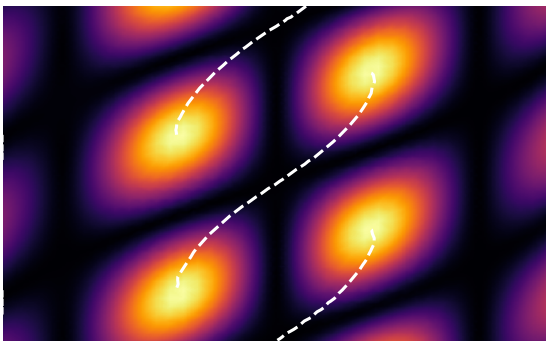
TÉCNICO LISBOA



Automatic characterization and calibration of a superconducting quantum processor capable of error correction

Victor Negîrneac

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisors: Prof. Dr. Leonardo DiCarlo
Prof. Dr. Pedro José Gonçalves Ribeiro

Examination Committee

Chairperson: Prof. Dr. Pedro Miguel Félix Brogueira
Supervisor: Prof. Dr. Leonardo DiCarlo
Member of the Committee: Prof. Dr. João Carlos Carvalho de Sá Seixas

January 2021

Dedicated to my brother

Acknowledgments

This work would have not been possible without the direct and indirect help and support of many people and institutions. I am forever grateful, thank you all! In the next paragraphs I do my best to honor the crossings of our paths. I wish I could make justice to all of you in these few lines.

Professor **Leo**, first of all, thank you for your teaching efforts! I am very glad I took the course on the Fundamentals of Quantum Information two years ago, where I met you for the first time. The passion and excitement you brought to the class inspired me to join the research efforts in your lab. I am very glad I had a chance to learn from a great scientist and become a better one myself. I have assimilated from you a good deal of scientific writing, rigor, high standards and integrity, among other learnings. At the same time, I was very pleased to see you becoming a greater group leader during my internship, your efforts do not pass by unnoticed and are much appreciated! A huge thank you for generously accepting the proposal for what become my first full time job. And of course, I could not be more grateful for the substantial efforts and resources that you allocated to the exploration of the Sudden Net-Zero! I wished my life circumstances allowed me to stay around for longer. Looking forward to the dance of a logical qubit in your lab! Do not let that guitar accumulate too much dust!

Professor **Pedro**, a big thank you for co-supervising my journey, revising my writing as well as encouraging my undertakings. I wish you success in the teaching activities and many breakthroughs in your research!

I would like to thank **the committee members** for taking the time to review my thesis. I did my best to deliver a self-contained, light, and yet rigorous manuscript. I hope you enjoy reading it.

I thank the MEFT coordinator, **Ilídio Lopes**, for authorizing my internship and recommending a great co-supervisor.

This work was developed with the support of the **DiCarlo Lab** (part of QuTech and the Kavli Institute of Nanoscience); and the **Erasmus+ programme** of the European Union. A big thank you to all the people that made this possible.

I was fortunate to develop my work along a numerous and diverse team. Dear lab mates and co-workers, thank you all for the warm welcoming and the support provided throughout my internship. I wished my personal deadlines and commitments allowed me to join you more often for coffee talks and social activities. As always, keep working smart the hardest you can and, remember, positive focus ;)

Hany, thank you for sharing this journey with me, for teaching me transmon basics, for the heads up to focus on the progress that was necessary in the moment, for the hard work and dedication towards our paper, for the company in late evenings, and for the kind words in tougher times. I hope you will get to work on the things you enjoy the most during your PhD. May well-targeted qubits come your way! **Ramiro**, it has been a lot of fun to share the Lamborghini ride with you. You tough me many things: the transmons in theory and in practice, the black magic of the HDAWGs, the inner workings of our setup and how to collect the easter eggs of PycQED. Thank you for the patience to explain so many things, so many times, always finding that extra minute in your "minutes bag" even when having a lot on your own plate. Many thanks for listening to and appreciating my ideas, encouraging me to explore them

and, at the right moment, giving me a hand to convey them in the best way possible! And thank you for introducing me to the Qblox folks, it has been a great joy and support for my life to join them during the last thesis-writing months. My best wishes for your future endeavors. I look forward to meet you again on a dance floor! **Adriaan**, I am very glad that you were still around when I joined the group. Thank you for the initial guidance, the time dedicated to ensure the transfer of essential know-how, and useful discussions along the way. One of those deserves a special mention. The coffee break where we both agreed it was not the safe path for me to try to overcome the core flaws of the Net-Zero calibration; and I did it anyway... As this work testifies now, it was worth digging deeper, thank you for sparking my interest and, of course, for all your previous work I was able to build upon. I wish you to thrive in the quantum computing ecosystem and cannot wait to see Quantify, the PycQED successor, empowering the measurements of kilo-qubit devices! **Francesco**, thank you for your work and support provided on the CZ simulations. Your flexibility and positive attitude throughout the SNZ challenges has been amazing. Miguel and Jorge, my Portuguese fellows we are slowly taking over the lab..., you made me feel closer to home. **Miguel**, it has been great to see you becoming proficient at calibrations and putting more structure to them. Keep Quantum Inspire accountable for its name with these new features! **Jorge**, could not help myself but still some of these beautiful analysis figures! Thank you for the useful physics discussions and keeping our readout sharp. I am pleased to know our transmons are in passionate hands! **Berend**, it has been nice to share the journey with you. I am glad your talent and skills will be invested in a PhD. Good luck in the research and the sports competitions! **Tumi**, these croissants at the early morning meetings were the best one could ask for. Keep the positive vibes up! **Thijs**, arguably the purest physicist among my lab mates, thank you for the experimental assistance on our setup. Take those nanowires to the next level! Dear fabrication team, your craft is at the core of our research. Many thanks for the continuous efforts to bring us the finest qubits! **Nandini**, it would have been amazing to share the PhD path with you. I wish you the best of luck in your work and future publications, do not let the competition intimidate you! **Ale**, keep the art flourishing, you made our chip look awesome! The quantum chips supply chain is waiting for you, good luck! **Christos**, hope you can master cooking quantum chips as good as the Greek cuisine!

Wouter, these reliable electronics make the day-to-day work a bliss. Keep advocating for good software engineering, we will thank you later! **Jordy**, the skilled man that never skips lunch and always invites his rebel peers (me), I am glad we got to work together. **Matt**, our climate control specialist, thank you for ensuring that the engines of our coolest qubits run smoothly. Your "homemade" products are up to the industry standards! **Chris**, I met you through your thesis. It was a worthwhile read and a precious "getting started" when I joined the lab. Good luck in your research!

Marja, a huge thank you for saving my internship application, for taking care of the necessary logistics upon my arrival and for the support with my accommodation issues when times were chaotic for everyone! **Ana Bela**, thank you for giving me a call on each of my silly emails and ensuring I would not screw up important paperwork at IST. My warmest thanks to **the people of the Mobility and International Cooperation Office** of IST for the excellent support and guidance provided once again during this second exchange program.

Jules and Niels, the Qblox early morning stand-up added a few hours to my thesis-writing days, many thanks for recruiting an yet to graduate part-time student! I hope Qblox becomes the most wanted dealer during the Quantum Revolution!

Diogo, thank you for sparking my initial interest for quantum computing and Delft in particular! **Mário**, I am eternally grateful for our long term friendship. Thank you for calling and reminding me of the important things in life! **Tiago**, my dear pal, that weekly boost of energy means a world to me. Next time I can finally say that the writing progress of my thesis is now matching the fidelity of my gates. Thank you for being in my corner! I cannot wait to see your future projects.

Finishing this degree was tough, but meeting great people along the way makes it worth no matter the adversities. To **all my friends**, please accept my deepest thanks for your precious company in this life! Many thanks to those who preordered a copy of my thesis many months ago, it motivated me into making it more comprehensive. I hope it lives up to your expectations.

My dear **family**, I am deeply grateful for your support in my career choices and for everything else you provided me with! Dad, thank you for the courage to look for a better life. Mom, thank you for turning my childhood laziness into the excellence of today. [*Dragă familie, sunt profund recunoscătoare pentru sprijinul vostru în alegerile mele de carieră și pentru tot ce mi-ați oferit! Tată, mulțumesc pentru curajul de a căuta o viață mai bună. Mamă, mulțumesc că mi-ai transformat lenea din copilărie în excelența de azi.*]

Resumo

A afinação simples de portas de dois *qubits* é essencial para a calibração de processadores quânticos. Neste trabalho, foram automatizadas duas rotinas de calibração imprescindíveis para essas portas numa plataforma de transmon *qubits* de fluxo ajustável. Foram implementados algoritmos de amostragem adaptável para acelerar as medições, juntamente com rotinas de análise para processar e visualizar os resultados.

