22)

[84] G. DELANGE, 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**, 127 002, 2015. (see page: 22)

[85] 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**, 070 502, 2014. (see pages: 22, 71, and 79)

[86] 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: 22)

[87] 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), 012 112, 2008. (see page: 22)

[88] F. YAN, S. GUSTAVSSON, A. KAMAL, J. BIRENBAUM, A. P. SEARS, D. HOVER, T. J. GUDMUNDSEN, D. ROSENBERG, G. SAMACH, S. WEBER, *et al. The flux qubit revisited to enhance coherence and reproducibility*. Nat. Comm., **7**, 12 964, 2016. (see page: 22)

[89] 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**, 180 504, 2012. (see page: 22)

[90] 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**, 240 501, 2015. (see page: 22)

[91] 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**, 011 001, 2016. (see pages: 22, 26, 38, 54, 57, and 71)

[92] 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: 22, 32)

[93] 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. Comm., **5**, 4015, 2014. (see page: 22)

[94] 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: 24, 33)

[95] 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**, 022 118, 2015. (see pages: 24, 54, 56, 57, and 61)

[96] E. MAGESAN, J. M. GAMBETTA, A. D. CÓRCOLES, and J. M. CHOW. *Machine learning for discriminating quantum measurement trajectories and improving readout*. Phys. Rev. Lett., **114**, 200 501, 2015. (see pages: 24, 54, 56, 57, and 61)

[97] 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**, 040 305, 2010. (see pages: 24, 40)

[98] M. REED. *Entanglement and quantum error correction with superconducting qubits*. PhD Dissertation, Yale University, 2013. (see page: 24)

[99] S. BOUTIN, C. K. ANDERSEN, J. VENKATRAMAN, A. J. FERRIS, and A. BLAIS. *Resonator reset in circuit QED by optimal control for large open quantum systems*. Phys. Rev. A, **96**, 042 315, 2017. (see page: 26)

[100] A. FRISK KOCKUM, L. TORNBERG, and G. JOHANSSON. *Undoing measurement-induced dephasing in circuit QED*. Phys. Rev. A, **85**, 052 318, 2012. (see pages: 28, 34, 56, 61, and 62)

[101] B. M. TERHAL. *Quantum error correction for quantum memories*. Rev. Mod. Phys., **87**, 307–346, 2015. (see pages: 29, 54, and 70)

[102] 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. Appl., **7**, 041 001, 2017. (see page: 29)

[103] 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**, 032 321, 2016. (see pages: 29, 38)

[104] E. JONES, T. OLIPHANT, P. PETERSON, *et al. SciPy: Open source scientific tools for Python*, 2001. [Online; accessed 06.04.2018]. (see pages: 32, 87)

[105] M. A. NIELSEN and I. L. CHUANG. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000. (see page: 34)

[106] S. BOIXO, S. V. ISAKOV, V. N. SMELYANSKIY, R. BABBUSH, N. DING, Z. JIANG, J. M. MARTINIS, and H. NEVEN. *Characterizing Quantum Supremacy in Near-Term Devices*. arXiv:1608.00263, 2016. (see page: 38)

[107] 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**, 062 304, 2016. (see page: 38)

[108] C. HORSMAN, A. G. FOWLER, S. DEVITT, and R. V. METER. *Surface code quantum computing by lattice surgery*. New J. Phys., **14** (12), 123 011, 2012. (see page: 38)

[109] J. M. MARTINIS. *Qubit metrology for building a fault-tolerant quantum computer*. npj Quantum Inf., **1**, 15 005, 2015. (see page: 38)

[110] 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**, 240 401, 2015. (see page: 38)

[111] D. J. EGGER and F. K. WILHELM. *Adaptive hybrid optimal quantum control for imprecisely characterized systems*. Phys. Rev. Lett., **112**, 240 503, 2014. (see page: 38)

[112] 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**, 240 504, 2014. (see pages: 38, 40, and 44)

[113] 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**, 034 008, 2016. (see pages: 39, 42, 54, 57, 70, and 71)

