THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Scaling Superconducting Quantum Processors

Coherence, Frequency Targeting and Crosstalk

Amr Osman

CHALMERS
UNIVERSITY OF TECHNOLOGY

Quantum Technology Laboratory
Department of Microtechnology and Nanoscience (MC2)
Chalmers University of Technology
Gothenburg, Sweden, 2024

Scaling Superconducting Quantum Processors
Coherence, Frequency Targeting and Crosstalk

Amr Osman
ISBN 978-91-8103-139-3

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr. 5597
ISSN 0346-718X

Quantum Technology Laboratory
Department of Microtechnology and Nanoscience (MC2)
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Phone: +46 (0)31 772 1000

Printed by Chalmers Reproservice
Gothenburg, Sweden, November 2024

To my beloved sister Heba

Your memory lives forever in my heart

This is for you

Abstract

The advancement of quantum computing hinges on the scalability and performance of quantum processors. Superconducting qubits require precision engineering to achieve long coherence times and high gate fidelities. However, their performance remains constrained by challenges such as fabrication uncertainty, imperfections in materials, and unintended signal crosstalk, all of which impose significant limitations on scalability and reliability.

A significant portion of this work investigates the reproducibility of Josephson junctions, essential components of superconducting qubits. Variations in these junctions lead to deviations in qubit frequencies, degrading gate fidelity. A streamlined fabrication process using Patch-Integrated Cross-Type (PICT) junctions reduces the steps for junction fabrication while maintaining reproducibility and qubit coherence. Further improvements are achieved by optimizing the fabrication process and using slightly larger junction sizes, leading to a qubit frequency reproducibility of 40 MHz (1%) on a planar chip.

Qubit coherence, essential for maintaining quantum states and enabling error-free operations, is investigated from a material perspective. Two-level systems (TLS) at material interfaces are identified as dominant sources of decoherence. Through TLS spectroscopy, we probe the density of defects in the vicinity of the qubit and their presence within the junction's tunnel barrier. Additionally, we evaluate the impact of fabrication steps on the performance of superconducting circuit and improve the qubit quality factor up to 5 million.

Crosstalk, arising from unintended drive-line interactions, degrades gate fidelity. On the first-generation 25-qubit flip-chip quantum processing unit (QPU), the average drive-line crosstalk was benchmarked at 40 dB. The second-generation QPU enables studies on parameter targeting, including the influence of chip-to-chip spacing on qubit frequencies and crosstalk's impact on gate fidelity. Advanced frequency allocation strategies are introduced to account for fabrication uncertainties while maximizing frequency separation between neighboring qubits. While these approaches effectively mitigate gate collisions, our results emphasize the need for further suppression of crosstalk and active mitigation techniques to achieve higher fidelities in large-scale quantum processors.

Keywords: Superconducting qubits, quantum computing, Josephson junction reproduciblity, TLS spectroscopy, 3D integration, quantum processor

List of Publications

This thesis is based on the following publications:

[A] **A. Osman**, J. Simon, A. Bengtsson, S. Kosen , P. Krantz , D. P. Lozano, M. Scigliuzzo , P. Delsing, J. Bylander and A. Fadavi Roudsari, "Simplified Josephson-junction fabrication process for reproducibly high-performance superconducting qubits". Appl. Phys. Lett. 118, 064002 (2021).

[B] **Amr Osman**, Jorge Fernández-Pendás, Christopher Warren, Sandoko Kosen, Marco Scigliuzzo, Anton Frisk Kockum, Giovanna Tancredi, Anita Fadavi Roudsari, and Jonas Bylander, "Mitigation of frequency collisions in superconducting quantum processors". Phys. Rev. Research 5, 043001 (2023).

[C] J. Biznárová, **A. Osman**, E. Rehnman, L. Chaynun, C. Križan, P. Malmberg, M. Rommel, C. Warren, P. Delsing, A. Yurgens, J. Bylander, and A. Fadavi Roudsari, "Mitigation of interfacial dielectric loss in aluminum-on-silicon superconducting qubits". npj Quantum Inf 10, 78 (2024).

[D] Lert Chayanun, Janka Biznárová, Lunjie Zeng, Per Malmberg, Andreas Nylander, **Amr Osman**, Marcus Rommel, Pui Lam Tam, Eva Olsson, Per Delsing, August Yurgens, Jonas Bylander, and Anita Fadavi Roudsari, "Characterization of process-related interfacial dielectric loss in aluminum-on-silicon by resonator microwave measurements, materials analysis, and imaging". APL Quantum 1, 026115 (2024).

[E] Christopher W. Warren, Jorge Fernández-Pendás, Shahnawaz Ahmed, Tahereh Abad, Andreas Bengtsson, Janka Biznárová, Kamanasish Debnath, Xiu Gu, Christian Križan, **Amr Osman**, Anita Fadavi Roudsari, Per Delsing, Göran Johansson, Anton Frisk Kockum, Giovanna Tancredi and Jonas Bylander , "Extensive characterization and implementation of a family of three-qubit gates at the coherence limit". npj Quantum Information 9, 44 (2023).

[F] Sandoko Kosen, Hang-Xi Li, Marcus Rommel, Daryoush Shiri, Christopher Warren, Leif Grönberg, Jaakko Salonen, Tahereh Abad, Janka Biznárová, Marco Caputo, Liangyu Chen, Kestutis Grigoras, Göran Johansson, Anton Frisk Kockum, Christian Križan, Daniel Pérez Lozano, Graham J Norris,

Amr Osman, Jorge Fernández-Pendás, Alberto Ronzani, Anita Fadavi Roudsari, Slawomir Simbierowicz, Giovanna Tancredi, Andreas Wallraff, Christopher Eichler, Joonas Govenius, and Jonas Bylander , "Building blocks of a flip-chip integrated superconducting quantum processor". Quantum Sci. Technol. 7, 035018 (2022).

[G] Sandoko Kosen, Hang-Xi Li, Marcus Rommel, Robert Rehammar, Marco Caputo, Leif Grönberg, Jorge Fernández-Pendás, Anton Frisk Kockum, Janka Biznárová, Liangyu Chen, Christian Križan, Andreas Nylander, **Amr Osman**, Anita Fadavi Roudsari, Daryoush Shiri, Giovanna Tancredi, Joonas Govenius, and Jonas Bylander, "Signal crosstalk in a flip-chip quantum processor". PRX Quantum 5, 030350 (2024).

Other publications by the author, not included in this thesis, are:

[H] Liangyu Chen, Simon Pettersson Fors, Zixian Yan, Anaida Ali, Tahereh Abad, **Amr Osman**, Eleftherios Moschandreou, Benjamin Lienhard, Sandoko Kosen, Hang-Xi Li, Daryoush Shiri, Tong Liu, Stefan Hill, Abdullah-Al Amin, Robert Rehammar, Mamta Dahiya, Andreas Nylander, Marcus Rommel, Anita Fadavi Roudsari, Marco Caputo, Grönberg Leif, Joonas Govenius, Miroslav Dobsicek, Michele Faucci Giannelli, Anton Frisk Kockum, Jonas Bylander, Giovanna Tancredi, "Fast unconditional reset and leakage reduction in fixed-frequency transmon qubits". arXiv:2409.16748 (2024).

[I] Marina Kudra, Mikael Kervinen, Ingrid Strandberg, Shahnawaz Ahmed, Marco Scigliuzzo, **Amr Osman**, Daniel Pérez Lozano, Mats O. Tholén, Riccardo Borgani, David B. Haviland, Giulia Ferrini, Jonas Bylander, Anton Frisk Kockum, Fernando Quijandría, Per Delsing, and Simone Gasparinetti, "Robust Preparation of Wigner-Negative States with Optimized SNAP-Displacement Sequences". PRX Quantum 3, 030301 (2022).

[J] Anita Fadavi Roudsari, Daryoush Shiri, Hampus Renberg Nilsson, Giovanna Tancredi, **Amr Osman**, Ida-Maria Svensson, Marina Kudra, Marcus Rommel, Jonas Bylander, Vitaly Shumeiko, Per Delsing, "Three-wave mixing traveling-wave parametric amplifier with periodic variation of the circuit parameters". Appl. Phys. Lett. 122, 052601 (2023).

[K] Liangyu Chen, Hang-Xi Li, Yong Lu, Christopher W. Warren, Christian J. Križan, Sandoko Kosen, Marcus Rommel, Shahnawaz Ahmed, **Amr Osman**, Janka Biznárová, Anita Fadavi Roudsari, Benjamin Lienhard, Marco Caputo, Kestutis Grigoras, Leif Grönberg, Joonas Govenius, Anton Frisk Kockum, Per Delsing, Jonas Bylander and Giovanna Tancredi , "Transmon qubit readout fidelity at the threshold for quantum error correction without a quantum-limited amplifier". npj Quantum Information 9, 26 (2023).

[L] K. Grigoras, N. Yurttagül, J.-P. Kaikkonen, E. T. Mannila, P. Eskelinen, D. P. Lozano, H.-X. Li, M. Rommel, D. Shiri, N. Tiencken, S. Simbierowicz, A. Ronzani, J. Hätinen, D. Datta, V. Vesterinen, L. Grönberg, J. Biznárová, A. Fadavi Roudsari, S. Kosen, **A. Osman**, M. Prunnila, J. Hassel, J. Bylander, and J. Govenius, "Qubit-Compatible Substrates With Superconducting Through-Silicon Vias". IEEE Trans. Quantum Eng. 3, 5100310 (2022).

[M] Phalgun Lolur, Mårten Skogh, Werner Dobrautz, Christopher Warren, Janka Biznárová, **Amr Osman**, Giovanna Tancredi, Göran Wendin, Jonas Bylander, Martin Rahm, "Reference-State Error Mitigation: A Strategy for High Accuracy Quantum Computation of Chemistry". J. Chem. Theory Comput. 2023, 19, 3, 783–789.

[N] Mohammed Ali Aamir, Claudia Castillo Moreno, Simon Sundelin, Janka Biznárová, Marco Scigliuzzo, Kowshik Erappaji Patel, **Amr Osman**, D. P. Lozano, Ingrid Strandberg, and Simone Gasparinetti, "Engineering Symmetry-Selective Couplings of a Superconducting Artificial Molecule to Microwave Waveguides". Phys. Rev. Lett. 129, 123604 – (2022).

[O] Claudia Castillo-Moreno, Kazi Rafsanjani Amin, Ingrid Strandberg, Mikael Kervinen, **Amr Osman**, Simone Gasparinetti, "Dynamical excitation control and multimode emission of an atom-photon bound state". arXiv:2404.05547 (2024).

[P] Mårten Skogh, Werner Dobrautz, Phalgun Lolur, Christopher Warren, Janka Biznárová, **Amr Osman**, Giovanna Tancredi, Jonas Bylander and Martin Rahm, "The electron density: a fidelity witness for quantum computation". Chem. Sci., 2024, 15, 2257-2265 (2023).

Acknowledgments

I would like to express my heartfelt gratitude to several individuals who have played a pivotal role in my journey to this point.

First, I am grateful to my supervisor, Jonas Bylander. Thank you for the opportunity to be part of your research group and for trusting me to work on projects that inspire me. Your support in making things possible rather than imposing obstacles has been invaluable. I look forward to continuing working with you in the future.

A heartfelt thank you to Anita. You have been more than just a co-supervisor; you have been a mentor from the moment I joined QTL six years ago. Your unlimited patience in guiding me through my first steps in the cleanroom, your constant availability for discussions, and your meticulous feedback on my manuscripts have been truly invaluable. I am deeply grateful for all that you have done for me, and I truly hope we can maintain a strong connection beyond this PhD journey.

I would also like to thank my co-supervisor, Avgust Jurgens, for his valuable insights on materials and fabrication, my examiner, Per Delsing, for his support and unmistakable interest in discussing new ideas, and the current head of our division, Witlef Wieczorek, for ensuring healthy and efficient work environment.

During the second half of my PhD, I had the privilege of working closely with Giovanna Tancredi, the current PI of our quantum computing effort. Your hard work and dedication, from ensuring everyone can run their experiments to discussing even the minor details, have been incredible. I also appreciate your valuable feedback on this manuscript.

A special thank you to Philip Krantz. Before coming to Chalmers for my second year of master's as part of the Erasmus Mundus program, I was planning to go to Dresden instead. But after your lecture in Leuven about the quantum computing effort at Chalmers, I changed my mind just two weeks before the decision deadline—and I believe it was the right choice. You later helped me land a master's thesis project in the group, and that's where it all began.

The achievements in this thesis and in our division over the past five years would not have been possible without our alumni, who laid a strong foundation of knowledge that provided a head start for my research. Thank you to Andreas Bengtsson, Marco Scigliuzzo, David Niepce, Gustav Andersson and Marina Kudra.

In every experimental group, it's great to have that go-to experimentalist who, for some reason, knows more about the theory than the rest—and even better, is willing to drop whatever they're doing to explain it to you. Chris, you have been a great companion throughout my PhD journey. Thank you for the countless discussions (sometimes arguments as you referred to them) and the probably over 500 hundred lunch meals we had together.

Christian Križan, thank you for always being there to help at work and outside.

Linda Brånell, you make our lives so much easier. We're really lucky to have you! I had the privilege of working with and learning from many incredible individuals throughout this journey, without whom, a lot of the work in this thesis wouldn't have been possible. Thank you to Sandoko (and your great sense of humor), Hang-Xi, Michele, Liangyu, Daryoush, Marcus, Andreas, Daniel Pérez, Janka, Lert, Emil, Halldor, Lefty, Stefan, Jorge, and Anton.

Our collaboration does not stop at shared projects, but it extends to discussions over lunch, seminars or casual office chats. Thank you to Simone, Tom, Anuj, Nermin, Robert, Lunjie, Irshad, Anna, Filip, Tangyou, Tong, Olga, Achintya, Aditya, Marti, Claudia, Simon, Aamir, Claire, and Maurizio. Thank you to all current and previous members of QTL and AQP.

A huge portion of this thesis revolves around fabrication, and it wouldn't have been possible without the pioneering cleanroom infrastructure we have at Myfab Chalmers. Thank you to Henrik, Johan, Niclas, Mattias, Marcus, Ruggero, Mahdad, and Petra for your invaluable support. Thank you to Lars for skillfully and reliably machining our packages and sample boxes.

Now, outside work ...
To Eliott, Pierluigi, Alessio, Timo, Erik, Maria, Anton, Anna, Camilla, Chiara, Valeria, Jonas, Emma, Andreas and all those I was blessed with their friendship and companionship. Obviously, you guys had no contribution whatsoever to this thesis, but you gave me something far greater—joy, laughter, sadness, silly and serious conversations, sports, celebrations, and navigating tough pandemic times. That, without a doubt, helped me keep going with my work. Thank you!

I am deeply grateful to my parents who never spared an effort to give me opportunities they never had, and for always having my best interest at heart. I could write a whole thesis about that, but I will keep it brief. I am also lucky to have my sister Salma, with her unmatched love, care, strength and sense of responsibility. Thank you to all my family.

Lena, there are no words to describe my gratitude to you. Since I met you, no one has supported me through tough times the way you have. No one understands me as deeply as you do. I consider myself fortunate in life, but I am without a doubt the luckiest for having you in it. You are my partner and my best friend, and I am blessed with our love. I am also deeply thankful to your family for welcoming me with open arms and making me feel like part of their own. Their kindness and support have meant so much to me.

To my beloved sister Heba, to whom I dedicate this thesis. You have always been the closest human to my heart. Losing you was an immeasurable sorrow, but your memory never fades. This memory has been a constant source of strength and inspiration.

Contents

Part I

Thesis

CHAPTER 1

Introduction

Technology evolves as a direct response to human needs and the challenges we encounter. Over time, these needs have transformed from basic pursuits, such as developing tools for farming and food preservation millennia ago, to more intricate ones, like powering large cities and establishing telecommunication networks. Today, we grapple with even more sophisticated challenges, including drug development, combating climate change, cryptography, and the search for materials with unique properties. Drug development, for instance, is a lengthy process that may take over a decade to identify molecules that exhibit the required medicinal activity against a biological target [1]. While the advent of computers and computational chemistry transformed drug discovery from an *in vitro* process with a low success rate to an accelerated, more efficient process, classical computations still rely on approximate and comparative modeling. Accurate simulation of protein folding and chemical interactions between large molecules can be prohibitively resource-intensive [2, 3]. Similarly, finding new efficient materials for batteries and photovoltaic cells, or as catalysts in chemical reactions, relies on approximate solutions that can be computationally infeasible [4–7]. A few decades ago, quantum computing emerged as a possible solution to these bottlenecks. By leveraging new logic, beyond that which is allowed classically, a quantum computer can implement highly efficient and accurate processing of complex computational tasks.

1.1 From classical to quantum

Classical computations rely on encoding discrete information in bits, where each bit is either 0 or 1. Think of it as a light switch; it can either be on or off. A collection of bits forms a number, a character, or a logical operation within an algorithm. A quantum computer, similarly, encodes information in a quantum bit or a qubit. Unlike classical bits, the state of a qubit can exist in a superposition of $|0\rangle$ and $|1\rangle$. The Bloch sphere in Fig. 1.1(a) is a standard representation of such a superposition, where the blue vector on that sphere represents the qubit state

$$|\psi\rangle = \cos\frac{\theta}{2}\,|0\rangle + e^{-j\phi}\sin\frac{\theta}{2}\,|1\rangle, \tag{1.1}$$

where θ and ϕ can take any arbitrary values. Scaling this argument, a set of 3 classical bits can only be in one of 8 possible permutations: 000, 001, ..., or 111, while a set of 3 qubits can be in all 8 permutations at the same time with different probabilities [Fig. 1.1(b)]. To generalize, an N-qubit computer spans over 2^N possible states.

To further expand on this, for a quantum computer with N qubits, a classical computer would need to store 2^N numbers that represent the state of the quantum computer. Each of these numbers consists of m classical bits, depending on the desired precision, resulting in a total storage requirement of $m \times 2^N$ classical bits. This exponential scaling demonstrates the inherent limitations of classical computers in handling quantum systems. This realization further inspired the idea of using quantum computers to simulate the quantum mechanical systems of nature [8].

> "And therefore, the problem is, how can we simulate the quantum mechanics? There are two ways that we can go about it. We can give up on our rule about what the computer was, we can say: Let the computer itself be built of quantum mechanical elements which obey quantum mechanical laws. Or we can turn the other way and say: Let the computer still be the same kind that we thought of before."
>
> — **Richard Feynman**, 1981

To illustrate with real numbers: If we want to store the spin state of 40 electrons in a molecule, a quantum computer would require 40 qubits. A

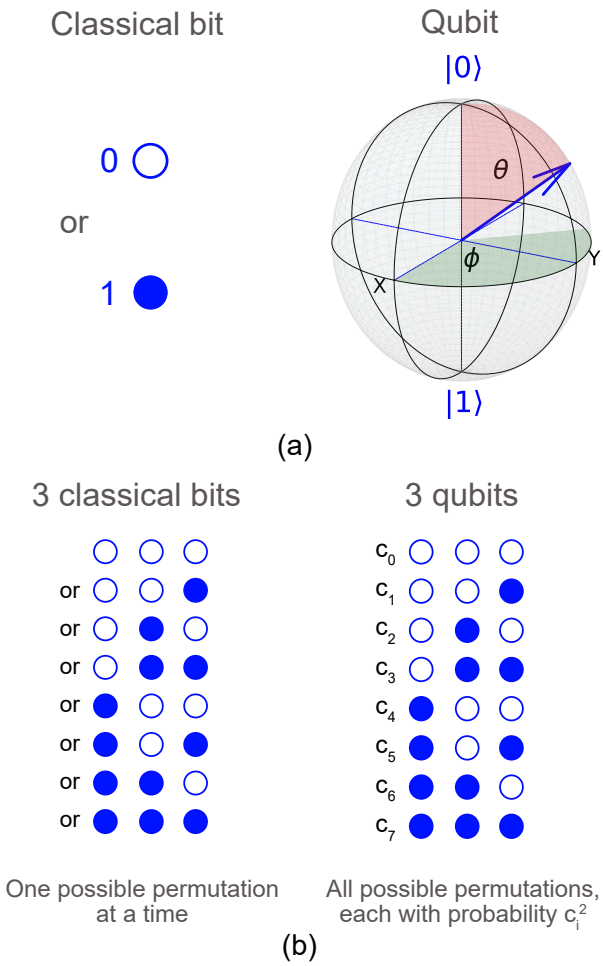


Figure 1.1: Comparison between classical and quantum bits. (a) Representation of a single classical bit and a quantum bit (qubit). A classical bit can exist in one of two states, 0 or 1, while a qubit state, represented by a unit vector on the Bloch sphere, can exist in an arbitrary superposition of $|0\rangle$ and $|1\rangle$. (b) Extension of this comparison to three classical bit and three qubits. Three classical bits can exist in one of 2^3=8 distinct states, whereas three qubits can exist in a superposition of all 2^3=8 states simultaneously.

classical computer, however, would require 2^{40} numbers; a total of 4 TB of memory for a single-precision number (32 bits). Just doubling the number of electrons to 80 would require $5{\times}10^{12}$ TB of memory. That is almost a hundred times more than the total volume of digital data stored and consumed worldwide in 2020 [2, 9, 10].

Now, how do we make a qubit? When envisioning a qubit, the first idea might be to use a component that is inherently quantum, such as a nucleus of an atom or an electron. In 1995, the first quantum logic gate was demonstrated with trapped ions at NIST [11]. In 1997 and 1998, the quantum gates were demonstrated at MIT and Oxford using nuclear magnetic resonance (NMR), where magnetic fields manipulate nuclear spins within molecules [12–16]. A year later, in Japan, researchers demonstrated that superconducting circuits could create an artificial atom, effectively functioning as a qubit [17]. This approach quickly gained traction, as the well-established fabrication techniques for superconducting circuits could later leverage advancements from the CMOS industry, which offered a robust technological foundation. As a result, superconducting qubits became one of the most promising and widely adopted technologies for building qubits, forming the architecture chosen by leading research groups and companies such as Google and IBM. Superconducting qubits are also the platform we adopt in our research group and are the focus of this thesis. Alongside superconducting qubits, other technologies such as trapped ions [18], quantum dots [19], and photonic qubits [20] are actively explored as alternative pathways.

1.2 Current state of quantum computers

It is essential to recognize that quantum computers are not expected to solve all computational problems, nor will they replace classical computers. Quantum computers rely on classical computers and electronics to control and interrogate them. The real advantage of quantum computers lies in their ability to handle a specific set of problems more efficiently than classical computers. When we say “more efficiently”, the difference can be staggering: a problem that would take thousands to millions of years to solve on a classical computer might be solvable by quantum computers in just hours [21].

So, where do quantum computers stand today? To answer this question, perhaps, it is useful to divide quantum algorithms into two broad classes.

The first class, fault-tolerant algorithms [22], assumes that qubits are error-free during the algorithm runtime. This class includes Shor's algorithm [23], intended for RSA decryption, Grover's algorithm [24], which can speed up database search exponentially, and Quantum Fourier Transform [25], which is essential for simulating the dynamics of quantum systems in material science and drug development.

The second class encompasses Noisy Intermediate-Scale Quantum (NISQ) algorithms [26], which are designed to run on quantum processors with noisy, error-prone qubits. These algorithms are heuristic, aiming to provide approximate solutions that are closer to the exact solution than what their classical counterparts can achieve. Examples of these algorithms include Quantum-Approximate Optimization Algorithms (QAOA) [27], and Variational Quantum Eigensolvers (VQE) [28].

Returning to the question of where we are now, a fault-tolerant quantum computer remains a distant goal. Boston Consultancy Group predicts that such a machine could become a reality around the year 2040, with a forecast economic value of $450 billion to $850 billion from quantum computing [29]. Meanwhile, the potential of NISQ algorithms to outperform classical algorithms and artificial intelligence is still uncertain. While the ultimate goal is to achieve fault tolerance, NISQ algorithms will offer useful insights and applications as we advance toward this goal.

1.3 Why is it hard to build a quantum computer?

Building quantum computers is a remarkably challenging endeavor, with most of our current hurdles rooted in engineering. To start, qubit decoherence (the general term for loss of information in a quantum computer) poses a constant challenge across most qubit platforms. Qubits are surrounded by and interacts with a noisy environment. This environment makes it difficult for qubits to maintain their exotic superposition state long enough to execute quantum algorithms effectively. Superconducting qubits typically decohere within a few hundred microseconds—insufficient to implement fault-tolerant algorithms. For instance, implementing Shor's algorithm to factor a 2048-bit RSA integer, which requires hundreds of seconds, becomes infeasible with such short-lived coherence times.

Error correction codes, however, offer a solution by mitigating decoher-

ence [30]. They encode a single logical qubit into hundreds to thousands of physical qubits [31, 32]. Despite the substantial overhead—potentially requiring up to 20 million qubits and over 8 hours of runtime to factor a 2048-bit RSA integer [21]—error correction remains the only feasible approach to implement these fault-tolerant algorithms.

Beyond coherence, we must also control these qubits by applying quantum gates to shift their state by a specific angle around the Bloch sphere or mediate interactions between them. The quality of such quantum gates is captured by the gate fidelity, which quantifies the accuracy of the target operation. Eventually, we must read out the state of the qubit to extract the results of our computation. The accuracy of reading out a qubit state that reflects its true state is quantified by the readout fidelity.[1]

High qubit coherence, gate fidelity, and readout fidelity are three of the main objectives toward the goal of building a quantum computer. Both fault-tolerant and NISQ algorithms benefit from their improvement. Higher qubit coherence and quantum gate fidelity for fault tolerance mean fewer physical qubits required for every logical qubit, while fast and high-fidelity readout is crucial for accurate and fast error detection, hence, shorter algorithm runtime. For NISQ algorithms, high qubit coherence, gate fidelity, and readout fidelity facilitate more complex algorithms with more quantum operations and reliable measurement outcomes.

1.4 In this thesis

Addressing these challenges and scaling a quantum processing unit (QPU) toward fault tolerance, or quantum advantage with NISQ algorithms demands interdisciplinary engineering efforts, each contributing from a unique perspective to the solution. In this thesis, we present modest yet critical steps in this direction by addressing specific engineering challenges at the level of single qubits and a 25-qubit QPU, and the potential scalability to a larger number of qubits. The following subsections explore these engineering perspectives, emphasizing the challenges that are central to this thesis.

[1] Readout fidelity is not a topic that is addressed further in this thesis.

1.4.1 Frequency collision and crosstalk

To run a quantum algorithm on a QPU, we implement sequences of single- and two-qubit gates. Single-qubit gates are microwave pulses with a frequency equal to the qubit transition frequency (typically 3—6 GHz). To implement two-qubit gates, we parametrically modulate the frequency of a mediating element (a coupler) between two qubits at a frequency corresponding to various two-qubit transitions (typically 50—1000 MHz). These control pulses travel from room-temperature electronics down through signal lines patterned on the QPU, which is housed in a dilution refrigerator at 10 mK. Achieving high gate fidelities requires precise calibration of the pulse amplitude, frequency, duration, and phase. This is feasible when controlling individual qubits. However, simultaneous control of multiple qubits introduces the risk of unintended pulses reaching neighboring qubits, altering their states—a phenomenon known as signal crosstalk, which degrades gate fidelity.

One solution to mitigate crosstalk is to design a qubit frequency allocation scheme that maximizes the frequency spacing between neighboring qubits. In this arrangement, even if a signal leaks to a neighboring qubit, it remains off-resonance and therefore minimally affects the unintended qubit. However, fabrication uncertainties lead to deviations in qubit frequencies from their design targets. These frequency shifts are primarily due to variations in the Josephson junction, a fundamental component of superconducting qubits.

Josephson junction fabrication and improving the reproducibility of qubit frequency are discussed in detail in Chapter 2, focusing on planar chips and uncoupled qubits. Chapter 4 delves into crosstalk characterization on our first-generation 25-qubit QPU and the frequency allocation strategies adopted to mitigate crosstalk. In the same chapter, I introduce our flip-chip architecture, an essential technology for scaling up to tens of qubits on a single QPU. In Chapter 6, I characterize our second-generation 25-qubit QPU, assessing both the frequency targeting and the impact of crosstalk on single-qubit gate fidelity.

1.4.2 Qubit coherence

Improving qubit coherence is an ongoing pursuit. No matter how high coherence times reach, the quest for further enhancement will continue. Although sources of decoherence arise from various channels, fabrication quality and

material purity remain the primary bottlenecks for superconducting qubits. Material interfaces are often riddled with defects, known as two-level system (TLS) defects. These defects interact with the qubit, leading to energy dissipation and loss of coherence. Probing these TLS defects and their impact on the performance of superconducting circuitry is discussed in Chapter 3, along with the methods we employ to enhance qubit quality.

1.4.3 Design and simulation

The design of a QPU begins with defining polygons in computer-aided design (CAD) software. The dimensions of these polygons and their proximity determine the QPU parameters, such as the qubit frequency, the couplings between neighboring qubits, as well as the couplings between the qubit and its readout and control elements. To define the optimal dimensions, circuits are simulated using microwave simulation tools, with a parametrized sweep of device geometry, until the simulation converges to the target circuit parameters. It is crucial to capture all circuit parameters that influence the simulation, while at the same time keeping the simulation as resource-efficient as possible.

Additionally, design and simulation do not always match the real circuit parameters, and therefore, there must always be a constant feedback loop between design, fabrication, and measurement. As qubit counts and circuit complexity increase, optimizing qubit placement and routing of control and signal lines becomes even more critical. Chapter 4 and Chapter 5 discuss some of the design aspects of the QPU, while Chapter 6 evaluates the accuracy of parameter targeting.

CHAPTER 2

Josephson Junctions

The Josephson junction (JJ) stands as a fundamental building block in the construction of qubits within superconducting circuits. The qubit's distinct states, $|0\rangle$ and $|1\rangle$, emerge from the nonlinearity of the junction's inductance—an essential factor that disrupts the otherwise evenly spaced energy levels typical of linear harmonic oscillators. Due to its small size, Josephson junction fabrication presents significant challenges, particularly in superconducting qubit circuits.

When designing a quantum processor, an intricately developed qubit frequency crowding scheme is crucial for executing quantum gates, while minimizing crosstalk between qubits. Achieving the target qubit frequencies strongly depends on the reproducibility of Josephson junction parameters. Over the past five years, considerable efforts have been dedicated to advancing various aspects of Josephson junction fabrication, including improvements in lithography, junction deposition, and oxidation techniques [33–38], along with strategies to compensate for size drifts arising from angular evaporation [39]. Beyond fabrication advancements, notable progress has been made in post-fabrication tuning of junction resistance [40–44].

Section 2.1 provides an overview of the Josephson effect, its significance in

superconducting qubits, and its influence on the reproducibility of the transmon qubit frequency. Section 2.2 outlines the JJ fabrication process, emphasizing the simplifications introduced through the Patch-Integrated Cross-Type (PICT) junctions. The reproducibility of the JJ's normal-state resistance, R_N, and its impact on qubit frequency reproducibility, is discussed in Sections 2.3 and 2.4. Finally, Section 2.5 delves into preliminary studies on various aspects of Josephson junctions, including aging effects, and the dependence of normal-state resistance on temperature.

2.1 Josephson junctions in transmon qubits

In this section, I present the theoretical framework underpinning the work discussed in this chapter. Without delving into detailed derivations, I focus on key equations and concepts that motivate the use of Josephson junctions in transmon qubits and help understand key aspects, especially their impact on qubit frequency targeting. The following subsections address the Josephson effect, the transmon qubit, and the factors influencing qubit frequency stability. This provides essential context for the experimental results discussed later.

2.1.1 The Josephson effect

In 1962, 22-year-old graduate student Brian D. Josephson made a groundbreaking theoretical prediction in superconductivity [45, 46]. Josephson postulated that a supercurrent I_s flows between two superconducting electrodes separated by a thin insulating barrier (or a weak link) in the absence of applied voltage. This supercurrent is described by the relation

$$I_s = I_c \sin \varphi, \tag{2.1}$$

where φ is the phase difference between the cooper pair wave functions in both electrodes, and I_c is the **critical current** of the junction. This is known as the DC Josephson effect. He further postulated that applying a voltage V across this junction leads to evolution of φ according to the relation

$$\frac{\partial \varphi}{\partial t} = \frac{2\pi}{\Phi_0} V, \tag{2.2}$$

where $\Phi_0 = h/2e$ is the magnetic flux quantum. (e is the electron charge, and h is Planck's constant.) This leads to an alternating I_s (Eq. 2.1) with frequency $2eV/h$, resulting in what is known as the AC Josephson effect.

From these relations, it can be shown that the current-voltage relation of the junction has the characteristics of an inductor [47], where

$$V = L(\varphi)\frac{\partial I}{\partial t}. \tag{2.3}$$

$L(\varphi) = \Phi_0/(2\pi I_c \cos\varphi)$ is the junction's kinetic inductance, where the phase-independent term is referred to as the Josephson Inductance $L_J = \Phi_0/(2\pi I_c)$. This kinetic inductance is associated with stored energy $E(\varphi)$ in the junction analogous to that of a coil inductor. $E(\varphi)$ is calculated by integrating the electric power $\int I_s V dt = \int I_s(\Phi_0/2\pi)\ d(\varphi)$ of the circuit, eventually leading to

$$E(\varphi) = E_J(1 - \cos\varphi), \tag{2.4}$$

where

$$E_J = \frac{\Phi_0 I_c}{2\pi} \tag{2.5}$$

is known as the **Josephson Energy**, an important parameter that will be referred to frequently throughout this thesis.

I conclude this introductory subsection with the **Ambegaokar-Baratoff relation**, which states that

$$I_c R_N = \frac{\pi\Delta}{2e}\tanh\frac{\Delta}{2k_B T}, \tag{2.6}$$

where R_N is the **normal state resistance** of the junction, Δ is the material-dependent superconducting gap, k_B is the Boltzman constant, and T is the temperature [48]. For practical values of Δ and typical operating temperature of junctions at $\sim$10 mK, the tanh term approaches one. The Ambegaokar-Baratoff relation is monumental for any technology that utilizes Josephson junctions as it allows one to infer I_c through relatively simple room temperature resistance measurements, as well be shown in the remainder of this chapter. Although there may be an offset in R_N from the measured resistance at room temperature, it can be compensated for as will be presented in Section 2.5.

2.1.2 Transmon qubits

To effectively prepare and manipulate quantum information, a quantum system with well-defined, discrete energy levels is required [49]. Quantum harmonic oscillators are exemplary systems that can be initialized in their ground state and possess equally spaced energy levels. However, this equal spacing renders them inefficient for the task, as distinct energy level differences are necessary for controlled quantum operations. To address this, a non-linear element is introduced to break the equal spacing between energy levels. The Josephson junction serves as an ideal element for this purpose due to its non-linear dependence of I_c, L_J, and E_J on the phase difference φ, as discussed in the previous section.

When incorporated into a harmonic oscillator circuit, the Josephson junction transforms the system into an anharmonic oscillator. Figure 2.1(a)-top illustrates the circuit schematic of the transmon anharmonic oscillator [17, 50, 51]. The Hamiltonian governing this circuit is given by:

$$H = 4E_C(\hat{n} - n_g)^2 - E_J \cos\hat{\varphi}, \tag{2.7}$$

where $E_C = e^2/2C_\Sigma$ is the capacitive charging energy, with C_Σ representing the total capacitance, composed of the shunt capacitance C_s and the junction capacitance C_J, i.e., $C_\Sigma = C_s + C_J$. (Notably, the junction also behaves as a parallel plate capacitor.) Different types of superconducting qubits operate in different E_J/E_C regimes, and a characteristic feature of the transmon qubit is the large ratio $E_J/E_C \gg 1$, which significantly enhances its robustness against charge noise [50]. Typical E_J/E_C ratios for the transmon qubit are greater than 50.