A variante súbita (SNZ) do esquema Net-Zero foi introduzida realizando portas de dois *qubits* do tipo *controlled*-Z (CZ) através de pulsos de fluxo que controlam a frequência de transição de um transmon. As portas SNZ operam no limite de velocidade do acoplamento transversal entre um estado computacional e um estado não computacional, maximizando a fuga intermédia para o estado não computacional. A principal vantagem das portas SNZ é a simplicidade de afinação, devido à estrutura regular da fase condicional e da fuga para o estado não computacional em função de dois parâmetros de afinação.

As portas SNZ foram realizadas num conjunto de quatro transmons, com resultados ao nível do estado da arte no melhor par, em termos de fidelidade de $99.93\,\%$ [$99.54(27)\,\%$ em média] e a fuga de $0.10\,\%$ [$0.18(4)\,\%$ em média]. A porta SNZ é compatível com esquemas escaláveis para correção de erros quânticos e adaptável a portas de fase condicional generalizadas (CPHASE) úteis em aplicações de escala intermédia. Tirando partido da simplicidade de afinação, as portas SNZ já são empregues num processador quântico disponível publicamente por meio da plataforma QuTech Quantum Inspire.

Palavras-chave:

processador quântico, transmon, porta CZ, automação, calibração, amostragem adaptável

Abstract

Simple tuneup of two-qubit gates is essential for the scaling of quantum processors. In this work two key two-qubit gate calibration routines for flux-tunable transmon qubits were automated. Custom adaptive samplers were implemented to speed up measurements, along with analysis routines to process and visualize results.

The sudden variant (SNZ) of the Net-Zero scheme was introduced realizing (two-qubit) controlled-Z (CZ) gates by baseband flux control of transmon frequency. The SNZ CZ gates operate at the speed limit of the transverse coupling between a computational and a non-computational state by maximizing intermediate leakage to the latter. The key advantage of the SNZ is tuneup simplicity, owing to the regular structure of conditional phase and leakage as a function of two control parameters.

SNZ CZ gates were realized in a four-transmon patch, matching the state of the art on the best pair with fidelity of $99.93\,\%$ [$99.54(27)\,\%$ average] and leakage of $0.10\,\%$ [$0.18(4)\,\%$ average]. The SNZ is compatible with scalable schemes for quantum error correction and adaptable to generalized conditional-phase (CPHASE) gates useful in intermediate-scale applications. Taking advantage of the tuning simplicity, SNZ CZ gates are already employed in a quantum processor publicly available via the QuTech Quantum Inspire platform.

Keywords:

quantum processor, transmon, CZ gate, automation, calibration, adaptive sampling

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In recent years Quantum Computers made their way into the mass media with several private initiatives announcing their latest progress and performances on the quest to master the art of building a device with capabilities beyond the modern information technologies [1, 2]. In 2017, alone, quantum technology firms, raised from private investors almost 300 millions US$, in total [3].

Even though quantum computers are not a panacea for all computational problems, they are the best candidate, if not the only one, to tackle problems that are out of reach of digital computers. Research in Quantum Computation aspires to build an information processing device by exploiting the laws of physics governed by quantum mechanics. This turns the same effects that negatively affect the operation of classical transistors on scales under $10\,\mathrm{nm}$ [4] into an asset. While we keep dreaming of a Quantum Supremacy for a few decades (does a random number generator count? [1, 5, 6]), the possible future applications seem endless [7, 8]: Artificial Intelligence (AI) [9], Molecular Modeling [10], Cryptography [11] and Particle Physics [12], to mention a few.

Although, the theory of Quantum Computation has evolved significantly in the last two decades, the physical implementations are still struggling to match the pace. The number of qubits is approaching three figures [5], isolated single qubit and two qubit operations are somewhat mature [13], however, several fundamental issues persist. Qubits lifetime is significantly limited, quantum operations are imperfect and, as the recent Quantum Supremacy device [5] demonstrates, when trying to orchestrate a few dozen of qubits the error rates rise to levels incompatible with long computations. The near-term quantum chips are called Noisy Intermediate-Scale Quantum (NISQ) devices and have restricted applications [14]. The holy grail to address those issues, at the end of the Fault-Tolerant Quantum Computation (FTQC) [15, 16] roadmap is to perform useful computation with imperfect quantum hardware. In an nutshell, it consists of using multiple imperfect qubits to encode redundantly a smaller amount of logical qubits that can tolerate the faulty hardware [17]. The amazing milestone of a logical qubit with longer lifetime than physical qubits has been achieved this year [18]. Yet, two essential checkpoints lie on the road, operations on a single logical qubit followed by algorithms on multiple logical qubits.

During this fierce race, the work presented in this thesis aimed to take the automation of the characterization and calibration of superconducting quantum processors as far as possible within the DiCarlo

Lab, QuTech & Kavli Institute of Nanoscience. Given the state of the art in the laboratory, the focus was on 7-qubit devices and, more specifically, two-qubit operations (gates) – an essential step towards practical large scale quantum processors.

The automation toolbox developed was promptly used to calibrate two-qubit gates for a NISQ applications, a Quantum Approximate Optimization Algorithm (QAOA) [19]. Besides the work on the automation, the research carried throughout this work led to an important contribution on physics grounds. An alternative approach to the implementation of the CZ[1] two-qubit gate was proposed and experimentally realized [Chapter 5 and Ref. 20], resulting in a significantly simplified calibration path, besides registering, to best of our knowledge, the highest CZ gate fidelity[2] in a transmon-type quantum processor. Owing to the calibration simplicity, these gates are already employed in a publicly available quantum processor via the QuTech Quantum Inspire platform [21].

1.1 Motivation

It is still very relevant to research and work on improving the fundamentals of superconducting qubits. Even though the FTQC roadmap takes into account imperfect qubits and operations, improving them as much as possible means that less physical qubits would be required for redundant information encoding, and thus we would achieve useful quantum computation sooner.

Unfortunately, the yield of functional devices from the clean rooms with parameters accurately matching the desired design are fairly low [22]. Therefore, characterizing and testing new devices that explore different designs, techniques, batches, etc. is a common and recurrent procedure, yet considerably complex. It requires to precisely and accurately calibrate and orchestrate multiple instruments on nanosecond time scales – a mundane task for a human but a perfect one for a (classical) machine that can reliably perform the same tasks over and over again while, at the same time, managing staggering levels of complexity.

Characterizing a device is not the only repetitive process involved. Once a promising chip has been characterized, a series of other calibrations need to be carried out in order to bring the system to a specification compatible with quantum computations. Moreover, over time, the system will inevitably drift out of spec requiring recurrent re-calibrations to maintain its maximum performance.

We can discern at least three highly impacted resources when performing all these procedures manually that can be spent much more efficient by putting efforts in automation routines.

Manpower As an example, the Google's supremacy device require several hours for a qualified PhD to manually perform a single calibration procedure (out of many such procedures required for tens of qubits) [23]. The more we free up highly specialized and qualified researchers to tackle other issues, the faster we can move forward the developments in the field. Furthermore, the complexity involved, which can only increase when scaling up, surpasses the human capacity of monitoring the entire device.

[1] CZ stands for Conditional Z or Controlled Z quantum gate, a particular case of a Conditional Phase (CPHASE) gate.
[2] Extracted from a two-qubit interleaved randomized benchmarking (2QIRB) experiment.

Time Automation allows for fast chip development and characterization cycles, as well as autonomous maintenance of the calibration state of a processor. This is a often a one-time development effort that pays off over and over again. Just recovering the state of the experimental setup from software or hardware malfunctioning might take several hours with manual initialization which is inherently prone to human errors.

Cost The two resources above are translatable in financial costs. However there are yet, more factors to consider. Maintaining in the fridge the fresh vacuum, colder than outer space, costs on the order of 10,000 US$ (per month!). Wasting experiment time, is wasting money. In the long term there is a big financial incentive to build a quantum computer and will likely require significant investments [24] which would be recovered by providing computing time in a similar fashion it is currently done for supercomputers. Being a scarce resource, all the down time for calibration routines will have significant financial impacts and therefore is of high value to automate such processes and make them as time efficient as possible.

When considering these issues it becomes clear that automation routines are a must for quantum devices, being far from a mere commodity.

1.2 Automation: state of the art

Experimental framework In the beginning of this work in the DiCarlo Lab, the automation was mature at the level of individual experiments and calibrations preceding the two-qubit gates. Some work towards higher level of automation was in progress. The experiments are performed using the PycQED[3] software framework [25] – a python based measurement environment developed in-house on top of the QCoDeS data acquisition framework [26]. PycQED is an extensible, easy-to-use and automatable toolbox for experimentalists to perform measurements and analyze results, however, it is not scalable on its own. The (published) research targeting the automation problem on the scale a multi-qubit quantum chip has been limited. Kelly et al. [27] proposed to solve the problem by organizing the calibration dependencies on a directed acyclic graph[4] (DAG) in an automatable and extensible manner. Inspired by this work, Rol et al. [28] developed the AutoDepGraph [28, see also Fig. 1.1]. This tools has been used to automatically characterize and calibrate 7-qubit quantum devices up to single qubit gates[5]. For more details and discussion on the PycQED framework and graph-based calibrations we point the interested reader to chapter 8 in Ref. 29.