[114] 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: 38)

[115] M. H. DEVORET and R. J. SCHOELKOPF. *Superconducting circuits for quantum information: An outlook*. Science, **339** (6124), 1169–1174, 2013. (see page: 38)

[116] 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: 38)

[117] D. RISTÈ and L. DICARLO. Digital feedback in superconducting quantum circuits. In *Superconducting devices in quantum optics*. Springer, 2015. (see page: 38)

[118] 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. Comm., **8** (14485), 2017. (see page: 38)

[119] 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: 40, 51)

[120] E. MAGESAN, J. M. GAMBETTA, and J. EMERSON. *Scalable and robust randomized benchmarking of quantum processes*. Phys. Rev. Lett., **106**, 180504, 2011. (see page: 40)

[121] E. MAGESAN, J. M. GAMBETTA, and J. EMERSON. *Characterizing quantum gates via randomized benchmarking*. Phys. Rev. A, **85**, 042311, 2012. (see page: 40)

[122] 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 page: 42)

[123] J. A. NELDER and R. MEAD. *A simplex method for function minimization*. The Computer Journal, **7** (4), 308–313, 1965. (see page: 43)

[124] E. Magesan, private communication. (see page: 44)

[125] 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: 45)

[126] 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**, 150401, 2018. (see page: 45)

[127] 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 page: 45)

[128] M. ROL, C. DICKEL, S.ASAAD, N. LANGFORD, C. BULTINK, R. SAGASTIZABAL, N. LANGFORD, G. DE LANGE, X. FU, S. DE JONG, F. LUTHI, and W. VLOTHUIZEN. *PycQED*, 2016. (see page: 45)

[129] A. JOHNSON, G. UNGARETTI, *et al. QCoDeS*, 2016. (see page: 45)

[130] E. NIELSEN, T. SCHOLTEN, K. RUDINGER, and J. GROSS. *pyGSTi: Version 0.9.1 beta*, 2016. (see page: 51)

[131] D. P. DIVINCENZO. *Fault-tolerant architectures for superconducting qubits*. Physica Scripta, **2009** (T137), 014020, 2009. (see page: 54)

[132] A. A. CLERK, M. H. DEVORET, S. M. GIRVIN, F. MARQUARDT, and R. J. SCHOELKOPF. *Introduction to quantum noise, measurement, and amplification*. Reviews of Modern Physics, **82** (2), 1155–1208, 2010. 0810.4729. (see page: 54)

[133] M. A. CASTELLANOS-BELTRAN, K. D. IRWIN, G. C. HILTON, L. R. VALE, and K. W. LEHNERT. *Amplification and squeezing of quantum noise with a tunable Josephson metamaterial*. Nat. Phys., **4** (12), 929, 2008. (see page: 54)

[134] R. VIJAY, M. H. DEVORET, and I. SIDDIQI. *Invited review article: The Josephson bifurcation amplifier.* Rev. Sci. Instrum., **80**, 111 101, 2009.

[135] N. BERGEAL, F. SCHACKERT, M. METCALFE, R. VIJAY, V. E. MANUCHARYAN, L. FRUNZIO, D. E. PROBER, R. J. SCHOELKOPF, S. M. GIRVIN, and M. H. DEVORET. *Phase-preserving amplification near the quantum limit with a Josephson ring modulator*. Nature, **465** (7294), 64–68, 2010.

[136] 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), 263 513, 2014.