The energy difference, E_{01}, between the two lowest energy levels of the transmon qubit, crucial for quantum operations, is given by

$$E_{01} = hf_{01} = \sqrt{8E_J E_C} - E_C, \tag{2.8}$$

where f_{01} is the qubit's resonant frequency. For a single-junction transmon, f_{01} is fixed. However, if two junctions are connected in a loop, as shown in Fig. 2.1(a)-bottom, this frequency becomes tunable by threading a magnetic

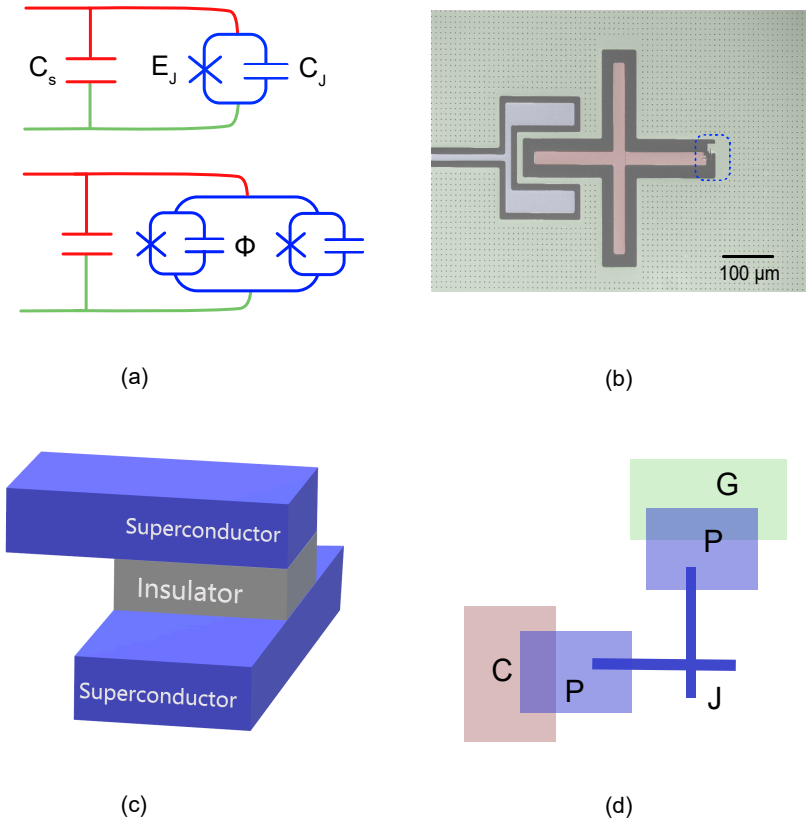


Figure 2.1: Transmon qubit. (a) Circuit schematic of the transmon qubit. The Josephson junction (blue) has a Josephson energy E_J and self-capacitance C_J and is shunted to ground (green) through a shunt capacitor (red) with capacitance C_S. The top schematic is a fixed-frequency version of the transmon, while the bottom schematic is a tunable-frequency version with the flux Φ threading the loop of the two Josephson junctions. (b) SEM image of a transmon qubit, with false colors to match the circuit schematic. The black dashed square highlights where the Josephson junction is. The blue region is the coupling arm of a coplanar waveguide resonator (not shown in (a)). (c) Illustration of a Josephson junction with the superconductor layers typically made of aluminum, and the insulating layer made of aluminum oxide. (d) A Josephson junction (J), galvanically connected by the patches (P) to the capacitor (C) and ground plane (G).

flux Φ which tunes E_J based on the relation

$$E_J(\Phi) = E_{J\max} \cos\left(\frac{\pi\Phi}{\Phi_0}\right) \sqrt{1 + d^2 \tan^2\left(\frac{\pi\Phi}{\Phi_0}\right)}, \tag{2.9}$$

where $E_{J\max}$ is the sum of the Josephson energies of the two junctions, and $d = (E_{J2} - E_{J1})/E_{J\max}$ is the junctions asymmetry. This relation will be useful when I address tunable qubits in Section 3.3.

2.1.3 Uncertainty in qubit frequency

Equation 2.8 shows that f_{01} depends on both E_C and E_J. Contributions to E_C arise from both C_J and C_s. In a transmon qubit, C_J is typically around 1–4 fF, significantly smaller than C_s, which is approximately 100 fF. Using Eqs. 2.5 and 2.6, E_J can be expressed in terms of R_N as

$$E_J = \frac{\Phi_0 \Delta}{4eR_N}. \tag{2.10}$$

The precise control of qubit frequency demands a precise control over C_s, C_J, and R_N. The latter two are properties of the junction, and their reproducibility depends on the accurate control of the junction area, as well as the thickness uniformity and stoichiometry of the oxide barrier. In contrast, C_s is determined by a large planar capacitor that is 5–6 orders of magnitude larger than the junction (on the scale of $\sim$0.5 mm), as shown in Fig. 2.1(b). Consequently, the reproducibility of the qubit frequency is primarily limited by the Josephson junction.[1] Its small size makes it highly sensitive to dimensional variations, and the dependence of its properties on the uniformity of an oxide layer, approximately 2 nm thick, further complicates the matter. In this thesis, I focus solely on the reproducibility of R_N for two main reasons. Firstly, both R_N and C_J are governed by the same junction parameters, meaning that achieving reproducible R_N inherently ensures reproducibility in C_J. Secondly, the small C_J/C_s ratio implies that any variations in C_J have a negligible impact on the total capacitance (see Appendix B of **Paper B**).

If we ignore variations in E_C, the coefficient of variation (CV) in the qubit

[1] In Chapter 6, we will see that in the flip-chip architecture, variations in C_s play a non-negligible role in qubit frequency reproducibility, due to variations in the chip-to-chip spacing.

frequency can be derived as follows. Using Eqs. 2.8 and 2.10, the qubit frequency can be expressed as

$$hf_{01} = \sqrt{2\frac{\Phi_0\Delta}{eR_N}E_C} - E_C. \tag{2.11}$$

Using error-propagation rule, the standard deviation in f_{01}, $\sigma_{f_{01}}$, and the standard deviation in R_N, σ_{R_N}, have the relation

$$\sigma_{f_{01}} \approx \frac{1}{2}\frac{f_{01}}{R_N} \,.\, \sigma_{R_N}. \tag{2.12}$$

Finally, the coefficient of variation in qubit frequency $CV_{f_{01}} = \sigma_{f_{01}}/f_{01}$ relates to $CV_{R_N} = \sigma_{R_N}/R_N$ via the relation

$$CV_{f_{01}} \approx \frac{1}{2}CV_{R_N}. \tag{2.13}$$

Equation 2.13 serves as our foundation for characterizing the reproducibility of qubit frequency through room temperature characterization of the junction resistance. Although we ignored variations in C_s, with the level of variation in R_N achieved in our work, the qubit's total capacitance begins to play a non-negligible role.[2] This aspect will be briefly addressed in this thesis.

2.2 Fabrication process

As mentioned earlier, the Josephson junction is essentially two superconductors separated by a week link. The most commonly used type of junctions in superconducting qubits is a sandwich of aluminum/aluminum-oxide/aluminum [Fig. 2.1(c)]. Such a trilayer structure is relatively easy to fabricate with metal deposition and oxidation tools that were developed several decades ago. Simply, deposit a layer of aluminum, oxidize it (or let native oxide grow), then deposit another layer of aluminum. The difficulty, however, is manifested in three caveats. Firstly, both superconducting electrodes must be electrically accessible, separately. Secondly, we need to have precise control over the

[2] Additionally considering variations in E_C, $CV_{f_{01}}$ becomes

$$\sqrt{0.25\, CV_{R_N}^2 + 0.25\, CV_{E_C}^2 + 0.0625\, CV_{R_N}^2 CV_{E_C}^2}$$

area of the junction. Finally, the oxide layer thickness in-between the two superconducting layers must be controlled as well. Several techniques have been developed to tackle the first issue [36, 38, 52–56]. The second and third conditions are criteria for precise control over E_J.

The most commonly used junction fabrication techniques have been the Dolan bridge and the Manhattan style. The former was developed in 1977 [52], while the latter was developed in 1999 [53].[3] However, other techniques have recently been developed to avoid angular evaporation and lift-off [36, 55, 56]. In our research group, we adopted the Manhattan-style junction, and so are all junctions in this work. Figure 2.2 illustrates the process of making Manhattan junctions utilizing shadow evaporation for selective deposition of the two junction electrodes with *in-situ* controlled oxidation. Our choice of the Manhattan junction over the Dolan junction is motivated by its relative simplicity. While in the former, variations in the junction area are caused by lithographic uncertainty and sidewall metal deposition, in the latter, the thickness of the bridge, its height and the possibility of sagging constitute additional sources of variations. Lithographic uncertainties include those attributed to electron-beam lithography exposure, resist development and descumming. Both techniques, however, suffer from systematic drift in their junction size across the wafer, due to the point-source nature of the evaporated metal. More details on sources of deviations in the Manhattan junctions are discussed in Section 2.3.3

2.2.1 Wiring layer

Our junctions are typically fabricated after the wiring layer is fully ready. On a qubit wafer, the wiring layer is the base metal, which may include features like coplanar waveguide resonators, qubit capacitors and a ground plane. On a test wafer for calibrating the Josephson junction normal resistance, the wiring layer is typically the contact pads used for four-point probe measurements as will be explained later.

The wafers used in this work are either 2 or 3-inch intrinsic (100) silicon wafers, with resistivity of 10–20 kΩ cm. Before any processing, wafers undergo

[3] The Dolan-Bridge Josephson junction is named after its inventor, Gerald J. Dolan. In the original work, Manhattan-style junctions were never called by this name. The name was given retrospectively due to their step-like structure, which resembles the grid layout of Manhattan streets.

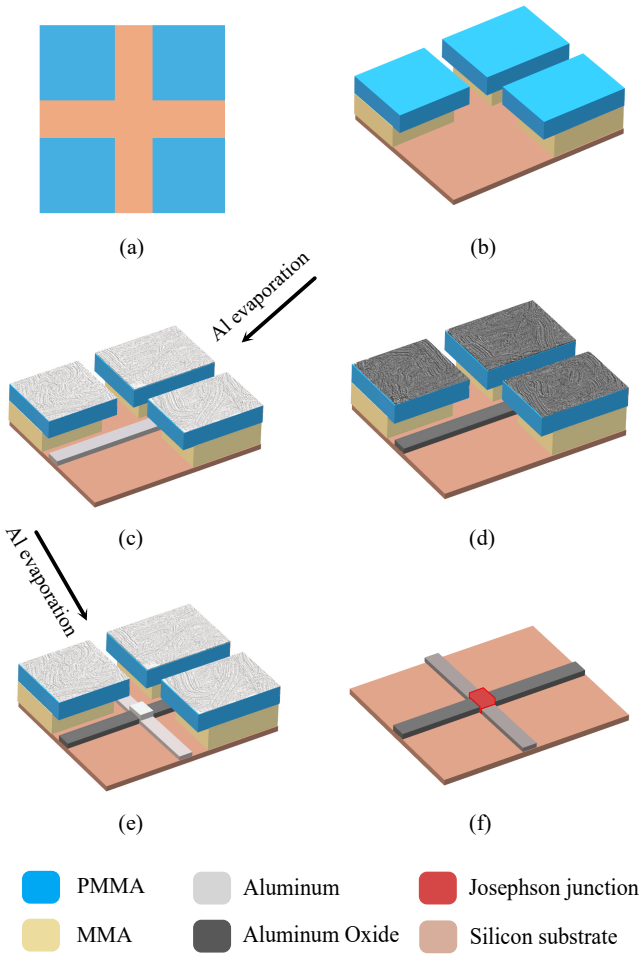


Figure 2.2: Illustration of the fabrication process of Manhattan-style Josephson junctions. (a) Top view of a pattern stack of PMMA/MMA on a silicon substrate after exposure and development. Isometric view of (b) developed stack of PMMA/MMA, (c) evaporation of the first (bottom) junction electrode at 45°, (d) *in-situ* oxidation of the bottom electrode, (e) evaporation of the second (top) junction electrode at 45° tilt and 90° planar rotation with the respect to the first electrode, and (f) Josephson junctions after oxidation of the top electrode and lift-off. The junction is formed at the overlap between the two electrodes. Note that part of the PMMA/MMA stack is removed from (b)-(e) for better illustration.

the standard cleaning (SC–1) process, where they are dipped in a bath of ammonium hydroxide (NH_4OH) and hydrogen peroxide (H_2O_2) and Deionized (DI) water (1:1:5), heated to 80° for 10 minutes, and then rinsed in DI water.[4] To ensure that our aluminum is grown on a clean silicon substrate, we chose to pattern our wiring layer through etching instead of lift-off. However, before depositing aluminum, the wafer is dipped in a 2% hydrofluoric acid (HF) solution for 1 minute, then rinsed in DI water. This step etches away the native-oxide grown on the silicon substrate. The wafer is then quickly loaded into an aluminum deposition tool (Plassys MEB 550s), where it is pumped to $\sim 10^{-6}$ mbar in about 10 minutes. The wafer holder is then heated to 300° for 10 minutes to dehydrate the silicon surface and prevent the regrowth of native oxide. After about four hours and when the pressure drops to $\sim 8 \times 10^{-8}$ mbar, we deposit (150—500 nm) aluminum film at a rate of 1 nm/s. Finally, we chose to controllably oxidize our base metal *in-situ*. This ensures a cleaner aluminum surface than if it was left to oxidize natively in air. In chapter 3, I will discuss the implications of the different aforementioned treatments on the quality factors of our quantum circuits.

To pattern the wiring layer, we spin-coat a layer of photoresist (typically AZ 1512), which we pattern using a laser-writer tool (DWL 2000 or MLA 150). After resist developing and descumming, aluminum is then wet-etched in Transene-A, a standard wet-etchant of aluminum, containing phosphorice, nitric acid, acetic acid, and DI water. After rinsing in DI water, the photoresist is then stripped using our standard resist stripping and lift-off process, which consists of remover 1165 at 80°C,[5] sonication for 5 minutes in the same remover, sonication for 3 minutes in methanol, and finally sonication for 3 minutes in isopropyl alcohol (IPA).

2.2.2 Two-step fabrication process

Our two-step fabrication process of Manhattan junctions is laid out in Fig. 2.2, where junctions are fabricated in one lithography step and a bandage/patch layer is fabricated in an additional lithography step. The patch forms a galvanic connection between the junction electrodes and the rest of the wiring layer [Fig. 2.1(d)] by removing the interfacial oxide, which is known to cause qubit losses [57].

[4] We skip the SC–2 process as we observed no effect on the quality of our devices.
[5] This is done for at least 10 minutes but may take longer in the case of the lift-off process.

As shown in Fig. 2.2, Manhattan junctions are fabricated using angular evaporation and lift-off process. To ensure a smooth lift-off, sidewall deposition during metal evaporation must not form a continuous film with the junction electrodes. Therefore, it is crucial to form an undercut profile. To achieve such an undercut, we use a resist stack of PMMA (poly methyl methacrylate) on top and MMA (methyl methacrylate) below. This positive resist stack is commonly used for lift-off of metalic structures. Both resists are composed of long polymer chains of carbon atoms, with PMMA having a higher molecular weight than MMA. When the junction features are exposed using an electron beam, the low molecular weight of the MMA compared to PMMA makes it more sensitive to the electron beam breaking down its polymer chains, rendering it easily soluble during the development process. This contrast in sensitivity between the two stack layers creates the necessary undercut profile required for a clean lift-off process.

The process starts with spin-coating of MMA El 12 at 3000 rpm for 1 minute, which is then baked on a hotplate at 160°C for 5 minutes, resulting in a $\sim$ 700-nm-thick layer. Then PMMA 950 A6 is spincoated at 6000 rpm for 1 minute, which is then backed on a hotplate at 160°C for 5 minutes, which results in a $\sim$ 250-nm-thick layer [58]. Features are exposed in EBL - JEOL JBX 9300FS or Raith EBPG 5200 electron-beam lithography (EBL) systems; both with 100 kV acceleration voltage. Since PMMA is less sensitive than MMA to electron beam, it controls the minimum required dose for exposure. The base dose used is 1600 μC/cm^2 on average after implementing proximity error correction (PEC), which accounts for electron scattering during the exposure. Junctions are exposed with a beam current of 2–10 nA.

The exposed features are then developed in a solution of MIBK:IPA 1:1 for 90 seconds, rinsed in IPA for 30 seconds, then immediately blow dried with N_2. Before depositing the junctions, the exposed features are descummed with oxygen plasma in a reactive-ion etching (RIE) system at low power, calibrated to remove $\sim$ 10 nm of resist. Removing the resist residues before junction deposition is crucial for qubit coherence, and to avoid junction aging [57, 59].

The wafer is subsequently loaded to Plassys for junction formation. The first junction electrode is created by depositing 50 nm of aluminum at a 0° planetary rotation and 45° tilt. Accounting for the tilt, the final thickness of the bottom electrode is $45 \times \cos(45°) = 35$ nm. To form the tunnel barrier, we oxidize the bottom electrode under a specific pressure and time, which I

will discuss in Section 2.3.4. The top junction electrode is then deposited at 90° planetary angle and 45° tilt with a final thickness of 78 nm. The standard deposition rate of junction electrodes is 1 nm/s. Finally, the surface of the two junction electrodes is controllably oxidized *in-situ*, similar to the wiring layer's base aluminum. Junctions then undergo the standard lift-off process in remover 1165.

Patches are subsequently fabricated using the same lithography process and similarly deposited in Plassys. However, before deposition, the aluminum oxide (AlO_x) on the junction electrodes and the wiring layer is milled away *in-situ* using a normally incident Ar^+ beam that is neutralized before hitting the substrate to avoid charge accumulation. More than 10 nm of AlO_x/Al is removed to ensure a galvanic connection between the junction electrodes and wiring layer when depositing the patches. The milling calibration for our process can be found in [60]. The patches are typically 200-nm thick.

Angular evaporation imposes limitations on the junction sizes that can be formed. These limitations depend on the resist height s and tilt angle ϕ [Figs. 2.3(a) and 2.3(b)]. The maximum junction electrode width d is given by $s/\tan(\phi)$. Since $\phi = 45°$ and $s = 950$ nm, the maximum junction electrode width is 950 nm. In practice, this limit is well below $s/\tan(\phi)$ to avoid aluminum deposition in the undercut region. For a typical undercut of 200 nm, the maximum electrode width that avoids this residual deposition is 750 nm. Typical junction sizes used in our qubit devices are below this limit.

2.2.3 Patch-integrated cross-type junctions

As mentioned earlier, a galvanic and superconducting connection between the junction electrode and the wiring layer was proven to be crucial for qubit coherence [57]. However, making patches in an additional lithography step is a lengthy process that consumes valuable resources, increases the probability of error, and exposes the loss-sensitive junction area to additional processing. Therefore, we invented the patch-integrated cross-type (PICT) junction to mitigate those three issues, which is presented in **Paper A**.

The idea behind the integrated patch initiates from the Manhattan-style junction itself. We again utilize the selectivity of angular evaporation to different patterned trenches, except that this time, we add a third trench at a 45° tilt between the two junction electrodes to form the patches. The process steps are illustrated in Fig. 2.3. The 45° fringes overlap with the junction

electrodes and the wiring layer.

The lithography steps of patterning the PICT junctions are identical to the process explained in Section 2.2.2. Deposition of the junction electrodes [Figs. 2.3(b) and 2.3(c)] is also identical to that explained in Fig. 2.2. For the PICT process, we subsequently mill away AlO_x/Al at 45° planetary rotation and 45° tilt using an Ar+ beam [Fig. 2.3(d)]. At the same angles, aluminum patches are deposited in the fringes to form the galvanic connection, followed by a final protective oxide [Fig. 2.3(e)]. PICT junctions are similarly lifted-off in remover 1165. Fig. 2.3(f) shows an SEM image of the PICT junction after lift-off.

Similar to the two-step process, PICT junctions are limited in size. However, the limitation here is more stringent due to the addition of the fringes. The maximum junction electrode width that avoids undesired aluminum deposition on the junction electrodes trenches during milling and deposition of the patches is $d{=}s{\times}\cos(\zeta)/\tan(\phi)$, where ζ is the angle between the fringes and the two junction electrodes. Since ζ is always 45° and ϕ is set to 45° in our case, the maximum d is 670 nm. Accounting for an undercut of 200 nm, the maximum d becomes 470 nm. Finally, I would like to note that, although we use three fringes per junction electrode, it is possible to pattern only one fringe, as was replicated in [61] and [62]. This may alleviate concerns about superconducting loops trapping flux near the qubit.

2.3 Reproducibility of R_N

The focus of this section is on the reproducibility of Josephson junctions R_N and its impact on qubit frequency targeting. To ensure robust statistical analysis, all studies discussed here are conducted at the wafer scale. First, I start with laying out the standard designs and measurement infrastructure for characterizing R_N. Then, I present our baseline PICT junctions reproducibility and investigate the potential sources of variations. I end this section with our lowest achieved wafer-level standard deviation of 2% in R_N.

2.3.1 R_N characterization setup

The junction resistance R_N is measured using a 4-point-probe setup. A standard wafer layout is shown in Fig. 2.4(a). Each wafer contains 40 5×5 mm^2

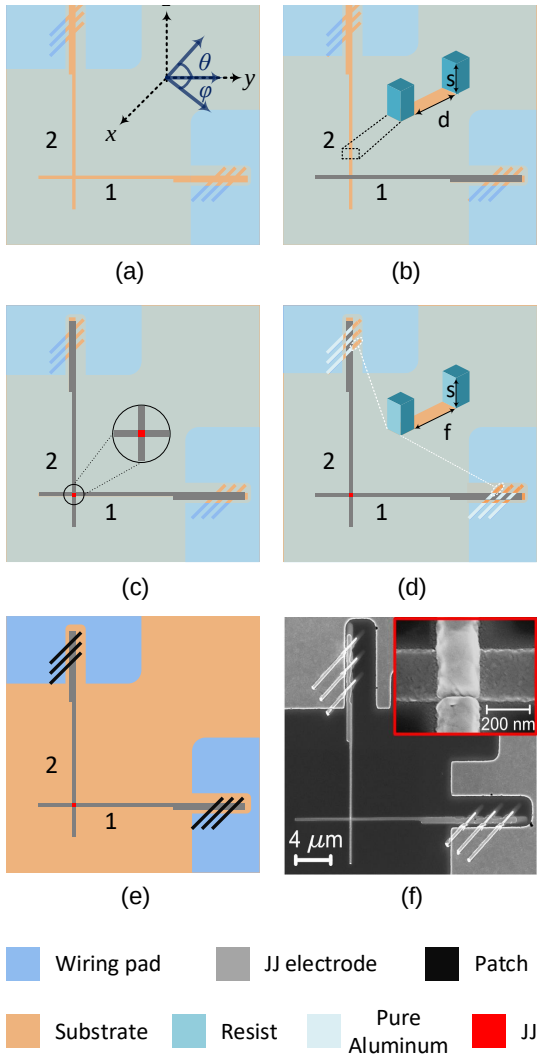


Figure 2.3: PICT junction process steps. (a)-(e) Schematic of the PICT process flow, showing the layout after (a) development, (b) first electrode deposition and oxidation to form the tunnel barrier, (c) second electrode deposition and oxidation, (d) ion milling of the aluminum oxide at the fringes, and (d) deposition of the patches. θ and ϕ are the planetary and tilt angles of the sample holder, respectively. Deposition of Al on top of the resist is not shown for clarity. (f) SEM image of a fabricated junction with the patch layer.

dies, and each die contains 100 Devices Under Test (DUT). Each of the four pads of a DUT is 150×150 μm^2, and the junction lies in the middle. In total, each wafer contains 4000 test devices, with each row of a die containing 10 different junction sizes, ranging from 100×100 to 400×400 nm^2. All junctions in this thesis are formed by two perpendicular electrodes with the same nominal design width.

Junctions are then characterized in an automated probe-station MPI TS2000 [Fig. 2.4(b)], where a probe card containing 80 pins (4 pins for each DUT) measures R_N of 20 DUTs during a single touchdown [Fig. 2.4(c)]. Each vertical DUT pair is measured simultaneously with two different SMUs of a single Keithly 2612b sourcemeter, that iterate over each pair sequentially via a Keithley 3706A-S system switch. After measuring each block, the probe card moves to the following block of 20 devices. All measurements are current biased, where the current is swept between -1μA to 1μA through a vertical pair of pins on a DUT, and the voltage drop is measured across the other pair of pins. For each DUT, an I-V curve has 20 points that are fit linearly to extract the slope R_N. For each R_N, there is an associated r^2 representing the coefficient of determination (or goodness of fit). DUTs with r^2 below 0.99 are excluded from our statistics.

2.3.2 PICT-junctions reproducibility

To qualify PICT as a standard process for fabricating superconducting qubits, we characterized both qubit coherence and R_N reproducibility. Coherence is discussed in Chapter 3. Here, I present our study on R_N reproducibility.

The dependence of $\overline{R_N}$ (mean R_N) on junction area is shown in Fig. 2.5(a). R_N is given by R_J/A_J, where R_J is the junction's resistance per unit area, and A_J is the junction area. To account for the lithography bias in the junction area, R_N can be expressed in terms of the junction width $d_J = d_n - d_0$ as $R_N = R_J/(d_n - d_0)^2$, where d_n is the nominal design width and d_0 is the bias in junction width, to account for the deviation from the nominal. To better illustrate R_J and d_0, Fig. 2.5(b) shows a fit of $1/\sqrt{\overline{R_N}}$ versus d_n. The slope s of such a fit reveals R_J as $1/s^2$, while the intersection with the x-axis is d_0. The extracted R_J for the junctions shown in the figure is 260 $\Omega \cdot \mu$m^2, while d_0 is -28 nm. [6]

[6] In **Paper B**, we reported R_J to be 240 $\Omega \cdot \mu$m^2; however, that value, incorrectly, did not account for the bias d_0.

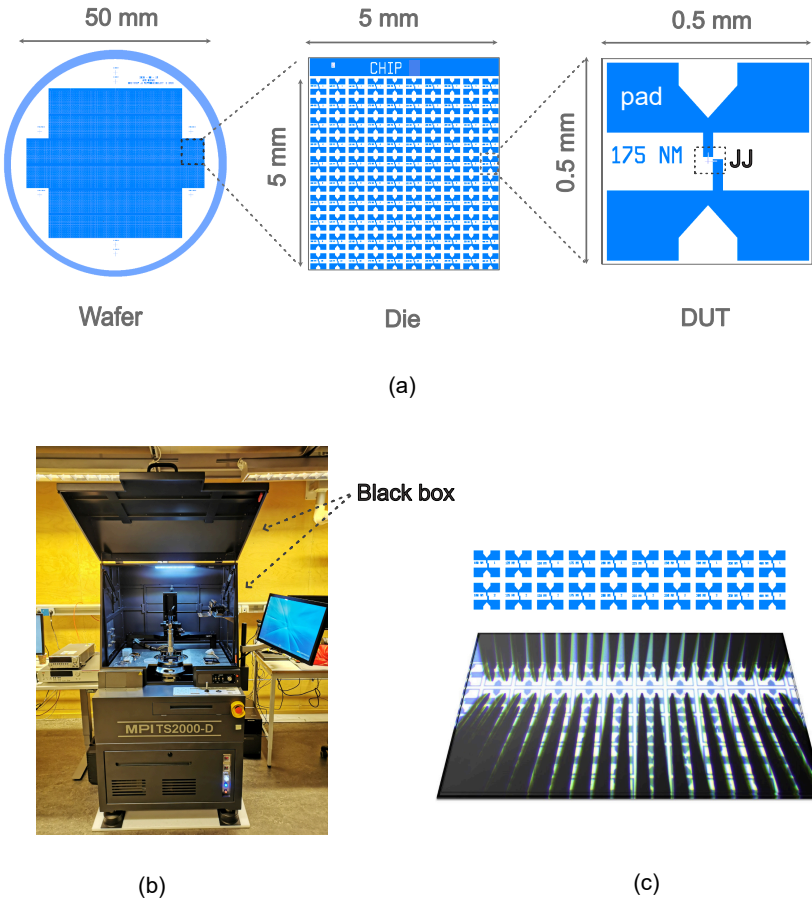


Figure 2.4: Josephson junction characterization design and setup. (a) Layout of the wafer, die and DUT. (b) A picture of the MPI TS2000 automated probe station, equipped with a black box to shield the DUTs from light during measurements. (c) A block of 10×2 DUTs measured by an 80-pin probe card.

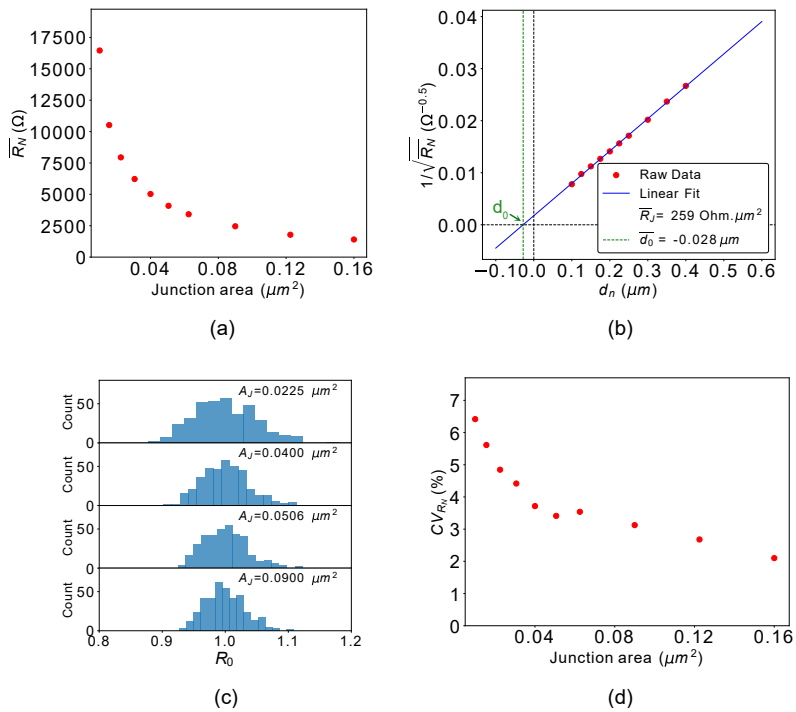


Figure 2.5: Wafer-level reproducibility of PICT junctions. (a) Mean R_N as a function of the junction area. (b) $1/\sqrt{\overline{R_N}}$ as a function of junction nominal design width d_n. The junction unit area resistance R_J is given by $1/s^2$, where s is the slope of the linear fit. (c) Histograms of the normalized resistance $R_0 = R_N/\overline{R_N}$ for four different junction areas. (d) Dependence of R_N's coefficient of variation (CV_{R_N}) on junction area A_J.

Histograms of the normalized resistance, $R_0 = R_N/\overline{R_N}$, are shown in Fig. 2.5(c), resembling normal distributions with Kolmogorov-Smirnov p-values of 0.52, 0.83, 0.44 and 0.21 from top to bottom, respectively [63]. One clearly notices the narrowing of the distribution of these histograms from smaller to larger junction sizes, meaning that larger junctions are more reproducible in their resistance. The coefficient of variation in R_N, CV_{R_N} (also referred to as the relative standard deviation), is shown in Fig. 2.5(d) as a function of the junction area A_J. Such reproducibility levels were on par with the best reported R_N reproducibility levels (see Table II in **Paper A**).

2.3.3 Decoupling sources of variations in R_N

The spread in R_N is attributed to both non-uniformity in oxide-barrier thickness, and variations in the junction area A_J. The latter affects R_J and it depends on the *in-situ* oxidation conditions, surface morphology, and grain size of the bottom electrode. Variations in A_J are attributed to lithographic uncertainties in exposure, development, descumming and sidewall deposition, and angular evaporation of aluminum electrodes. Figures. 2.6(a) and 2.6(b) show wafer heatmaps of the measured R_N for two different junction sizes.

To decouple these two sources of variations, the heatmaps in Figs. 2.6(c) and 2.6(d) show the extracted R_J and d_0, respectively. The inter-chip standard deviation in R_J is 3.3 $\Omega \cdot \mu$m^2 (CV_{R_J}=1.3%), while d_0 has a standard deviation of 4 nm. While the gradient in d_0 can be explained by sidewall deposition and angular evaporation [39], the gradient in R_J, albeit small, is not readily understood. Since R_J depends on the surface morphology, the surface of the bottom electrodes on the left side of the wafer could lead to a thicker tunnel-barrier oxide, hence, higher R_J. The rougher surface on the left side of the wafer could be attributed to the steeper deposition angle compared to the right side [34].

Another method of decoupling the two sources of uncertainties is to fit the CV_{R_N} data in Fig. 2.5(d) to the model proposed in **Paper A**, where

$$CV_{R_N}^2 = CV_{R_J}^2 CV_{A_J}^2 + CV_{R_J}^2 + CV_{A_J}^2. \tag{2.14}$$

Since $CV_{A_J} = \sigma_{A_J}/A_J = 2\sigma_{d_J}/\sqrt{A_J}$, one extracts CV_{R_J} of 1.8%, and $\sigma_{d_J} = 4$ nm. This comes to a close agreement with the previous method, with the higher value of 1.8% instead of 1.3% in CV_{R_J}. One important difference

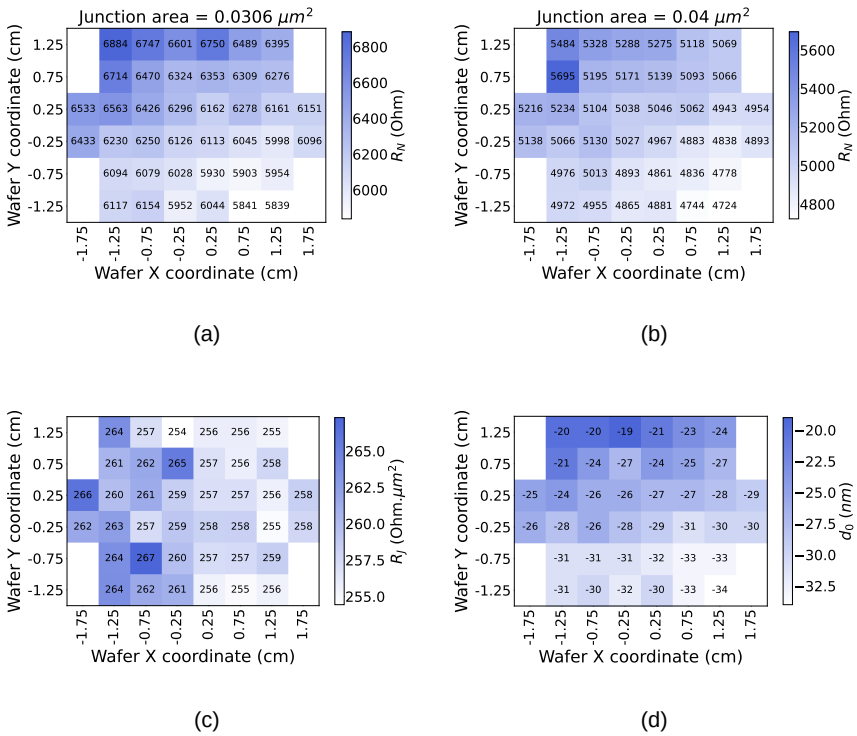


Figure 2.6: Wafer-level heatmaps of junction resistance. (a) and (b) Wafer heatmaps of the measured R_N for two different junction sizes. (c) and (d) Wafer heatmaps of the extracted junction's resistance per unit area R_J, and the bias in junction width d_0, respectively.

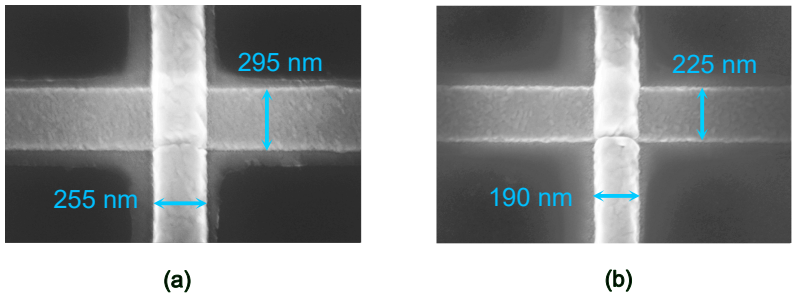


Figure 2.7: Difference in electrode width due to sidewall deposition. SEM images of two junctions, designed to be nominally the same width, where the junction in (a) is near the center of the wafer, and the junction in (b) is near the edge. The difference in their measured widths is an artifact of the angular evaporation.

between the two methods is that in the second method, CV_{R_J} is calculated between all individual junctions on the wafer, leading to a higher spread than taking the mean R_J of each chip and then calculating the standard deviation.

Although both methods provide an insightful understanding of the two sources of uncertainty in R_N, we consider them qualitative models since they rely on two imperfect assumptions. The first assumption is that R_J is independent of A_J. The second assumption is that the junctions have square areas. While junctions are designed to be so, the deposition of the first (bottom) electrode causes sidewall deposition on the resist wall of the second (top) electrode, resulting in a narrower top junction electrode [37]. Specifically, near the center of the wafer, the width of the top electrode is $t_b \times \cos(45°)$ narrower than the bottom electrode, where t_b is the nominal thickness of the bottom electrode (50 nm in our case). This results in the top electrode being approximately 35 nm narrower than the bottom one. This difference varies as we move away from the center since the deposition angle deviates from 45°. Figure 2.7 shows SEM images of two junctions, which were designed to be nominally the same, but in practice, their width differs.

For accurate separation of the two error sources, a comprehensive wafer-

scale SEM analysis of junction dimensions is essential. This process can be greatly accelerated by utilizing automated SEM machines, which are capable of scanning thousands of junctions within a few hours. To better understand the uncertainty in R_J, a deeper comprehension of the uniformity of oxide barrier thickness and its relationship with junction area is required. However, achieving this is more complex, as conducting statistically significant TEM studies on the tunnel barrier is resource-intensive. Furthermore, linking oxide barrier thickness to R_J requires modeling the thickness distribution within the barrier [64].