Optimization algorithms Many calibrations involve finding optimal values for parameters that are correlated in nontrivial ways requiring optimization algorithms for efficient automation. In DiCarlo Lab, for non-trivial calibrations and optimizations, it is common [30, 31] to employ algorithms such as Nelder-Mead [32] or CMA-ES [33]. However, the former is know to not be noise resilient (a requirement for

[3] PycQED stands for Python circuit Quantum ElectroDynamics.

[4] A DAG is a mathematical structure used to model pairwise relations between objects, in this case calibrations. The graph is acyclic because each relation is unilateral, i.e. one calibration depends on previous calibration(s).

[5] Unpublished work, initially developed by T. van Abswoude, and later by A. Brandwijk and H. Ali.

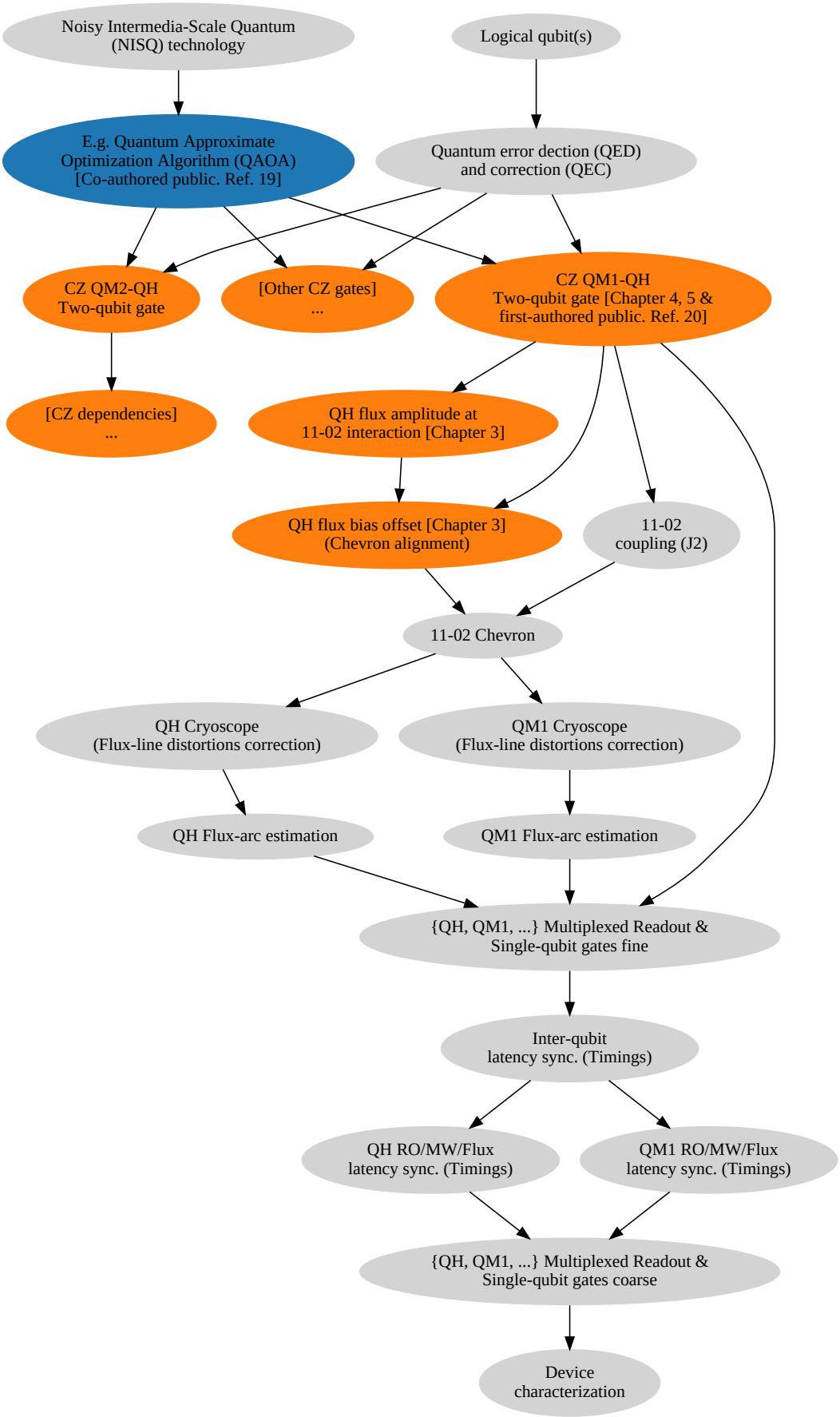


Figure 1.1: Thesis context on a pseudo-dependency graph of a transmon quantum processor. The direction of an arrow indicates "depends directly on". Gray nodes at the top are long-term goals/applications in the (universal) quantum computing field. The contributions to Ref. 19 (blue node) are presented in Chapter 4. The orange nodes encapsulate the main contributions of this thesis. The CZ approach introduced in Ref. 20 (Chapter 5) was conceived within this thesis. The first authorship is shared with H. Ali. Calibration routines corresponding to the rest of the gray nodes were used for day-to-day operation of the quantum processors. Figure generated with the AutoDepGraph [28].

in-experiment optimization), while the latter needs a significant number of iteration for good results. Besides that, an experimentalist is often interested in gaining a general understanding of the parameters space which is usually not easy with these algorithms.

Two-qubit gate There are several implementation of two-qubit entangling gates reported in the literature [34–38]. In the DiCarlo Lab the quantum chip architecture employs the CZ gate [13]. Its implementation relied on the parametrization proposed by Martinis and Geller [34] with recent gate repeatability and leakage[6] suppression improvements by Rol et al. [39] based on a Net-Zero (NZ) integral flux-pulse. However, at the beginning of this work, the experimental calibration was not automated, yet. Excluding a preceding (one-time) pulse distortion correction procedure [31], the calibration of a single two-qubit gate required from an experienced PhD under optimistic conditions: $\sim 1\,\mathrm{h}$ of configuring hardware (e.g. timings between pulses) and preparing code; $\sim 1\,\mathrm{h}$ of sanity checks based on single parameter sweeps; before starting an optimizer that took $\sim 1\,\mathrm{h}$ per parameter, which translated into $2\,\mathrm{h} - 3\,\mathrm{h}$ of extra unsupervised optimization. This lengthy process is unsustainable even for a 7-qubit device, let alone hundreds of qubits.

1.3 Objectives

The routines developed within this thesis target this gap and correspond to nodes intended to later be integrated in a full-processor automatic calibration. Figure 1.1 illustrates this on a high level. Along the chapters, the reader might find useful to revisit this figure. The main targets of the work developed are listed below. As a starting point, it is assumed that the infrastructure for single qubit gates and measurement readout calibrations are in place. All are to be implemented for and tested on 7-qubit devices, even though most of the work is expected to be easily extensible to devices comprising higher number of qubits.

- Systematize calibration routines required for the two-qubit CZ gate and evaluate the necessity/priority of automation.
- Automate sampling/optimization and analysis of 2D calibration landscapes for two-qubit CZ gates (compatible with experiment and simulations).
- Implement flexible routines usable by an experimentalist and ready to be embed in a dependency graph of automatic calibrations (similar to Fig. 1.1). This process requires outlining the dependencies of the previous steps and an understanding of the overall chip characterization and calibration.
- Explore non-adiabatic CZ gate pulse parameterizations in search of calibration simplification and gate performance improvement. The adiabatic CZ gate is the one being used in the DiCarlo Lab. An adiabatic two-qubit gate is implemented by changing slowly the two-qubit Hamiltonian such that the system remains in the Hamiltonian's instantaneous eigenstate at all times. This approach

[6] Leakage occurs when a two-qubit system ends up in a non-computational state (an inherent issue of the CZ).

makes the gates longer and more prone to incoherent errors[7]. A non-adiabatic approach aims to perform the gate as fast as possible to minimize such errors while preserving the computational state of the two qubits.

An equilibrium between exploring alternative options and moving the developments forward was adopted given the time constraints of the project. At a lower level the targets were achieved through the integration of multiple Python modules, development of specialized code and re-implementation of inefficient or ineffective code. Both CZ simulation and experimental sampling/optimization within the parameters space are time consuming. To alleviate this adaptive sampling techniques were used.

1.4 Thesis Outline

Chapter 2 introduces the reader to core concepts of quantum computation followed by a theoretical and experimental background overview of superconducting transmon qubits with focus on the chip architecture employed in this work. This chapter is intended to build an intuition on how the quantum devices are operated, namely single-qubit gates, two-qubit gates, qubits readout and control synchronization.

In Chapter 3, we describe the automation of a key routine preceding the sampling of CZ calibration landscapes ("flux bias offset" and "flux amp. at $|11\rangle - |02\rangle$ interaction" nodes in Fig. 1.1), along with the associated measurement analysis and the inner workings of the 1D adaptive sampling algorithm developed to this end.