[137] C. EICHLER, Y. SALATHE, J. MLYNEK, S. SCHMIDT, and A. WALLRAFF. *Quantum-limited amplification and entanglement in coupled nonlinear resonators.* Phys. Rev. Lett., **113**, 110 502, 2014. (see page: 54)

[138] 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 pages: 54, 60, and 79)

[139] M. R. VISSERS, R. P. ERICKSON, H. S. KU, L. VALE, X. WU, G. C. HILTON, and D. P. PAPPAS. *Low-noise kinetic inductance traveling-wave amplifier using three-wave mixing*. Appl. Phys. Lett., **108**, 012 601, 2016. (see page: 54)

[140] 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**, 042 318, 2006. (see pages: 54, 56)

[141] R. VIJAY, D. H. SLICHTER, and I. SIDDIQI. *Observation of quantum jumps in a superconducting artificial atom.* Phys. Rev. Lett., **106**, 110 502, 2011. (see page: 54)

[142] M. HATRIDGE, S. SHANKAR, M. MIRRAHIMI, F. SCHACKERT, K. GEERLINGS, T. BRECHT, K. SLIWA, B. ABDO, L. FRUNZIO, S. GIRVIN, R. SCHOELKOPF, and M. DEVORET. *Quantum back-action of an individual variable-strength measurement*. Science, **339** (6116), 178–181, 2013. (see page: 54)

[143] Q. LIU, M. LI, K. DAI, K. ZHANG, G. XUE, X. TAN, H. YU, and Y. YU. *Extensible 3d architecture for superconducting quantum computing*. Appl. Phys. Lett., **110** (23), 232 602, 2017. (see page: 54)

[144] M. REAGOR, C. B. OSBORN, N. TEZAK, A. STALEY, G. PRAWIROATMODJO, M. SCHEER, N. ALIDOUST, E. A. SETE, N. DIDIER, M. P. DA SILVA, E. ACALA, J. ANGELES, A. BESTWICK, *et al. Demonstration of universal parametric entangling gates on a multi-qubit lattice*. Science Advances, **4** (2), 2018. https://advances.sciencemag.org/content/4/2/eaao3603.full.pdf.

[145] M. TAKITA, A. W. CROSS, A. D. CÓRCOLES, J. M. CHOW, and J. M. GAMBETTA. *Experimental demonstration of fault-tolerant state preparation with superconducting qubits*. Phys. Rev. Lett., **119**, 180 501, 2017. (see page: 70)

[146] 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. https://science.sciencemag.org/content/360/6385/195.full.pdf. (see page: 54)

[147] T. P. HARTY, D. T. C. ALLCOCK, C. J. BALLANCE, L. GUIDONI, H. A. JANACEK, N. M. LINKE, D. N. STACEY, and D. M. LUCAS. *High-fidelity preparation, gates, memory, and readout of a trapped-ion quantum bit*. Phys. Rev. Lett., **113**, 220 501, 2014. (see page: 70)

[148] C. J. BALLANCE, T. P. HARTY, N. M. LINKE, M. A. SEPIOL, and D. M. LUCAS. *High-fidelity quantum logic gates using trapped-ion hyperfine qubits.* Phys. Rev. Lett., **117**, 060 504, 2016. (see page: 70)

[149] 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**, 120 502, 2019. (see pages: 70, 71, and 80)

[150] H. BOMBIN and M. A. MARTIN-DELGADO. *Topological computation without braiding*. Phys. Rev. Lett., **98**, 160 502, 2007. (see page: 70)

[151] N. C. BROWN and K. R. BROWN. *Comparing zeeman qubits to hyperfine qubits in the context of the surface code: $^{174}Yb^{+}$ and $^{171}Yb^{+}$*. Phys. Rev. A, **97**, 052 301, 2018. (see page: 70)

[152] R. W. ANDREWS, C. JONES, M. D. REED, A. M. JONES, S. D. HA, M. P. JURA, J. KERCKHOFF, M. LEVENDORF, S. MEENEHAN, S. T. MERKEL, A. SMITH, B. SUN, A. J. WEINSTEIN, *et al. Quantifying error and leakage in an encoded Si/SiGe triple-dot qubit*. Nat. Nanotech., **14**, 747–750, 2019. (see page: 70)

[153] P. ALIFERIS and B. M. TERHAL. *Fault-tolerant quantum computation for local leakage faults*. Quantum Info. Comput., **7** (1), 139–156, 2007. (see pages: 70, 77, and 92)