2.3.4 Larger, thick-oxide Josephson junctions

To scale beyond a few tens of qubits in a quantum processor, better reproducibility of junction fabrication is required compared to the values reported in Section 2.3.2.[7] To achieve a lower spread in R_N, in **Paper B**, we proposed fabricating junctions with larger areas. As Fig. 2.5(d) shows, the larger the junction area is, the lower CV_{R_N} becomes, due to the lower sensitivity to lithographic errors. There, the low spread in R_N for larger junctions was not utilized, as qubit junctions had small areas between 0.02–0.04 μm^2. This size range is limited by the target R_N required to satisfy a typical transmon qubit frequency range between 4 and 6 GHz.

To increase the junction area, while maintaining the same R_N (or qubit frequency f_{01}) range, the oxide thickness must increase. We first conducted an experiment to study the effect of junction oxidation time and pressure on R_J. Fig. 2.8(a) shows the extracted $\overline{R_J}$ for six different oxidation doses. We define the oxidation dose as the product of oxidation pressure and oxidation time. The thin-oxide dose (used for junctions in Section 2.3.2) is 40 mbar·min, with an oxidation pressure of 2 mbar over a duration of 20 minutes. For larger-area, thick-oxide junctions, we selected a dose of 600 mbar·min, with an oxidation pressure of 10 mbar and a time of 60 minutes. R_J was extracted to be 926 $\Omega \cdot \mu$m^2 for the latter dose, compared to 260 $\Omega \cdot \mu$m^2 for the former. The reasoning behind choosing the 600 mbar·min dose is threefold. Firstly, junction areas that produce the target frequency range become large enough such that CV_{R_N} is at a nearly saturating level. Secondly, selecting a higher dose does not provide a significant gain in reproducibility without increasing

[7] In Chapter 4, I present simulation and analysis of required junction reproducibility levels to achieve a low number of collisions between quantum gates in a QPU.

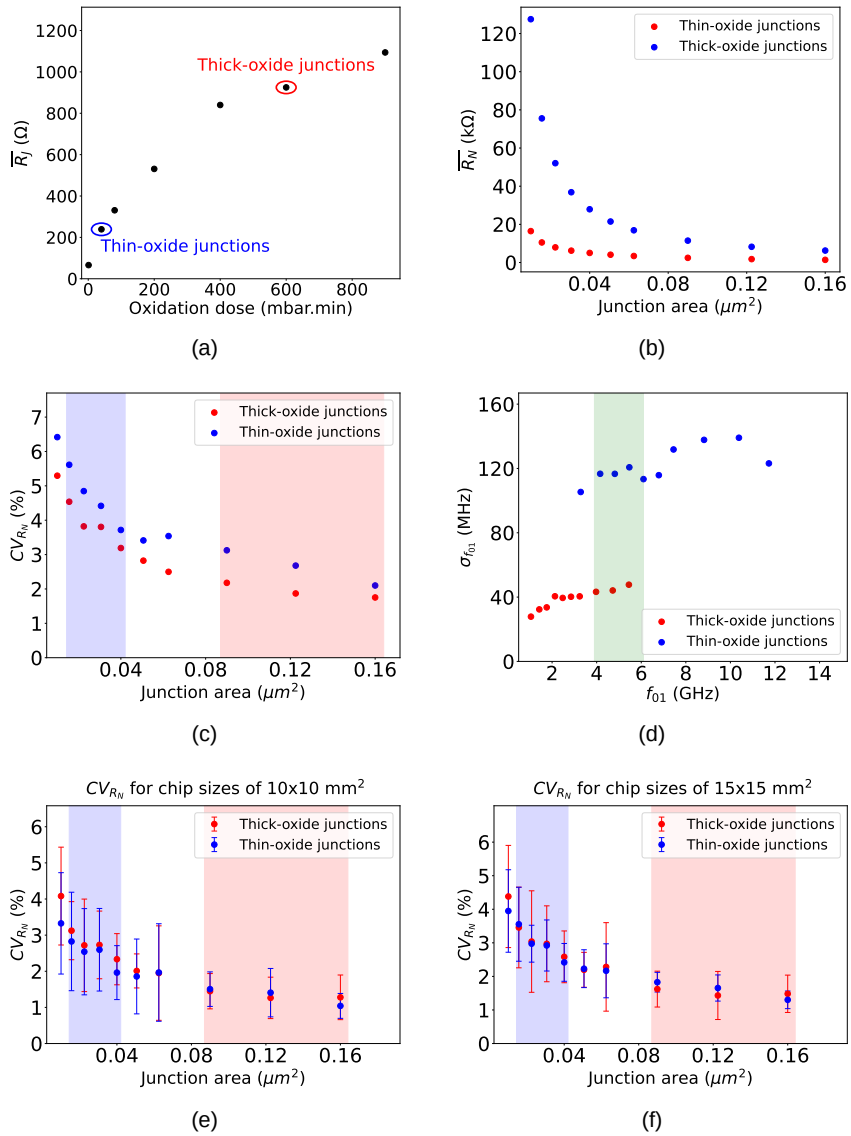


Figure 2.8: Wafer-scale comparison between thin- and thick-oxide junctions. (a) Dependence of junction resistance on oxidation conditions, highlighting the two oxidation conditions presented in the rest of the figures: 40 mbar·min for thin-oxide junctions and 600 mbar·min for thick-oxide junctions. (caption continued on next page)

Figure 2.8: (b) $\overline{R_N}$ as a function of the junction area. (c) CV_{R_N} as a function of the junction area. (d) Expected $\sigma_{f_{01}}$ using Eq. 2.13 for the two types of junctions, where the green region highlighes the standard frequencies of our transmons. (e) and (f) Average chip-level CV_{R_N} as a function of the junction area for chip areas of 10×10 mm^2 and 15×15 mm^2, respectively.

the risk of hosting two-level-system (TLS) defects within the tunnel barrier (see Chapter 3). Finally, larger junctions require a thicker resist stack, which may adversely affect R_N reproducibility, as patterning high-resolution features becomes more challenging.

To study the reproducibility of the thick-oxide junctions, I fabricated and characterized a test wafer, and ran a similar analysis to that in Section 2.3.2. Figure 2.8(b) shows $\overline{R_N}$ versus junction area for both thin- and thick-oxide junctions highlighting the increase in $\overline{R_N}$ for thick-oxide junctions. Figure 2.8(c) shows CV_{R_N} as a function of the junction area. Most notably, thicker-oxide junctions do not show higher uncertainties. This is an essential milestone to verify that the advantage of fabricating a larger junction is not canceled out by a degraded tunnel barrier uniformity. In fact, thicker-oxide junctions show lower CV_{R_N}; however, we attribute this improvement mostly to a more careful descumming [33]. The blue(red)-shaded region on Fig. 2.8(c) highlights typical junction areas and their corresponding CV_{R_N} used with thin(thick)-oxide junctions, with up to threefold reduction in CV_{R_N}. The wafer-level CV_{R_N} for junction sizes of interest is ∼2%. To understand the implications of this on the qubit frequency reproducibility, Fig. 2.8(d) shows the anticipated wafer-level standard deviation in qubit frequency $\sigma_{f_{01}}$, using derivations in Section 2.1.3. Larger, thicker-oxide junctions show near threefold reduction in $\sigma_{f_{01}}$– down to 45 MHz from an anticipated ∼120 MHz for the smaller, thin-oxide junctions.

To examine chip-level variations in R_N, I plotted the average CV_{R_N} for areas of 10×10 mm^2 and 15×15 mm^2 in Figs. 2.8(e) and 2.8(f), respectively. The chip areas were analyzed by considering all possible combinations of 5×5 mm^2 windows, resulting in 27 combinations of 10×10 mm^2 areas and 16 combinations of 15×15 mm^2 areas. On a QPU with a square qubit grid and an average spacing of 2 mm between qubits, a 10×10 mm^2 area can accommodate 25 qubits, while a 15×15 mm^2 area can hold 49 qubits.

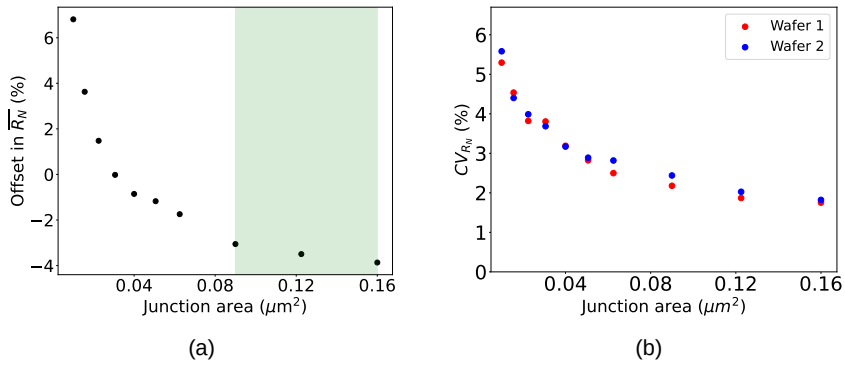


Figure 2.9: A study of two test wafers for R_N reproducibility. (a) Offset in wafer-level $\overline{R_N}$ from wafer 1 to wafer 2. (b) CV_{R_N} showing identical values for the two wafers.

Finally, we confirmed the reproducibility of the CV_{R_N} values by fabricating an additional wafer (wafer 2) using the same fabrication process as the thick-oxide-junction wafer presented earlier in this section (wafer 1). Figure 2.9(a) illustrates the percentage offset in $\overline{R_N}$ of wafer 2 from wafer 1. Notably, this offset is size-dependent, as the sensitivity to an offset in the junction width differs across junction areas. The extracted R_J for wafer 2 was 6.5% lower than that of wafer 1. For junction sizes of interest, the offset in R_N between the two wafers is within 3–4% (green-shaded area in Fig. 2.9(a)). This level of R_N shift is acceptable for our QPU architecture, as the frequency shift across all qubits would be similar. For instance, qubits with f_{01} in the 4 to 5 GHz range would experience a shift of approximately 80–100 MHz for a 4% offset in R_N. More importantly, we observed that CV_{R_N} remained nearly identical between the two wafers, demonstrating that the low uncertainty in R_N could be consistently achievable.

It is important to note that wafer-to-wafer reproducibility in Josephson junction fabrication depends not only on the specific process steps but also on the cleanroom environment, tool conditions, and chemicals used. Thus, a comprehensive study on reproducibility requires careful control of these factors, as well as the analysis of tens of wafers to obtain statistically significant results. However, this study is beyond the scope of this thesis.

2.4 Qubit frequency reproducibility

While characterizing R_N at room temperature provides a useful metric for the reproducibility of f_{01}, it was crucial to confirm that the reproducibility levels inferred in Fig. 2.8(d) hold at low temperature. Various factors, such as shifts in R_N, Δ (the superconducting gap), and E_C (the charging energy), may come into play at low temperature, potentially affecting the device performance. Therefore, low-temperature verification was essential to ensure consistent reproducibility.

To characterize the reproducibility of f_{01} at low temperature, I fabricated and characterized 32 transmon qubits over an area of 10×14 mm^2. This area was comprised of four identical 5×7-mm^2 adjacent chips. Figure 2.8(a) shows the layout of a representative chip, containing 8 qubits, each capacitively coupled to a readout resonator, which in turn is inductively coupled to a feedline. The nominal design area of all qubit junctions were identical with $d_n = 0.33$ μm. The chips were packaged in copper boxes and cooled down to $\sim$10 mK in a dilution refrigerator.

The frequencies and lifetimes of the 32 qubits were characterized. In this section, I will focus on the reproducibility of qubit frequencies, while in Chapter 3, I will address qubit lifetimes and quality factors. The distribution of the measured f_{01} is depicted in Fig. 2.10(b), where the mean f_{01} of the 32 qubits is 4.596 GHz, with a standard deviation $\sigma_{f_{01}}$ of 39 MHz (0.8%), excluding three outliers.[8] Without excluding the outliers, $\sigma_{f_{01}}$ increases to 49 MHz (1.1%).

To disentangle the contributions of E_C and E_J to the frequency variations, we measured the qubit anharmonicity and extracted both parameters using the Cooper pair box Hamiltonian[50]. Figures 2.10(c) and 2.10(d) show histograms for the two parameters, respectively. σ_{E_C} is 1.7 MHz (0.9%) after excluding three outliers and increases to 2.9 MHz (1.5%) when no exclusions are made. This non-negligible contribution of E_C to frequency variations highlights the need to optimize the patterning process of the wiring layer, which we expect to be an easier task than improving the reproducibility of junctions, given the relatively large capacitor size. As for E_J, the calculated σ_{E_J} is 0.27 GHz (1.8%). Using Eq. 2.10, we extract R_N, where σ_{R_N} is 0.17 k.Ω ($CV_{R_N} = 1.8\%$, with no outliers excluded). This comes to a very close agree-

[8] Outliers are defined as values more than two standard deviations away from the mean, though they are still shown on the histograms.

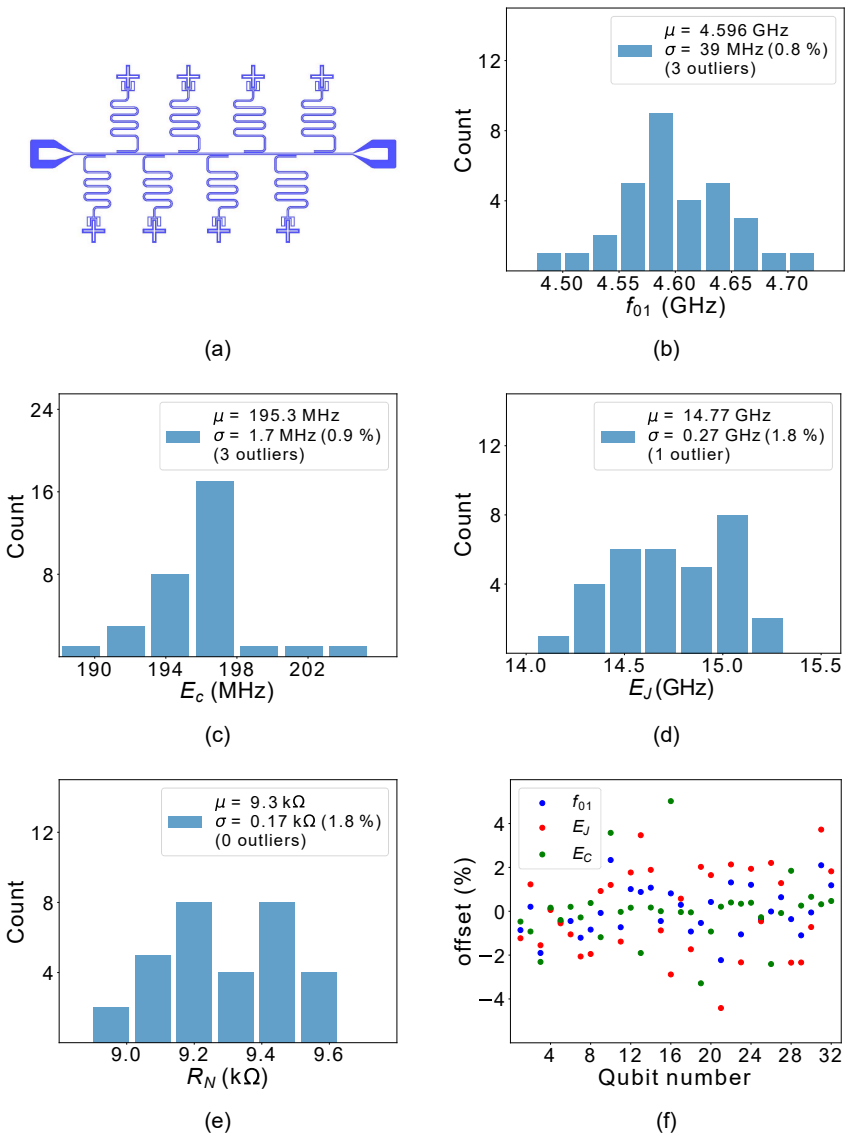


Figure 2.10: Qubit frequency reproducibility. (a) Layout of a representative chip from the four characterized chips, where each chip contains 8 qubits. (b), (c), (d), and (f) Histograms of the measured f_{01}, extracted E_C, E_J, and R_N, respectively. (f) The offset around the mean for the measured f_{01} and extracted E_C and E_J.

ment with the anticipated chip-level CV_{R_N} shown in Figs. 2.8(e) and 2.8(f) for junctions with the same A_J as the 32 qubits (0.1089 μm^2) and on a similar chip area.[9] The offset around the mean measured f_{01} and extracted E_C and E_J for all 32 qubits is illustrated in Fig. 2.10(f). This provides further insight into the source of deviation in f_{01} for each qubit. Finally, it is important to note that the extraction of R_N from E_J, using Eqs. 2.5 and 2.6, assumes that Δ is uniform across all junctions.

2.5 Further insight into Josephson junctions

In this section, I present preliminary studies conducted on our junctions, each of which requires further extensive investigation, which I highlight at the end of each subsection.

2.5.1 Aging of Josephson junctions

R_N is known to increase over time from the moment the junctions are fabricated, a process referred to as Josephson junction aging [59, 65, 66]. This phenomenon is caused by the diffusion of oxygen into the tunnel barrier or the redistribution of oxygen within the barrier. The aging follows a logarithmic trend, with a sharp increase in R_N shortly after fabrication that gradually slows and saturates over a longer period. The extent and duration of this increase in R_N depend on several factors. Firstly, stronger oxidation conditions (in terms of time and pressure) are more likely to saturate the oxidation of aluminum, as the process is self-limiting, thereby reducing the likelihood of further aging. Secondly, the conditions under which the junctions are stored after fabrication, such as pressure, humidity, temperature, and the ambient gas or liquid, also play a significant role in the aging process.

In this study, we investigated the aging of Josephson junctions and its dependence on junction size. R_N values from the same wafer, as discussed in Section 2.3.4, were first measured on the second day following junction lift-off (day 0), then again after 7 days, and finally after 21 days from the initial characterization. Figure 2.11(a) illustrates the change in R_N over these three

[9] In **Paper B**, we used algebraic manipulation to extract CV_{R_N} from both f_{01} and E_C and their CV. However, the method presented here is more accurate as it extracts R_N directly.

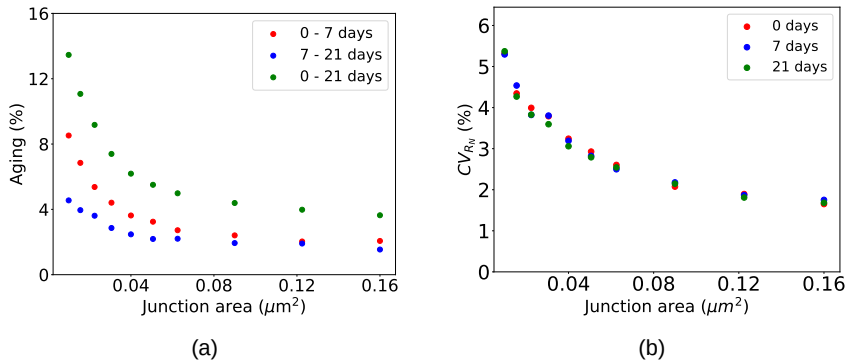


Figure 2.11: Aging of Josephson junctions. (a) Aging of Josephson junction's $\overline{R_N}$ over a wafer after 7 days (red) and 21 days (green), and after 21 days compared to after 7 days (blue). Aging is defined as the increase in $\overline{R_N}$ after n days from day 0, divided by $\overline{R_N}$ at day 0 in percentage. (b) CV_{R_N} of the same wafer showing no change over time.

points in time. Firstly, we observe that aging is size-dependent; smaller junctions exhibit greater aging than larger ones. Secondly, the aging rate decreases over time; the change in R_N over the 7 days from day 0 to day 7 is larger than the change over the 14 days from day 7 to day 21. However, longer-term studies are required to investigate the aging rate in more detail.

To understand whether the reproducibility of R_N degrades over time, CV_{R_N} at the three points in time is shown in Fig. 2.11(b). With no noticeable change in CV_{R_N}, we concluded that the oxide redistribution or diffusion, due to aging, has not significantly impacted the uniformity of the junction resistance. This conclusion is significant. An increase in R_N over time can be accounted for in design or subsequent chip selection; however, an increase in CV_{R_N} cannot be mitigated, except with post-fabrication tuning techniques [40, 42].

A more comprehensive study on junction aging could encompass extended timescale characterization of R_N for junctions fabricated under varying oxidation conditions. Furthermore, investigating the impact of storage environments on aging, such as comparing normal office space conditions, cleanroom storage, and liquid nitrogen preservation, would provide valuable insights into the factors influencing long-term stability.

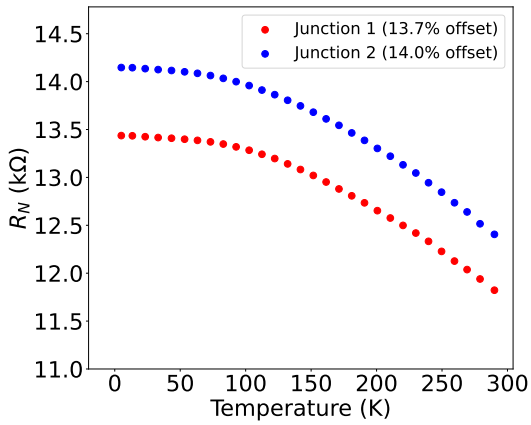


Figure 2.12: R_N dependence on temperature. R_N of two thin-oxide junctions with nominal area of 170×170 nm^2 measured using a PPMS system at different temperatures from 300 K down to 5 k.

2.5.2 Low temperature vs Room temperature R_N

One important characteristic of Josephson junctions is the increase in measured R_N from room temperature to low temperature. This is attributed to two likely factors. The first is the shunting effect, where the finite sheet resistance of the silicon substrate forms a shunt for charge carriers to bypass the tunnel barrier. This shunt resistance depends on the bare silicon wafer property, temperature of the substrate and light. With regard to the latter, we have consistently observed that the measured R_N drops by up to 10% depending on R_N if the substrate is exposed to light, which generates more charge carriers in the silicon's conduction band.[10] The second factor is the quantum tunneling and its temperature dependence. At room temperature, a higher number of charge carriers has sufficient energy to tunnel through the barrier. The number of carriers drops at low temperature which increases the barrier resistance [67].

Figure 2.12 shows R_N measurements for two junctions characterized using a Physical Property Measurement System (PPMS), where the temperature is swept from 300 K down to 5 K. Both thin-oxide junctions have a nominal

[10] All wafer-scale R_N studies are conducted inside a dark box to minimize the effect of light [Fig. 2.4(b)].

area of 170×170 nm^2. We found that both junctions increased in resistance by 14%. Normally, we take into account this 14% whenever we design qubit junctions to achieve a certain low-temperature R_N, hence, E_J.

A further study on R_N dependence on temperature should investigate this dependence for different junction sizes. Additionally, this study could be conducted on junctions made with different oxidation conditions, as one would expect a different behavior of tunneling due to the different tunnel barrier heights.

CHAPTER 3

Qubit Coherence

The performance of superconducting qubits and our ability to run quantum computations on them are highly constrained by their coherence. Coherence is the ability of a qubit to maintain a well-defined phase between its quantum states in a superposition over time. These constraints are imposed by the qubit's coupling to the environment, such as two-level system (TLS) defects, quasi-particle (QP) tunneling, radiation, magnetic field noise, and Purcell decay. Protecting the qubit from these channels of loss requires careful circuit design, process development, and shielding from the radiative and magnetic environment. In Section 3.1, I start with a brief introduction to some of the loss channels. In Section 3.2, I present coherence characterization of transmon qubits focusing on a comparison between the two-step and PICT-junction qubits and later large-junction qubits, which I discussed in Chapter 2. In Section 3.3, I present TLS spectroscopy measurements on small and large-junction tunable qubits. Finally, in Section 3.4, I present our work on mitigating interfacial loss in superconducting circuits.

3.1 Loss channels

A qubit's coherence is defined by the characteristic time T_2, given by:

$$\frac{1}{T_2} = \frac{1}{2T_1} + \frac{1}{T_\phi}, \tag{3.1}$$

where T_1 is the relaxation time from $|1\rangle$ to $|0\rangle$ state; in other words, loss of energy, while T_ϕ is the pure dephasing, i.e., the phase loss without loss of energy.[1] Considering the different loss channels affecting T_2, it can be expressed as:

$$\frac{1}{T_2} = \sum_i \frac{1}{T_2^i}, \tag{3.2}$$

where T_2^i is the coherence time limited by the loss channel i. One can readily see that the shortest time constant dictates the maximum coherence time attainable, emphasizing the necessity to mitigate these losses altogether. These loss channels can be QP [68–74], TLS [75–92], Purcell decay [50, 93–98], radiation, and magnetic field noise [99–104]. In the remainder of this section, I give a brief overview of the relevant dynamics of TLS and the Purcell decay. However, other sources of coherence loss are not discussed in this thesis.

3.1.1 Two-level systems

TLSs have long been a source of mystery in the field of superconducting qubits and other areas of research. While their existence and detrimental effects on qubit coherence are well-studied, the exact origins of TLS are still not well-understood. However, several hypotheses have been proposed, including OH bonds in amorphous materials [75], interstitial hydrogen [76], hydrogenated aluminum vacancies [77], delocalized oxygen atoms [78], tunneling electrons [79, 80], surface spins and magnetic impurities [81–83, 87, 88]. There is a consensus that the highest density of TLS defects occurs at the interfaces, such as metal-air, substrate-air, and substrate-metal interfaces, largely due to their amorphous structure. Despite the uncertainty surrounding their origins, the standard tunneling model (STM) is widely accepted as a phenomenological approach to treating TLS in these different systems [105, 106]. In this

[1] Note that sometimes, I refer to the decay rate $\Gamma_x = 1/T_x$ instead of the decay time T_x throughout the thesis.

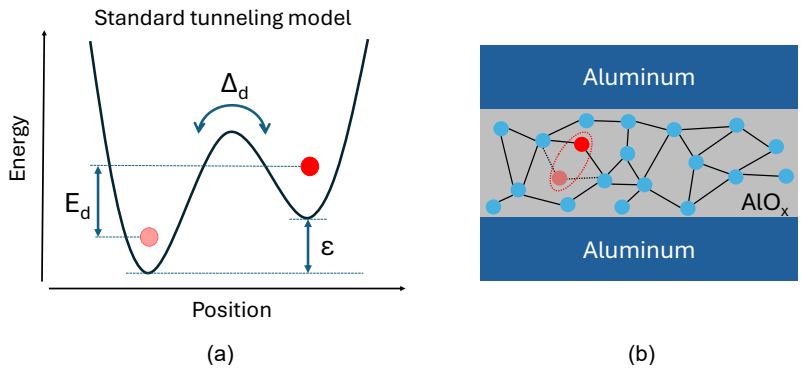


Figure 3.1: TLS model. (a) Standard-tunneling-model representation of a TLS, where a particle (red) is trapped in a double-well potential with a ground state asymmetry of ε and a tunneling energy between the two sides of Δ_d. The energy eigenstates of such a system are separated by E_d [105, 106]. (b) Example of a TLS within a junction tunnel barrier. In this example, the TLS is an atom (red) that can switch its position between two energetically similar configurations.

model, a particle is trapped in a double-well potential [Fig. 3.1(a)], where it can sit in one of two parabolic sides, with a non-zero probability of tunneling between them. Due to this tunneling, the two lowest eigenstates in the left and right sides of the potential well hybridize to form two energy eigenstates with energy difference

$$E_d = hf_d = \sqrt{\varepsilon^2 + \Delta_d^2}, \tag{3.3}$$

where f_d is the TLS defect frequency, ε is the energy asymmetry between the two sides of the double well, and Δ_d is the tunneling energy between them. An example of such a TLS is shown in Fig 3.1(b), where it lies within the junction tunnel barrier's amorphous AlO_x. In this example, a bistable atom (red) can switch between two energetically similar positions.

Such a TLS can be treated as an electric dipole that couples to the qubit's (or resonator's) electric field $\mathbf{E}$, with strength $g' = \mathbf{p}\cdot\mathbf{E}$, where $\mathbf{p}$ is the electric dipole moment of the TLS. This suggests that defects in regions with strong electric field, such as within the junction's tunnel barrier of a transmon qubit, exhibit a stronger coupling to the qubit than those in regions with weaker electric field away from the metal edges of the transmon capacitor and the junction electrodes. For a transmon qubit, $\mathbf{E}$ across the junction tunnel barrier is ~ 3000 V·m^{-1}, which drops to a few V·m^{-1} about 200 nm away from the metal edges [84, 85]. The qubit-TLS coupling g_d is $g'\Delta_d/\sqrt{\varepsilon^2 + \Delta_d^2}$. When a qubit interacts with a TLS, the relaxation rate of the qubit Γ_1 exhibits a Lorentzian behavior that is given by:

$$\Gamma_1 = \frac{2g_d^2\Gamma}{\Gamma^2 + \delta_d^2} + \Gamma_{1,q}. \tag{3.4}$$

$\Gamma = \Gamma_{1,q}/2 + \Gamma_{\phi,q} + \Gamma_{1,d}/2 + \Gamma_{\phi,d}$ is the total decoherence rate of the qubit-TLS system, where $\Gamma_{1,d}$ and $\Gamma_{\phi,d}$ are the TLS relaxation and dephasing rates, respectively [84]. This equation will be useful when I discuss TLS spectroscopy in Section 3.3.

3.1.2 Purcell decay

Another channel of qubit's energy relaxation is the decay due to the Purcell effect, where a qubit loses a photon via spontaneous emission to the readout resonator or the qubit drive line. In the case of the readout resonator, the

qubit decay rate due to the Purcell effect can be approximated by

$$\Gamma_1^P \approx \kappa_r \frac{g_{qr}^2}{\Delta_{qr}^2}, \tag{3.5}$$

where κ_r is the resonator linewidth, g_{qr} is the qubit-resonator coupling strength, and Δ_{qr} is the detuning between the qubit and the resonator. While it seems easy to mitigate this decay by either increasing Δ_{qr} or reducing g_{qr}, the qubit state readout time becomes slower as fast readout requires high g_{qr}. Instead, another circuit component, referred to as the Purcell filter, is used to protect the qubit from such a decay while maintaining a high enough coupling and a low enough detuning [94–98]. While the Purcell effect is not a discussion topic of this thesis, Eq. 3.5 will be used later in this chapter to decouple Purcell losses from TLS losses and will be revisited in Chapter 5 when I discuss design considerations of the QPU.

3.2 Coherence characterization

Qubit characterization usually involves measurements of T_1 and T_2. All qubits characterized in this section have the same design, where a transmon qubit is capacitively coupled to a readout resonator, which is inductively coupled to a feedline as shown in Fig. 3.2(a). Only the junction sizes are varied to achieve the required E_J. To characterize T_1, a π-pulse at the qubit frequency excites the qubit to the $|1\rangle$ state. After variable delays, the qubit state is read out through the readout resonator. The measured population of the $|1\rangle$ state decays with an exponential e^{-t/T_1}, where t is the delay time. To characterize T_2, a $\pi/2$-pulse, slightly detuned from the qubit frequency, excites the qubit to the equator of the Bloch sphere, then we wait for variable delays for the qubit to acquire some phase, then another $\pi/2$-pulse is applied to the qubit with a readout pulse immediately after. The measured population of the $|1\rangle$ state can be modeled as a decaying sinusoid, $e^{-t/T_2} \sin(2\pi\delta_{01})$, where δ_{01} is the detuning of the drive pulse from the qubit frequency. Figures 3.2(b) and 3.2(c) show the pulse sequences and exemplary traces for T_1 and T_2 measurements. Note that for these devices, the qubit is driven through the readout resonator.

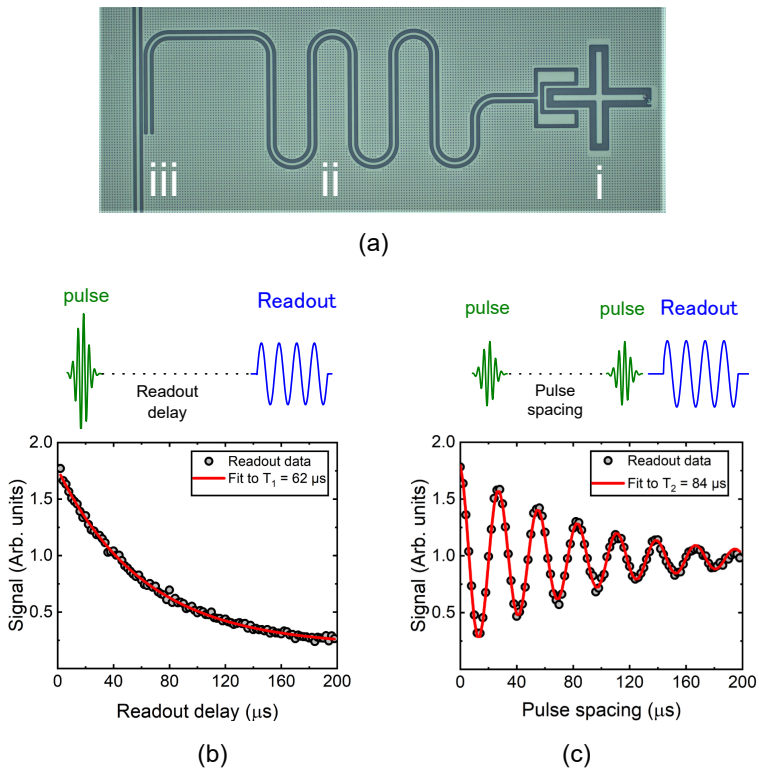


Figure 3.2: Coherence characterization. (a) Micrograph of a transmon qubit (i), coupled to a readout resonator (ii), which is inductively coupled to a feedline (iii). (b) and (c) Pulse sequence and exemplary traces of T_1 and T_2 measurements, respectively.

3.2.1 PICT transmons

As mentioned in Chapter 2, qualifying the PICT process entails characterization of both R_N reproducibility and the qubit coherence. Here, I present our coherence study of qubits made with PICT junctions in comparison to the standard (two-step) junction. The study includes two qubit chips: one with PICT junctions and the other with standard junctions. Both chips are identical in design and fabrication, except for the junctions. Each chip contains three qubits, named S-X1, S-X2, and S-X3 for the standard qubits and P-X1, P-X2, and P-X3 for the PICT qubits. Figure 3.3 shows histograms of the measured T_1 and T_2 for all qubits.[2] Both qubit types show comparable coherence values. Thus, we concluded that the PICT qubits maintained the quality factors ($Q{=}2\pi f_{01}T_1$ $>1.5{\times}10^6$) achieved with the two-step process. Beyond this study, several tens of qubits were later fabricated with the PICT process, which showed high quality factors, such as those qubits in **Paper B** (Section 3.2.2) and **Paper C** (Section 3.4).

3.2.2 Larger-junction transmons

In Chapter 2, I demonstrated that larger junctions lead to higher reproducibility in R_N and f_{01}. However, one concern with using larger-junction qubits is the potentially increased presence of TLS defects within the junction's tunnel barrier [75, 85, 107]. In [85], experiments probing these TLSs by tuning their frequencies—and thereby their interaction with the qubit—revealed that the density of these TLSs is about 1.5 $\mathrm{GHz}^{-1}\cdot\mu\mathrm{m}^{-2}$ for a 2-nm-thick dielectric.

To investigate whether this renders the large-junction qubits prohibitively prone to losses compared to the small-junction qubits, we measured T_1 values of the 32 qubits presented in Section 2.4. The average quality factor of all qubits was $1.65{\times}10^6$ [Fig. 3.4(a)], comparable to the quality factors reported in the previous section. Notably, we found that T_1 for most of these qubits had a non-negligible effect from Purcell decay. Figure 3.4(b) shows the scaling of the measured T_1 with T_1^P. Similarly, Fig. 3.4(c) shows the scaling of Q with

[2] Coherence data shown in Table I of **Paper A** is limited to only 15 hours so that all qubits matched in their measurement time span, while those shown in Fig. 3.3 include all measured coherence data.