Next, in Chapter 4, we dive into the technicalities built into the analysis of the CZ 2D calibration landscapes, followed by an overview of the N-dimensional adaptive sampler created through the generalization of its 1D counterpart (Chapter 3).

Chapter 5 presents one of the most important contributions of this work. The SNZ flux pulses implementing two-qubit CZ gates are introduced and contrasted with the previous state-of-art conventional NZ pulses (CNZ). The experimental performance is evaluated among four SNZ CZ gates. We close the chapter inspecting dominant error sources though numerical simulation for both SNZ and CNZ pulses.

Finally, in Chapter 6 we first summarize the conclusions of the work in the previous chapters. Based on the achieved results and the accumulated hands-on experience we provide an outlook for next steps in tackling the calibration problem

[7] Incoherent errors, as opposed to coherent errors, have stochastic nature and cannot be mitigated by gate design and calibration of non-ideal factors.

Chapter 2

Background

The research field of Quantum Computing (QC) extends from the quantum information fundamentals and pure quantum algorithms to the entire system stack of a specific physical implementation on a quantum chip. This chapter is a brief introduction to key concepts required to understanding the context, problems and solutions presented in this work. Despite my efforts, it is impossible to compress and include in this chapter all the aspects of a field with accumulated knowledge over more than two decades. The theoretical overview on quantum computation and information is much inspired from one of the bibles in quantum computation and quantum information – the book by Nielsen and Chuang [40]. References 29, 41–44 are good introductions specific to the transmons world which guided the writing of this chapter.

Chapter outline First the reader is introduced to core concepts of quantum mechanics in the scope of quantum information processing based on quantum circuits, followed by the physical requirements for a quantum computer. Second, we overview the transmon theory from Cooper-pair box to the fundamentals underlying the qubit readout, the single-qubit gates and the flux-based two-qubit gates. Third, we highlight important aspects when the theory meets the physical world and what that means for a practical quantum computer of today and near future. Finally, as link to this work, we close the chapter with an experimental overview of transmons, giving insight into the architecture of the physical chips of the lab and how it looks like to operate such a device on the day-to-day basis regarding single-qubit gates, readout, two-qubit gates and synchronization of pulse-generating instruments. Some minor contributions to the corresponding calibration routines and analyses are not detailed for the sake of brevity.

2.1 Quantum computation

2.1.1 Qubits

The term *qubit* was coined to designate the quantum analog of the *bit* – the classic digital information unit, even though a qubit behaves and is operated in fundamentally different ways. While a bit can have either of two discrete states **0** or **1**, the state of a qubit lives in a continuous two-dimensional Hilbert space spanned by two states that form the base of that space. The canonical base, represented in the

ket notation introduced by Dirac, is $\{|0\rangle, |1\rangle\}$. An arbitrary pure[1] single qubit quantum state $|\psi\rangle$ takes the form

$$|\psi\rangle = \beta\,|0\rangle + \varepsilon\,|1\rangle = \begin{pmatrix} \beta \\ \varepsilon \end{pmatrix}, \tag{2.1}$$

where the amplitudes β and ε are arbitrary complex numbers that must obey the normalization constraint of unit probability $|\beta|^2 + |\varepsilon|^2 = 1$. The linear combination of the two base states in Eq. (2.1) is referred to as *superposition* and is at the core of language abuses equivalent to "being in two states at the same time". This notation extends naturally to multi qubit systems. For example, for a two-qubit system the state would be represented by $|\psi\rangle = \beta\,|00\rangle + \varepsilon\,|01\rangle + \gamma\,|10\rangle + \eta\,|11\rangle$, again constrained by the normalization $|\beta|^2 + |\varepsilon|^2 + |\gamma|^2 + |\eta|^2 = 1$. In general, 2^n complex numbers are required to represent a n-qubit state. Note that a trivial state $|\psi\rangle = |01\rangle$ can also be represented as two separate single qubit states $|\psi\rangle_A = |0\rangle$ for qubit A and $|\psi\rangle_B = |1\rangle$ for qubit B. However, that is not always possible. Such an example is the state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}|00\rangle + \frac{1}{\sqrt{2}}|11\rangle$. This is a Bell state and illustrates the *entanglement* concept, more specifically this is a maximally entangled state. The curious behavior of such a state is that, if we measure only one of the two qubits, say in the $|0\rangle$ state, than the second qubit will have unity probability of being found in the same state[2]. For a single qubit state, or more generally for the superposition of two orthogonal quantum states, an intuitive geometrical representation exists denominated the Bloch sphere.

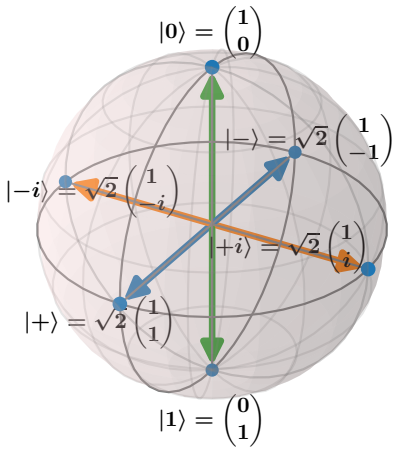


Figure 2.1: Representation of quantum states on the Bloch sphere. Orthogonal states are presented by parallel vectors pointing in opposite directions. Generated with [45].

Figure 2.1 shows this representation including the particular states at the intersection of the axes with the unitary sphere. Such representation arises from a parametrization that, besides the normalization, takes into account that a global arbitrary quantum phase has no observable effects. Therefore, the superposition of two orthogonal states has the equivalent definition

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

in terms of two real parameters (angles), θ and φ, which specify the direction of a unit vector $\vec{r}$ – the Bloch vector. It can also be defined in terms of Pauli matrices as $\vec{r} = \langle\psi|\,\vec{\sigma}\,|\psi\rangle$, where we take the expectation value of the Pauli vector given by $\vec{\sigma} = (\hat{X}, \hat{Y}, \hat{Z}) = (\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z)$.

2.1.2 Operations

In order to perform computations using qubits we need to be able to act on their states. The states of the qubits are modified by quantum operations. We can distinguish three types of operations (single

[1] *Pure* here is to distinguish from a *mixed* state that can only be represented through a more general formalism - the density matrix formalism - that accounts for probabilistic states, e.g. a state that was prepared either in $|1\rangle$ with probability $99.9\,\%$ or in $|0\rangle$ with probability $0.1\,\%$ which is often the reality of the limitations of a physical implementation.

[2] Amazingly, the two qubits can be separated by an arbitrary distance, when measured.

and multi-qubit): quantum gates, initial state preparation and measurements. All three have an analog in the realm of classical digital circuits where the last two are taken for granted, not requiring any special attention, which is not the case in the quantum realm.

Quantum gates

Ideally, quantum gates are reversible (there is always an inverse that allows to recover the previous state), non-destructive (there is no loss of quantum information) and unitary operators that are applied to qubits. Mathematically, in the notation introduced above, n-qubit quantum gates are represented by a $2^n \times 2^n$ matrix. Such operators are chained sequentially (matrix multiplication) to compose more complex algorithms. The only requirement for a valid quantum gate is to be unitary. Moreover, any unitary is a valid quantum gate and can be approximated to an arbitrary extent with a finite sequence of gates from some particular non-unique set of gates, named a universal quantum gate set for quantum computation. Nevertheless, such task is generically hard and many implementations support larger gate sets.

Measurement

After running a quantum algorithm we are interested in observing (measuring) the system and retrieving some information about its outcome, however, the laws of quantum mechanics put physical limitations on the information that can be obtained. Observation in quantum mechanics is an invasive procedure that typically changes the state of the system. It is non-reversible[3] and destructive in its nature[4]. Therefore, the measurement is also a state preparation procedure in some implementations.

Formally, a general quantum measurement is postulated to be described by a collection $\{M_m\}$ of measurement operators. The index m stands for the possible results that a measurement may have in experiment. Note that the collection of operators associated with a measurement does not predict the outcome of a specific measurement, but allows to calculate the probability that result m occurs. For a state $|\psi\rangle$ immediately before the measurement these probabilities are given by

$$P(m) = \langle\psi| M_m^\dagger M_m |\psi\rangle \tag{2.3}$$

and the post-measurement state is given by:

$$|\psi'\rangle = \frac{M_m |\psi\rangle}{\sqrt{\langle\psi| M_m^\dagger M_m |\psi\rangle}}. \tag{2.4}$$

The probabilities of all possible outcomes sum to 1. An ubiquitous example of a measurement is the measurement of a single qubit in the *computational basis* defined by the measurement operators $M_0 = |0\rangle\langle 0|$ and $M_1 = |1\rangle\langle 1|$. For a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the probability of observing the measurement outcome $m = 0$ is $P(0) = \langle\psi| M_0^\dagger M_0 |0\rangle = |\alpha|^2$. A special case of the measurement formalism just described is particularly relevant for this work, the *projective measurements*.