[154] A. G. FOWLER. *Coping with qubit leakage in topological codes*. Phys. Rev. A, **88**, 042 308, 2013. (see page: 75)

[155] J. GHOSH and A. G. FOWLER. *Leakage-resilient approach to fault-tolerant quantum computing with superconducting elements*. Phys. Rev. A, **91**, 020 302, 2015. (see pages: 77, 92)

[156] 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: 70, 77, and 92)

[157] N. C. BROWN, M. NEWMAN, and K. R. BROWN. *Handling leakage with subsystem codes*. New Journal of Physics, **21** (7), 073 055, 2019. (see page: 70)

[158] Z. CAI, M. A. FOGARTY, S. SCHAAL, S. PATOMÄKI, S. C. BENJAMIN, and J. J. L. MORTON. *A Silicon Surface Code Architecture Resilient Against Leakage Errors*. Quantum, **3**, 212, 2019. (see page: 70)

[159] Y. LIU, S. SHANKAR, N. OFEK, M. HATRIDGE, A. NARLA, K. M. SLIWA, L. FRUNZIO, R. J. SCHOELKOPF, and M. H. DEVORET. *Comparing and combining measurement-based and driven-dissipative entanglement stabilization*. Phys. Rev. X, **6**, 011 022, 2016. (see pages: 70, 71, and 73)

[160] 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.

[161] R. HARPER and S. T. FLAMMIA. *Fault-tolerant logical gates in the ibm quantum experience*. Phys. Rev. Lett., **122**, 080 504, 2019. (see page: 70)

[162] L. HU, Y. MA, W. CAI, X. MU, Y. XU, W. WANG, Y. WU, H. WANG, Y. P. SONG, C.-L. ZOU, S. M. GIRVIN, L.-M. DUAN, and L. SUN. *Quantum error correction and universal gate set operation on a binomial bosonic logical qubit*. Nature Physics, 2019. (see page: 70)

[163] V. NEGNEVITSKY, M. MARINELLI, K. K. MEHTA, H.-Y. LO, C. FLÜHMANN, and J. P. HOME. *Repeated multi-qubit readout and feedback with a mixed-species trapped-ion register*. Nature, **563** (7732), 527–531, 2018. (see pages: 71, 73, and 77)

[164] E. KNILL. *Quantum computing with realistically noisy devices*. Nature, **434**, 39, 2005. (see page: 71)

[165] L. BAUM and T. PETRIE. *Statistical inference for probabilistic functions of finite state markov chains*. Ann. Math. Stat., **37** (6), 1554–1563, 1966. (see pages: 74, 84, and 87)

[166] H. AKAIKE. A new look at the statistical model identification. IEEE Transactions on Automatic Control, **19** (6), 716–723, 1974. (see page: 75)

[167] D. M. GREEN and J. A. SWETS. *Signal Detection Theory and Psychophysics*. John Wiley & Sons, New York, 1966. (see page: 75)

[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**, 060 502, 2018. (see pages: 77, 92)

[169] 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 page: 79)

[170] C. C. BULTINK, B. TARASINSKI, N. HAANDBAEK, S. POLETTO, N. HAIDER, D. J. MICHALAK, A. BRUNO, and L. DICARLO. *General method for extracting the quantum efficiency of dispersive qubit readout in circuit qed*. Appl. Phys. Lett., **112**, 092 601, 2018. (see page: 79)

[171] 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**, 080 505, 2012. (see page: 80)

[172] M. NEWVILLE, T. STENSITZKI, D. B. ALLEN, and A. INGARGIOLA. *LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python¶*, 2014. (see page: 83)

[173] 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**, 062 329, 2013. (see page: 93)

[174] X. WAINTAL. *What determines the ultimate precision of a quantum computer*. Phys. Rev. A, **99**, 042 318, 2019. (see page: 93)

[175] The quantumsim package can be found at `http://github.com/brianzi/quantumsim`. (see page: 97)

[176] B. VARBANOV. *Characterization, detection and mitigation of leakage in simulations of a superconducting-based implementation of Surface-17*. Master's thesis, TU Delft, 2019. (see pages: 102, 103)