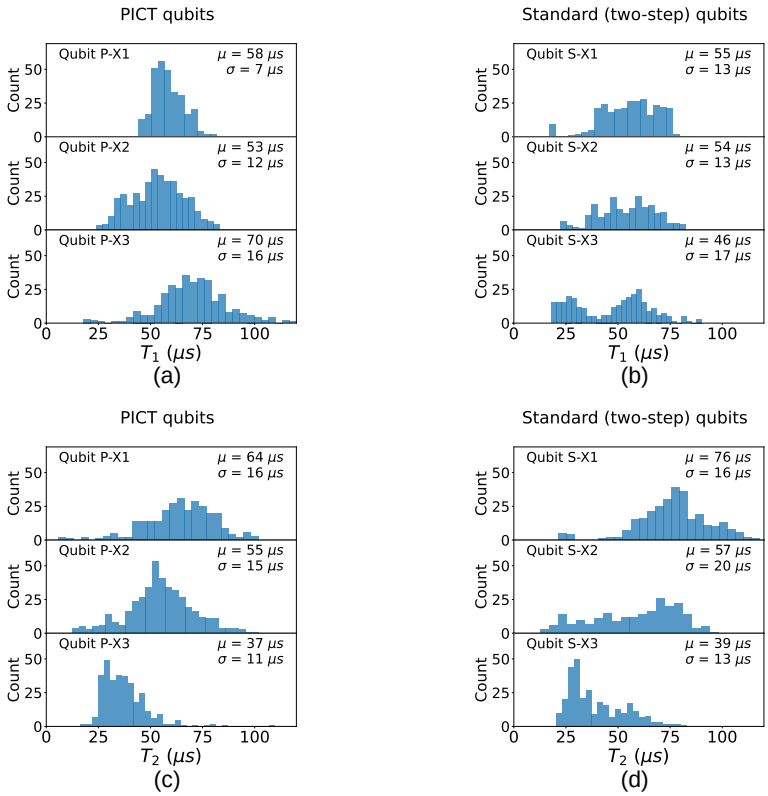


Figure 3.3: PICT-junction qubit coherence. (a) and (b) Histograms of the measured T_1 on three qubits of the PICT-junction and two-step-junction qubits, respectively. (c) and (d) Histograms of the measured T_2 of the same qubits in (a) and (b), respectively. The mean μ and standard deviation σ of all histograms are stated on the plots.

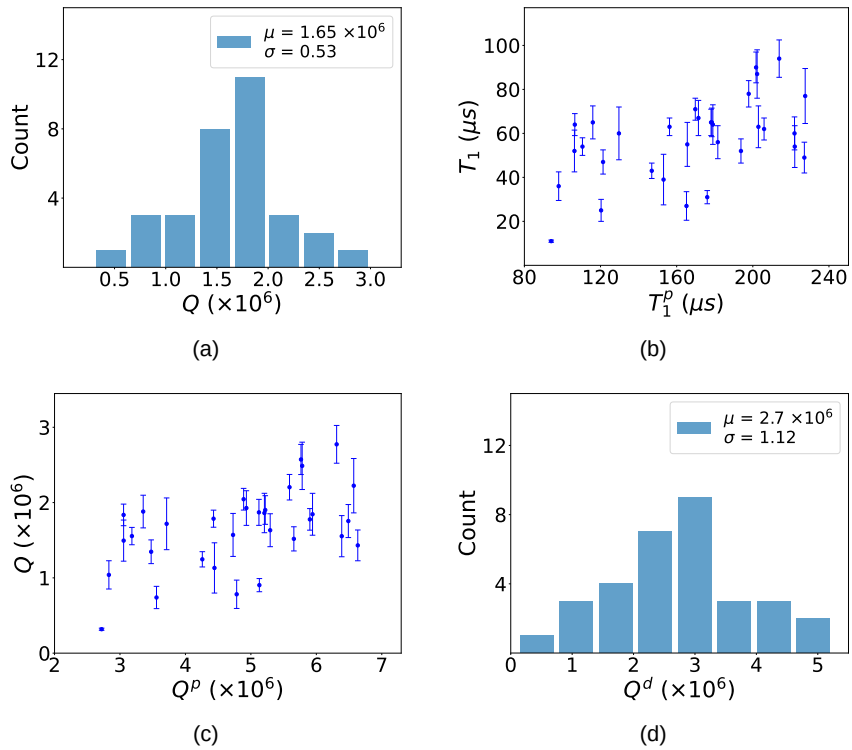


Figure 3.4: Quality factors and lifetimes of large-junction qubits. (a) Histogram of the measured quality factors of 32 large-junction qubits. (b) scaling of T_1 with the Purcell-limited lifetime T_1^P showing that some qubits were Purcell-limited. (c) Scaling of Q with the Purcell-limited quality factors Q^P. (d) Histogram of the TLS-limited quality factors Q^d. Error bars on (b) and (c) represent one standard deviation away from the mean.

Q^P (the limit on Q imposed by Purcell decay). Assuming

$$\frac{1}{Q} = \frac{1}{Q^d} + \frac{1}{Q^P}, \tag{3.6}$$

the average Q^d (the limit on Q imposed by TLS defect losses) is about 2.7×10^6 [Fig. 3.4(d)].[3] These quality factors show that there is no noticeable degradation in qubit performance when fabricated with large junctions. In Section 3.3, I present further investigation on the quality of larger-junction qubits via TLS spectroscopy measurements.

3.3 TLS spectroscopy

TLS spectroscopy is considered an essential tool for characterizing the quality of the qubit material and design. This spectroscopy technique detects TLSs by tuning the qubit frequency to or near resonance with these TLSs. The qubit-TLS interaction results in a sharp increase in the measured qubit decay rate according to Eq. 3.4. Fig. 3.5(a)-bottom shows a simulation of Γ_1 in the presence of three TLS with different couplings and decoherence rates. To mediate such interaction while probing the qubit's decay rate, a tunable qubit is used, together with the pulse sequence shown in Fig. 3.5(a)-top is applied to the qubit. A π-pulse excites the qubit to the $|1\rangle$ state, then a flux pulse tunes the qubit frequency for a duration T_s, with a readout pulse immediately after. If the flux pulse tunes the qubit into resonance with a TLS, a larger decrease in the $|1\rangle$ state population is measured in comparison to the no-near-resonance-TLS case, due to the qubit energy decay into the TLS. More details on the pulse sequence and extracting Γ_1 are in Appendix C of **Paper B**.

Here, we implement TLS spectroscopy as a means of characterizing the large-junction qubits in comparison to the small-junction ones. Two tunable qubits were fabricated for each of the two types. Tunability of these qubits was achieved by means of asymmetric SQUIDs, where the junction sizes in the SQUID loop were 0.012 and 0.032 μm^2 for the small-junction qubits, and 0.035 and 0.097 μm^2, aiming for an asymmetry $d = 0.5$ (Eq. 2.9). Qubits were tuned via a DC flux line coupled to the SQUID loop as shown in Fig. 3.5(b). Fig. 3.6(a) shows the qubit frequency as a function of the applied flux on qubit

[3] Here, the quality factors limited by other loss channels were neglected, assuming they are much higher than Q^d and Q^P.

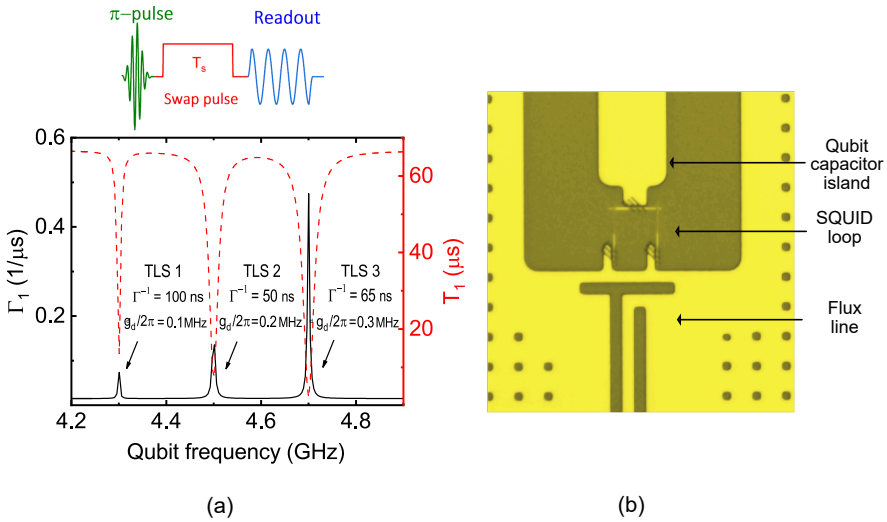


Figure 3.5: TLS spectroscopy design and simulation. (a)-top TLS swap spectroscopy pulse sequence, where a π-pulse excites the qubit to the $|1\rangle$ state, then a flux pulse with different amplitudes tunes the qubit to different frequencies to interact with TLSs in the vicinity, and finally, the $|1\rangle$ state population is readout through the resonator. (a)-bottom Simulation of Γ_1 as a function of qubit frequency in the presence of three TLSs with different frequencies, couplings g_d, and coherence rates Γ_d. (b) Micrograph of a tunable qubit with a SQUID loop, where the qubit frequency is tuned via a flux line.

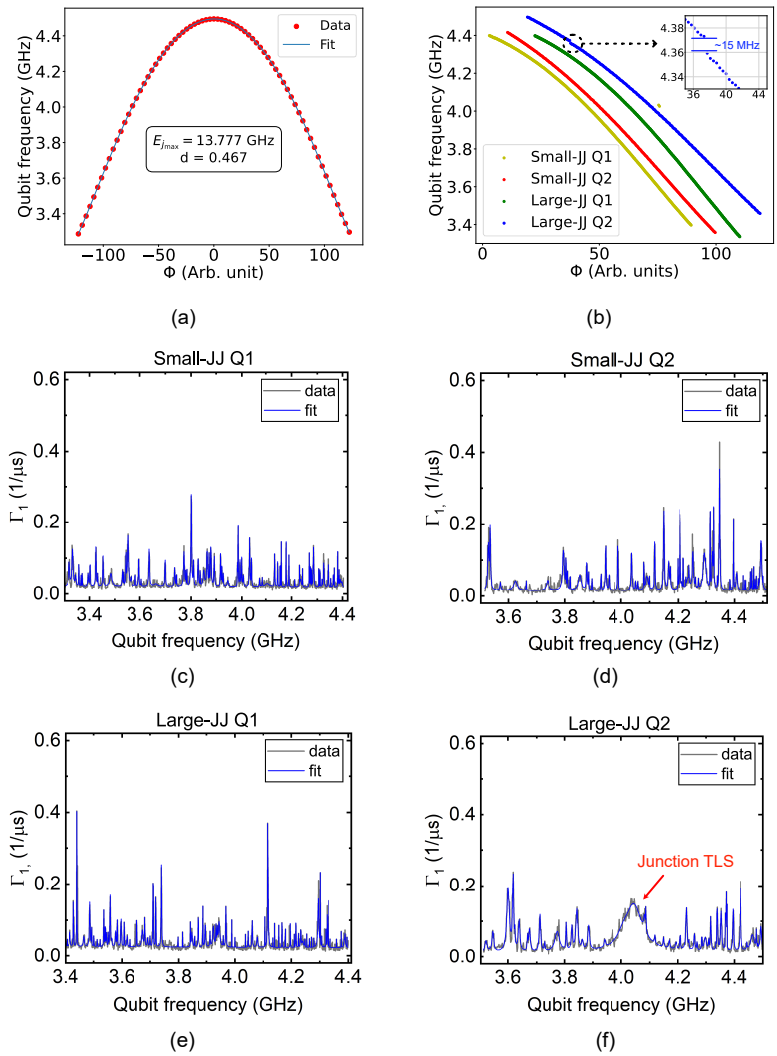


Figure 3.6: TLS spectroscopy measurements. (a) Tunability of qubit frequency via the applied flux, along with a fit to Eqs. 2.8 and 2.9. (b) TLS scan (using qubit spectroscopy) versus applied flux for four qubits: two with small junctions (Small-JJ Q1 and Small-JJ Q2), and two with large junctions (Large-JJ Q1 and Large-JJ Q2). Qubit Large-JJ Q2 exhibits an avoided-level crossing around 4.365 GHz, due to its interaction with a junction TLS, with a frequency splitting of $\sim$ 15 MHz. (c)-(f) Exemplary traces of Γ_1 as a function of qubit frequency for the four different tunable qubits.

Large-JJ Q1. By fitting this curve to Eq. 2.8 and Eq. 2.9, we extract both E_{Jmax} and d. We also see that $d = 0.467$, close to the expected $d = 0.5$. Since we are interested in extracting Γ_1 as a function of qubit frequency, using this fit, we calculate the expected qubit frequency as a function of the applied flux [x-axes of Fig. 3.6(c)-(f)].

3.3.1 TLS characterization

Before applying the TLS spectroscopy sequence, an initial characterization is conducted by sweeping the flux on the qubit and measuring its frequency, searching for any detectable avoided-level crossings. Those crossings are signs of the qubit strongly interacting with a TLS. Fig. 3.6(b) shows such a scan for the four qubits. No avoided-level crossings were observed, except for qubit Large-JJ Q2 [inset of Fig. 3.6(b)]. The splitting S in the qubit frequency reveals the coupling strength g_d between the qubit and the TLS, which we extract to be $g = S/2 \approx 7.5$ MHz. Observing such a strongly coupled TLS indicates that the TLS lies within the junction tunnel barrier [85]. The presence of this TLS was also confirmed by the TLS spectroscopy experiment.

Next, we implemented the TLS-spectroscopy experiment on the four qubits. Each qubit was scanned for TLS over a 1 GHz span, acquiring over 30 traces per qubit over a time span of approximately 15 hours. Fig. 3.6(c)-(f) show exemplary traces of Γ_1 as a function of the qubit frequency for the four qubits, along with fits to Eq. 3.4. The primary aim of this experiment is to determine if the large-junction qubits exhibit a significant increase in the number of detected TLS compared to the small-junction ones. To assess this, we first count the number of TLSs.

Given a TLS density of 1.5 $\mathrm{GHz}^{-1} \cdot \mu\mathrm{m}^{-2}$, assuming a dielectric thickness of 2 nm, the metal-air and substate-air interface of our qubit island is expected to host over 70 thousand TLSs. However, the TLS-spectroscopy technique is sensitive only to TLS with g_d above a certain threshold, set by the experimental resolution [85]. In our experiment, this threshold is $g_d/2\pi \approx 0.09$ MHz. We therefore consider only TLS above this limit to ensure a fair comparison across the four qubits. Those TLSs that are coupled more strongly lie within a few hundred nanometers from the metal edges [85]. Fig. 3.7 shows the number of well-fitted TLSs in each trace for the four qubits.[4] No significant difference

[4] Variations in the number of TLSs within each qubit over different traces are attributed to the robustness of the fitting routine rather than the variation in the number of TLSs.

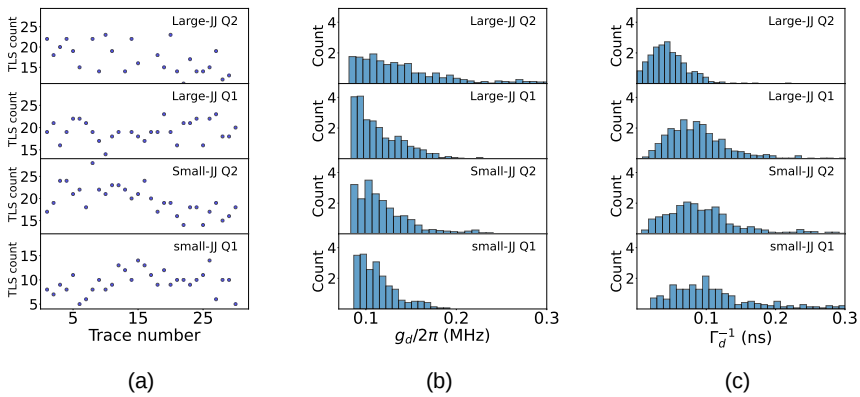


Figure 3.7: TLS parameters for small- and large-junction qubits. (a) Number of well-fitted TLSs for the four different qubits. (b) and (c) Histograms of the fitted TLS $g_d/2\pi$ and Γ_d^{-1}, respectively.

was observed between the two types of qubits, where the maximum number of well-fitted TLS varied between 14 and 29 for all qubits.

We do observe, however, a strongly coupled TLS in qubit Large-JJ Q2, indicating that this TLS resides within the junction tunnel barrier [Fig. 3.6(f)]. This confirms the strong avoided crossing we observe in Fig. 3.6(b). The difference in this TLS frequency between the two figures (4.365 GHz in Fig. 3.6(b) and 4.040 GHz in Fig. 3.6(f)) is likely because the two measurements were taken in two different cooldowns of the cryostat, causing the TLS to move in frequency due to thermal cycling (see Appendix C of **Paper B**).

What was the likelihood of observing a junction TLS? Given a total tunnel barrier area of 0.044 and 0.132 μm^2 for the two small-junction and the two large-junction qubits, we expect, on average, 0.13 and 0.4 junction-TLS, respectively. Therefore, there was a relatively high probability of observing a junction TLS in the large-junction qubit. To robustly quantify the number of TLS between the two junction size regimes, a statistically significant number of tunable qubits must be characterized, which is outside the scope of this thesis. However, in the next section, I present simulations of the expected number of junction TLSs in a QPU with hundreds of qubits.

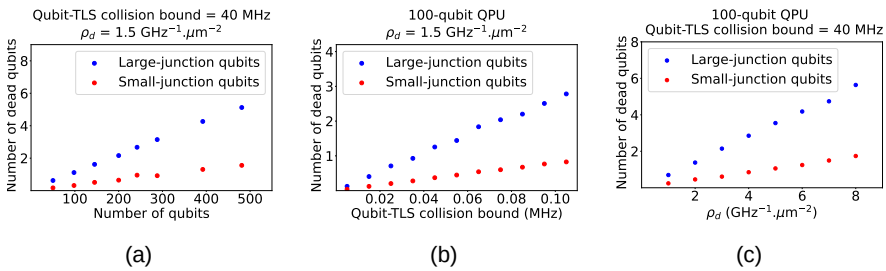


Figure 3.8: Simulation of junction-TLS impact on QPU. Scaling of the number of dead qubits as a function of the (a) number of qubits in a QPU, (b) qubit-TLS collision bound, and (c) density of junction TLS within the tunnel barrier.

To provide further insight into the TLS, Figs. 3.7(b) and 3.7(c) show histograms of $g_d/2\pi$ and Γ_d^{-1}, respectively. The distribution of $g_d/2\pi$ almost fades completely beyond 0.3 MHz and is consistent with g_d of TLSs near the film edges. The coherence time of the TLS is around 50–200 ns.

3.3.2 Implications on the QPU

Although large-junction qubits still remain within the regime of small junction sizes, with a low number of junction TLSs (compared, for example, to phase qubit junctions [75]), a QPU containing tens to hundreds of qubits incorporates multiple of such junctions. As the number of qubits increases, so does the likelihood of a qubit colliding with a junction TLS. Therefore, it becomes valuable to simulate the expected number of these collisions. For a QPU with a total number of qubits N_Q and an individual junction area A_J, the total number of junction TLSs N_d^J within a given span of 1 GHz is:

$$N_d^J = \rho_d N_Q A_J, \tag{3.7}$$

where ρ_d is the defect density. We ran a Monte Carlo simulation and randomly assigned N_d^J TLSs with different frequencies over the QPU qubits. Qubits were assigned frequencies between 4.3 and 5.3 GHz, based on the two-frequency group lattice (see Chapter 4). A qubit is considered "dead" if its

frequency is closer than a certain collision bound to a TLS frequency.

Figure 3.8(a) shows the expected number of dead qubits for small- and large-junction QPUs with different number of qubits. A_J was chosen to be 0.036 and 0.109 μm^2, respectively. As expected, the number of dead qubits on a large-junction QPU is three times larger than that of a small-junction QPU, since A_J is three times larger. On a 100-qubit large-junction QPU, we expect one qubit colliding with a junction TLS for $\rho_d = 1.5\ \text{GHz}^{-1} \cdot \mu\text{m}^{-2}$ and collision bound of 40 MHz. N_d^J also scales with the qubit-TLS collision bound, which depends on g_d [Fig. 3.8(b)]. Finally, since ρ_d depends on the quality of the tunnel barrier oxide, it is also valuable to simulate N_d^J for different ρ_d emphasizing the value of maintaining a low density of junction TLSs regardless of the junction size [Fig. 3.8(c)].

3.4 Mitigating loss in aluminum-on-silicon qubits

Given that TLS defects lie within the different dielectric interfaces of a superconducting circuit, it is crucial to minimize their volume and their participation ratio of the qubit's electric field. In **Paper D**, we conducted a systematic investigation of how different process steps influence the mitigation of TLS-induced loss. Additionally, in **Paper C**, we demonstrated improvements in the quality factors of our resonators and qubits by increasing the aluminum film thickness from 150 to 300 and 500 nm. This section summarizes the key findings from these studies. More details can be found in the appended papers.

3.4.1 Systematic study of the process steps

In Section 2.2, I provided an overview of the fabrication process steps of our superconducting circuits. In **Paper D**, we reported a systematic study of these steps, assessing their impact, or lack thereof, on the CPW resonators, which serve as a proxy for the qubit quality factors. Table I of that paper summarizes these effects, which I highlight here.

Figure 3.9 (top) shows the internal quality factors Q_i of the CPW resonators under different treatments as a function of the photon number $\langle n \rangle$. Each Q_i point is obtained from microwave forward transmission (S_{21}) measurements in a notch configuration of a CPW resonator coupled to a feedline. By fitting

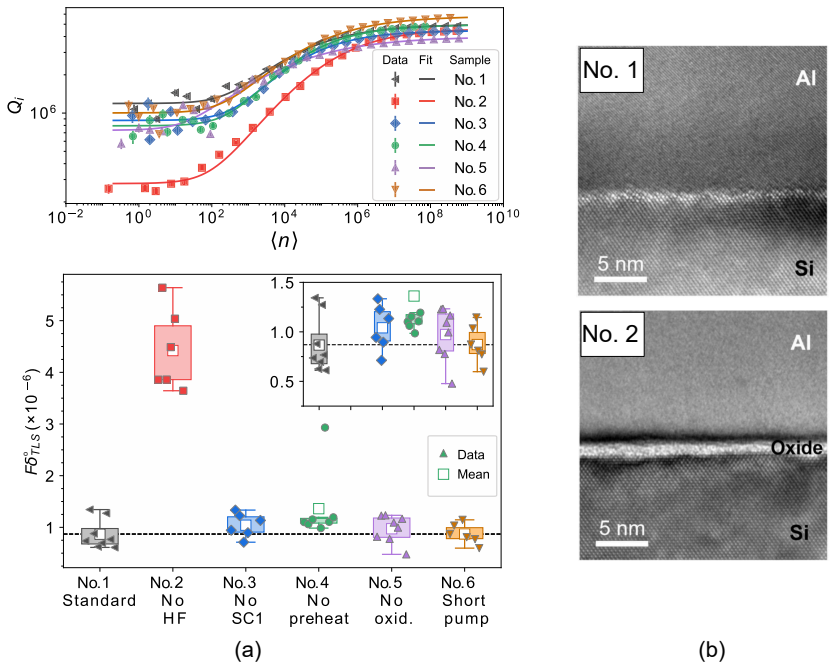


Figure 3.9: Effect of different fabrication process steps on resonator quality factor [108]. (a)-bottom Extracted TLS loss $F\delta^o_{\text{TLS}}$ of resonators under different fabrication process treatments. Sample No. 1 is the standard sample fabricated based on the process presented in Section 2.2. In each of the other 5 samples, a different process step was skipped/modified. These $F\delta^o_{\text{TLS}}$ values were extracted from fitting the internal quality factor Q_i as a function of photon number $\langle n \rangle$ shown in (a)-top to Eq. 3.8. (b) TEM images of the silicon-aluminum interface for the standard sample (No. 1) with HF treatment, and sample No. 2 with no HF treatment, before aluminum film deposition.

the S_{21} data to a circle fit [109], we extract Q_i for each photon number in the resonator. Q_i has contributions from both TLS loss, $\delta_{\text{TLS}} = 1/Q_{\text{TLS}}$, and other loss channels, $\delta_{\text{other}} = 1/Q_{\text{other}}$, such that $1/Q_i = 1/Q_{\text{TLS}} + 1/Q_{\text{other}}$.

The TLS-induced δ_{TLS} can be expressed in terms of the resonator frequency f_r, temperature T, and critical photon number n_c as

$$\delta_{\text{TLS}} = F\delta^o_{\text{TLS}} \frac{\tanh\left(hf_r/2k_BT\right)}{(1 + \langle n\rangle/n_c)^\beta}, \tag{3.8}$$

where the filling factor F is the ratio of the electric field stored in the TLS material to the total electric field, and δ^o_{TLS} is the intrinsic TLS loss. The TLS saturation rate with photon power, β, is ≤ 0.5 [110–113]. By fitting Q_i in Fig. 3.9(a)-top to Eq. 3.8, we extracted $F\delta^o_{\text{TLS}}$ as shown in Fig. 3.9(a)-bottom.

We found that the most crucial step, in those investigated, was the native oxide removal of the silicon substrate using an HF dip. TLS losses of resonators increased five-fold from 0.87×10^{-6} to 4.4×10^{-6} when that step was skipped. This comes as no surprise as the metal-substrate interface has the highest energy participation ratio of all interfaces [114].[5] TEM imaging of the silicon-aluminum interface confirms the presence of a 1.5-nm-thick interfacial oxide layer in the No-HF sample (No. 2), which is not present in the standard process sample (No. 1). Using x-ray photoelectron spectroscopy (XPS), we found that this layer is composed of AlOx instead of the supposedly native SiO_x, which we attributed to the oxidation-reduction reaction between aluminum and SiO_x [115, 116].

Another step that showed a non-negligible impact on TLS loss is the preheating of the silicon substrate before the wiring layer deposition, aimed at desorbing moisture from the surface. Skipping the preheating step in sample No. 4, resulted in a 61% increase in $F\delta^o_{\text{TLS}}$. We found no noticeable increase in oxygen concentration in the time-of-flight secondary ion mass spectrometry (ToF-SIMS), which implies that either the concentration of oxygen is below the observable limit, or desorption of oxidizing species is not the reason for the increased performance of the standard devices.

Other treatment steps had little to no impact on TLS loss. For example, $F\delta^o_{\text{TLS}}$ increased by 15% when the standard cleaning (SC1) step was skipped (sample No. 3). Similarly, sample No. 5 with no *in situ* oxidation showed

[5] We also show this in the participation ratio simulation section of **Paper C**.

a small increase in $F\delta^o_{\text{TLS}}$ of 10%, although these samples showed voids in the aluminum wiring layer. A potential cause for these voids is the lattice mismatch between aluminum and silicon, generating stress on the aluminum film and leading to its defect formation. Oxidizing this film *in situ* may protect the aluminum film from forming these voids by keeping the grains together. However, this is inconclusive. In sample No. 6, where we pump for 4 hours instead of overnight pumping before depositing the aluminum film, we see no effect on the resonators' quality factors. This is an important conclusion as overnight pumping is costly.

3.4.2 Thicker aluminum films for lower loss

As mentioned earlier, our standard aluminum film of the wiring layer (ground plane) has a thickness of 150 nm, where the average qubit T_1 was around 70-80 μs. In **Paper C**, we demonstrated that increasing the film thickness to 300 nm or 500 nm increased the average qubit quality factors by twofold. The best qubit T_1 was observed in the 500-nm-thick film with an average of 270 μs ($Q = 5.1\times10^6$) and a best single value of 501 μs [Fig. 3.10(a)]. However, on average, increasing the film thickness beyond 300 nm did not show significant improvement. Similarly resonators fabricated with the thicker films showed a reduction in TLS loss, where $F\delta^o_{\text{TLS}}$ dropped from 1×10^{-6} on average for 150-nm films to 8×10^{-7} and 5×10^{-7} in the 300-nm and 500-nm films, respectively. However, resonators made with the 500-nm-thick films showed a frequency-dependent increase in non-TLS loss, δ_{other}, which was not observed in qubits. Therefore, we adopted the 300-nm-thick films as a new standard thickness for our fabrication process.

At first, the reason for the observed reduction in TLS loss was not readily clear. However, using TEM and ToF-SIMS analysis, we found that such improvement is attributed to the larger grain sizes of thicker aluminum films. ToF-SIMS depth profile analysis of the different film thicknesses showed that the oxygen concentration at the silicon-aluminum interface was lower for thicker films as indicated with the black arrows in Fig 3.10(b). That, alone, explains the reduction in TLS loss, however, the reason for such a drop in oxygen concentration at the interface was still unclear. Therefore, we conducted a TEM study of the films' cross sections, where we found that thicker films are formed by larger grains than those in thinner films [Fig. 3.10(c)]. Larger grains mean fewer grain boundaries; thus, less diffusion of oxygen from

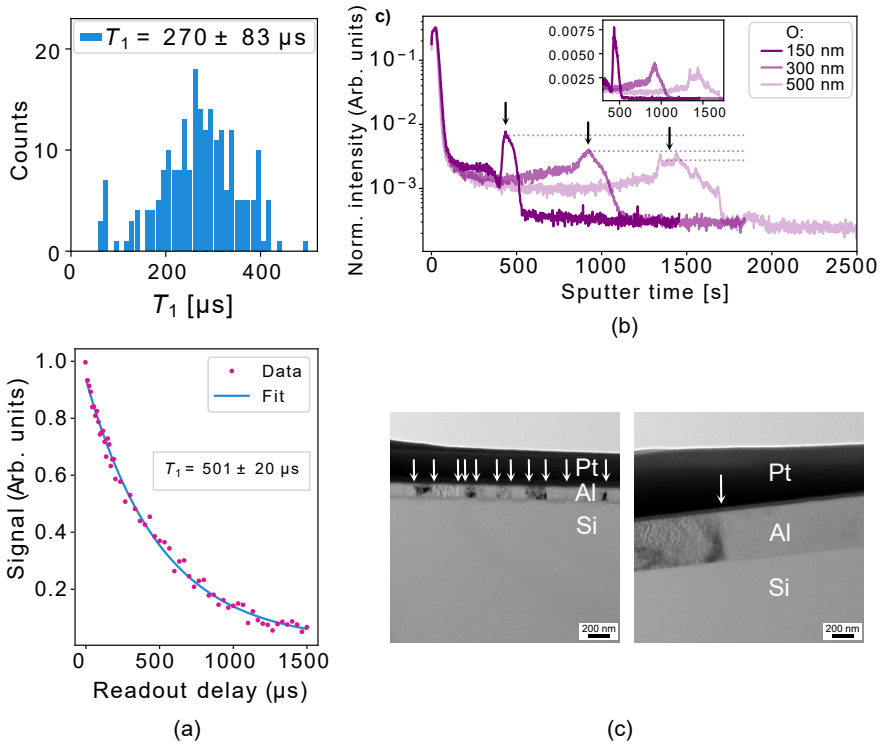


Figure 3.10: Reduction of TLS loss with increased aluminum film thickness [117]. (a)-top T_1 histogram of the best-performing qubit with 500 nm film thickness. (a)-bottom best T_1 value of that qubit. (b) Tof-SIMS depth profile of three different film thicknesses showing the concentration intensity of oxygen along the etched depth, where thicker films show lower concentration marked by the black arrows. (c) TEM images of cross sections of two metal films with 150 nm (left) and 500 nm (right), emphasizing the larger grain size of the thicker film. The platinum (Pt) layer is added during sample preparation for imaging.

the ambient to the silicon-aluminum interface through these boundaries where oxygen diffusivity is higher than that through pure aluminum [118].

CHAPTER 4

Crosstalk and frequency collisions in a QPU

The focus of this thesis so far has been on enhancing the quality of single qubits, with Chapter 2 addressing the reproducibility of qubit frequency and Chapter 3 addressing qubit coherence. Moving forward, the emphasis shifts to scaling QPUs to accommodate several qubits. In this chapter, I lay the groundwork for Chapters 5 and 6 where I discuss the evolution of QPUs from a planar to a 3D (or 2.5D) architecture. Section 4.1 introduces our first QPU that extends beyond two qubits, featuring the airbridge circuit element. In Section 4.2, we explore the flip-chip architecture, which is essential for scaling QPUs. Finally, Section 4.3 details the layout of our first-generation 25-qubit QPU within a flip-chip architecture, with an emphasis on crosstalk. This sets the stage for the next two chapters, where I discuss the second generation of this QPU.

4.1 5-qubit planar QPU

Quantum-information processing relies on the ability to control and couple qubits, enabling the implementation of quantum gates that perform the target computations. Depending on the QPU architecture, this coupling can be

achieved through a capacitive or inductive element, or a resonator-mediated interaction [119]. We adopted the parametric-gate architecture [120–123], where coupling between two transmon qubits is mediated by a transmon coupler, that is tunable via its SQUID loop. By parametrically driving this coupler, we implement either the iSWAP or the CZ gate between the two transmons. In [124], we demonstrated the use of the CZ gate by implementing a quantum approximate optimization algorithm on the exact cover problem. Our first QPU with more than two qubits was a 5-qubit QPU presented in **Paper E**. There, we implemented a three-qubit gate by simultaneously applying two-qubit operations, realizing a three-body interaction [125]. The implementation of such a gate is not a topic of this thesis. Here, I discuss one of the challenges encountered with scaling to 5 coupled qubits, which is signal routing and shielding.

Figure 4.1(a) shows a portion of the QPU, focusing on three of the qubits: q_1, q_2 and q_3, where couplers $cp_{(1,2)}$ and $cp_{(1,3)}$ couple qubits q_1 to q_2 and q_1 to q_3, respectively. The corresponding circuit schematic is shown in Figure 4.1(b). Each qubit is capacitively coupled to a control (or charge) line xy_i, where i is the qubit index. Each coupler is inductively coupled to a control (or flux) line $z_{i,j}$, where i, j is the coupler's index. Biasing the flux line with current induces a magnetic field threading the SQUID loop of the coupler, enabling the modulation of its frequency, necessary to implement a two-qubit gate.

With an increasing number of control lines (9 in total), in addition to the 5 readout resonators and the feedline, it becomes essential to provide sufficient signal shielding and suppress the excitation of parasitic slotline modes in these discontinuous transmission lines [126]. This can be achieved via crossover structures, also known as airbridges. For this reason, we developed a fabrication process for airbridges, following the process in [127]. Figure 4.1(c) shows a micrograph of airbridges fabricated on the 5-qubit QPU, where some airbridges were used for signal routing, crossing the central conductor of a transmission line crossing over an intersecting transmission line. Figure 4.1(d) shows an SEM image of several airbridges fabricated on a different wafer. These airbridges can be extended to provide nearly complete shielding for transmission lines as will be shown in Section 4.3. This proved to be especially useful for suppressing DC-flux crosstalk when flux lines are closer to each other than a certain distance, as will be discussed in Section 4.3.2.

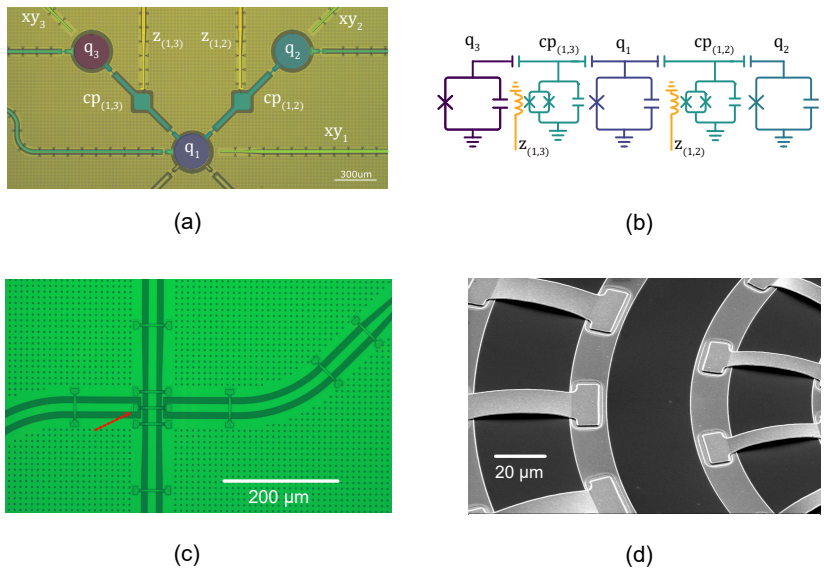


Figure 4.1: 5-qubit QPU in a planar architecture. (a) A portion of the QPU showing three of the five qubits, with each qubit q_i coupled to a dedicated charge line xy_i. Couplers $cp_{i,j}$ mediate coupling between qubits i and j. Each coupler is coupled to a flux line z_i. (b) The circuit schematic of the device shown in (a). (c) Micrograph of airbridges used for grounding both sides of the transmission lines. The airbridge indicated by the red arrow is used for signal routing. (d) SEM image of airbridges fabricated on a different wafer.

4.2 Flip-chip architecture

When we scale to tens of qubits on a QPU, planar architectures, such as the one presented in Section 4.1 become cumbersome due to the high density of routing lines required for driving quantum gates and reading out the qubit state. It is possible to mitigate this issue by fabricating airbridges to provide a crossover for signals, routing them over other routing lines. However, this solution is infeasible with an increasing number of qubits, where managing signal routing while mitigating crosstalk becomes a dilemma. Therefore, the flip-chip architecture, widely used in microprocessors and electronics packaging, was proposed for superconducting quantum processors [128–131]. In the flip-chip architecture we adopted, a qubit chip (Q-chip) is bonded onto a control chip (C-chip) [Figs. 4.2(a) and 4.2(c)]. The Q-chip hosts qubits and couplers, while the C-chip hosts readout resonators and signal routing lines (or transmission lines). This gives significantly larger space for routing lines, maximizing the spacing between them, which minimizes crosstalk.