[3] We don't have the necessary knowledge to reverse the operation, but there is a unitary corresponding to that operation.
[4] The upper bound on the amount of information accessible from a quantum state is know as the Holevo's bound.

This special class of measurements is also known as von Neumann measurements and is of primary interest for many applications. A projective measurement is described by an observable, M - a Hermitian operator on the state space of the system being observed. The spectral decomposition of the observable is given by

$$M = \sum_m m\mathcal{P}_m, \tag{2.5}$$

where $\mathcal{P}_m$ is the projector onto the eigenspace of M with eigenvalue m. The eigenvalues, m, of the observable are the possible outcomes of a measurement, with probability

$$P(m) = \langle\psi| \mathcal{P}_m |\psi\rangle \tag{2.6}$$

of being measured. The state of the system immediately after the measurement is given by

$$|\psi'\rangle = \frac{\mathcal{P}_m |\psi\rangle}{\sqrt{\langle\psi| \mathcal{P}_m |\psi\rangle}}. \tag{2.7}$$

One of the nice properties of the projective measurements is the ease to calculate average values for projective measurements: $\langle M\rangle = \langle\psi| M |\psi\rangle$, which is very handy considering that the average value is usually the only information that can be retrieved in experiment.

There are two widely used nomenclatures for projective measurements. The first one is to simply give the complete set of orthogonal projectors $\mathcal{P}_m$, e.g. $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$, that satisfy $\sum_m \mathcal{P}_m = I$ and $\mathcal{P}_m\mathcal{P}_{m'} = \delta_{mm'}\mathcal{P}_m$. Here the implicit observable is $M = \sum_m m\mathcal{P}_m$. The second equivalent way of referring to a projective measurement is to say "measure in a basis $|m\rangle$", where $|m\rangle$ form an orthogonal[5] basis with the projectors being given by $\mathcal{P}_m = |m\rangle\langle m|$. More generally, single-qubit projective measurements can be described by the observable

$$M = \vec{n}\cdot\vec{\sigma} = n_1\hat{\sigma}_x + n_2\hat{\sigma}_y + n_3\hat{\sigma}_z, \tag{2.8}$$

which is often referred to as "measurement of spin along the $\vec{n}$ axis" due to historical reasons associated with spin measurements. When all axioms of quantum mechanics are taken into account, projective measurements augmented by unitary operations are completely equivalent to general measurements.

Initial state preparation

One of the most important requirements to perform useful computation is to start with a known initial state, even in classical digital circuits. Preparation of fiducial initial states falls more into the experiment rather then theory of quantum computation, but it is essential to cover it as it is a core requirement.

Generally a quantum system will exist in the continuum of the corresponding Hilbert space in an unknown state and it might be very difficult to bring the system into a desired state, depending on the qubit physical realization, even though it is only necessary to (repeatedly) generate one specific state with high fidelity (which can then be transformed to any other target state by a unitary). Beside being

[5] An orthogonal basis is a requirement in order to reliably distinguish quantum states.

able to prepare a state with high fidelity it is important that the state being prepared is a useful one., i.e. ideally, it should be a pure state with zero entropy, as it becomes impossible to extract the answer of a computation with the rise of entropy.

2.1.3 Quantum phase

In the context of quantum mechanics "phase" can have several distinct meanings. Two important cases are: *global phase* and *relative phase*. It is said that the states $e^{i\theta}|\psi\rangle$ and $|\psi\rangle$ are equal "up to a global phase factor" $e^{i\theta}$, where θ is a real number. There is no experiment that can distinguish between the two states based on the statistics of the measurement outcomes when performing the same measurement on the same initial states. Therefore, global phases are irrelevant to observations. Taking as an example the states $|\psi\rangle = (|0\rangle + i\,|1\rangle)/\sqrt{2}$ and $|\varphi\rangle = (|0\rangle - i\,|1\rangle)/\sqrt{2}$, we say that the amplitudes of the state $|1\rangle$ "differ by a relative phase" because there is a real number θ such that $-i = e^{i\theta}i$. Note, that the relative phase concept is basis-dependent, unlike global phase. As a result, measuring in a different basis makes the states $|\psi\rangle$ and $|\varphi\rangle$ (that differ only by a relative phase) give rise to measurable differences.

2.1.4 Quantum circuit model and quantum circuits

The most common Quantum Computing Model is called Quantum Circuit or Quantum Gate model. The quantum computation is seen as a sequence of quantum gates applied to a qubit register – a composite quantum memory element. The computations usually end with measurements that allow information to be retrieved. Quantum algorithms are described as quantum circuits and are generally hardware-agnostic.

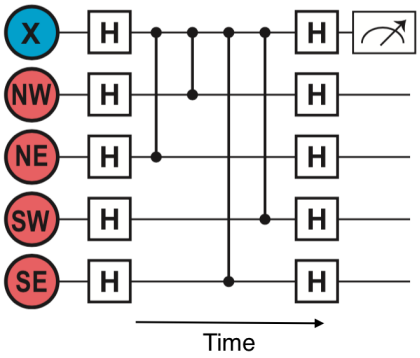


Figure 2.2: Representation of an exemplary quantum circuit: X-type stabilizer circuit for a surface code. Colored circles on the left represent the qubits. **H** represents the Hadamard single-qubit gate, vertical lines connecting two qubits represent CZ (Controlled-Z) two-qubit gates. Adapted from [17].

Figure 2.2 shows an exemplary quantum circuit. On the left, the colored circles disposed on the vertical axis represent the qubit register. The horizontal lines connected to each qubit denote the qubit at each point in time, with the time flowing from left to right. Symbols overlaying these lines represent operations applied to qubits at the respective (relative) moment in time corresponding to the position of the symbol on the horizontal axis. The time axis is usually omitted for convenience in the context of the quantum circuit model, but timing is of utmost importance from the technical perspective of the physical implementation. For instance, the freedom of arbitrary parallel operations performed on distinct qubits might not be possible due to hardware limitations. The vertical line symbol connecting two qubits represents a two-qubit gate, in this case a CZ, which stands for Controlled-Z gate – whose calibration is the main focus of this work. The measurements are commonly represented by a meter. For the particular circuit in this figure (only a small portion of the computation is shown) the mid-computation measurement is intended for quantum error detection for the qubits represented in red. Generally, at the end of a quantum algorithm all qubits

will be measured.

In quantum circuits we are physically prohibited to copy a quantum state from one qubit to another, this is know as the *no cloning theorem*. As a result, a quantum state will live most of the time in the same physical system. We are allowed to swap quantum states between qubits, but not duplicate arbitrary quantum states. Nevertheless, quantum circuits have the remarkable advantages over classical digital circuits: *superposition* and *entanglement*. The former is at the core of the exponential speed-ups of certain algorithms, e.g. factoring prime numbers, and the later having promising applications in quantum communication, for instance secure communication channels guaranteed by the laws of physics.

2.1.5 Decoherence

Decoherence is the term often used to refer to two distinct effects, energy relaxation and dephasing. Energy relaxation describes the decay of the qubit into the ground state, while the dephasing describes the loss of quantum coherence, i.e. the loss of the definite phase relation between different states. The two processes follow exponential decays with a time constant frequently referred to as T_1, for energy relaxation, and T_2, for dephasing. Both limit the total time available to perform quantum computations before the decoherence effect becomes significant and retrieving useful information is no longer possible. Qubit coherence times can range from nanoseconds to few minutes [17, 46]. Characterizing and minimizing the contributions to decoherence remains an active field of research with novel techniques still being developed [47].

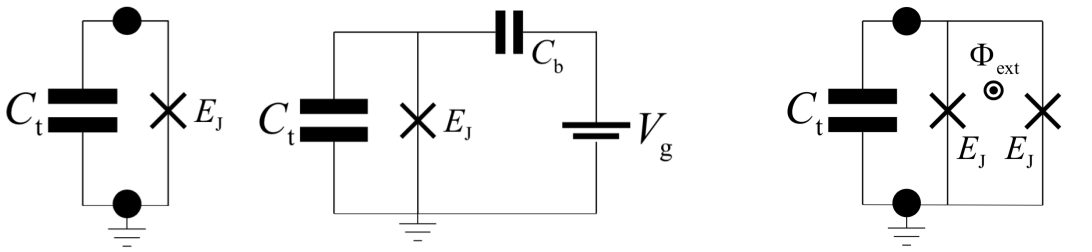
2.1.6 Physical implementation requirements

Regarding qubits and quantum gates as mathematical objects is convenient as it maintains the theory general and independent of a specific implementation allowing for a consensus of the research community to communicate on these common topics. Yet, a useful quantum computer must be physical. A famous statement of the requirements an experimental setup must satisfy to successfully implement quantum algorithms is the DiVincenzo criteria[6] [48]:

1. A scalable physical system with well characterized qubits
2. The ability to initialize the state of the qubits to a simple fiducial state such as $|000...\rangle$
3. Long relevant decoherence times, much longer than the gate operation time
4. A "universal" set of quantum gates
5. A qubit-specific measurement capability

Most criteria are self-explanatory, but building a device that matches those requirements well has proven to be a taught challenge. Several implementations for a quantum computer are being researched: superconducting [49], trapped ions [50] and nitrogen-vacancy centers [51], to name a few, with distinct underlying technologies, even though each comes in many more flavors. All approaches meet those

[6] The rest of the list, number 6 and 7, concern quantum communication.