[177] B. VARBANOV, B. TARSINSKI, V. OSTROUKH, C. BULTINK, and L. DICARLO. M38 update of logiq feasibility study: The qusurf logical qubit and its simulated performance with current physical-qubit parameters. M38 Update of LogiQ Feasibility Study. (see pages: 102, 103)

[178] M. W. BEEKMAN. *Superconducting Transmon Qubit Chip Design and Characterization through Electromagnetic Analysis*. Master's thesis, TU Delft, 2018. (see page: 103)

CURRICULUM VITÆ

Cornelis Christiaan BULTINK

Dec. 31, 1986	Born in Bergen op Zoom, the Netherlands.

Education

1999–2005	Secondary School - Gymnasium RSG 't Rijks, Bergen op Zoom, the Netherlands
2005-2012	Bachelor of Science and Master of Science in Applied Physics Delft University of Technology, Delft, the Netherlands *BSc. Thesis:* Analyse van de richtingskarakteristiek van Multi-Actuator-panels als basis voor het ontwerpen van filters voor golfveldsynthese *Supervisors:* Dr. ir. D. de Vries and Dr.-ing. L. Hörchens *MSc. Thesis:* First applications of feedback control of transmon qubits: fast reset and real-time detection of quasiparticle tunneling *Supervisors:* Dr. ir. D. Ristè and Prof. dr. L. DiCarlo
2015-2019	Ph.D. in experimental Physics Delft University of Technology, Delft, the Netherlands *Thesis:* Protecting quantum entanglement by repetitive measurement *Promotors:* Prof. dr. L. DiCarlo and Prof. dr. L. M. K. Vandersypen

Professional Experience

2011–2012	Member of the Dutch National Think Tank (de Nationale Denktank), Amsterdam, the Netherlands
2012–2015	Experimental Physicist at Mapper Lithography, Delft, the Netherlands
2018–present	Co-founder and CEO of Qblox, Delft, the Netherlands

Grants

3. Zuid-Holland, MIT R&D-samenwerkingsproject, development of a quantum control highway (2020)

2. NWO - take-off grant, Qblox - controlling a 1000-qubit quantum computer (2019)

1. Casimir - NanoFront Seed Money (2019)

Awards

2. Zurich Instruments Pioneer Award, (1500 CHF prize money) (2018)

1. Best paper Award-IEE/ACM microarchitecture, 2017 (MICRO-50) (2017)

Scientific Publications

10. M. A. Rol, L. Ciorciaro, F. K. Malinowski, B. M. Tarasinski, R. E. Sagastizabal, **C. C. Bultink**, Y. Salathe, N. Haandbaek, J. Sedivy, L. DiCarlo *Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor* , Applied Physics Letters, **116** 054001 (2020).

9. **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, L. DiCarlo, *Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements*, Science Advances, **6** 12 eaay3050 (2020).

8. 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, *A fast, low-leakage, high-fidelity two-qubit gate for a programmable superconducting quantum computer*, Physical Review Letters, **123** 120502 (2019).

7. 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*, Physical Review A, **100** 010302 (2019).

6. **C. C. Bultink**, B. Tarasinski, N. Haandbaek, S. Poletto, N. Haider, D. J. Michalak, A. Bruno and L. DiCarlo, *A general method for extracting the quantum efficiency of dispersive qubit readout in circuit QED*, Applied Physics Letters, **112** 092601 (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. and 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. 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*, Nature Communications **4**, 1913 (2013).

1. D. Ristè, **C. C. Bultink**, K. W. Lehnert, and L. DiCarlo, *Feedback control of a solid-state qubit using high-fidelity projective measurement*, Physical Review Letters **109**, 240502 (2012).

Non-Scientific Publications

1. J. C. van Oven, **C. C. Bultink**, *Op weg naar quantumvoordeel: qubits aansturen met duizend tegelijk*, Bits & Chips **4**, 32-33 (2019).