In **Paper F**, we presented our flip-chip architecture, achieving average coherence times exceeding 90 μs, with single- and two-qubit gate fidelities above 99.9% and 98.6%, respectively. This performance was comparable to our planar devices, demonstrating that our flip-chip processing does not degrade device performance — a significant milestone toward scalability. A photograph of the flip-chip device bonded to a printed circuit board (PCB) that is mounted into a microwave package is shown in Fig. 4.2(c), alongside an illustrated 3D model in Fig. 4.2(d), where the Q-chip is rendered transparent. Figure 4.2(e) shows a schematic of the device design, featuring two fixed-frequency transmon qubits (q_1 and q_2) with a flux-tunable coupler on the Q-chip. xy_1 and xy_2 lines are used to drive the qubits state transitions. They are open-ended and capacitively coupled to the qubit as depicted in the leftmost inset of Fig. 4.2(e). The coupler's z line forms a short-ended loop concentric with the coupler's SQUID as shown in the middle inset of Fig. 4.2(e). The quarter-wavelength readout resonators (readout 1 and readout 2) are each capacitively coupled to their respective qubit as shown in the rightmost inset of Fig. 4.2(e), and are also coupled to a feedthrough transmission line. xy_1, xy_2, and z lines, along with the readout resonators and the feedthrough transmission line are located on the C-chip.

Fabrication of the C-chip and the Q-chip independently follows from the fabrication process laid out in Section 2.2. The two chips are connected together

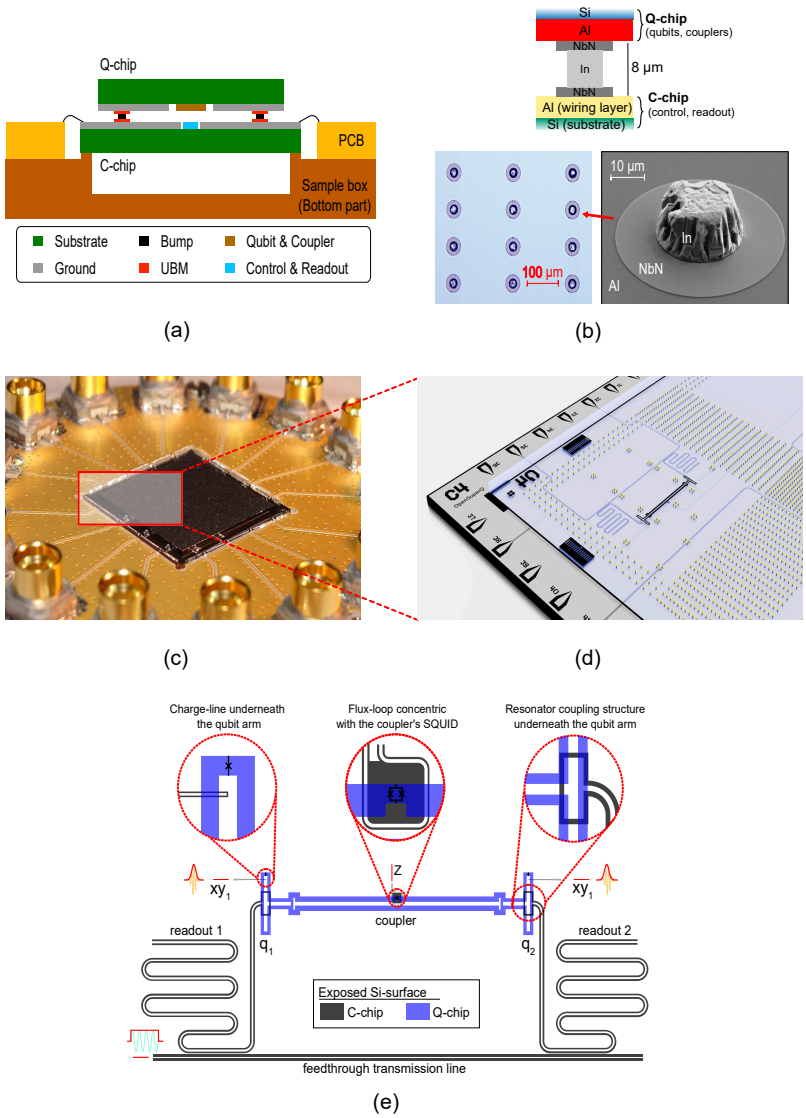


Figure 4.2: 2-qubit QPU in a flip-chip architecture [132]. (a) Simplified illustration of the flip-chip module. A Q-chip is bonded onto a C-chip via indium (In) bumps, with an Under-Bump Metallization (UBM) layer between each bump and the Al ground plane of both chips. The Q-chip includes qubits and couplers, while the C-chip includes control lines and readout resonators. (caption continued on next page)

Figure 4.2: The control-line traces and ground plane of the C-chip are wirebonded to a PCB, which is glued together with the module to a microwave package (sample box). (b)-top Schematic of an In bump structure, where In is separated from the Al ground plane of both chips by an NbN UBM. (b)-bottom-left Micrograph of an array of In bumps. (b)-bottom-right SEM image of an In bump. (c) Photograph of the QPU. (d) 3D illustration of a portion of the flip-chip module with the Q-chip rendered transparent.

by bump-bonding compressible pillars of indium (In) as shown in Fig. 4.2. To prepare the two chips for bump-bonding, after the ground plane deposition on both chips, a 50-nm thin under-bump metallization (UBM) layer of niobium nitride (NbN) is sputtered onto patterned resist, which is then lifted off. UBMs have circular shapes that are 50-μm wide, and act as a diffusion barrier between Al and In, preventing the formation of an Al-In intermetallic state [133]. The planar fabrication of both chips proceeds according to the standard process.

To form the bumps, In pillars are evaporated on both chips on patterned resist, which is then lifted off.[1] The deposited In is 8-μm-thick, with each pillar forming a 25-μm-wide cylinder. The two chips are flip-chip bonded by compression at room temperature, with a target chip-to-chip spacing of 8 μm after a 50% compression factor of the In bumps. A superconducting electrical contact is created between the ground planes of both chips since both In and NbN are superconducting.

This architecture is used for all our flip-chip QPUs, with different variations in qubits, couplers (and their control lines), and readout resonators to meet target circuit parameters. Signal routing schemes are also adapted based on the number of qubits and signal lines in the QPU. We will see this later in Section 4.3 and Chapter 5.

4.3 25-qubit flip-chip QPU

After demonstrating high-performance qubits and quantum gates on the flip-chip architecture, our next natural step was to increase the number of qubits.

[1] Fabrication of In pillars and flip-chip bonding is done at VTT Technical Research Center of Finland.

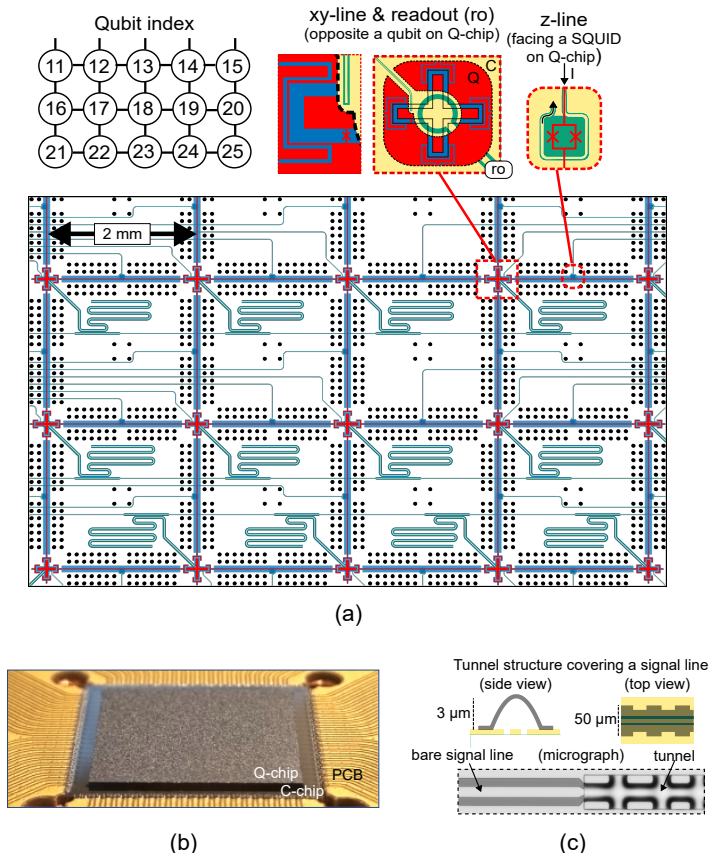


Figure 4.3: 25-qubit QPU architecture [134]. (a) Overlapped layout of the Q-chip (qubits and couplers) and C-chip (control lines, readout resonators, and feedthrough transmission lines) of a portion of the QPU including qubits 11 to 25 ($q_{11} - q_{25}$). Qubit indices are shown on the top left. The left inset shows the coupling points of a qubit to its charge xy line and readout (ro) resonator. The right inset shows the coupling point of a coupler's SQUID loop to its flux z line. The red (blue) elements represent Al (Si) on the Q-chip, while the yellow (green) elements represent Al (Si) on the C-chip. (b) Photograph of the QPU. (c) Tunnel structures covering transmission lines on one version of the QPU.

This will be the focus of this section and the next two chapters, where I discuss design and characterization results of our 25-qubit QPU. Figure 4.3(a) shows the layout of a portion of the device (qubits 11 to 25, demonstrated as q_{11} to q_{25}), where the insets show the coupling points to the qubit and the coupler. Qubits are arranged in five rows with five qubits each, forming a 5×5 grid, with a 2-mm pitch between qubits in each row (or column). Each qubit is capacitively coupled to an xy line, and also coupled to two, three or four neighboring qubits, depending on its position. Coupling between qubits is mediated by a tunable coupler as explained in Section 4.2. Each qubit is also capacitively coupled to a readout resonator, which is coupled to a feedthrough transmission line. Each feedthrough transmission line is coupled to five readout resonators of one qubit row.

The qubit capacitor design was modified from that presented in Section 4.2 to an Xmon-shaped qubit to facilitate coupling to the neighboring qubits [84]. Signal routing was designed to be as identical as possible for the inner three rows, i.e. q_6-q_{10}, q_{11}-q_{15} and q_{16}-q_{20}. Control lines are always routed from either side of the corridor above the respective qubits row, serving qubits and couplers of that row, along with the vertical couplers above.

Readout resonators are placed in the corridor below the qubits. The first and last rows have different routing configurations to adapt to the distribution of the signal launch pads at the square perimeter of the C-chip. Each control line is routed all the way from its qubit (or coupler) to the edge of the C-chip, where it is terminated by a launch pad that is wirebonded to a dedicated trace on the PCB hosting the QPU [Fig. 4.3(b)]. Each trace is then electrically connected to a dedicated microwave line to be routed to the control electronics at room temperature. The ground plane between each two neighboring launch pads is also wirebonded to the ground of the PCB.

In the following two subsections, I will present the relevant results on crosstalk characterization. The interested reader can find more details in **Paper G**. Details on the circuit parameters and the motivation behind them will be presented in Chapter 5, where I focus on the second generation of the 25-qubit QPU. I also note that our crosstalk study was conducted on two versions of the QPU: one with airtunnels covering the control lines as shown in Fig. 4.3(c), and one without them.

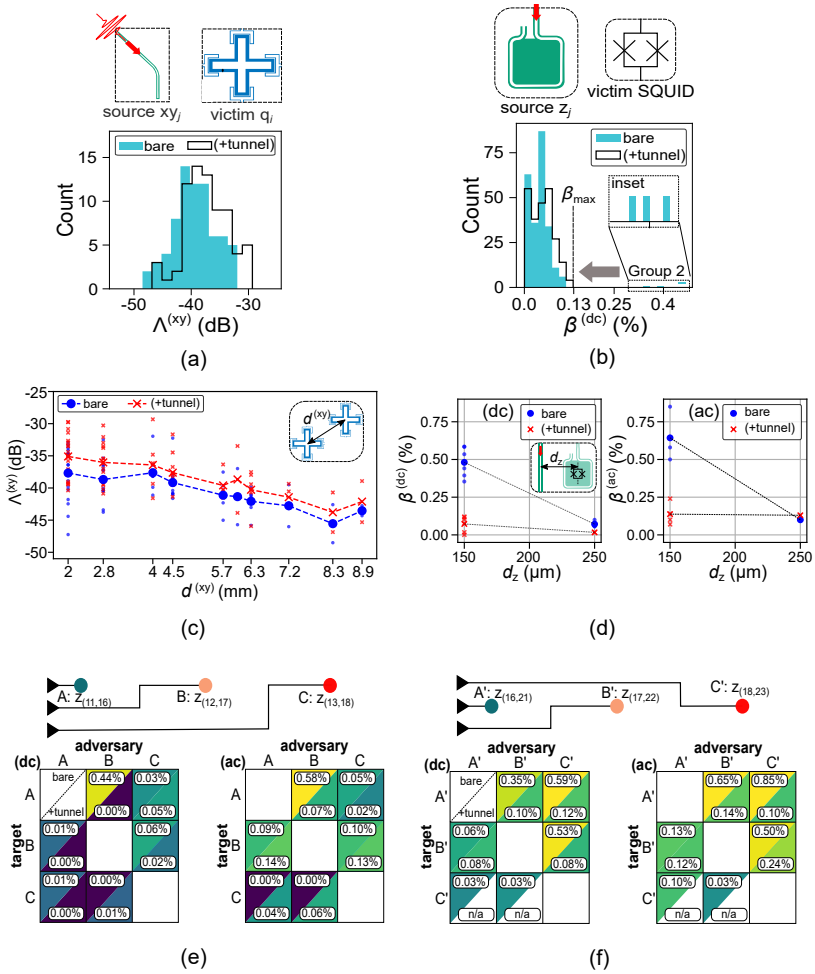


Figure 4.4: Crosstalk in the 25-qubit QPU [134]. (a) Histogram of XY crosstalk $\Lambda^{(xy)}$ between 72 source-victim pairs in the QPU with and without tunnels. (b) Histogram of DC flux crosstalk $\beta^{(dc)}$ between 240 source-victim pairs. The inset shows $\beta^{(dc)}$ for outlier (high-flux-crosstalk) couplers with configurations shown in (e) and (f). (c) Scaling of $\Lambda^{(xy)}$ with the distance $d^{(xy)}$ between a source and a victim with an average reduction of 1 dB/mm. (d) $\beta^{(dc)}$ and $\beta^{(ac)}$ as a function of the distance d_z between the loop of a victim coupler's z line and the nearest portion of a source z line. (e) and (f) Two configurations of coupler sets: (A, B, and C) and (A', B', and C'), respectively, with observed high flux crosstalk, along with their DC- and AC-flux crosstalk matrices.

4.3.1 XY crosstalk

To control the quantum state vector of any qubit, we transmit a signal through its xy line at the qubit's resonant frequency. However, this signal can unintentionally reach a neighboring qubit, causing an undesired rotation of its state vector. The qubit being intentionally driven is referred to as the *source* qubit (q_j) with a *source* charge line (xy_j), while the qubit receiving the unintended drive is referred to as the *victim* qubit (q_i). To quantify this unintended drive in practice, we compare the Rabi frequency of q_i, $\Omega_{i,j}$, to that of q_j, $\Omega_{j,j}$, when xy_j is driven with the same signal amplitude. In reality, q_i and q_j may be detuned in frequency, and driving xy_j at the resonant frequency of q_j has a weakened effect on q_i state vector due to the off-resonant drive. However, to characterize XY crosstalk, we drive xy_j with the resonant frequency of q_j when measuring $\Omega_{j,j}$, and the resonant frequency of q_i when measuring $\Omega_{i,j}$. This approach captures the naturally low crosstalk level, as we are interested in quantifying this crosstalk regardless of the qubit frequency arrangement in the QPU. The ratio of the two Rabi frequencies in dB is the XY crosstalk $\Lambda_{i,j}^{(xy)}$ and is defined as:

$$\Lambda_{i,j} = 10 \times \log_{10} \left(\frac{\Omega_{i,j}}{\Omega_{j,j}} \right)^2, \tag{4.1}$$

where a lower magnitude of $\Lambda_{i,j}$ indicates lower XY crosstalk, meaning that xy_j is more effective at selectively driving q_j.

Figure 4.4(a)-bottom shows a histogram of $\Lambda^{(xy)}$ for 72 different source-victim pairs, with an average and standard deviation of -39.4 ± 3.7 dB for the QPU version without tunnels, and -37.4 ± 3.9 dB for the version with tunnels. Those 72 pairs correspond to qubits that were functional on both versions of the QPU. $\Lambda^{(xy)}$ varied between -56 dB and -27 dB, comparable with the lowest reported crosstalk values in [135, 136]. This shows that adding airtunnels to control lines does not decrease the average crosstalk.

Despite the densely routed signal lines on the C-chip, the observed crosstalk values could not be explained by direct capacitive coupling between the signal lines. Our simulations indicated that direct capacitive coupling between the source xy line and the victim q alone accounts for XY crosstalk values ranging from -49 dB to -150 dB, much lower than the measured values. Additionally incorporating microwave crosstalk between signal lines on the

PCB, wirebonds and launch pads sets an upper bound of −47 dB for neighboring xy lines. The discrepancy between this upper bound and the highest measured XY crosstalk of −27 dB (or the average −36 dB)[2], suggests that additional factors are at play, requiring further investigation. A key finding of our crosstalk characterization is the relationship between the crosstalk parameter $\Lambda^{(xy)}$ and the distance $d^{(xy)}$ between the victim qubit and the source xy line shown in Fig. 4.4(c). For an architecture to be scalable, $\Lambda^{(xy)}$ must decrease with increasing $d^{(xy)}$, thereby reducing the cumulative crosstalk error from several source xy lines. This allows us to focus only on crosstalk effects from neighboring qubits, while neglecting contributions from more distant ones, simplifying the frequency allocation scheme needed for low crosstalk errors (see Section 4.4).

Fig. 4.4(c) suggests such a favorable behavior, where the average $\Lambda^{(xy)}$ drops at a rate of 1 dB/mm. However, our simulation showed that for a realistic frequency allocation scheme on a QPU with simultaneous driving of neighboring qubits, a drop rate of 1.5 dB/mm (or 2 dB/mm) in $\Lambda^{(xy)}$ is required to maintain single-qubit gate errors below 0.1% in a QPU with more than 100 (or 1000) qubits, if the maximum crosstalk from a nearest neighbor is below −50 dB (see Appendix L of **Paper G**). Additionally, for the worst case $\Lambda^{(xy)}$ of −27 dB, a detuning of more than 28 MHz in frequency between a victim and a source qubit is necessary to achieve single-qubit gate fidelity above 99.9%, accounting for the effect of only one source xy line. Therefore, further efforts are necessary to suppress XY crosstalk below the current measured level. Moreover, optimizing the frequency allocation scheme is needed to maximize detuning between neighboring quantum gates. This is the topic of Section 4.4.

4.3.2 Z crosstalk

Analogous to XY crosstalk, Z crosstalk occurs when the magnetic flux threading the SQUID loop of a victim coupler cp_i changes with $\Delta\Phi_i$ when a neighboring coupler cp_j is tuned via its source z_j line by $\Delta\Phi_j$. In such a case, Z crosstalk $\beta_{i,j}$ is defined as:

[2]This accounts for all measured source-victim pairs, and not just the 72 pairs, functional on both QPU versions. See Appendix F of **Paper G** for the full data sets.

$$\beta_{i,j} = \left| \frac{\Delta\Phi_i}{\Delta\Phi_j} \right|. \tag{4.2}$$

A tunable coupler cp_i has both a DC- and an AC-flux component, such that $\Phi_i = \Phi_i^{(dc)} + \Phi_i^{(ac)} \cos(\omega_i^{(ac)} t + \theta_i)$. Thus, we characterized DC-flux crosstalk $\beta_{i,j}^{(dc)}$ and AC-flux crosstalk $\beta_{i,j}^{(ac)}$ separately.

Figure 4.4(b) shows a histogram of $\beta_{i,j}^{(dc)}$, where the average DC-flux crosstalk is 0.05%. Similarly to the AC crosstalk, these values are comparable with the lowest reported values in [137, 138]. However, the QPU without tunnels covering the z lines included a subset of coupler pairs where $\beta_{i,j}^{(dc)}$ increased significantly to values between 0.3 and 0.6%. These outlier pairs corresponded to couplers whose z lines were arranged in a specific way.

Figure 4.4(e) and 4.4(f) show the two coupler set configurations to which the outlier $\beta_{i,j}^{(dc)}$ values belong, along with their DC-flux and AC-flux crosstalk matrices. For the configuration shown in Fig. 4.4(e), $\beta_{A,B}^{(dc)}$ corresponding to victim coupler A when source z_B is biased, has a high value of 0.44%. On the other hand, for the configuration shown in Fig. 4.4(f), $\beta_{A',B'}^{(dc)}$, $\beta_{A',C'}^{(dc)}$ and $\beta_{B',C'}^{(dc)}$ also have high values of 0.35%, 0.59% and 0.53%, respectively. Interestingly, the two DC-flux crosstalk matrices show an asymmetric behavior; that is, the high crosstalk values are not observed when the roles of the source and the victim are reversed. For example, $\beta_{B,A}^{(dc)}$ is 44 times lower than $\beta_{A,B}^{(dc)}$. AC-flux crosstalk for both configurations showed a similar behavior.

Observation of such an asymmetry indicates that the high flux crosstalk arises from the proximity of the current loop of a victim z line and the nearest section of a source z line. We define this distance as d_z. Let us take the configuration in Fig 4.4(e) as an example. When d_z increases from 150 μm, as is the case between victim cp_A and source z_B, to 250 μm, as is the case between victim cp_B and source z_C, $\beta^{(dc)}$ and $\beta^{(ac)}$ drop by almost a factor of 10. Figure 4.4(d) shows $\beta^{(dc)}$ (left) and $\beta^{(ac)}$ (right) for the two different d_z values of 150 μm and 250 μm for identical configurations. Notably, the high flux crosstalk was significantly suppressed for victim-source pairs with $d_z = 150$ μm in the QPU where z lines were covered with tunnels, whereas pairs with $d_z = 250$ μm were relatively unaffected by the incorporation of tunnels. This indicates that it is necessary for these coupler configurations to either cover z lines with tunnels or ensure that d_z is greater than 250 μm. In the second generation of the QPU presented in Chapters 5 and 6, we adopted

the latter approach.

4.4 Frequency allocation on a QPU

When we allocate qubit frequencies in a QPU, it is essential to keep the crosstalk between single and two-qubit quantum gates in mind. Neighboring gates should be separated in frequency as much as possible, adhering to a specific collision threshold set by the crosstalk level and target gate fidelities (see Appendix D of **Paper B**). Even after careful allocation of frequencies by design, uncertainties in qubit frequencies arising from the fabrication process can eventually render quantum gates closer to each other in frequency than intended. This highlights the importance of maximizing frequency separation between neighboring gates to account for fabrication-induced variations and to maintain the desired gate performance.

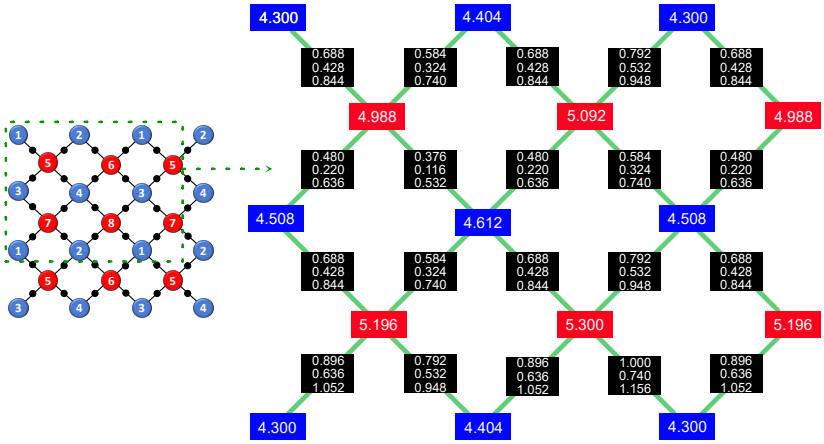
To minimize the effect of crosstalk, in **Paper B**, we proposed a frequency allocation scheme for the parametric-gate architecture shown in Fig.4.6(a)-left, where qubit frequencies were divided into two groups: a lower-frequency group A [1 to 4] (blue) and a higher-frequency group B [5 to 8] (red). Two qubit gates are implemented exclusively on couplers (black) connecting qubits from group A with neighboring qubits from group B. Collectively, these eight qubits form a unit cell of the QPU lattice. To determine what these frequencies are, we consider the qubit gates implemented in our architecture which are the iSWAP, CZ(02), and CZ(20). The coupler modulation frequencies corresponding to these gates between two neighboring qubits are given by the set of equations:

$$f^{jk}_{\text{iSWAP}} = |f^{j}_{01} - f^{k}_{01}|, \tag{4.3a}$$

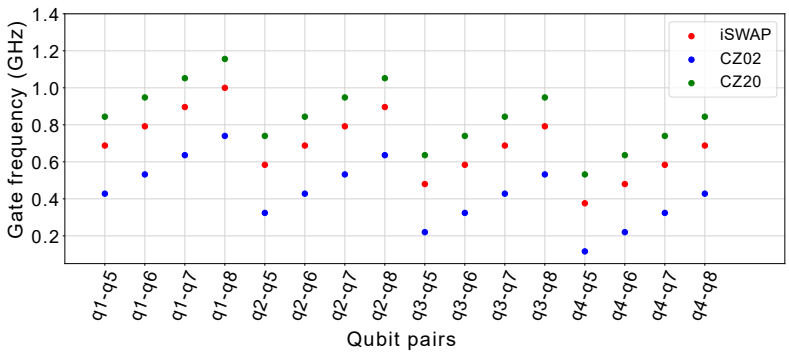
$$f^{jk}_{\text{CZ(20)}} = |f^{k}_{01} - f^{j}_{01} - \alpha^{j}|, \tag{4.3b}$$

$$f^{jk}_{\text{CZ(02)}} = |f^{j}_{01} - f^{k}_{01} - \alpha^{k}|. \tag{4.3c}$$

These equations indicate that to avoid collision between neighboring two-qubit gates, detuning $\Delta^{jk} = |f^{j}_{01} - f^{k}_{01}|$ between neighboring qubits j and k and their anharomnicities α^{j} and α^{k}, respectively, must be taken into account. Δ^{jk} sets f^{jk}_{iSWAP} for each pair, where $f^{jk}_{\text{CZ(20)}}$ and $f^{jk}_{\text{CZ(02)}}$ are offset by α^{j} and α^{k}, respectively.



(a)



(b)

Figure 4.5: Two-group frequency allocation in a Square lattice. (a)-left Schematic of a square lattice of a QPU with two frequency groups: group A in blue and group B in red. Two-qubit gates are implemented between qubits from each group on the tunable coupler (black). (a)-right Single-qubit gate frequencies (inside the blue and red boxes) and two-qubit gate frequencies (inside the black boxes) for a portion of the square lattice (green dashed square). Two-qubit gate frequencies are for the iSWAP, CZ02 and CZ20, ranked from top to bottom, respectively. All frequencies are expressed in GHz. (b) Two-qubit gate frequencies for all possible qubit pairs.

Table 4.1: Allocation of qubit frequency f_{01} and anharmonicity α in a square lattice of QPU with eight-qubit unit cells.

Qubit	1	2	3	4	5	6	7	8
f_{01}(GHz)	4.3	4.404	4.508	4.612	4.988	5.092	5.196	5.3
α (MHz)	−156	−156	−156	−156	−260	−260	−260	−260

Table 4.1 shows f_{01} and α of the eight qubits. Figure 4.5-right shows the qubit frequency allocation (red and blue squares) on a portion of the lattice, and the corresponding f^{jk}_{iSWAP}, $f^{jk}_{\text{CZ(02)}}$ and $f^{jk}_{\text{CZ(20)}}$ in the black squares from top to bottom, respectively. Such a frequency allocation scheme aims to ensure that the neighboring single and two-qubit gates are as evenly spaced as possible within the available bandwidth, maximizing the separation. For single qubit gates, this requires a fixed f_{01} pitch, Δ_Q, between qubits in each group. For two-qubit gates, additionally, α^A and α^B (i.e. anharmonicities of qubits in group A and B, respectively) are fixed for all four qubits within each group. We set

$$\alpha^A = 1.5\Delta_Q, \tag{4.4a}$$

$$\alpha^B = 2.5\Delta_Q. \tag{4.4b}$$

This guarantees that any two-qubit gate frequency is equidistant from the two closest neighboring two-qubit gates by 0.5 Δ_Q. For example, let us consider the iSWAP gate between q4 and q5, $f^{45}_{\text{iSWAP}} = 0.376$ GHz. The two closest neighboring two-qubit gate frequencies are the CZ(02) between q4 and q7, $f^{47}_{\text{CZ(02)}} = 0.324$ GHz, and the CZ(02) between q4 and q8, $f^{48}_{\text{CZ(02)}} = 0.428$ GHz.[3]

Equations 4.4 are general conditions for our frequency allocation scheme. However, determining the values of f_{01} and Δ_Q depends on several factors, such as qubit coherence, detunings between the qubit and readout resonator, limitations of the control electronics and the target range of coupler modulation frequency. In our case, we define three additional conditions through which we derive those values. Firstly, the minimum E_J/E_C ratio is 60, which sets a specific higher limit for qubit T_2. Increasing α (or E_C) increases the bandwidth over which we can assign gate frequencies [Eq. 4.3]; hence, lower

[3]Here, qz refers to qubit z, where z is the qubit index [1, 2, 3, ..., 8].

chance of collision. However, we set a second condition where we cap the maximum qubit frequency at 5.3 GHz to maintain a relatively high T_1. Considering these two conditions, we choose $\alpha^B = 260$ MHz, which means that $\alpha^A = 156$ MHz and $\Delta_Q = 104$ MHz. Finally, the bandwidth over which f_{01} are assigned is set to 1 GHz. This also sets the limit on the maximum two-qubit gate frequency, which is $f^{18}_{\mathrm{CZ}(20)} = 1$ GHz $+\alpha^1$, where α^1 is α^A. This limitation is mainly set by the bandwidth of the control electronics. Therefore, the lowest f_{01} is 4.3 GHz. Then f_{01} of the other 6 qubits are populated accordingly.

We finally note that our strategy of frequency allocation is not necessarily optimal, and there could be other logical procedure through which we arrive at nearly equispaced frequency crowding. However, this strategy is simple and ensures the tileability of the QPU lattice.

4.5 Simulating frequency collisions on a QPU

The frequency allocation scheme presented in the previous section ensures no collisions between neighboring gates under ideal conditions. However, as previously mentioned, deviation in the assigned frequencies due to fabrication uncertainties can cause collisions. In this section, I present a Monte Carlo simulation to model quantum-gate collisions on a QPU based on the target gate fidelities F_{target}, AC crosstalk level X_{AC} and the standard deviation in qubit frequency from the fabrication process $\sigma_{f_{01}}$.

In this simulation, a square-lattice QPU is populated with qubits following the frequency allocation in Fig. 4.5(a) and Table. 4.1. However, f_{01} are assigned according to a normal distribution with a mean equal to the target f_{01} in Table 4.1 and a standard deviation $\sigma_{f_{01}}$. Based on these assignments, single and two-qubit gates are recalculated. A collision occurs when two neighboring quantum gates fall closer in frequency than a certain bound, which is determined by X_{AC} and F_{target}. Derivations of these bounds can be found in Appendix D of **Paper B**. Here, we define $g_X = X_{AC} g_{wcs}$, where g_X is the gate strength on one qubit if a neighboring qubit is driven, and g_{wcs} is the worst case scenario of g_X when X_{AC} is 100%.

Fig. 4.6(a) shows the average number of collisions as a function of the number of qubits in a QPU for F_{target}=99.9% and $X_{AC} = 2\%$. A QPU fabricated with large-junction qubits, where $\sigma_{f_{01}} = 40$ MHz, has almost three times lower

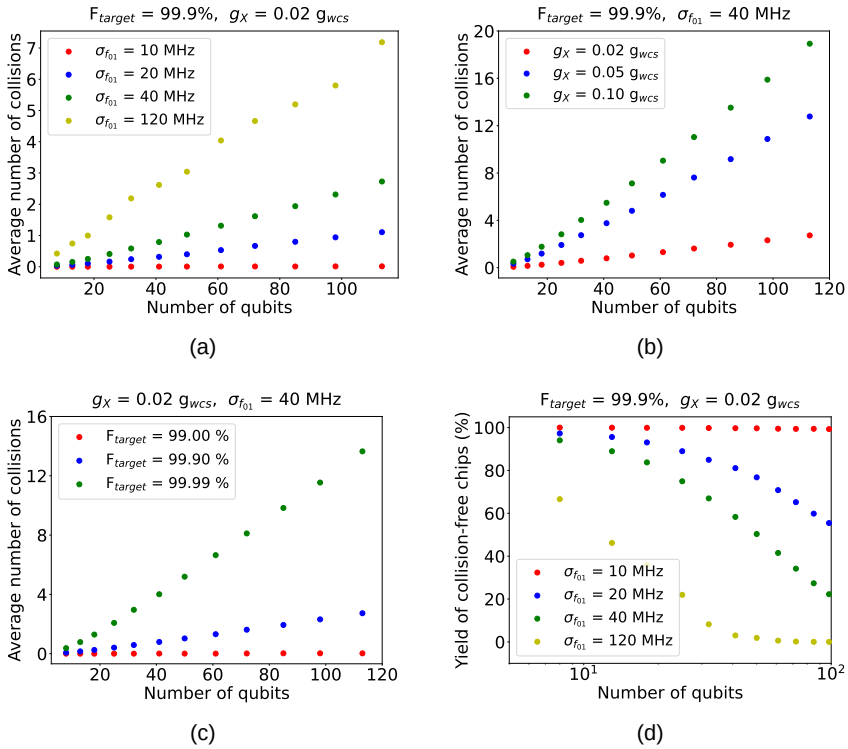


Figure 4.6: Frequency collisions in a QPU based on the parametric-gate architecture. Average number of collisions as a function of the number of qubits in a QPU for (a) different standard deviation in qubit frequency, σ_{f01}, for a 2% AC crosstalk and 99.9% target gate fidelity, (b) different AC crosstalk for σ_{f01}= 40 MHz and 99.9% target gate fidelity, and (c) different target gate fidelities for σ_{f01}= 40 MHz and 2% AC crosstalk. (d) Yield of fabricating collision-free QPU for different σ_{f01} when the target gate fidelity is 99.9%, and AC crosstalk is 2%.

number of gate collisions than a QPU made with the small-junction qubits, where $\sigma_{f_{01}} = 120$ MHz. Figures 4.6(b) and (c) show the scaling of the number of collisions for different X_{AC} and F_{target}, respectively when $\sigma_{f_{01}}$ is 40 MHz. These results highlight the importance of maintaining a low crosstalk level while investigating the feasibility of achieving high-fidelity quantum gates. Finally, Fig. 4.6(d) shows the yield of fabricating a collision-free QPU for different $\sigma_{f_{01}}$. While it is impossible to fabricate a collision-free 100-qubit QPU with a $\sigma_{f_{01}}$ of 120 MHz, there is a 20% chance of achieving this with a $\sigma_{f_{01}}$ of 40 MHz. This yield increases significantly if the tolerance increases from zero to one collision (not shown). More importantly, this simulation shows that in order to achieve a 100% yield of collision-free QPU, the standard deviation in qubit frequency must lie below 10 MHz, emphasizing the need for further optimization of junction parameters in the fabrication process.

We note that in this simulation, we assume that we are only interested in one of the CZ gates, allowing us to disregard collisions between the iSWAP and the other CZ gate. Additionally, We do not account for qubit decoherence, control pulse imperfections, or simultaneous driving of gates when we calculate the collision bounds necessary to achieve a specific fidelity. Nevertheless, the simulation offers valuable insight into the scaling of collisions for different target fidelities, given the available crosstalk suppression and fabrication process qualities. It is also important to clarify that the presence of a collision does not render the QPU unusable. A collision means that the target gate fidelity is not achievable for certain neighboring gates. In such case, one could either accept the lower fidelities for those gates while minimizing their use in a given algorithm. Alternatively, if a collision occurs between two neighboring two-qubit gates, one could explore the possibility of implementing one of the other two available gates without collision.

CHAPTER 5

QPU Design

In this chapter, I present the design parameters and fabrication aspects of the second-generation QPU, incorporating five key modifications. (1) Qubit frequencies and anharmonicities are assigned according to the frequency allocation scheme described in Section 4.4. (2) Larger-area junctions are used in qubit fabrication to achieve the qubit frequency reproducibility demonstrated in Section 2.4. (3) The Q-chip ground plane is fabricated with a 300-nm-thick aluminum film to achieve higher qubit coherence, as presented in Section 3.4. (4) Purcell decay to both the qubit drive line and the readout resonator is minimized to utilize the anticipated coherence improvement from the thicker ground plane. However, this is not guaranteed to be achieved in the flip-chip architecture, since the energy-participation ratio of the amorphous interfaces is different from that of planar architectures. (5) The proximity of coupler flux loops to neighboring coupler control lines is maintained at a minimum distance of 250 μm to reduce flux crosstalk, as discussed in Section 4.3.2.