(a) CPB (b) Voltage biased CPB (c) Flux biased CPB/Transmon

Figure 2.3: Cooper-pair box models. Adapted from [44].

requirements with varying degrees of success. The most mature technology so far are superconducting qubits, with semiconductor qubits catching up. Discrepancies between device design and fabrication, scalability, cross-talks and decoherence are still big concerns in most technologies.

2.2 Transmon theory

The transmon is one the most promising technologies for building a quantum computer. The term was coined in 2007 in the paper "Charge-insensitive qubit design derived from the Cooper-pair box" [41]. This type of qubit is the focus of this work, in particular two-qubit interactions of such qubits.

2.2.1 From Cooper-pair box to Transmon

The work reported in that publication builds upon the "Cooper-pair box" (CPB). The CPB can be modeled as a parallel combination of a capacitor and a Josephson junction, as depicted in Fig. 2.3(a). Fundamentally, the transmon is a CPB with a clever choice of the capacitor and the Josephson junction parameters. The quantum Hamiltonian can be written, in terms of the charge across the capacitance, Q_t, and the flux across the tunnel junction, Φ_t, as [41, 44, 52, 53]:

$$\hat{H} = \frac{\hat{Q}_t}{2C_t} - E_J \cos\left(2\pi \frac{\hat{\Phi}_t}{\Phi_0}\right); \qquad \text{Commutation relation: } [\hat{\Phi}_t, \hat{Q}_t] = i\hbar \tag{2.9}$$

where E_J is the Josephson coupling energy and Φ_0 is the superconducting flux quantum. If we couple a voltage source, as shown in Fig. 2.3(b), the Hamiltonian becomes:

$$\hat{H} = \frac{\left(\hat{Q}_t + C_b V_g\right)^2}{2\left(C_t + C_b\right)} - E_J \cos\left(2\pi \frac{\hat{\Phi}_t}{\Phi_0}\right) \tag{2.10}$$

Changing variables to $\hat{\varphi}_t \equiv 2\pi \frac{\hat{\Phi}_t}{\Phi_0}$ (phase operator), $\hat{N}_t \equiv -\frac{\hat{Q}_t}{2e}$ (Cooper-pair number operator), $N_g \equiv \frac{C_b V_g}{2e}$ (number offset) and $E_C \equiv \frac{e^2}{2(C_t+C_b)}$ (charging energy), we can express the Hamiltonian as:

$$\hat{H} = 4E_C \left(\hat{N}_t - N_g\right)^2 - E_J \cos(\hat{\varphi}_t); \qquad \text{Commutation relation: } [\hat{\varphi}_t, \hat{N}_t] = -i. \tag{2.11}$$

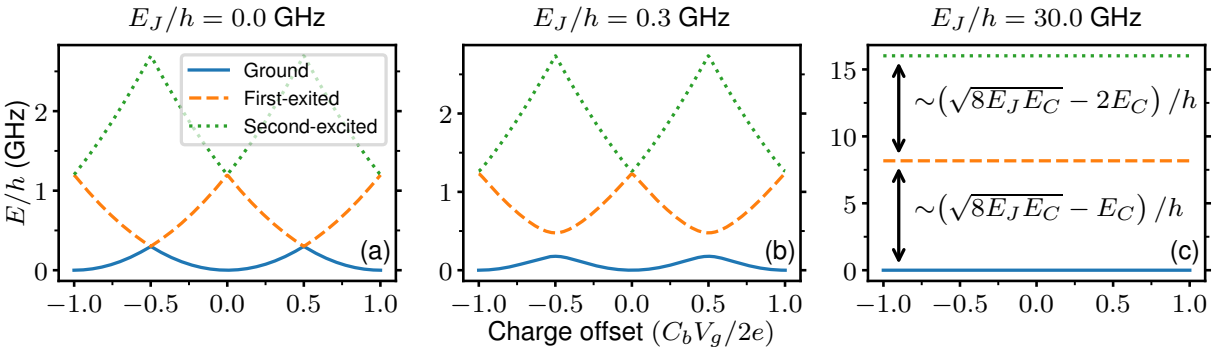


Figure 2.4: First three eigenenergies of the CPB [Eq. (2.12)] as a function of the offset charge for different values of E_J with fixed $E_C = 0.3$ GHz. The zero point of energy in each case is chosen as the minimum of the ground state.

Writing both operators in the charge basis (a base of quantum states corresponding to integer Cooper-pair charges, N, on the Josephson junction island), the Hamiltonian takes the shape:

$$\hat{H} = \sum_{N=-\infty}^{\infty} 4E_C \left(N - N_g\right)^2 |N\rangle \langle N| - \frac{E_J}{2} \left(|N\rangle \langle N+1| + |N+1\rangle \langle N|\right) \tag{2.12}$$

where N_g quantifies a charge offset. The value of E_J defines two distinct regimes: the CPB regime and the Transmon regime. In the CPB regime [Fig. 2.4(b)], $E_C \sim E_J$, N_g could be a knob to implement a single-qubit gate [54], but at the same time it makes the CPB sensitive to the environmental charge noise. The CPB transitions into the transmon regime for $E_J/E_C \gg 1$, experimentally $\gtrsim 30$ is already enough [44]. In this regime [Fig. 2.4(c)], the system becomes insensitive to the charge offset, i.e. the dependence of the energy levels on N_g is negligible in practice, which is a desired property for a physical qubit. The transition energies between the first three energy levels are depicted in Fig. 2.5 as a function of E_J. In the transmon regime, they are given approximately by

$$E_{0\to1} \simeq \sqrt{8E_JE_C} - E_C \qquad E_{1\to2} \simeq \sqrt{8E_JE_C} - 2E_C, \tag{2.13}$$

This gives rise to a weak anharmonicity, $\alpha \equiv E_{1\to2} - E_{0\to1} \sim -E_C$, which is enough to be able to address independently only the first transition of the qubit for the implementation of single-qubit gates [42]. This is essential in order to be able to use the transmon as a two level system in quantum computations. Furthermore, it is remarkable that the energy-level spectrum is mainly defined by the parameters of circuit elements, and therefore, with enough control over fabrication, they are configurable through chip design.

2.2.2 The flux degree of freedom

If we add one more Josephson junction in parallel and apply a magnetic field (by applying a current through a nearby wire) such that there will be an external magnetic flux Φ_{ext} in the closed loop formed by the two junctions (called SQUID loop) as shown in Fig. 2.3(c), we obtain a flux dependent Josephson energy, $E_J(\Phi_{\text{ext}})$. Assuming symmetric junctions, i.e. $E_{J1} = E_{J2}$, the dependence on the external flux

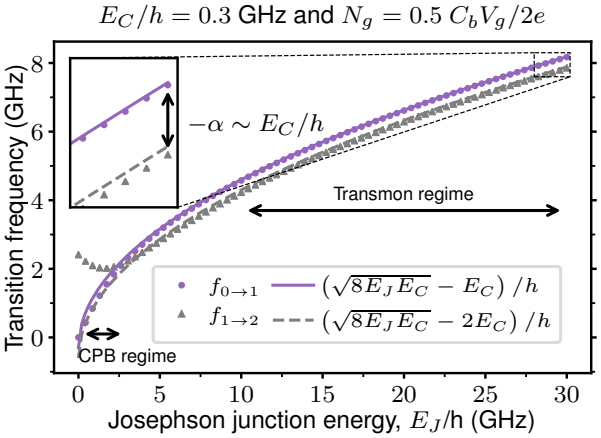


Figure 2.5: Transmon first two transitions frequencies as a function of E_J and analytical approximation.

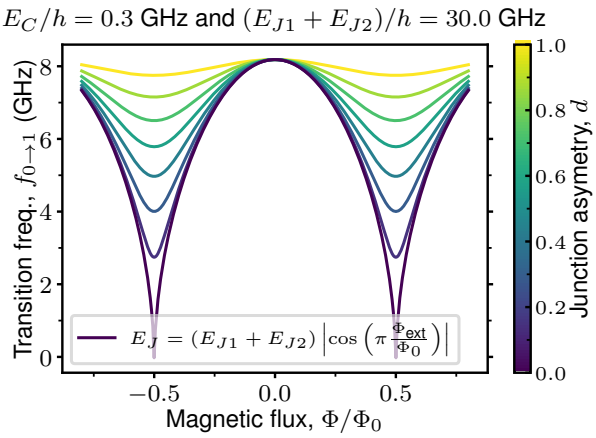


Figure 2.6: Transmons transition frequency as a function of the magnetic flux Φ in the SQUID loop and the junctions asymmetry d.

is given by $E_J(\Phi_{\text{ext}}) = (E_{J1} + E_{J2})\cos(\pi\Phi_{\text{ext}}/\Phi_0)$ (Fig. 2.6, dark curve) [41].