5.1 Design overview

The design and simulation flow of the QPU followed from that presented in **Paper G**. IBM `QISKIT METAL` was used for generating layouts and rendering them to `ANSYS` electromagnetic simulation software [139, 140]. Additionally, `L-EDIT` layout editor was used in the final stages of design tape-out. The simulation model consisted of two 280-μm-thick silicon chips (C-chip and Q-chip), with their ground planes facing each other (Fig. 4.2). The ground planes were modeled as planar sheets with zero thickness. Readout resonator frequencies were simulated using `ANSYS` HFSS eigenmode solver, while capacitances were simulated using `ANSYS` Maxwell electrostatic solver. To achieve a target circuit parameter in simulation, a parameterized sweep of the component's geometry was implemented. For example, the gap between the Xmon metallic cross and the ground plane was swept over different values until the target qubit capacitance was achieved. More details on simulation parameters can be found in the supplementary material of **Paper F**.

In the remainder of this section, I present our adaptation of the frequency allocation scheme on the 5×5 QPU grid. Then, I elaborate on some of the design aspects and the motivation behind selecting them.

5.1.1 Qubit lattice

In Section 4.4, I presented the frequency allocation scheme we adopted in the second-generation 25-qubit QPU. There are different possibilities of mapping those frequencies to the 5×5 grid shown in Fig. 4.3. The two-frequency group assignment was designed to prevent collisions between nearest-neighbor qubits and couplers in the grid. However, we should further account for the proximity of their control lines, as parasitic coupling between them may also cause collisions.

As explained in Section 4.3, control lines of qubits and couplers in rows 2, 3, and 4 (along with the vertical coupler above each row) are routed through the corridor above [Fig. 4.3(a)]. Since some farther-than-nearest-neighbor quantum gates have identical frequencies in our frequency allocation strategy, control lines for such gates must not be routed through the same side of the same corridor. With that in mind, the frequency allocation in a 5×5 grid is mapped as shown in Fig. 5.1, where frequencies shown in blue (red) boxes correspond to the $0-1$ transitions of group A (group B) qubits. Assigning

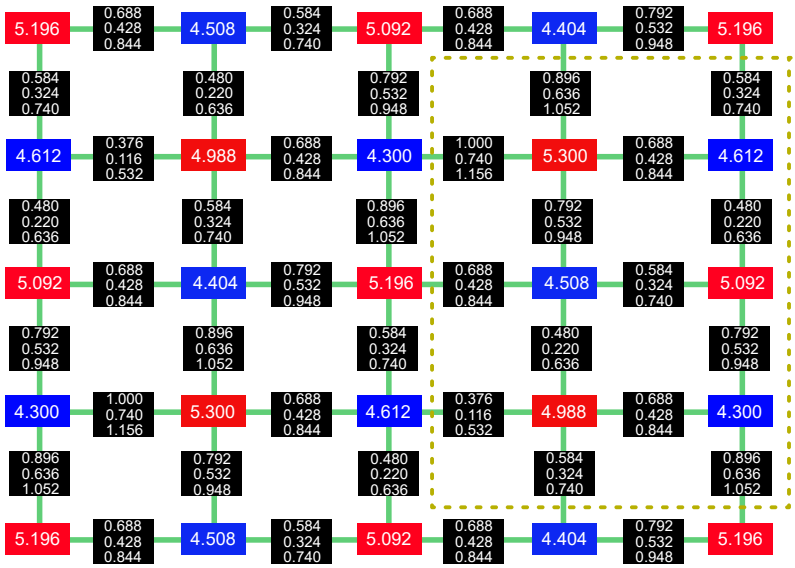


Figure 5.1: Frequency allocation over the second-generation QPU. A 5×5 square-lattice grid showing single-qubit gate frequencies for qubits of group A (blue) and group B (red), where two-qubit gate frequencies between neighboring qubits are shown in the black squares (iSWAP, CZ02, and CZ20 from top to bottom, respectively). Frequencies are reported in GHz.

anharmonicities of 156 and 260 MHz for group A and group B, respectively, we extract iSWAP and CZ02(20) two-qubit gate frequencies as shown in black boxes. Qubits and couplers in the gold dashed box have their control lines routed from the right side, while the rest of the qubits and couplers in the same three rows have their control lines routed from the left side. Control lines for the first and fifth rows are routed from the top and bottom, respectively, and are therefore less critical.

5.1.2 Qubit design

Each qubit in the QPU lattice has a target f_{01} and α value. To achieve this by design, we simulate the Cooper-pair-box Hamiltonian and solve for E_J and E_C that result in the target qubit energy levels [50]. As discussed in Chapter 2, E_J depends on the Josephson junction's R_N, while E_C depends on the qubit capacitance C_q. For an Xmon qubit, this capacitance is controlled by the size of the inner metallic part of the capacitor, and its gap to the ground plane. Figure 5.2(a) shows an Xmon design from group A, and its coupling points to different circuit components. Table 5.1 contains the simulated design parameters for each qubit. In the remainder of this section, I elaborate on some of the design parameters.

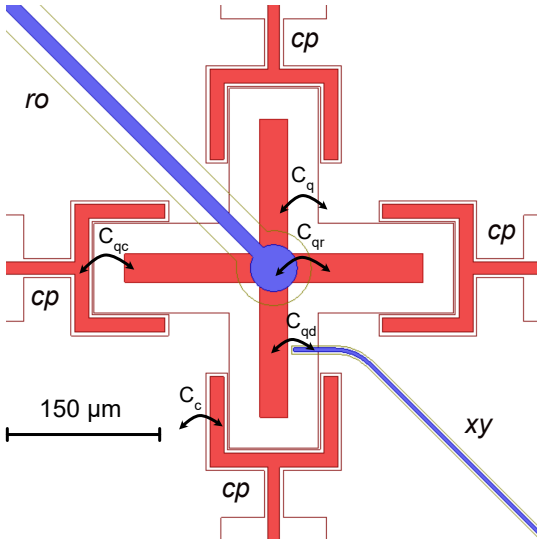
5.1.3 Qubit control and readout

The coupling strength of a qubit to its control line determines its Rabi frequency, Ω_q, when driven at resonance with a signal amplitude V_d, such that

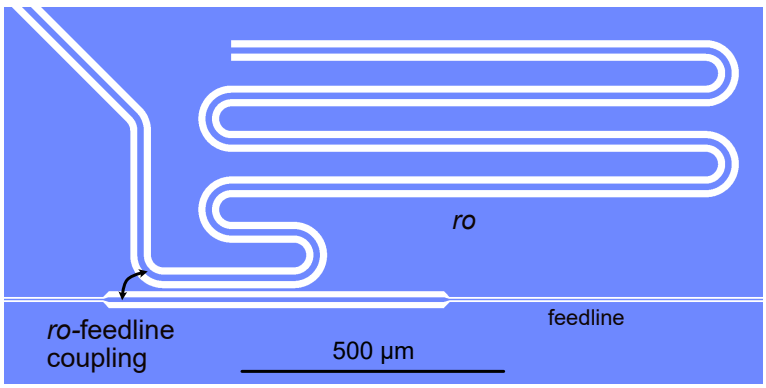
$$\Omega_q = \frac{e}{\hbar}\frac{C_{qd}}{C_\Sigma}\sqrt{\frac{E_C}{2E_J}}V_d, \tag{5.1}$$

where C_{qd} is the coupling capacitance of the qubit to its control line, and $C_\Sigma = C_q + C_{qd}$ [141]. On one hand, to implement a single-qubit gate in time $\tau_{sg} = 1/(2\Omega_q)$ within the available V_d range, a lower bound is set on C_{qd}. On the other hand, a qubit can lose energy to the control line by spontaneous emission, where the T_1 limit due to the Purcell decay to the drive line, T_1^{dP}, is given by:

$$T_1^{dP} \approx \frac{C_\Sigma}{(2\pi f_{01})^2 C_{qd}^2 Z_d}, \tag{5.2}$$



(a)



(b)

Figure 5.2: Qubit and resonator design. (a) The design of a qubit from group A and its capacitive coupling points to the readout resonator (*ro*), control (or drive) line *xy*, and coupler arm *cp*. Metal parts on the Q-chip and C-chip are shown in red and blue, respectively. The ground planes of both chips are rendered transparent for clarity. (b) Design of the readout resonator and its coupling to the feed-through transmission line (feedline).

where Z_d is the impedance of the load connected to the line (50 Ω). This sets an upper bound on C_{qd}, as a higher coupling capacitance degrades T_1.

To readout the qubit state, the qubit is capacitively coupled to a CPW resonator with frequency f_r and linewidth κ_r, where the qubit-resonator coupling strength is g_{qr}.[1]

$$g_{qr} = f_r \frac{C_{qr}}{C_q} \left(\frac{E_J}{2E_C} \right)^{1/4} \sqrt{\frac{\pi Z_r}{R_K}}, \tag{5.3}$$

where Z_r is the resonator's characteristic impedance, and $R_K = h/e^2 \sim 25.8$ kΩ is the resistance quantum [51]. The linewidth κ_r depends on the coupling quality factor of the readout resonator to the feedline, $Q_c = f_r/\kappa_r$. When a transmon qubit changes its state from $|0\rangle$ to $|1\rangle$, it induces a frequency shift of 2χ on the readout resonator [50], where

$$\chi = \frac{g_{qr}^2}{\Delta_{qr}} \frac{\alpha}{\alpha + \Delta_{qr}}. \tag{5.4}$$

Achieving fast, high-fidelity readout necessitates high g_{qr} and κ_r, along with a low qubit-resonator frequency detuning $\Delta_{qr} = f_{01} - f_r$. However, this conflicts with the goal of maintaining a high qubit T_1 limit due to the Purcell decay into the readout resonator, T_1^{rP}, which is given by:

$$T_1^{rP} \approx \frac{1}{2\pi\kappa_r} \frac{\Delta_{qr}^2}{g_{qr}^2}. \tag{5.5}$$

Since no Purcell filters are implemented in this generation of the QPU, a compromise is necessary regarding the acceptable Purcell decay limit.

Resonators in each row are assigned f_r values over an 800 MHz span with Δ_{qr} of 2250 MHz on average. This span is constrained by the 1 GHz bandwidth available for each readout module dedicated to a qubit row. Resonator frequencies are spaced as evenly as possible within this bandwidth to account for deviations in the spacing between the C-chip and Q-chip [142]. Figure 5.2(b) shows the design of a CPW readout resonator, coupled to a feedline, where the distance between the resonator and the feedline determines Q_c, and consequently, κ_r.

In Table 5.1, T_1^P represents the total Purcell-limited T_1, such that $1/T_1^P =$

[1] From here on, frequencies, coupling strengths, linewidths ... etc., are only defined in Hz.

$1/T_1^{rP} + 1/T_1^{dP}$. For a maximum qubit Q of $5{\times}10^6$ (Section 3.4), T_1 decreases by approximately 30% due to Purcell decay. For an average Q of $3{\times}10^6$, T_1 decreases by about 20%. Further increasing the Purcell decay limit would render that readout challenging. This emphasizes the need for designing Purcell filters to protect the qubit from decay [93–98].

5.1.4 Qubit-qubit coupling

In the parametric-gate architecture, coupling between two qubits (q_1 and q_2) is mediated by a tunable coupler, where the coupler flux Φ_c is modulated with a frequency corresponding to the target two-qubit gate (e.g. CZ or iSWAP gate) [120, 122, 141]. The two-qubit gate strength depends on the effective coupling between the two qubits g_{eff}, where

$$g_{\text{eff}} \approx g_{12} + \frac{g_1 g_2}{2}\left(\frac{1}{f_1 - f_c} + \frac{1}{f_2 - f_c}\right). \tag{5.6}$$

g_{12} is the direct coupling between q_1 and q_2, while g_1 (g_2) is the individual qubit-coupler coupling strength g_{qc} for q_1 (q_2). The frequencies f_1, f_2 and f_c are the f_{01} transitions of q_1, q_2 and the coupler, respectively [141]. The individual qubit-coupler coupling strength is given by:

$$g_{qc} = \frac{C_{qc}}{C_q C_c}\frac{e^2}{\hbar}\left(\frac{E_{J_q} E_{J_c}}{4E_{C_q} E_{C_c}}\right)^{1/4}, \tag{5.7}$$

where C_{qc} is the capacitance between the qubit and coupler, and C_c is the coupler's capacitance. E_{J_q} (E_{J_c}) and E_{C_q} (E_{C_c}) are the qubit's (coupler's) E_J and E_C, respectively.[2] In our design, the direct capacitance between any two qubits is about four orders of magnitude smaller than C_{qc}; hence, g_{12} is ignored.

All couplers on the QPU have E_{J_c} of 126 GHz and E_{C_c} of 58 MHz (C_c = 333 fF), resulting in an f_c of 7.6 GHz. Group A and group B qubits have C_{qc} values of 2.4 and 4.2 fF; thus, g_{qc} of about 48 and 60 MHz, respectively. This lies at the low end of typical g_{qc} values of 50-100 MHz for our parametric gates, which is limited by the maximum achievable C_{qc}, due to qubit and coupler arm dimension constraints [Fig. 5.2(a)]. However, the gate strength

[2] See Fig. 4.1(b) for a circuit schematic of the parametric-gate architecture.

also depends on the qubit-coupler frequency detuning Δ_{qc} (Eq. 5.6) and the coupler's DC bias point and AC modulation amplitude [120, 122]. This allows for compensation of the slightly low values of g_{qc}. Nevertheless, other factors beyond design must be taken into account when calibrating a CZ or iSWAP gate. We will see a glimpse of this in Chapter 6.

This page intentionally left blank.

Table 5.1: Design parameters of the second-generation 25-qubit QPU

Qubit	1	2	3	4	5	6	7	8	9	10	11	12
f_r (MHz)	7190	6794	6989	6609	7414	6792	7209	6591	7391	6993	6989	6609
f_{01} (MHz)	5199	4502	5094	4409	5199	4608	4997	4302	5298	4608	5094	4408
Δ_{qr} (MHz)	-1991	-2292	-1895	-2200	-2215	-2185	-2211	-2289	-2093	-2385	-1895	-2201
α (MHz)	-260.6	-155.1	-261.3	-155.4	-260.6	-154.7	-261.1	-155.7	-259.2	-154.8	-261.3	-155.4
E_J (MHz)	15925	18849	15319	18095	15925	19715	14808	17265	16544	19715	15319	18095
d_J (μm)	0.339	0.371	0.332	0.363	0.339	0.38	0.326	0.354	0.346	0.38	0.332	0.363
E_C (MHz)	232.5	143.4	232.5	143.5	232.5	143.4	231.9	143.4	231.9	143.4	232.5	143.4
C_q (fF)	83.1	134.7	83.1	134.7	83.1	134.7	83.3	134.7	83.3	134.7	83.1	134.7
E_J/E_C	68.5	131.4	65.9	126.1	68.5	137.5	63.9	120.4	71.3	137.4	65.9	126.2
C_{qr} (fF)	2.4	4.2	2.4	4.2	2.5	4	2.6	4.5	2.5	4.2	2.4	4.2
g_{qr} (MHz)	47	56	46	54	52	54	50	57	50	59	46	54
χ (MHz)	0.13	0.088	0.133	0.088	0.126	0.09	0.122	0.092	0.134	0.088	0.132	0.088
Q_c (kHz)	17.2	23	16.5	22.5	19.5	21.5	21.1	24.5	17.6	23.7	16.5	22.5
κ_r (MHz)	0.417	0.296	0.424	0.294	0.38	0.315	0.342	0.269	0.419	0.295	0.424	0.294
T_1^{rP} (μs)	677	888	651	888	774	814	896	942	656	889	651	889
C_{qd} (aF)	41.4	62	41.4	62	41.4	61.7	41.4	62	41.5	61.8	41.2	61.9
T_1^{dP} (μs)	907	876	946	915	909	843	986	960	872	840	953	918
T_1^{P} (μs)	388	441	385	451	418	414	469	475	374	432	387	451
C_{qc} (fF)	2.4	4.4	2.4	4.4	2.4	4.4	2.4	4.4	2.4	4.4	2.4	4.4

Qubit	13	14	15	16	17	18	19	20	21	22	23	24	25
f_r (MHz)	7413	6794	7211	6591	7391	6990	7209	6811	7211	6794	6989	6609	7413
f_{01} (MHz)	5192	4502	5094	4302	5297	4608	4997	4302	5197	4502	5094	4408	5198
Δ_{qr} (MHz)	-2220	-2292	-2117	-2289	-2094	-2382	-2212	-2509	-2014	-2292	-1895	-2201	-2214
α (MHz)	-259.8	-155	-261.3	-155.7	-259.2	-154.8	-261.1	-155.7	-260.4	-155	-261.3	-155.3	-260.5
E_J (MHz)	15925	18849	15319	17265	16544	19715	14808	17265	15925	18849	15319	18095	15925
d_J (μm)	0.339	0.371	0.332	0.354	0.346	0.38	0.326	0.354	0.339	0.371	0.332	0.363	0.339
E_C (MHz)	231.9	143.4	232.5	143.4	231.9	143.4	231.8	143.4	232.3	143.4	232.5	143.4	232.5
C_q (fF)	83.3	134.7	83.1	134.7	83.3	134.7	83.3	134.7	83.2	134.7	83.1	134.7	83.1
E_J/E_C	68.7	131.4	65.9	120.4	71.3	137.5	63.9	120.4	68.6	131.4	65.9	126.2	68.5
C_{qr} (fF)	2.5	4.2	2.5	4.5	2.5	4.2	2.6	4.7	2.4	4.2	2.4	4.2	2.5
g_{qr} (MHz)	51	56	50	57	50	59	50	62	47	56	46	54	52
χ (MHz)	0.125	0.088	0.128	0.092	0.134	0.088	0.122	0.09	0.128	0.088	0.132	0.088	0.126
Q_c (kHz)	19.4	23	19.2	24.5	17.6	24.2	21.1	27.3	17.2	23	16.4	22.5	19.4
κ_r (MHz)	0.382	0.296	0.376	0.269	0.419	0.289	0.342	0.25	0.42	0.296	0.427	0.294	0.382
T_1^{rP} (μs)	776	888	770	942	656	906	895	1042	685	889	646	889	769
C_{qd} (aF)	41.4	61.9	41.4	61.9	41.3	61.9	41.5	62.1	41.4	61.9	41.3	62	41.3
T_1^{dP} (μs)	916	880	948	961	880	840	981	957	910	878	952	915	914
T_1^P (μs)	420	442	425	476	376	436	468	499	391	442	385	451	418
C_{qc} (fF)	2.4	4.4	2.4	4.4	2.4	4.4	2.4	4.4	2.4	4.4	2.4	4.4	2.4

Table 5.2: Simulated capacitance offset in percentage for group A (group B) qubits on the QPU due to variations in chip-to-chip spacing d_c from 8 μm to 7 μm and 9 μm. The simulation is performed on one qubit from each group, but a similar behavior is expected for all qubits.

d_c	7 μm	9 μm
C_q	4 (3.5)	-3 (-2.4)
C_{qr}	13.5 (13)	-11.5 (-10)
C_{qd}	13 (11)	-10 (-9)
C_{qc}	-0.4 (-1)	1.3 (0.4)

5.1.5 Parameter sensitivity to chip-to-chip spacing

Design parameters extracted in Table 5.1 are based on a chip-to-chip spacing d_c of 8 μm between the C-chip and Q-chip. However, variations in d_c arise from the uncertainty in the flip-chip bonding process and chip thickness non-uniformity. Therefore, it is useful to simulate the effect of such variations on design parameters. Table 5.2 shows the offset in percentage in C_q, C_{qr}, C_{qd} and C_{qc} when d_c is 7 μm and 9 μm. The highest simulated deviation is in C_{qr} and C_{qd} with up to 13.5%, while C_{qc} is almost insensitive to variations in d_c within this 2 μm range. Crucially, C_q can deviate by up to 4%, which leads to $\sim$2% deviation in qubit frequency. In [142], simulation of a similar device showed that a 1 μm offset in d_c can lead to more than 100 MHz offset in resonator frequency. As we will see in Section 5.2 and Chapter 6, deviations in d_c indeed led to large deviations in resonator frequencies and small deviations in qubit frequencies.

5.2 Flip-chip module selection

Fabrication of the flip-chip module follows the process laid out in Section 2.2 and Section 4.2. As shown in Table 5.1, Josephson junction width d_J varied between 0.326 and 0.38 μm depending on the target E_J (hence, junction R_N).[3] Two wafers (one with four C-chips and one with four Q-chips), were fabricated and flip-chip bonded [Fig. 5.3(a)]. On the Q-chip wafer, junction

[3] See Chapter 2 for the relation between d_J and E_J.

test structures, similar to those presented in Section 2.3.1, surrounded the four Q-chips, and were utilized for junctions R_N characterization of the nearby Q-chips.

To select the optimal module out of the four, two factors were considered: the target junctions' R_N and the measured chip-to-chip spacing d_c. A module with a large gradient in d_c (even if average d_c is on target) is less tolerable in terms of frequency collisions than a module with a global offset. The selected module had a d_c gradient ranging from 5.8 to 6.7 μm.[4] Figure 5.3(b) shows a 3D schematic of the flip-chip module, illustrating a tilt. Based on our simulations (Table 5.2), this would lead to about an 8% increase in capacitance and, consequently, an 8% drop in E_C. However, measurement of junction test structures near that module showed that R_N was expected to be 3% below target; hence, 3% higher E_J than the target. This slightly compensates for the drop in E_C, causing qubit frequencies to decrease by only 2% globally (Eq. 2.8). This module combined Q-chip 1 and C-chip 1, in which case junction test structures on the top and left side of the wafer were used to predict E_J. As I will show in Chapter 6, the measured reduction in E_C was lower than expected from the simulation, and the measured qubit frequencies were close to the target.

[4] Chip-to-chip spacing measurements were conducted at VTT using SEM. The 5.8 to 6.7 μm gradient excludes outlier measurements, likely caused by chipping at the edge of the Q-chip.

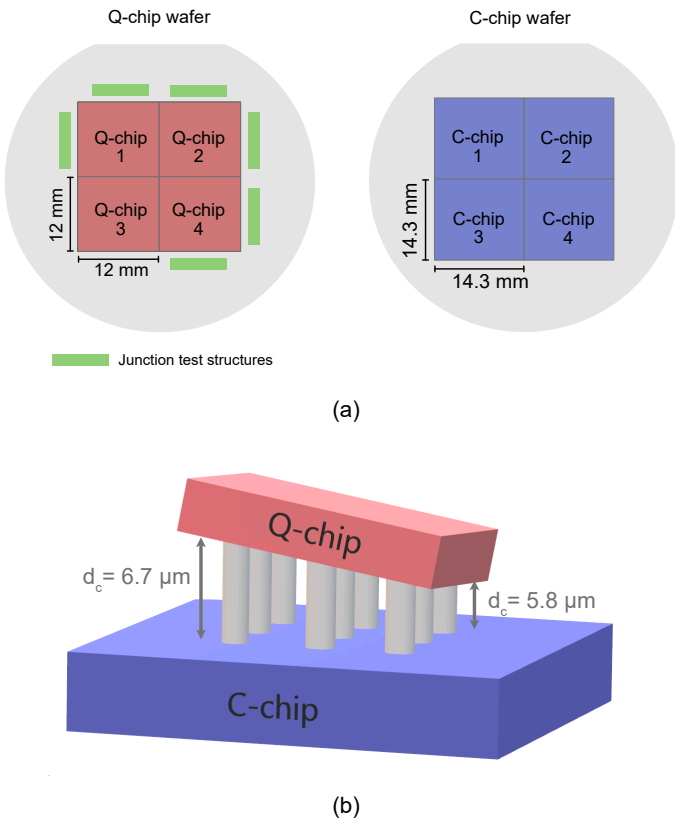


Figure 5.3: Fabrication and flip-chip bonding. (a) Schematics of the Q-chip wafer and C-chip wafer, each containing four chips. The Q-chip wafer contains junction test structures (green) surrounding the Q-chips for post-fabrication R_N characterization and optimal module selection. (b) 3D illustration of the selected flip-chip module for measurements, where the chip-to-chip spacing d_c deviates from the target value due to uncertainties in the bonding process. Chip sizes, d_c, and bumps are not drawn to scale, and the tilt is exaggerated for clarity.

CHAPTER 6

QPU Benchmarking

As processors grow in complexity, benchmarking QPUs is critical for evaluating and advancing quantum computing technologies. Key benchmarking metrics include parameter targeting, coherence times, crosstalk, and gate fidelities. These metrics determine the feasibility of various quantum algorithms, identify sources of errors, and guide the optimization of fabrication processes.

In this chapter, I present benchmarking results of our second generation of the 25-qubit QPU, whose design was presented in Chapter 5. At the time of writing this thesis, this QPU is still undergoing further characterization. Therefore, I only present results that have been accomplished thus far. Section 6.1 provides an overview of the measurement setup. In section 6.2, I lay out the calibration steps followed to extract the relevant parameters. In section 6.3, I present an analysis of parameter targeting, the most critical metric in this work. In section 6.4, I discuss qubit coherence. In section 6.7, I analyze the quantum-gate performance, focusing on single-qubit gate fidelities in isolation and during simultaneous operations to assess crosstalk effects. Section 6.6 provides an example of a two-qubit gate implementation on the QPU. However, automatic calibration and characterization of two-qubit gate

performance is an ongoing work. Finally, in Section 6.7, I present a list of future characterizations to be implemented on the QPU, along with potential experiments.

6.1 Measurement setup

After flip-chip bonding, the QPU module is glued to a microwave package, where signal launch pads on the C-chip are wire-bonded to the corresponding traces on the PCB. The packaged module is installed at the mixing chamber stage of the dilution refrigerator. A schematic of the measurement setup is shown in Fig. 6.1(a), and a photograph of the assembled package, connected to SMPM cables, is shown in Fig. 6.1(b).

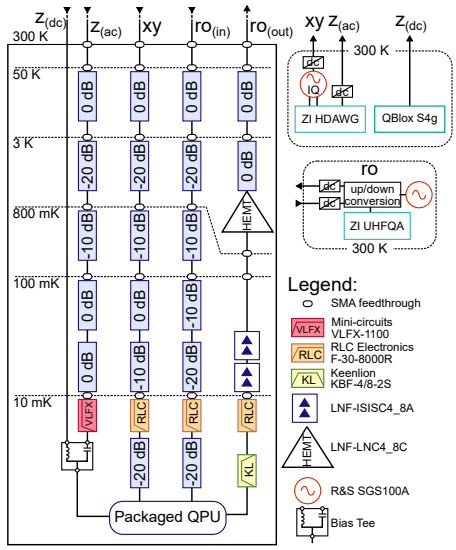
Gate pulses on the xy and $z_{(ac)}$ signal lines are generated by Qblox QCM-RF modules, supporting single-qubit gates and coupler AC-flux modulation for two-qubit gates, respectively. Qubit readout is performed by Qblox QRM-RF modules. Each module is equipped with a dedicated local oscillator (LO), and thus capable of generating RF signals with frequencies between 2 and 18.5 GHz. Additionally, a down-converter with an LO fixed at 4.4 GHz is used to down-convert the two-qubit gate pulse generated by the QCM-RF modules to the parametric-gate frequency regime.

DC currents for the $z_{(dc)}$ lines, dedicated to coupler DC biasing, are supplied by Qblox SPI racks. Attenuators at different temperature stages of the cryostat, along with filters, reduce the number of thermal photons and infra-red radiation reaching the QPU [143–146]. I note that our current measurement setup can only accommodate 20 qubits. Therefore, the first row of 5 qubits is excluded from the characterization.

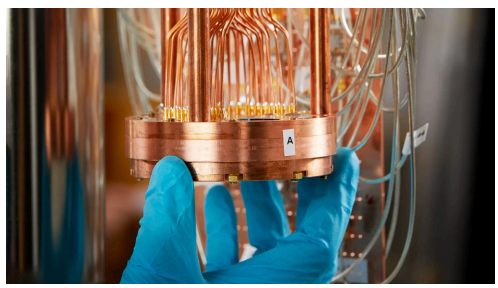
6.2 Device calibration

The standard calibration flow of single-qubit gates and readout calibration, along with their pulse sequences, is shown in Fig. 6.2. [1] Before starting this sequence, we first extract resonator frequencies using continuous-wave spectroscopy via a Vector Network Analyzer (VNA). QPU calibration and char-

[1] Calibration of single-qubit gates follows a method similar to that described in Appendix D of Ref. [125], which includes exemplary plots.



(a)



(b)

Figure 6.1: QPU measurement setup. (a) Schematic of equipment and circuit components at different temperature stages. (b) Photograph of the assembled microwave package connected to SMPM cables.

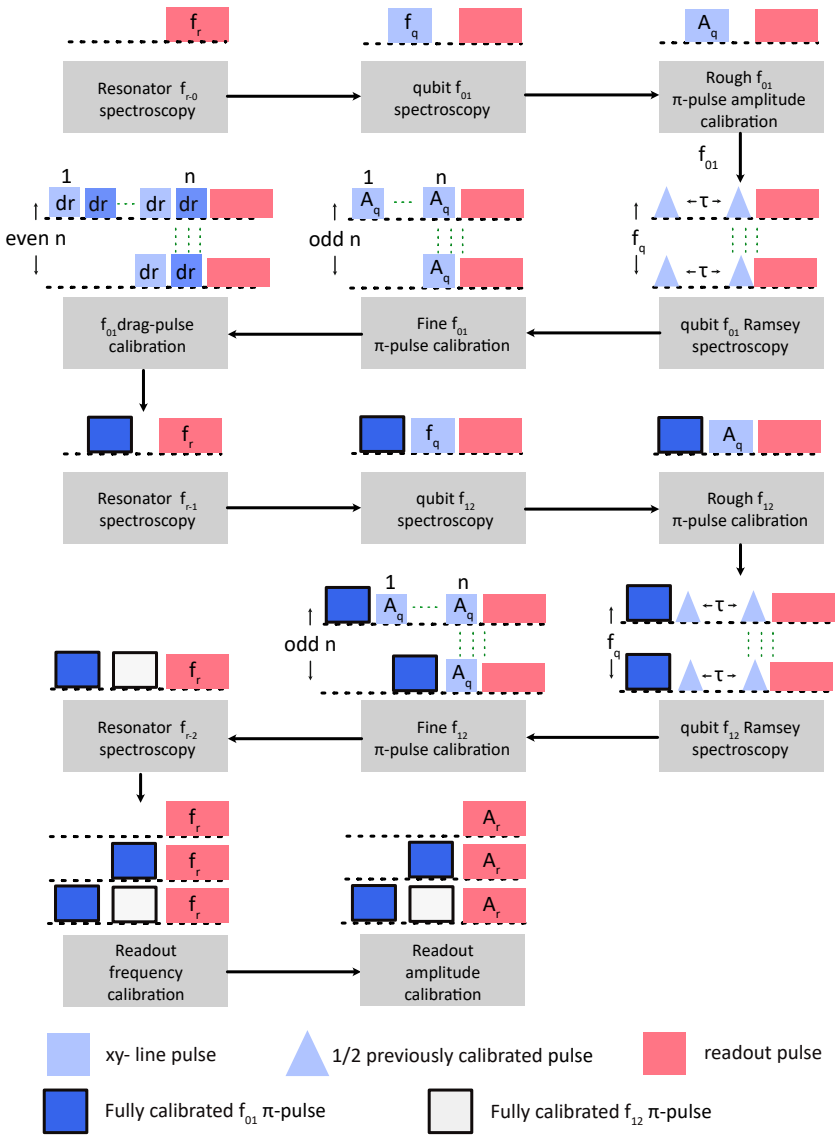


Figure 6.2: Standard calibration flow of single-qubit gates and readout. Calibration steps are indicated in the gray boxes. f_{r-0}, f_{r-1} and f_{r-2} are the resonator frequencies when the qubit is in the $|0\rangle$, $|1\rangle$ and $|2\rangle$ state, respectively. (caption continued on next page)

Figure 6.2: Readout pulses are shown as red boxes, while xy (or qubit-control) pulses are shown as blue and white boxes/triangles. The parameter(s) varied in each step are indicated in black text. f_r and A_r are the readout pulse frequency and amplitude, respectively. f_q and A_q are the qubit-control pulse frequency and amplitude, respectively. τ is the delay between control pulses. dr is the DRAG parameter that minimizes leakage to higher qubit energy states [147]. n is the number of pulses. Note that for the drag-pulse calibration, the dark blue control pulse is a 180° out of phase from the light blue control pulse.

acterization are then performed using `Tergite Autocalibration`, an open-source software tool developed at Chalmers [148].

Figures 6.3(a) and 6.3(b) show color maps of the measured 20 qubit frequencies f_{01} and anharmonicities α, respectively. These two color maps are shown in a blue-red scale to highlight the two frequency groups: A in blue and B in red (Fig. 5.1). Resonator frequencies f_r are shown in Fig. 6.3(c). We fit the resonators S_{21} transmission measurements to the complex resonator model proposed in [149], and extract the loaded quality factor Q_l, from which we extract $\kappa_r = f_r/Q_l$.[2] The measured resonators dispersive shift χ is shown in Fig. 6.3(e), based on which, we extract the qubit-resonator coupling strength g_{qr} in Fig. 6.3(f), from Eq. 5.4.

In the next section, I compare the extracted QPU parameters with the target design parameters, presented in Section 5.1.2.

6.3 Parameter Targeting

A QPU is designed to have ideal parameters, such as frequencies and coupling strengths, tailored to meet target performance goals. However, these parameters deviate from the target due to uncertainties in the fabrication process and packaging, or simulation inaccuracies. In the following, I quantify our ability to meet the target design parameters.

[2] In Table 5.1, κ_r values are only based on Q_c. Here, we assume $Q_i \gg Q_c$, hence, $Q_c \approx Q_l$. For accurate comparison between measurement and design parameters, Q_c should also be extracted in future calibrations.

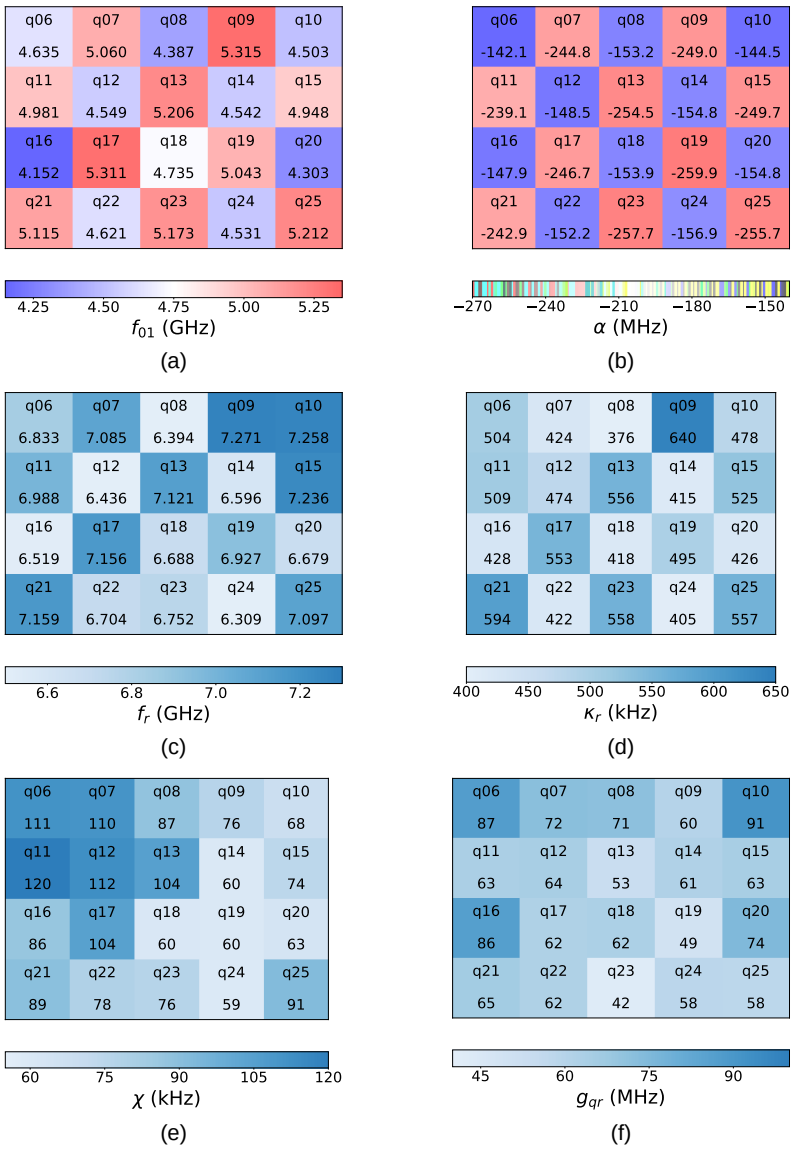


Figure 6.3: Measurement of the QPU's qubits and resonators parameters. Heatmaps of (a) qubit frequency f_{01}, (b) qubit anharmonicity α, (c) resonator frequency f_r, (d) resonator linewidth κ_r, (e) dispersive shift χ, and (f) qubit-resonator coupling g_{qr} for 20 qubits in the QPU.

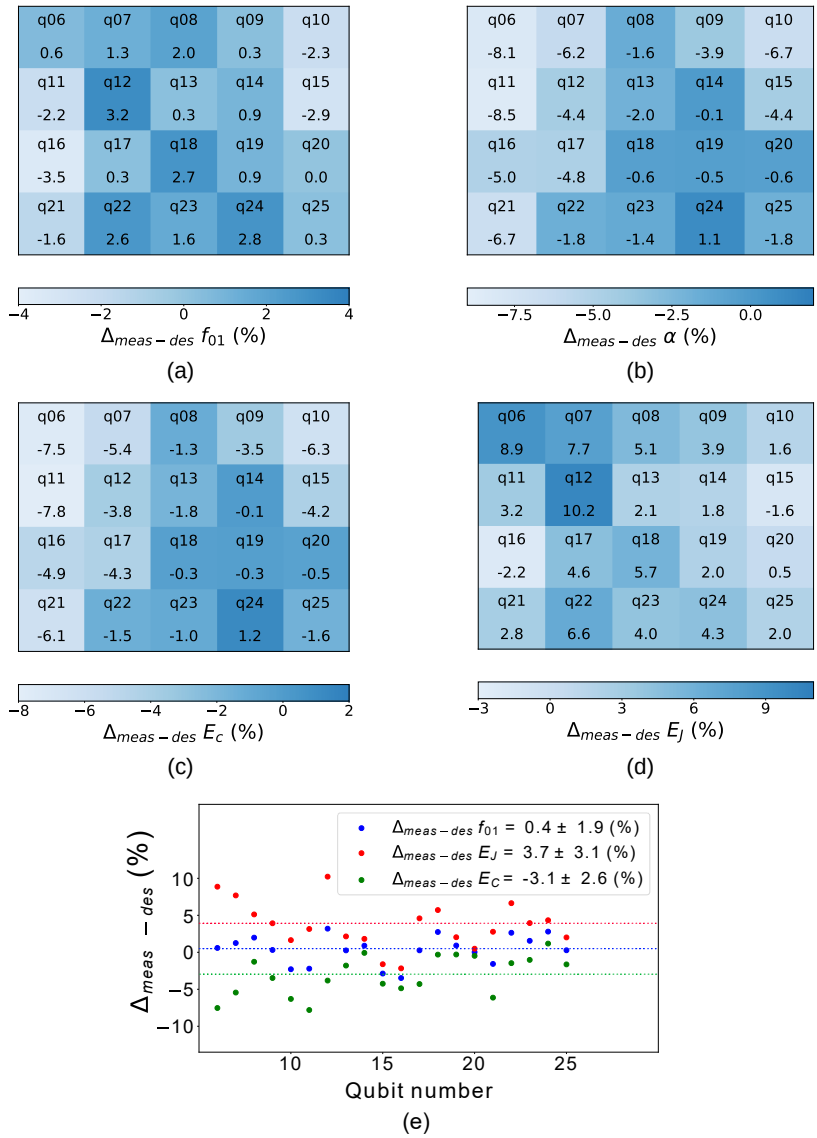


Figure 6.4: Qubit parameter targeting. Heatmaps of the measurement-design offset $\Delta_{\text{meas-des}}$ in (a) qubit frequency f_{01}, (b) qubit anharmonicity α, (c) charging energy E_C, and (d) Josephson energy E_J. (e) Scatter plot of $\Delta_{\text{meas-des}}$ in f_{01}, E_J, and E_C for all 20 qubits. The dashed lines correspond to the mean $\Delta_{\text{meas-des}}$ of each parameter. The mean and standard deviation in $\Delta_{\text{meas-des}}$ are stated in the legend.

6.3.1 Qubit frequency

Given that a primary objective of this thesis is to enhance qubit frequency targeting on a QPU, I begin by presenting an analysis of qubit frequencies and their deviations from the allocated frequencies. Figure 6.4(a) shows a heatmap of the measurement-design offset, $\Delta_{\text{meas-des}}$, in f_{01}. Measured qubit frequencies are on average 0.4% higher than the target, with a standard deviation of 1.9% around that offset [Fig. 6.4(e)]. The 0.4% increase in f_{01} quantifies our ability to target the mean f_{01}, while the 1.9% quantifies the deviation around that mean. Such a high precision in targeting the mean qubit frequency is a key performance metric. However, as discussed in Sections 2.3.4 and 4.4, a global offset in the mean f_{01} of a few percent is tolerable, while the standard deviation around that mean is the main source of collisions between quantum gates. To our knowledge, no studies have reported qubit frequency targeting in flip-chip architectures. Thus, it is valuable to analyze the sources of the 1.9% standard deviation in qubit frequency, and its deviation from the achieved 1% on planar chips, as presented in Section 2.4.

Targeting the qubit anharmonicity, α, is also of crucial importance, although less emphasis is put on it, due to the relative simplicity of targeting E_C of the large qubit capacitor in a planar architecture. However, in a flip-chip architecture, deviations in the chip-to-chip spacing, d_c, lead to deviations in E_C and consequently in α. Figure 6.3(b) shows a heatmap of $\Delta_{\text{meas-offset}}$ in α. One can readily see a gradient in $\Delta_{\text{meas-des}}$ with a low offset at the bottom-right corner to a high offset at the left side and top-right corner of the chip. This is consistent with the measured d_c tilt shown in Figs. 5.3(c) and 6.5(b), where d_c varied between 5.8 and 6.7 μm. A lower d_c leads to a higher capacitance and hence lower E_C and α.

To disentangle the sources of deviations in f_{01} and α, we determine E_J and E_C, by numerically solving the Cooper-pair-box Hamiltonian. Figure 6.4(c) shows a heatmap of E_C, emphasizing the gradient observed in α. However, as shown in Fig. 6.4(d), E_J did not show an obvious gradient in its offset, which is to be expected in a small Q-chip area of 12×12 mm^2, where the gradient in R_N is small. Also, d_c has no impact on E_J. Figure 6.4(e) shows a scatter plot of $\Delta_{\text{meas-des}}$ in f_{01}, E_J and E_c, with means and standard deviations of 0.4 ± 1.9 %, 3.7 ± 3.1 % and -3.1 ± 2.6 %, respectively. As mentioned in Section 5.2, we selected a flip-chip module where the mean E_J is expected to be 3% above target, to compensate for the anticipated decrease of 4–8%

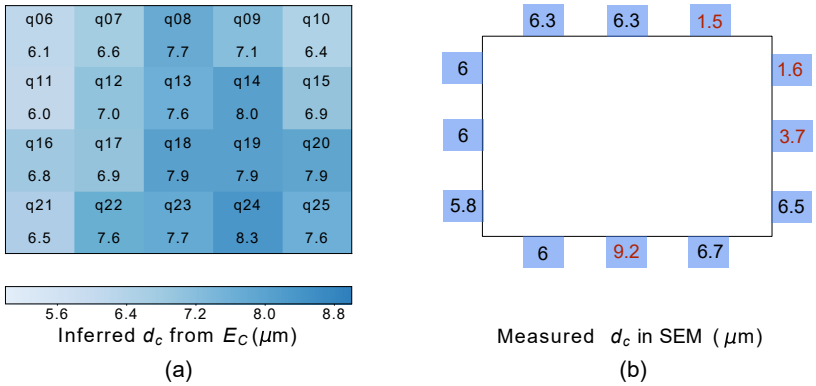


Figure 6.5: Chip-to-chip spacing d_c. (a) Inferred d_c from the measured charging energy E_C. (b) Measured d_c using SEM, at VTT, where values written in red are considered outliers and are likely due to inaccurate measurements due to chipping at the edge of the qubit chip.

in E_C, based on d_c measurements. The 3.7% mean offset in E_J is close to the 3% we anticipated. However, the standard deviation in this offset (3.1%), which is the same as the chip-level standard deviation in R_N, is higher than the 1.8% standard deviation in E_J measured on planar chips (see Section 2.4).

Interpreting the standard deviation of such a small sample size of 20 qubits is challenging. One can use statistical methods for an insight into the standard deviation of a larger qubit population. For example, the 95% confidence interval of the standard deviation in E_J is between 2.4 and 4.6%. Additionally, small sample sizes are more sensitive to outliers. Excluding q_{06}, q_{07} and q_{12}, this standard deviation drops to 2.3%, with a 95% confidence interval between 1.8 and 3.6%. Alternatively, the higher standard deviation could be explained by the additional fabrication steps introduced by the flip-chip processing.

The standard deviation in E_J alone should still lead to only 1.55% standard deviation in f_{01}. However, E_c also has a similar contribution, leading to the total 1.9% standard deviation in f_{01}. The offset in E_c was lower than we anticipated, but worked to our advantage. Simulations showed that E_c of group A (group B) qubits decreases by 4 (3.5)% per 1-μm reduction in d_c from 8 μm. Accordingly, we infer d_c from the measured E_c and plot its heatmap in

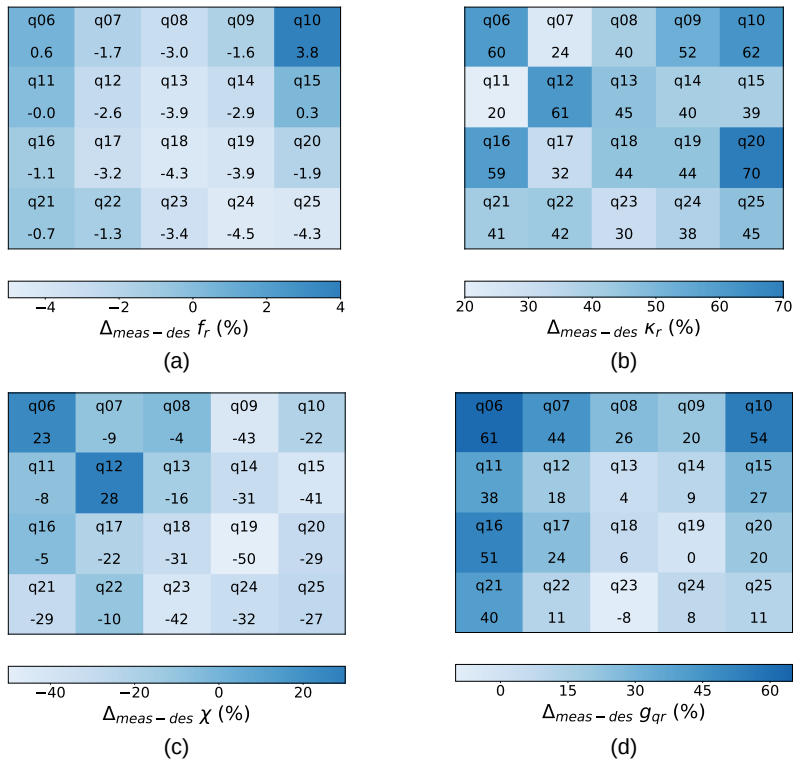


Figure 6.6: Resonator-qubit parameter targeting on the QPU. Heatmaps of the measurement-design offset $\Delta_{\text{meas-des}}$ in (a) resonator frequency f_r, (b) resonator linewidth κ_r, (c) dispersive shift χ, and (d) qubit-resonator coupling g_{qr} for 20 qubits.

Fig. 6.5(a), where it varies between 6 and 8.3 μm, with a mean and standard deviation of 7.2 $\pm$ 0.65 μm. SEM measurements of d_c at the four edges of the Q-chip are shown in Fig. 6.5. The discrepancy between the inferred and measured d_c could be attributed to inaccuracy in SEM measurements of d_c, due to chipping at the Q-chip edge or alignment of the flip-chip module inside the SEM tool chamber. Additionally, bows in the Q-chip could also introduce variations in d_c that are not captured at the edge of the Q-chip. Furthermore, inferring d_c from E_C measurements does not account for lithography variations in the planar capacitor. This could explain the local fluctuations in the inferred d_c between neighboring qubits.

6.3.2 Resonator-qubit coupling

Another benchmarking metric on this QPU generation is targeting resonator frequencies f_r and qubit-resonator coupling g_{qr}. This metric is less crucial as improving readout fidelity is not one of the main goals of this QPU, provided the qubits are not strongly affected by the Purcell decay through the resonator. Figure 6.6(a) shows a heatmap of $\Delta_{\text{meas-des}}$ in f_r with a similar gradient to E_C, but with an opposite sign, since f_{01} increases with smaller d_c [142]. However, in regions where d_c is inferred to be 8 μm, f_r is $\sim$4% lower than the target, indicating that our simulation of f_r is off by as much.

Resonator linewidth, κ_r, is a more sensitive parameter to d_c [142]; its $\Delta_{\text{meas-des}}$ heatmap is shown in Fig. 6.6(b). I note that inaccuracy in the measured κ_r offset may arise from the imperfection of the circle fit, and the assumption that $Q_l = Q_c$ (see footnote 2). Similarly, $\Delta_{\text{meas-des}}$ in χ and g_{qr} are significantly high for some qubits, as these two parameters depend on f_r, f_{01}, α and the coupling capacitance between the qubit and the resonator, c_{qr}.

Although both κ_r and g_{qr} are on average higher than their target values, we observed no noticeable effect on T_1 from the Purcell decay through the resonator, as I will show in the next section.

6.4 Qubit coherence

A second key metric of the QPU is qubit coherence. Figures 6.7(a), 6.7(c) and 6.7(e) show heatmaps of the mean T_1, T_2 and T_2^{echo}, respectively. The mean values, with error bars representing $\pm$1 standard deviation, are shown in

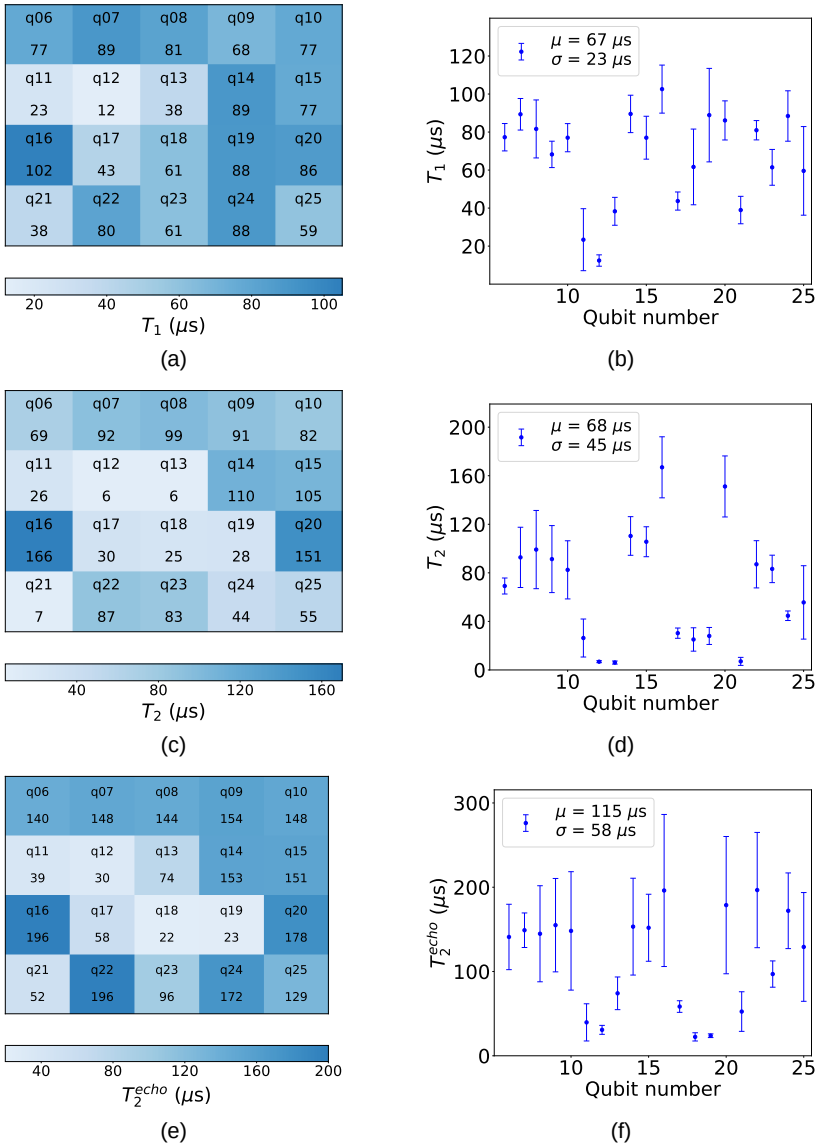


Figure 6.7: Qubit coherence. Heatmaps of the measured (a) T_1, (c) T_2, and (e) T_2^{echo}. Scatter plots of each parameter are shown in (b), (d), and (f), respectively, where each data point represents 200 measurements over a time span of 16 hours. The mean μ and standard deviation σ of each parameter over all qubits are stated in the legend of each plot.

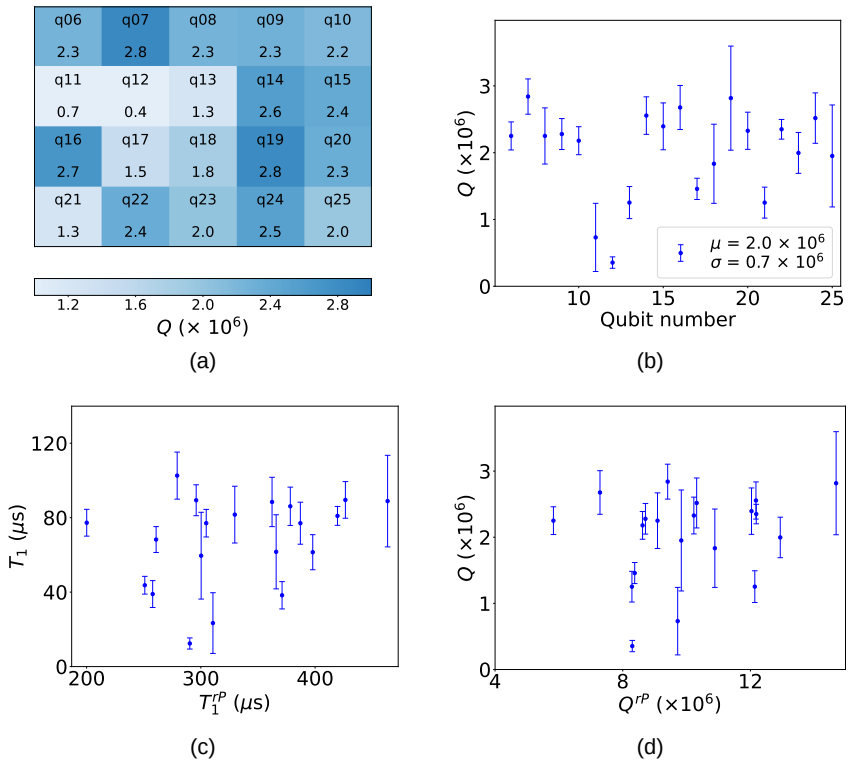


Figure 6.8: Quality factors and Purcell decay of qubits. (a) Heatmap of the measured quality factors Q. (b) Scatter plot of Q, where its mean μ and standard deviation σ over all qubits are stated in the legend. (c) T_1 versus its Purcell limit through the readout resonator T_1^{rP}. (d) Q versus its Purcell limit Q^{rP}. Neither T_1 nor Q scale with their Purcell limit, indicating that qubits are not Purcell-limited.

scatter plots in Figs. 6.7(b), 6.7(d) and 6.7(f). (Each data point represents 200 measurements, over a time span of 16 hours.) The mean μ and the standard deviation σ in each parameter over all 20 qubits are shown in the legend of each plot. The mean T_1, T_2 and T_2^{echo} are 67, 68 and 115 μs, respectively.[3] The significant increase in T_2^{echo} compared to T_2, approaching the $2T_1$ limit, highlights the presence of low-frequency noise that is canceled by the echo pulse [150, 151].

For a fair assessment of the qubit quality on the QPU, we must scale their T_1 to their quality factors, Q. Figures 6.8(a) and 6.8(b) show a heatmap and a scatter plot of Q, respectively. The average quality factors of the 20 qubits is 2×10^6; almost a factor of two improvement compared to the average Q of 1.1×10^6 measured on the first generation of QPU without tunnels, discussed in Section 4.3. To understand whether the qubits T_1 and Q are limited by the Purcell decay through the resonator, it is useful to plot both parameters against T_1^{rP} and Q^{rP}, as shown in Figs. 6.8(c) and 6.8(d), respectively. We observe no obvious scaling of T_1 and Q with their Purcell limit and conclude that their Purcell decay has no significant effect on the qubit performance.[4]

6.5 Crosstalk and gate fidelity

Errors in quantum-gate implementation can be divided into two types of errors, which prevent us from implementing perfect operations [31, 32, 152–159]. The first type, coherent errors, arises from systematic imperfection of the control, resulting in deterministic but incorrect manipulation of the qubits. Examples of coherent errors are gate miscalibration, pulse distortions, leakage and crosstalk. The second type, incoherent errors, arises from random processes that degrade the qubit state over time. Examples of incoherent errors are decoherence and thermal excitations. Different benchmarking techniques are used to quantify the different types of error.

One of the main goals of this thesis is to mitigate the effect of crosstalk on quantum-gate fidelity. Therefore, characterization of single- and two-qubit gates is required to assess our frequency allocation and to determine future steps to improve those fidelities. At the time of writing this thesis, this char-

[3] The value of T_1 of q_{11}, q_{12} and q_{13} decreased from 62, 31, and 51 μs, respectively, in a previous cooldown, which is probably due to thermal redistribution of TLSs.

[4] See Fig. 3.4 for the case where scaling of T_1 and Q with the Purcell limit is obvious.

acterization has only been tentatively conducted on single-qubit gates, which will be the focus of this section. An example of two-qubit gate implementation is presented in Section 6.6.

6.5.1 Randomized Benchmarking

Another key benchmarking metric, that I present in this thesis, is the effect of crosstalk on quantum gates. We quantify this effect by extracting the gate fidelity on a victim qubit, while other qubits are being driven in parallel. The gate fidelity on the victim qubit is then compared to when it is driven in isolation. One commonly used method of extracting the average quantum-gate fidelity, F_{avg}, is the randomized benchmarking (RB) [160–163]. In an RB sequence, a qubit is prepared in its initial state, typically the $|0\rangle$ state, and then a random sequence of m unitary operations (gates) is applied on the qubit, with a final inversion operation, such that the total sequence of operations should eventually maintain the initial state of the qubit. The random sequence of unitaries is sampled from the Clifford group. To extract F_{avg}, the RB sequence is implemented with increasing m, then the population of the initial state $P(|0\rangle)$ is measured and fit to the exponential decay

$$P(|0\rangle) = Ap^m + B. \tag{6.1}$$

A and B are constants that account for the state-preparation and measurement (SPAM) errors, and p is the depolarizing parameter. The depolarizing parameter summarizes the cumulative effects of all errors in the system and reports them into a single figure of merit. The average gate fidelity is then given by

$$F_{avg} = \frac{1}{2}(1 + p), \tag{6.2}$$

and the corresponding average gate error ε_{avg} is $1 - F_{avg}$.

Figure 6.9(a) shows an example of an RB measurement, where populations of the $|0\rangle$, $|1\rangle$ and $|2\rangle$ states are plotted as a function of the number of Clifford gates. $P(|0\rangle)$ is fit to Eq. 6.1, where the fit parameter p is shown on the plot legend. Figure 6.9(b) shows a scatter of the extracted ε_{avg} for 18 qubits, where 12 qubits have $\varepsilon_{avg} < 0.1\%$. The two qubits not shown are q_{23}, where $\varepsilon = 0.625\%$, and q_{12}, where the significantly low readout fidelity at the time of measurements prohibited us from extracting a good fidelity fit. Further

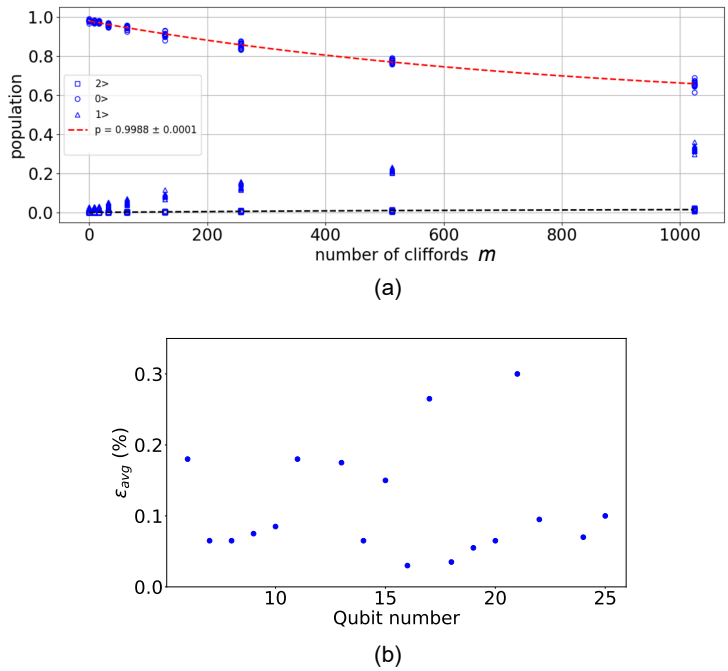


Figure 6.9: Randomized Benchmarking and average gate error ε_{avg}. (a) Example of RB measurement, where the qubit's $P(|0\rangle)$ decays exponentially with the number of applied Clifford gates m, and a fit to Eq. 6.1. (b) ε_{avg} of 18 qubits when characterized in isolation. The two qubits not shown are q_{23}, where $\varepsilon_{avg} = 0.625\%$, and q_{12}, where the significantly low readout fidelity at the time of measurements prohibited us from extracting a good fidelity fit.

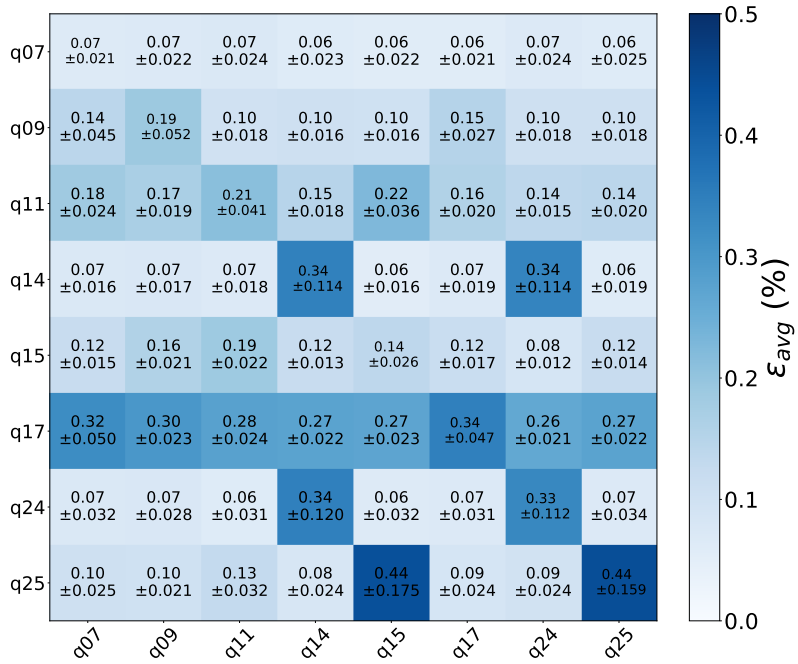


Figure 6.10: Average gate error matrix of 8 qubits on the QPU.

reduction of error requires further optimization of single-qubit gates. However, this is beyond the scope of this thesis, and instead, I focus on possible crosstalk between qubits.

To achieve this goal, we implement the RB sequence in three different forms: individual, pairwise, and simultaneous characterization. Initially, we attempted to conduct such characterization with all 20 qubits. However, this takes a prohibitively long time– especially in the absence of an active reset of the qubit state– such that gate calibration parameters may drift over the measurement.[5] Therefore, we limited the characterization, presented here, to qubits with the most noticeable fluctuations in ε_{avg} when driven in parallel with other qubits.

Figure 6.10 shows ε_{avg} matrix for 8 such qubits, where the diagonal elements

[5]Running such measurements on 20 qubits takes over 48 hours, with m=1024 and 36 randomness seeds.

are ε_{avg} when all 8 qubits are driven in parallel. Thus, the diagonal element is expected to have the highest ε_{avg} of the respective row elements. The off-diagonal elements in each row are ε_{avg} of this row's victim qubit when the qubit in the corresponding column is characterized in parallel. Among these 8 qubits, two pairs showed the strongest effect of crosstalk on the QPU: $q_{14}-q_{24}$ and $q_{15}-q_{25}$. For the $q_{14}-q_{24}$ pair, the crosstalk effect is symmetric, such that ε_{avg} of both qubits increases when characterized in parallel. We attribute this noticeable effect of crosstalk to the proximity of both qubits' f_{01}, with only 11 MHz detuning. As for $q_{15}-q_{25}$, the crosstalk effect was asymmetric, where only ε_{avg} of q_{25} increased when both qubits were characterized in parallel. We attribute this asymmetry to the proximity of q_{15}'s f_{01} (4.948 GHz) to q_{25}'s f_{12} (4.956 GHz), with only 8 MHz detuning between the two frequencies. This means that single-qubit gates on q_{15} induce $|1\rangle$ to $|2\rangle$ state transitions on q_{25}.

6.5.2 01 and 12 crosstalk

Based on these observations, it becomes useful to classify the crosstalk effects on single-qubit gates into two types. The first type is a symmetric "01-crosstalk", which occurs when the victim qubit's f_{01} is close in resonance to a source qubit's f_{01}. The second type is an asymmetric "12-crosstalk", which occurs when the victim qubit's f_{12} is close in resonance to a source qubit's f_{01}.

Besides the observed crosstalk on the two-qubit pairs mentioned in the previous subsection, several other *weaker* crosstalk effects were observed on other pairs. In Fig. 6.10, a "01-crosstalk" effect is observed between q_{09} and q_{17}, where both their ε_{avg} increased by 0.05% and 0.03%, respectively, compared to their isolated case. Additionally, we notice a "12-crosstalk" effect between the victim q_{17} and the source q_{07}, where ε_{avg} of q_{17} increased by 0.05%.

We find it useful to plot an f_{01} collision matrix, and an $f_{12}-f_{01}$ collision matrix for all qubits, which helps anticipate symmetric crosstalk (01-crosstalk) or asymmetric crosstalk (12-crosstalk). Figure 6.11(a) shows a matrix where each element is the absolute value of the measured detuning between f_{01} values of the two corresponding qubits. Figure 6.11(b) shows a matrix where each element is the absolute value of the measured detuning between the row qubit's f_{12}, and the column qubit's f_{01}. Elements that are bordered by red boxes in each matrix correspond to qubit pairs where a hint of crosstalk was observed across different characterization rounds. However, a more extensive

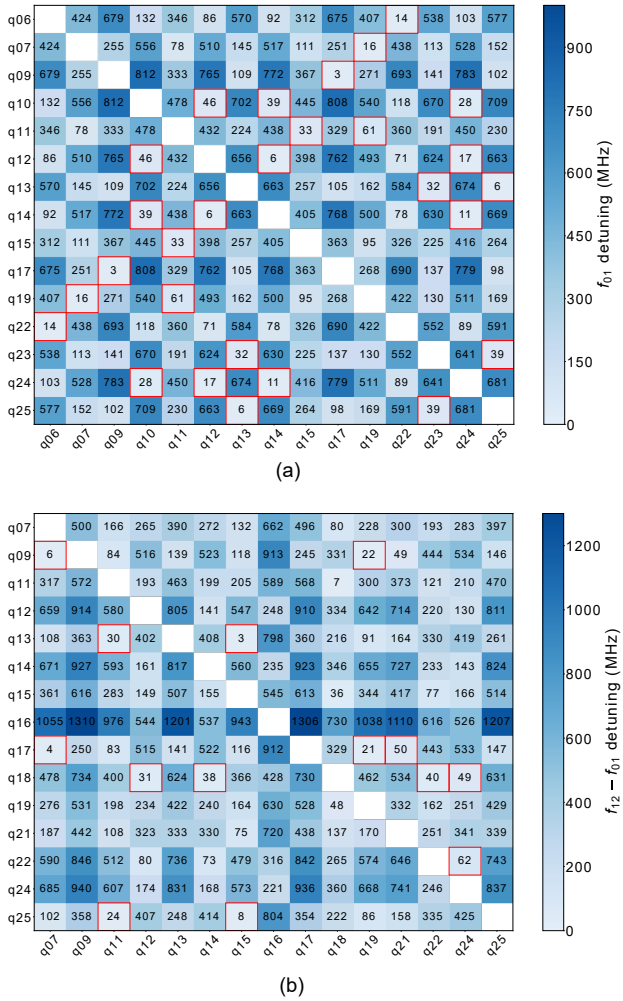


Figure 6.11: Collision matrix of some qubits on the QPU. (a) f_{01} collision matrix, where each element is the absolute value of the measured detuning between f_{01} values of the two corresponding qubits. This matrix is relevant for "01-crosstalk". (b) $f_{12} - f_{01}$ collision matrix, where each element is the absolute value of the measured detuning between the row qubit's f_{12}, and the column qubit's f_{01}. This matrix is relevant for "12-crosstalk". Qubits that are not close in resonance to any other qubits are removed to simplify the matrices.

characterization of these qubit pairs is needed. Note that on both matrices, qubits that are not close in resonance to any other qubits are removed to simplify the matrices.

6.5.3 Preliminary conclusions

There are three important conclusions to make out of the single-qubit gate characterization. Firstly, a frequency allocation scheme that maximizes the spacing between qubit frequencies is crucial. While this has been achieved between most qubits on the QPU, deviations in qubit frequencies arising from uncertainties of the fabrication process render some of these qubits closer to each other in frequency than intended. Secondly, our frequency allocation of 8-qubit unit-cell is designed to mitigate crosstalk between nearest-neighboring qubits, in which case, qubit frequencies are, by design, repeated over the QPU lattice. For example, q_{09} and q_{17} are indeed designed to have the same frequency, so it is not surprising that their detuning is low. Thirdly, and consequently, it is of most importance to understand the sources of crosstalk and mitigate them. Although the cause of crosstalk between certain qubit pairs, as discussed in the previous two subsections, is not yet well understood, direct capacitive coupling—either between qubits themselves or between victim qubits and source xy lines—can be excluded as a potential cause. In each of these pairs, qubits are separated by at least one other qubit, and their xy lines are spatially well separated by at least 4 mm, with launch pads that are often on different sides of the C-chip. The direct capacitance between an xy line and its qubit is $40-60$ aF, which is more than seven orders of magnitude larger than the direct capacitance between an xy line and a next-nearest-neighbor qubit.

6.6 Two-qubit gate

Mitigating crosstalk between two-qubit gates is equally important. However, since our automatic calibration of two-qubit gates is still ongoing, I limit the discussion here to the fidelity of a CZ gate. When applied between two qubits,

the CZ gate has an effective unitary operation:

$$U_{\mathrm{CZ}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix},$$

with a truth table

Input $\lvert q_1 q_2\rangle$	Output $\lvert q_1 q_2\rangle$
$\lvert 00\rangle$	$\lvert 00\rangle$
$\lvert 01\rangle$	$\lvert 01\rangle$
$\lvert 10\rangle$	$\lvert 10\rangle$
$\lvert 11\rangle$	$-\lvert 11\rangle$

where the $|11\rangle$ state acquires a geometric π phase. Such an operation is capable of generating entanglement between qubits, and along with single-qubit gates, it forms a universal gate set, that is sufficient for implementing any quantum algorithm [164].

We implemented the CZ gate between q_{14} and q_{15}, where the gate frequency f_{cz} is 559 MHz, and the gate time τ_{cz} is 262 ns.[6] To characterize the CZ gate fidelity, F_{cz}, we use two-qubit RB and iRB (interleaved RB). In iRB, the gate under investigation (CZ) is inserted after applying every random Clifford gate of the RB [165]. The p parameter for RB and iRB (p_{RB} and p_{iRB}, respectively) is extracted using Eq. 6.1, where F_{cz} is given by:

$$F_{\mathrm{cz}} = \frac{3}{4}\left(1 - \frac{p_{iRB}}{p_{RB}}\right). \tag{6.3}$$

Figure 6.12(a) shows the population of the $|00\rangle$ state, $P|00\rangle$ as a function of the number of Cliffords for RB and iRB. The extracted gate fidelity is 97.4%, below the 98.6% achieved on our two-qubit flip-chip device (**Paper F**), and planar devices [124, 125], and below the 99% state-of-the-art CZ-gate fidelities reported for the same gate architecture [166].