This provides a knob to lower (tune down) the qubit transition frequency. For small flux, the dependence is quadratic which, in first order, provides resilience to flux noise at the maximum qubit transition frequency (known as *sweetspot* – maximum frequency point in Fig. 2.6). In practice, achieving symmetric junctions is hard if not impossible. The consequences can be accounted for as a function of the asymmetry $d \equiv |E_{J1} - E_{J2}|/(E_{J1} + E_{J2})$ [41]:

$$E_J(\Phi_{\text{ext}}) = (E_{J1} + E_{J2}) \left|\cos\left(\pi \frac{\Phi_{\text{ext}}}{\Phi_0}\right)\right| \sqrt{1 + d^2 \tan^2\left(\pi \frac{\Phi_{\text{ext}}}{\Phi_0}\right)} \tag{2.14}$$

Replacing E_J in the expression for the approximate transition energy [Eq. (2.13)] we obtain the dependence on external flux and junctions asymmetry as depicted in Fig. 2.6. On one hand, it shows that the asymmetry reduces the maximum frequency detuning possible. On the other hand, it creates a second sweetspot at points of minimum frequency. The transition frequency tuneability mechanism can be used in transmons to implement two-qubit gates, such as the CZ gate [13].

2.2.3 Transmon measurement: qubit readout

Similarly to the case of the CPB, the transmon can have an associated transmission line resonator that establishes mechanisms for coherent control and readout of the qubit state. This scenario is known as circuit Quantum Electrodynamics (cQED) and it is analogous to an atom in an optical cavity as depicted in Fig. 2.7. The circuit quantization of this scenario adds new terms to the Hamiltonian in Eq. (2.11) [41]:

$$\hat{H} = 4E_C(\hat{N}_t - N_g)^2 - E_J\cos(\hat{\varphi}_t) + \hbar\omega_r\hat{a}^\dagger\hat{a} + 2\beta e V^0_{\text{rms}}\hat{N}_t(\hat{a} + \hat{a}^\dagger), \tag{2.15}$$

with $\hat{a}$ ($\hat{a}^\dagger$), the annihilation (creation) operator for photons in the transmission line, $\omega_r = 1/\sqrt{L_r C_r}$, the resonator frequency, $\beta = \frac{C_g}{C_\Sigma}$ set by the geometry of the capacitances, and $V^0_{\text{rms}} = \sqrt{\hbar\omega_r/2C_r}$ the voltage of the local oscillator. Changing the basis to the uncoupled transmon states $|i\rangle$ we obtain the

generalized Jaynes-Cummings Hamiltonian:

$$\hat{H} = \sum_j \hbar\omega_j |j\rangle\langle j| + \hbar\omega_r \hat{a}^\dagger\hat{a} + \sum_{i,j} g_{ij} |i\rangle\langle j| (\hat{a} + \hat{a}^\dagger), \tag{2.16}$$

with coupling energies $g_{ij}\hbar = 2\beta e V^0_{\text{rms}} \langle i| \hat{N}_t |j\rangle$. Next several approximations and simplification are employed (for details see [41]) that include evaluating the asymptotic behavior of the coupling energies with a perturbative approach in the limit of large E_J/E_C followed by a rotating wave approximation (RWA) and the dispersive limit approximation. In the dispersive limit, defining the detuning $\Delta_i = \omega_{i\to i+1} - \omega_r$, we assume $g_{01}/|\Delta_0| \ll 1$ and $g_{01}/|\Delta_0 + \alpha| \ll 1$. In the end, an interesting effective Hamiltonian is obtained:

$$\hat{H}_{\text{eff}} = \frac{1}{2}\hbar\omega'_{0\to1}\hat{\sigma}_z + \hbar\omega'_r\hat{a}^\dagger\hat{a} + \hbar\chi\hat{\sigma}_z\hat{a}^\dagger\hat{a}, \tag{2.17}$$

where $\omega'_{0\to1} = \omega_{0\to1} + \chi_{01}$, $\omega'_r = \omega_r - (\chi_{12}/2)$ and $\chi = \chi_{01} - (\chi_{12}/2)$ is the effective *dispersive shift*, with $\chi_{ij} \equiv g^2_{ij}/(\omega_{i\to j} - \omega_r)$. In fact, the above equation has a form, known before the transmon, as the Stark-shift Hamiltonian for a CPB coupled to a transmission line. The important result is that the resonator-qubit interaction leads to a resonator whose frequency depends on the qubit state (and a shift of the qubit frequency depending on the number of photons in the resonator). In practice this makes possible to probe the transmission line at a fixed frequency and based on the transmission/reflection (phase and amplitude) infer the qubit state.

Remarkably, even though the charge dispersion in the transmon regime decreases exponentially, the coupling with the resonator cavity increases. This is essential in order to achieve a transmon-based qubit! Furthermore, because of the familiarity of the scientific community with Eq. (2.17), it was possible to transfer almost directly the readout and control methods developed for the CPB qubits. On the control side (i.e. single-qubit gates and more detailed coupling to other qubits), Ref. 42 is an excellent theory and engineering introduction.

2.2.4 Transmon single-qubit gates

There are two practical ways of implementing single-qubit gates. In case of flux-tunable transmons, the detuning from the maximum-frequency flux allows the implementation of $\hat{z}$-axis single qubit rotations (besides several types of two-qubit gates). However, that does not provide a universal quantum gate set and a second mechanism for transmon single-qubit gates is required: capacitively coupling a microwave drive line as depicted in Fig. 2.8. This approach provides the ability to perform single qubit rotations around axes on the $\hat{x}\hat{y}$-plane. This subsection highlights the steps that lead to this conclusion following references 42 and 44.

We start with the relatively simple model of the transmon [Fig. 2.8] for which the quantum Hamiltonian is given by [44]:

$$\hat{H}(t) = -E_J \cos\left(2\pi\frac{\hat{\Phi}_t}{\Phi_0}\right) + \frac{\left(\hat{Q}_t + C_c V_{ac}(t)\right)^2}{2\left(C_t + C_c\right)}, \tag{2.18}$$

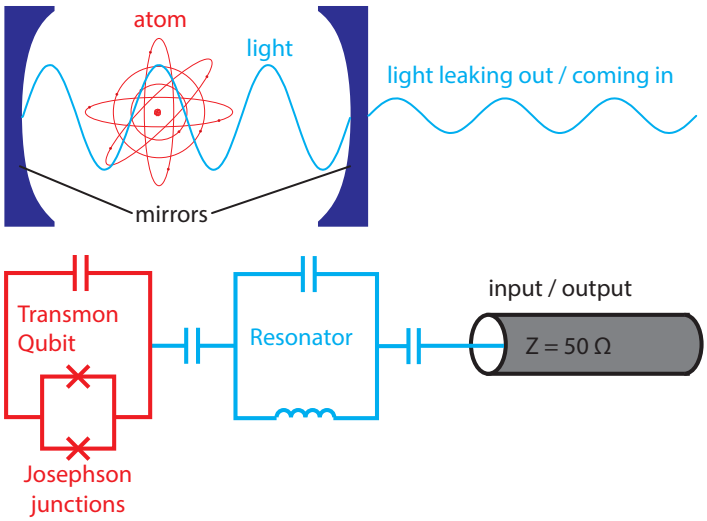


Figure 2.7: Analogy between Cavity Quantum Electrodynamics and circuit Quantum Electrodynamics. The transmon behaves like an artificial atom and the LC-like resonator plays the role of cavity. Reproduced with permission from Ref. 43.

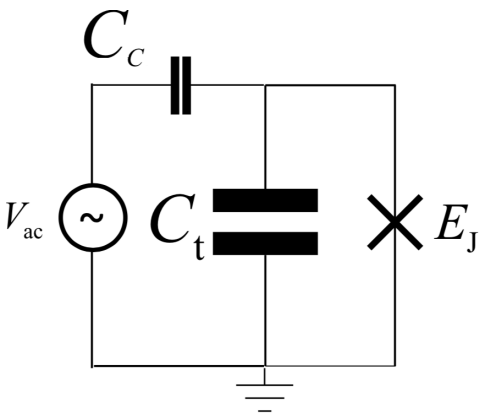


Figure 2.8: Microwave drive line – a time-dependent voltage source, $V_{\rm ac}(t)$ – capacitively coupled to a generic transmon. Adapted from Ref. 44.

which can be split in two terms:

$$\begin{aligned}\hat{H}(t) &= \hat{H}_{\text{transmon}} + \frac{C_c}{C_t + C_c} V_{ac}(t)\hat{Q}_t \\ &= \hat{H}_{\text{transmon}} + \hbar\varepsilon(t)\frac{\hat{Q}_t}{2e}; \qquad \varepsilon(t) \equiv \frac{2e}{\hbar}\frac{C_c}{C_t + C_c}V_{ac}(t)\end{aligned} \tag{2.19}$$

where we have written the driving term more conveniently. We now consider a drive that takes the form

$$\varepsilon(t) = \varepsilon_{\mathsf{I}}(t)\cos(\omega_d t) + \varepsilon_{\mathsf{Q}}(t)\sin(\omega_d t). \tag{2.20}$$

As the equation suggests we are composing a driving signal that is composed of two sinusoids in phase quadrature each shaped by an envelope, ε_{I} and ε_{Q}. I stands for *in-phase* and Q stands for *in-quadrature*.