We found that most of the gate error (2.6%) is attributed to qubit decoherence. To extract the coherence limit of the CZ gate fidelity, $F_{\mathrm{cz}}^{\mathrm{coh}}$, we DC-bias the coupler to the same bias point we use during gate implementation. Then

[6] Calibration of the CZ gate follows a method similar to that described in Appendix D of Ref. [125], which includes exemplary plots.

we measure T_1 and T_2, extract T_ϕ using Eq. 3.1, then calculate $F_{\text{cz}}^{\text{coh}}$ [125, 167], where

$$F_{\text{cz}}^{\text{coh}} = \left(\frac{1}{2}\frac{1}{T_1^{q_{14}}} + \frac{3}{10}\frac{1}{T_1^{q_{15}}} + \frac{61}{80}\frac{1}{T_\phi^{q_{14}}} + \frac{29}{80}\frac{1}{T_\phi^{q_{15}}}\right)\tau_{\text{cz}}. \tag{6.4}$$

We measure $T_1^{q_{14}} = 70$ μs, $T_1^{q_{15}} = 75$ μs, $T_\phi^{q_{14}} = 23$ μs, and $T_\phi^{q_{15}} = 22$ μs, resulting in $F_{\text{cz}}^{\text{coh}}$ of 98.4%. These T_ϕ values are significantly lower than T_ϕ values at the coupler's zero bias point. Figure 6.12(b) shows the measured T_1 and T_ϕ for both qubits and the corresponding $F_{\text{cz}}^{\text{coh}}$, for different applied currents to the coupler's z line. While T_1 is not affected, T_ϕ of q_{14} drops by an order of magnitude, and T_ϕ of q_{15} drops by a factor of three. This leads to a drop in $F_{\text{cz}}^{\text{coh}}$ from 99.6% at zero applied current to 98.4% at -0.33 Φ_0 (dashed line on Fig. 6.12(b)).[7]

The cause of reduced qubit coherence with increasing coupler bias is still under investigation. A potential cause is the increased sensitivity of the coupler to flux noise at -0.33 Φ_0, increasing its dephasing [168–171]. In turn, this degrades the coherence of the two qubits, which are dispersively coupled to the coupler.

6.7 Future work

Despite the progress made in understanding crosstalk in our quantum processor, further work is needed to fully characterize crosstalk, its effects, and contributing mechanisms. Here, I outline a list of experiments and studies that may enhance our understanding and aid in the mitigation of crosstalk in future devices:

- Expand the study of single-qubit gate fidelities to extract an error matrix, similar to that in Fig. 6.10, on all 25 qubits. For such an experiment, the duration of each measurement must decrease significantly to avoid temporal drifts in the calibration parameters. Active reset of the qubit state [172], instead of waiting for the qubit state to decay to $|0\rangle$, improves the characterization throughput.

[7] The coupler's -0.33 Φ_0 bias point corresponds to an applied current of -720 μA to the coupler's z line. At zero applied current, the flux bias of the coupler is $-0.109\Phi_0$, due to a pre-existing offset in the flux bias.

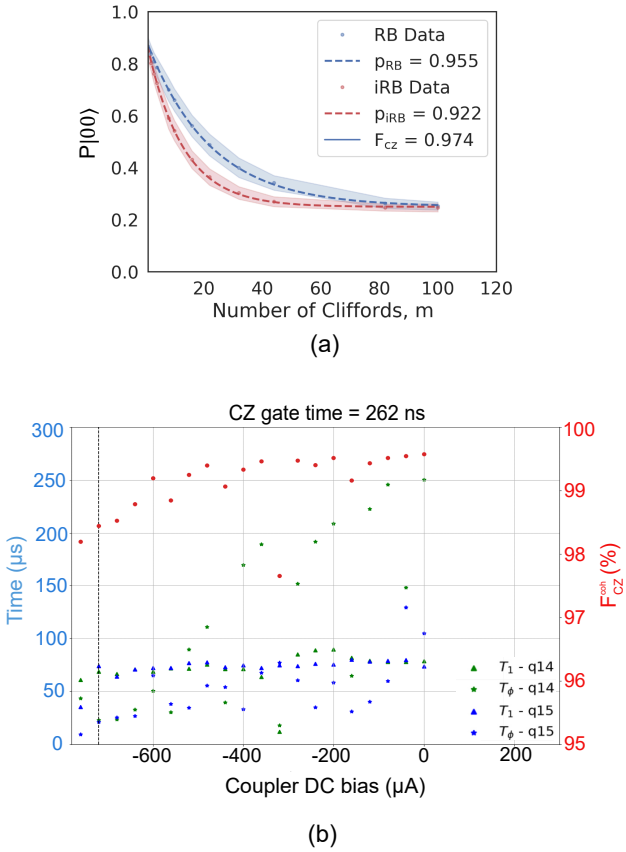


Figure 6.12: Fidelity of the CZ gate. (a) Two-qubit gate RB and iRB on qubits 14 (q14) and 15 (q15), where the population of the $|00\rangle$ state, P$|00\rangle$, decays exponentially with the number of applied Cliffords, m, and fit to Eq. 6.1. The CZ gate fidelity F_{cz} is calculated using Eq. 6.3. (b) Coherence limit of the CZ gate fidelity, $F_{\mathrm{cz}}^{\mathrm{coh}}$, based on T_1 and T_ϕ (left y-axis) of the two qubits at different coupler DC-bias. The dashed line represents the DC-bias point where the CZ gate is implemented, corresponding to an applied flux of $-0.33\Phi_0$.

- Implement purity benchmarking (PB) [173], to decouple coherent and incoherent errors, providing an insight into the nature of crosstalk and the possibility of compensating for its effect [158, 174, 175].
- Study the correlation between the strength of the crosstalk effect and the detuning between the source and victim qubits' frequencies.
- Integrate active crosstalk mitigation into the calibration of single-qubit gates to minimize crosstalk-induced errors [31].
- Expand the crosstalk investigation to two-qubit gates, analyzing both DC- and AC-flux crosstalk effects.

CHAPTER 7

Summary and Contributions

To summarize the work presented in this dissertation, we invented the PICT Josephson junction fabrication process for superconducting qubits, cutting fabrication time and cost, reducing the probability of error, while protecting the sensitive junction area from additional processing. PICT junctions maintained high resistance reproducibility and qubit quality factors. By fabricating transmon qubits with larger junction areas, we demonstrated highly reproducible qubit frequencies with a standard deviation of 40 MHz (i.e., 1%) on a planar chip. These qubits maintained high quality factors exceeding 2 million. We conducted an extensive study on the impact of different fabrication steps on the loss tangent of our coplanar-waveguide resonators. By increasing the ground plane thickness to 300 and 500 nm, we demonstrated an average qubit quality factor of 3 million, with the best qubit achieving an average T_1 of 270 μs.

To mitigate crosstalk on quantum processors, we proposed a scalable frequency allocation scheme that maximizes detuning between neighboring qubits. We measured a crosstalk level between qubits on our first-generation 25-qubit QPU, which is on par with the lowest reported values in the literature. However, we concluded that further reduction of crosstalk is essential to scale

QPUs beyond 100 qubits with low gate errors. We conducted an extensive study of parameter targeting on our second-generation QPU, where the standard deviation in qubit frequency was 1.9%, and found that uniformity of the chip-to-chip spacing in the flip-chip architecture is crucial for better accuracy in parameter targeting.

By characterizing single-qubit gate fidelities on the same QPU, we found that for the two pairs of qubits, impacted the most by crosstalk, the average gate error increased by $\sim$0.35%. Such an effect was only observed between next-nearest-neighbor qubits, highlighting the efficiency of our frequency allocation scheme in avoiding collisions between nearest-neighbor qubits, while emphasizing the importance of suppressing crosstalk below the current level. However, our crosstalk characterization requires further experimentation and analysis. We finally demonstrated a CZ gate on two qubits of the same QPU, with a gate fidelity of 97.4%, mostly limited by the low coherence when the coupler is biased close to the frequency of the two qubits.

In light of these conclusions, I would like to specify my contribution to this work. While most of the content in Chapters 2, 3, and 4 is based on already published papers, the work presented in Chapters 5 and 6 is yet to be published.

For the published work, the following were my contributions:

- **Paper A:** I co-invented the PICT process, developed the necessary infrastructure for the automated junction resistance characterization, designed the devices, fabricated and characterized thousands of Josephson junctions and dozens of transmon qubits, and devised the model to quantify the contributions of junction area fluctuations and oxide thickness variations to resistance reproducibility. Additionally, I authored the first draft of the paper.

- **Paper B:** I co-conceptualized the idea of making larger junctions for better qubit frequency reproducibility, developed the fabrication recipe, fabricated and characterized thousands of junctions and the transmon qubits investigated, calibrated and conducted the TLS-spectroscopy experiment, and performed the analysis. I contributed to the frequency allocation scheme proposed by our device theory colleagues in the Applied Quantum Physics division, by optimizing the frequency selection. I developed the Monte Carlo simulation, illustrating the impact of the frequency allocation and qubit frequency reproducibility on the perfor-

mance of QPUs. I also authored the first draft of the paper.

- **Paper C:** I contributed with some qubit characterization and data analysis. I also performed some fabrication steps and characterized surface roughness and aluminum critical temperature.
- **Paper D:** I collaborated in the development of the study, discussion of the fabrication process, and data analysis of SIMS, XPS, and TEM measurements.
- **Paper E:** I fabricated the 5-qubit chip, contributed to the design to achieve the target qubit frequencies, and co-developed the fabrication process of airbridges, which was essential for signal routing and suppression of crosstalk.
- **Papers F and G:** I contributed with process development and numerous discussions on fabrication and design.

For the unpublished work in Chapter 5 and Chapter 6, I did the following:

- defined the parameters of the second-generation QPU.
- designed the QPU and simulated the qubit chip, using a design code base, previously developed by colleagues for our first-generation QPU.
- fabricated the wiring layer and Josephson junctions of the QPU.
- characterized the resistance of the test junctions on the qubit wafer and selected the module to be measured.
- ran all the measurements presented in Chapter 6 and performed all the analysis on parameter targeting and crosstalk. For these measurements, I used `tergite automatic calibration` [148], developed by our software team and further enhanced with contributions from several colleagues in our research group.
- contributed to the `tergite automatic calibration` by improving and developing some calibration nodes of single- and two-qubit gates.

References

[1] Y. Cao, J. Romero, and A. Aspuru-Guzik, Potential of quantum computing for drug discovery, IBM Journal of Research and Development 62, 6:1–6:20 (2018).

[2] I. M. Georgescu, S. Ashhab, and F. Nori, Quantum simulation, Reviews of Modern Physics 86, 153–185 (2014).

[3] M. J. Abraham et al., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1-2, 19–25 (2015).

[4] N. Yao et al., Applying Classical, *Ab Initio* , and Machine-Learning Molecular Dynamics Simulations to the Liquid Electrolyte for Rechargeable Batteries, Chemical Reviews 122, 10970–11021 (2022).

[5] A. Delgado et al., Simulating key properties of lithium-ion batteries with a fault-tolerant quantum computer, Physical Review A 106, 032428 (2022).

[6] A. Giani and Z. Eldredge, Quantum Computing Opportunities in Renewable Energy, SN Computer Science 2, 393 (2021).

[7] Y. Cao et al., Quantum Chemistry in the Age of Quantum Computing, Chemical Reviews 119, 10856–10915 (2019).

[8] R. P. Feynman, Simulating physics with computers, International Journal of Theoretical Physics 21, 467–488 (1982).

[9] M. Hilbert and P. López, The World's Technological Capacity to Store, Communicate, and Compute Information, Science 332, 60–65 (2011).

[10] IDC, & Statista., Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025 (in zettabytes) [Graph], , URL: `https://www.statista.com/statistics/871513/worldwide-data-created/` (2021).

[11] C. Monroe et al., Demonstration of a Fundamental Quantum Logic Gate, Physical Review Letters 75, 4714–4717 (1995).

[12] D. G. Cory, A. F. Fahmy, and T. F. Havel, Ensemble quantum computing by NMR spectroscopy, Proceedings of the National Academy of Sciences 94, 1634–1639 (1997).

[13] D. G. Cory, M. D. Price, and T. F. Havel, Nuclear magnetic resonance spectroscopy: An experimentally accessible paradigm for quantum computing, Physica D: Nonlinear Phenomena 120, 82–101 (1998).

[14] J. A. Jones and M. Mosca, Implementation of a quantum algorithm on a nuclear magnetic resonance quantum computer, The Journal of Chemical Physics 109, 1648–1653 (1998).

[15] I. L. Chuang, N. Gershenfeld, and M. Kubinec, Experimental Implementation of Fast Quantum Searching, Physical Review Letters 80, 3408–3411 (1998).

[16] N. A. Gershenfeld and I. L. Chuang, Bulk Spin-Resonance Quantum Computation, Science 275, 350–356 (1997).

[17] Y. Nakamura, Y. A. Pashkin, and J. S. Tsai, Coherent control of macroscopic quantum states in a single-Cooper-pair box, Nature 398, 786–788 (1999).

[18] C. D. Bruzewicz et al., Trapped-ion quantum computing: Progress and challenges, Applied Physics Reviews 6, 021314 (2019).

[19] F. P. García De Arquer et al., Semiconductor quantum dots: Technological progress and future challenges, Science 373, eaaz8541 (2021).

[20] S. Slussarenko and G. J. Pryde, Photonic quantum information processing: A concise review, Applied Physics Reviews 6, 041303 (2019).

[21] C. Gidney and M. Ekerå, How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits, Quantum 5, 433 (2021).

[22] E. T. Campbell, B. M. Terhal, and C. Vuillot, Roads towards fault-tolerant universal quantum computation, Nature 549, 172–179 (2017).

[23] P. Shor, Algorithms for quantum computation: discrete logarithms and factoring, Proceedings 35th Annual Symposium on Foundations of Computer Science, 124–134 (1994).

[24] L. K. Grover, A fast quantum mechanical algorithm for database search, Proceedings of the twenty-eighth annual ACM symposium on Theory of computing - STOC '96, 212–219 (1996).

[25] D. Coppersmith, An approximate Fourier transform useful in quantum factoring, arXiv:quant-ph/0201067 (2002).

[26] J. Preskill, Quantum Computing in the NISQ era and beyond, Quantum 2, 79 (2018).

[27] E. Farhi, J. Goldstone, and S. Gutmann, A Quantum Approximate Optimization Algorithm, arXiv:1411.4028 (2014).

[28] A. Peruzzo et al., A variational eigenvalue solver on a photonic quantum processor, Nature Communications 5, 4213 (2014).

[29] J.-F. Bobier et al., The Long-Term Forecast for Quantum Computing Still Looks Bright, Boston Consulting Group (2024).

[30] J. Roffe, Quantum error correction: an introductory guide, Contemporary Physics 60, 226–245 (2019).

[31] Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, Nature 614, 676–681 (2023).

[32] R. Acharya et al., Quantum error correction below the surface code threshold, arXiv:2408.13687 [quant-ph] (2024).

[33] J. M. Kreikebaum et al., Improving wafer-scale Josephson junction resistance variation in superconducting quantum coherent circuits, Superconductor Science and Technology 33, 06LT02 (2020).

[34] D. O. Moskalev et al., Optimization of shadow evaporation and oxidation for reproducible quantum Josephson junction circuits, Scientific Reports 13, 4174 (2023).

[35] N. Muthusubramanian et al., Wafer-scale uniformity of Dolan-bridge and bridgeless Manhattan-style Josephson junctions for superconducting quantum processors, Quantum Science and Technology 9, 025006 (2024).

[36] J. Verjauw et al., Path toward manufacturable superconducting qubits with relaxation times exceeding 0.1 ms, npj Quantum Information 8, 93 (2022).

[37] T. Takahashi et al., Uniformity improvement of Josephson-junction resistance by considering sidewall deposition during shadow evaporation for large-scale integration of qubits, Japanese Journal of Applied Physics 62, SC1002 (2023).

[38] J. Van Damme et al., Advanced CMOS manufacturing of superconducting qubits on 300 mm wafers, Nature 634, 74–79 (2024).

[39] A. A. Pishchimova et al., Improving Josephson junction reproducibility for superconducting quantum circuits: junction area fluctuation, Scientific Reports 13, 6772 (2023).

[40] J. B. Hertzberg et al., Laser-annealing Josephson junctions for yielding scaled-up superconducting quantum processors, npj Quantum Information 7, 129 (2021).

[41] H. Kim et al., Effects of laser-annealing on fixed-frequency superconducting qubits, Applied Physics Letters 121, 142601 (2022).

[42] D. P. Pappas et al., Alternating-bias assisted annealing of amorphous oxide tunnel junctions, Communications Materials 5, 150 (2024).

[43] X. Wang et al., Precision frequency tuning of tunable transmon qubits using alternating-bias assisted annealing, arXiv:2407.06425 [quant-ph] (2024).

[44] M. Toselli. Post-fabrication Frequency Tuning in Superconducting Transmon Qubits. MA thesis. Gothenburg, Sweden: Chalmers University of Technology, 2024.

[45] B. Josephson, Possible new effects in superconductive tunnelling, Physics Letters 1, 251–253 (1962).

[46] M. Tinkham. Introduction to Superconductivity. 2nd. Includes comprehensive treatment of superconducting circuits, quantum effects, and Josephson junctions. Mineola, NY: Dover Publications, 1996.

[47] J. M. Martinis. Course 13 Superconducting qubits and the physics of Josephson junctions. en. In: Les Houches. 79. Elsevier, 2004, 487–520.

[48] V. Ambegaokar and A. Baratoff, Tunneling Between Superconductors, Physical Review Letters 10, 486–489 (1963).

[49] D. P. DiVincenzo, The Physical Implementation of Quantum Computation, Fortschritte der Physik 48, 771–783 (2000).

[50] J. Koch et al., Charge-insensitive qubit design derived from the Cooper pair box, Physical Review A 76, 042319 (2007).

[51] A. Blais et al., Circuit quantum electrodynamics, Reviews of Modern Physics 93, 025005 (2021).

[52] G. J. Dolan, Offset masks for lift-off photoprocessing, Applied Physics Letters 31, 337–339 (1977).

[53] A. Potts et al., Novel fabrication methods for submicrometer Josephson junction qubits, Journal of Materials Science: Materials in Electronics 12, 289–293 (2001).

[54] F. Lecocq et al., Junction fabrication by shadow evaporation without a suspended bridge, Nanotechnology 22, 315302 (2011).

[55] X. Wu et al., Overlap junctions for high coherence superconducting qubits, Applied Physics Letters 111, 032602 (2017).

[56] D. Wan et al., Fabrication and room temperature characterization of trilayer junctions for the development of superconducting qubits on 300 mm wafers, Japanese Journal of Applied Physics 60, SBBI04 (2021).

[57] A. Dunsworth et al., Characterization and reduction of capacitive loss induced by sub-micron Josephson junction fabrication in superconducting qubits, Applied Physics Letters 111, 022601 (2017).

[58] M. Corp, Nano PMMA and copolymer, datasheet.

[59] I. M. Pop et al., Fabrication of stable and reproducible submicron tunnel junctions, Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: Materials, Processing, Measurement, and Phenomena 30, 010607 (2012).

[60] A. Osman. Reliability and reproducibility of Josephson junction fabrication. en. MA thesis. Chalmers University of Technology, 2019.

[61] S. Shirai et al., All-Microwave Manipulation of Superconducting Qubits with a Fixed-Frequency Transmon Coupler, Physical Review Letters 130, 260601 (2023).

[62] Y. Zheng et al., Fabrication of Al/AlOx/Al junctions with high uniformity and stability on sapphire substrates, Scientific Reports 13, 11874 (2023).

[63] F. J. Massey, The Kolmogorov-Smirnov Test for Goodness of Fit, Journal of the American Statistical Association 46, 68–78 (1951).

[64] L. J. Zeng et al., Direct observation of the thickness distribution of ultra thin AlO_x barriers in Al/AlO_x/Al Josephson junctions, Journal of Physics D: Applied Physics 48, 395308 (2015).

[65] P. J. Koppinen, L. M. Väistö, and I. J. Maasilta, Complete stabilization and improvement of the characteristics of tunnel junctions by thermal annealing, Applied Physics Letters 90, 053503 (2007).

[66] J. K. Julin et al., Stability, sub-gap current, 1/f-noise, and elemental depth profiling of annealed Al:Mn-AlO_x-Al normal metal-insulator-superconducting tunnel junctions, AIP Advances 6, 125026 (2016).

[67] J. G. Simmons, Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film, Journal of Applied Physics 34, 1793–1803 (1963).

[68] G. Catelani et al., Decoherence of superconducting qubits caused by quasiparticle tunneling, Physical Review B 86, 184514 (2012).

[69] G. Catelani et al., Quasiparticle Relaxation of Superconducting Qubits in the Presence of Flux, Physical Review Letters 106, 077002 (2011).

[70] J. M. Martinis, M. Ansmann, and J. Aumentado, Energy Decay in Superconducting Josephson-Junction Qubits from Nonequilibrium Quasiparticle Excitations, Physical Review Letters 103, 097002 (2009).

[71] C. Wang et al., Measurement and control of quasiparticle dynamics in a superconducting qubit, Nature Communications 5, 5836 (2014).

[72] L. Sun et al., Measurements of Quasiparticle Tunneling Dynamics in a Band-Gap-Engineered Transmon Qubit, Physical Review Letters 108, 230509 (2012).

[73] D. Ristè et al., Millisecond charge-parity fluctuations and induced decoherence in a superconducting transmon qubit, Nature Communications 4, 1913 (2013).

[74] J. Aumentado, G. Catelani, and K. Serniak, Quasiparticle poisoning in superconducting quantum computers, Physics Today 76, 34–39 (2023).

[75] J. M. Martinis et al., Decoherence in Josephson Qubits from Dielectric Loss, Physical Review Letters 95, 210503 (2005).

[76] L. Gordon et al., Hydrogen bonds in Al2O3 as dissipative two-level systems in superconducting qubits, Scientific Reports 4, 7590 (2014).

[77] A. M. Holder et al., Bulk and Surface Tunneling Hydrogen Defects in Alumina, Physical Review Letters 111, 065901 (2013).

[78] T. C. DuBois et al., Delocalized Oxygen as the Origin of Two-Level Defects in Josephson Junctions, Physical Review Letters 110, 077002 (2013).

[79] K. Agarwal et al., Polaronic model of two-level systems in amorphous solids, Physical Review B 87, 144201 (2013).

[80] R. H. Koch, D. P. DiVincenzo, and J. Clarke, Model for 1 / f Flux Noise in SQUIDs and Qubits, Physical Review Letters 98, 267003 (2007).

[81] E. Paladino et al., 1/f noise: Implications for solid-state quantum information, Reviews of Modern Physics 86, 361–418 (2014).

[82] S. M. Anton et al., Magnetic Flux Noise in dc SQUIDs: Temperature and Geometry Dependence, Physical Review Letters 110, 147002 (2013).

[83] A. Jayaraman et al., Loss and decoherence in superconducting circuits on silicon: Insights from electron spin resonance, Physical Review Applied 22, 014030 (2024).

[84] R. Barends et al., Coherent Josephson Qubit Suitable for Scalable Quantum Integrated Circuits, Physical Review Letters 111, 080502 (2013).

[85] J. Lisenfeld et al., Electric field spectroscopy of material defects in transmon qubits, npj Quantum Information 5, 105 (2019).

[86] J. Lisenfeld et al., Measuring the Temperature Dependence of Individual Two-Level Systems by Direct Coherent Control, Physical Review Letters 105, 230504 (2010).

[87] S. E. De Graaf et al., Direct Identification of Dilute Surface Spins on Al 2 O 3 : Origin of Flux Noise in Quantum Circuits, Physical Review Letters 118, 057703 (2017).

[88] S. E. De Graaf et al., Suppression of low-frequency charge noise in superconducting resonators by surface spin desorption, Nature Communications 9, 1143 (2018).

[89] T. Lindström et al., Properties of superconducting planar resonators at millikelvin temperatures, Physical Review B 80, 132501 (2009).

[90] S. E. De Graaf et al., Two-level systems in superconducting quantum devices due to trapped quasiparticles, Science Advances 6, eabc5055 (2020).

[91] J. Burnett, L. Faoro, and T. Lindström, Analysis of high quality superconducting resonators: consequences for TLS properties in amorphous oxides, Superconductor Science and Technology 29, 044008 (2016).

[92] J. J. Burnett et al., Decoherence benchmarking of superconducting qubits, npj Quantum Information 5, 54 (2019).

[93] E. A. Sete, J. M. Martinis, and A. N. Korotkov, Quantum theory of a bandpass Purcell filter for qubit readout, Physical Review A 92, 012325 (2015).

[94] E. Jeffrey et al., Fast Accurate State Measurement with Superconducting Qubits, Physical Review Letters 112, 190504 (2014).

[95] Y. Sunada et al., Fast Readout and Reset of a Superconducting Qubit Coupled to a Resonator with an Intrinsic Purcell Filter, Physical Review Applied 17, 044016 (2022).

[96] H. Yan et al., Broadband bandpass Purcell filter for circuit quantum electrodynamics, Applied Physics Letters 123, 134001 (2023).

[97] N. T. Bronn et al., Reducing Spontaneous Emission in Circuit Quantum Electrodynamics by a Combined Readout/Filter Technique, IEEE Transactions on Applied Superconductivity 25, 1–10 (2015).

[98] L. Chen et al., Transmon qubit readout fidelity at the threshold for quantum error correction without a quantum-limited amplifier, npj Quantum Information 9, 26 (2023).

[99] D. A. Rower et al., Evolution of 1/f Flux Noise in Superconducting Qubits with Weak Magnetic Fields, Physical Review Letters 130, 220602 (2023).

[100] P. Krantz. Investigation of Transmon Qubit Designs. en. MA thesis. Chalmers University of Technology, 2010.

[101] I. Siddiqi, Engineering high-coherence superconducting qubits, Nature Reviews Materials 6, 875–891 (2021).

[102] L. Cardani et al., Disentangling the sources of ionizing radiation in superconducting qubits, The European Physical Journal C 83, 94 (2023).

[103] A. P. Vepsäläinen et al., Impact of ionizing radiation on superconducting qubit coherence, Nature 584, 551–556 (2020).

[104] T. Thorbeck et al., Two-Level-System Dynamics in a Superconducting Qubit Due to Background Ionizing Radiation, PRX Quantum 4, 020356 (2023).

[105] W. A. Phillips, Tunneling states in amorphous solids, Journal of Low Temperature Physics 7, 351–360 (1972).

[106] P. W. Anderson, B. I. Halperin, and C. M. Varma, Anomalous low-temperature thermal properties of glasses and spin glasses, Philosophical Magazine 25, 1–9 (1972).

[107] A. Bilmes et al., Probing defect densities at the edges and inside Josephson junctions of superconducting qubits, npj Quantum Information 8, 24 (2022).

[108] L. Chayanun et al., Characterization of process-related interfacial dielectric loss in aluminum-on-silicon by resonator microwave measurements, materials analysis, and imaging, APL Quantum 1, 026115 (2024).

[109] S. Probst et al., Efficient and robust analysis of complex scattering data under noise in microwave resonators, Review of Scientific Instruments 86, 024706 (2015).

[110] J. Gao. The Physics of Superconducting Microwave Resonators. en. Medium: PDF Version Number: Final. PhD thesis. California Institute of Technology, June 2008.

[111] J. Burnett et al., Low-Loss Superconducting Nanowire Circuits Using a Neon Focused Ion Beam, Physical Review Applied 8, 014039 (2017).

[112] P. Macha et al., Losses in coplanar waveguide resonators at millikelvin temperatures, Applied Physics Letters 96, 062503 (2010).

[113] D. P. Pappas et al., Two Level System Loss in Superconducting Microwave Resonators, IEEE Transactions on Applied Superconductivity 21, 871–874 (2011).

[114] C. Wang et al., Surface participation and dielectric loss in superconducting qubits, Applied Physics Letters 107, 162601 (2015).

[115] S. M. Goodnick et al., Effects of a thin SiO $_2$ layer on the formation of metal–silicon contacts, Journal of Vacuum Science and Technology 18, 949–954 (1981).

[116] F. Dadabhai et al., Reduction of silicon dioxide by aluminum in metal-oxide-semiconductor structures, Journal of Applied Physics 80, 6505–6509 (1996).

[117] J. Biznárová et al., Mitigation of interfacial dielectric loss in aluminum-on-silicon superconducting qubits, npj Quantum Information 10, 78 (2024).

[118] J. Jaseliunaite and A. Galdikas, Kinetic Modeling of Grain Boundary Diffusion: The Influence of Grain Size and Surface Processes, Materials 13, 1051 (2020).

[119] M. Kjaergaard et al., Superconducting Qubits: Current State of Play, Annual Review of Condensed Matter Physics 11, 369–395 (2020).

[120] D. C. McKay et al., Universal Gate for Fixed-Frequency Qubits via a Tunable Bus, Physical Review Applied 6, 064007 (2016).

[121] S. A. Caldwell et al., Parametrically Activated Entangling Gates Using Transmon Qubits, Physical Review Applied 10, 034050 (2018).

[122] M. Reagor et al., Demonstration of universal parametric entangling gates on a multi-qubit lattice, Science Advances 4, eaao3603 (2018).

[123] A. Bengtsson. Quantum information processing with tunable and low-loss superconducting circuits. en. PhD thesis. Sweden: Chalmers University of Technology, 2020.

[124] A. Bengtsson et al., Improved Success Probability with Greater Circuit Depth for the Quantum Approximate Optimization Algorithm, Physical Review Applied 14, 034010 (2020).

[125] C. Warren. Benchmarking and Metrology of Scaled Superconducting Quantum Processors. Doctoral thesis. Gothenburg: Chalmers University of Technology, 2024.

[126] G. Ponchak, J. Papapolymerou, and M. Tentzeris, Excitation of coupled slotline mode in finite-ground CPW with unequal ground-plane widths, IEEE Transactions on Microwave Theory and Techniques 53, 713–717 (2005).

[127] Z. Chen et al., Fabrication and characterization of aluminum airbridges for superconducting microwave circuits, Applied Physics Letters 104, 052602 (2014).

[128] D. Rosenberg et al., 3D integrated superconducting qubits, npj Quantum Information 3, 42 (2017).

[129] B. Foxen et al., Qubit compatible superconducting interconnects, Quantum Science and Technology 3, 014005 (2018).

[130] A. Gold et al., Entanglement across separate silicon dies in a modular superconducting qubit device, npj Quantum Information 7, 142 (2021).

[131] C. R. Conner et al., Superconducting qubits in a flip-chip architecture, Applied Physics Letters 118, 232602 (2021).

[132] S. Kosen et al., Building blocks of a flip-chip integrated superconducting quantum processor, Quantum Science and Technology 7, 035018 (2022).

[133] R. G. Gordon et al., Highly Conformal Diffusion Barriers of Amorphous Niobium Nitride, MRS Proceedings 564, 335 (1999).

[134] S. Kosen et al., Signal Crosstalk in a Flip-Chip Quantum Processor, PRX Quantum 5, 030350 (2024).

[135] P. A. Spring et al., High coherence and low cross-talk in a tileable 3D integrated superconducting circuit architecture, Science Advances 8, eabl6698 (2022).

[136] S. Krinner et al., Realizing repeated quantum error correction in a distance-three surface code, Nature 605, 669–674 (2022).

[137] X. Dai et al., Calibration of Flux Crosstalk in Large-Scale Flux-Tunable Superconducting Quantum Circuits, PRX Quantum 2, 040313 (2021).

[138] C. N. Barrett et al., Learning-Based Calibration of Flux Crosstalk in Transmon Qubit Arrays, Physical Review Applied 20, 024070 (2023).

[139] Z. K. Minev et al., Qiskit Metal: An Open-Source Framework for Quantum Device Design & Analysis, (2021).

[140] Ansys ®Electronics Desktop, Release 2023 R.

[141] R. Manenti and M. Motta. Quantum information science. eng. Oxford: Oxford University Press, 2023.

[142] H.-X. Li et al., Experimentally Verified, Fast Analytic, and Numerical Design of Superconducting Resonators in Flip-Chip Architectures, IEEE Transactions on Quantum Engineering 4, 1–12 (2023).

[143] S. Krinner et al., Engineering cryogenic setups for 100-qubit scale superconducting circuit systems, EPJ Quantum Technology 6, 2 (2019).

[144] J. M. Kreikebaum et al., Optimization of infrared and magnetic shielding of superconducting TiN and Al coplanar microwave resonators, Superconductor Science and Technology 29, 104002 (2016).

[145] A. Lukashenko and A. V. Ustinov, Improved powder filters for qubit measurements, Review of Scientific Instruments 79, 014701 (2008).

[146] K. L. Geerlings. Improving Coherence of Superconducting Qubits and Resonators. PhD thesis. Jan. 2013.

[147] F. Motzoi et al., Simple Pulses for Elimination of Leakage in Weakly Nonlinear Qubits, Physical Review Letters 103, 110501 (2009).

[148] CNL, Tergite Autocalibration, Apache License 2.0 (2024).

[149] M. S. Khalil et al., An analysis method for asymmetric resonator transmission applied to superconducting devices, Journal of Applied Physics 111, 054510 (2012).

[150] E. L. Hahn, Spin Echoes, Physical Review 80, 580–594 (1950).

[151] J. Bylander et al., Noise spectroscopy through dynamical decoupling with a superconducting flux qubit, Nature Physics 7, 565–570 (2011).

[152] J. S. Kelly. Fault-tolerant superconducting qubits. PhD thesis. University of California Santa Barbara, 2015.

[153] J. Kelly et al., Optimal Quantum Control Using Randomized Benchmarking, Physical Review Letters 112, 240504 (2014).

[154] P. Parrado-Rodríguez et al., Crosstalk Suppression for Fault-tolerant Quantum Error Correction with Trapped Ions, Quantum 5, 487 (2021).

[155] S. Sheldon et al., Characterizing errors on qubit operations via iterative randomized benchmarking, Physical Review A 93, 012301 (2016).

[156] F. Arute et al., Quantum supremacy using a programmable superconducting processor, Nature 574, 505–510 (2019).

[157] M. Sarovar et al., Detecting crosstalk errors in quantum information processors, Quantum 4, 321 (2020).

[158] K. Rudinger et al., Experimental Characterization of Crosstalk Errors with Simultaneous Gate Set Tomography, PRX Quantum 2, 040338 (2021).

[159] S. Krinner et al., Benchmarking Coherent Errors in Controlled-Phase Gates due to Spectator Qubits, Physical Review Applied 14, 024042 (2020).

[160] J. Emerson, R. Alicki, and K. Życzkowski, Scalable noise estimation with random unitary operators, Journal of Optics B: Quantum and Semiclassical Optics 7, S347–S352 (2005).

[161] C. Dankert et al., Exact and approximate unitary 2-designs and their application to fidelity estimation, Physical Review A 80, 012304 (2009).

[162] A. D. Córcoles et al., Process verification of two-qubit quantum gates by randomized benchmarking, Physical Review A 87, 030301 (2013).

[163] R. Barends et al., Superconducting quantum circuits at the surface code threshold for fault tolerance, Nature 508, 500–503 (2014).

[164] M. A. Nielsen and I. L. Chuang. Quantum Computation and Quantum Information: 10th Anniversary Edition. 1. Cambridge University Press, June 2012.

[165] E. Magesan et al., Efficient Measurement of Quantum Gate Error by Interleaved Randomized Benchmarking, Physical Review Letters 109, 080505 (2012).

[166] M. Ganzhorn et al., Benchmarking the noise sensitivity of different parametric two-qubit gates in a single superconducting quantum computing platform, Physical Review Research 2, 033447 (2020).

[167] T. Abad et al., Impact of decoherence on the fidelity of quantum gates leaving the computational subspace, arXiv:2302.13885 [quant-ph] (2024).

[168] S. M. Anton et al., Pure dephasing in flux qubits due to flux noise with spectral density scaling as $1/f^{\alpha}$, Physical Review B 85, 224505 (2012).

[169] F. Yoshihara et al., Decoherence of Flux Qubits due to 1/f Flux Noise, Physical Review Letters 97, 167001 (2006).

[170] G. Ithier et al., Decoherence in a superconducting quantum bit circuit, Physical Review B 72, 134519 (2005).

[171] J. M. Martinis et al., Decoherence of a superconducting qubit due to bias noise, Physical Review B 67, 094510 (2003).

[172] L. Chen et al., Fast unconditional reset and leakage reduction in fixed-frequency transmon qubits, arXiv:2409.16748v2 [quant-ph] (2024).

[173] J. Wallman et al., Estimating the coherence of noise, New Journal of Physics 17, 113020 (2015).

[174] V. Tripathi et al., Suppression of Crosstalk in Superconducting Qubits Using Dynamical Decoupling, Physical Review Applied 18, 024068 (2022).

[175] P. Murali et al., Software Mitigation of Crosstalk on Noisy Intermediate-Scale Quantum Computers, Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems, 1001–1016 (2020).