In order to understand more easily the effect of this signal on the transmon, it is common procedure in the literature to work in a frame that rotates at the frequency ω_d (denoted rotating frame or interaction frame):

$$|\tilde{\psi}(t)\rangle \equiv U(t)\,|\psi(t)\rangle\,; \qquad U(t) \equiv e^{iH_d t/\hbar}; \qquad H_d \equiv \hbar\omega_d \sum_{k_t=0}^{\infty} k_t\,|k_t\rangle\langle k_t|\,, \tag{2.21}$$

where $|k_t\rangle$ are the eigenstates of the transmon (i.e. $|0\rangle, |1\rangle$, etc.). Invoking the time-dependent Schrödinger equation, $\frac{d}{dt}|\tilde{\psi}\rangle = \frac{1}{i\hbar}\hat{\tilde{H}}\,|\tilde{\psi}(t)\rangle$, after a bit of math we derive the Hamiltonian in the rotating frame:

$$\hat{\tilde{H}} = \underbrace{\hat{H}_{\text{transmon}} - \hat{H}_d}_{\text{free-evolution}} + \underbrace{\frac{\hbar\varepsilon(t)}{2e}U(t)\hat{Q}_t U^{\dagger}(t)}_{\text{drive term}}; \qquad \hat{H}_{\text{transmon}} - \hat{H}_d = \sum_{k_t} E_{k_t}\,|k_t\rangle\langle k_t| - \sum_{k_t}\hbar\omega_d k_t\,|k_t\rangle\langle k_t|\,. \tag{2.22}$$

Introducing the dipole moment between transmon levels [7] k_t and l_t, $D_{kl} = \langle k_t|\,\hat{Q}_t\,|l_t\rangle = -2e\,\langle k_t|\,\hat{N}_t\,|l_t\rangle$,

[7] $\langle k_t|\,\hat{Q}_t\,|l_t\rangle = 0$ for k_t and l_t differing by even integer and $\langle k_t|\,\hat{Q}_t\,|l_t\rangle \neq 0$ for k_t and l_t differing by odd integer [41].

the driving term becomes:

$$\frac{\hbar\varepsilon(t)}{2e}U(t)\hat{Q}_t U^\dagger(t) = \frac{\hbar\varepsilon(t)}{2e}\sum_{k_t,l_t} U(t)D_{kl}\left|k_t\right\rangle\left\langle l_t\right| U^\dagger(t) = \frac{\hbar\varepsilon(t)}{2e}\sum_{k_t,l_t} e^{i(k_t-l_t)\omega_d t}D_{kl}\left|k_t\right\rangle\left\langle l_t\right| \tag{2.23}$$

Adopting the notation $\hat{\sigma}^+_{kl} = \left|k_t\right\rangle\left\langle l_t\right|$ and $\hat{\sigma}^-_{kl} = \left|l_t\right\rangle\left\langle k_t\right|$ we can write:

$$\frac{\hbar\varepsilon(t)}{2e}U(t)\hat{Q}_t U^\dagger(t) = \frac{\hbar\varepsilon(t)}{2e}\sum_{k_t>l_t} e^{i(k_t-l_t)\omega_d t}D_{kl}\hat{\sigma}^+_{kl} + e^{-i(k_t-l_t)\omega_d t}D^*_{kl}\hat{\sigma}^-_{kl} \tag{2.24}$$

$$= \frac{\hbar}{2e}\left(\varepsilon_I(t)\frac{e^{i\omega_d t}+e^{-i\omega_d t}}{2} + \varepsilon_Q(t)\frac{e^{i\omega_d t}-e^{-i\omega_d t}}{2i}\right)\sum_{k_t>l_t} e^{i(k_t-l_t)\omega_d t}D_{kl}\hat{\sigma}^+_{kl} + e^{-i(k_t-l_t)\omega_d t}D^*_{kl}\hat{\sigma}^-_{kl} \tag{2.25}$$

$$= \frac{\hbar}{4e}\left(\varepsilon_I(t)\sum_{k_t=l_t+1}\left(D_{kl}\hat{\sigma}^+_{kl} + D^*_{kl}\hat{\sigma}^-_{kl}\right) + i\varepsilon_Q(t)\sum_{k_t=l_t+1}\left(D_{kl}\hat{\sigma}^+_{kl} - D^*_{kl}\hat{\sigma}^-_{kl}\right)\right) + \text{fast osc. terms.} \tag{2.26}$$

The terms omitted above oscillate at frequencies $\pm\omega_d, \pm 2\omega_d, \pm 3\omega_d, ...$ which allows to discard them in what is know as the rotating wave approximation (RWA). Moreover, the dipole moments for the transmon levels as we have defined them are real. Hence, we arrive at more intuitive form of the driving term:

$$\frac{\hbar\varepsilon(t)}{2e}U(t)\hat{Q}_t U^\dagger(t) \approx \frac{\hbar}{4e}\left(D_{10}\underbrace{\left(\varepsilon_I(t)\hat{X}_{10} + \varepsilon_Q(t)\hat{Y}_{10}\right)}_{\text{first transition driving}} + D_{21}\underbrace{\left(\varepsilon_I(t)\hat{X}_{21} + \varepsilon_Q(t)\hat{Y}_{21}\right)}_{\text{undesired second transition driving}} + ...\right) \tag{2.27}$$

where $\hat{X}_{kl} \equiv \hat{\sigma}^+_{kl} + \hat{\sigma}^-_{kl}$ and $\hat{Y}_{kl} \equiv i\hat{\sigma}^+_{kl} - i\hat{\sigma}^-_{kl}$. As the equation above shows, the shape of the envelopes I and Q represent knobs to implement single qubit rotations. The relative amplitude between the two envelopes defines the rotation axis on the $\hat{x}\hat{y}$-plane (on the Bloch sphere) and the absolute amplitudes and the duration of the pulse will determine the angle of the rotation. A simple case is to have $\varepsilon_{\text{I}}(t) \neq 0$ and $\varepsilon_{\text{Q}}(t) = 0$ which implements a rotation around the $\hat{x}$-axis.

Unfortunately, in the second term in Eq. (2.27), the same envelops are also associated with operators that transfer (rotate in the corresponding subspace) state population between $|1\rangle$ and $|2\rangle$. This is an undesired effect for a transmon-based qubit, commonly denoted as leaking out of the computational subspace. Remarkably, practical solutions to overcome this inconvenience have been already developed easily yielding gate error rates below 10^{-2} [42]. This procedure and corresponding envelopes shaping is outlined later in Section 2.4.2. Going back to the considered signal, its driving frequency ω_d needs to be in resonance with the targeted transition (e.g. $\omega_d = \omega_{0\to1}$ for $\hat{X}_{10}/\hat{Y}_{10}$ rotations) [8]. When this is not the case it is not possible to achieve a complete population transfer between $|0\rangle$ and $|1\rangle$ [44].

2.2.5 Coupled transmons

In order to entangle two qubits, it is necessary to engineer an interaction Hamiltonian that connects their degrees of freedom. Several physical coupling mechanisms have been studied in the context of superconducting transmons [42], for example direct capacitive and/or inductive coupling, and even

[8] Note that it is possible to implement a "gate" that transfers population between $|1\rangle$ and $|2\rangle$ for $\omega_d = \omega_{1\to2}$ which is useful for leakage characterization.

couplings via a dedicated coupler, which is a dedicated quantum system (in some cases similar to the qubit itself) that mediates the interaction and in some designs the coupling is tunable. The transmons used in this work are coupled capacitively through a resonator cavity (which can be regarded as a harmonic oscillator), sometime referred to as transmission line resonator or quantum bus[9].

The transmon-resonator-transmon system can be modeled through the Tavis-Cummings Hamiltonian generalized to multi-level transmons [13]. It consists of three terms describing the individual systems and two terms expressing the interaction between each transmon ($\mathrm{Q_H}$ and $\mathrm{Q_L}$) and the resonator [29]:

$$\begin{aligned}\hat{H} = &\overbrace{\sum_i \hbar\omega_{0\to i}^{(\mathrm{Q_L})} \left|i\right\rangle_{\mathrm{Q_L}} \left\langle i\right|_{\mathrm{Q_L}}}^{\mathrm{Q_L}} + \overbrace{\sum_j \hbar\omega_{0\to j}^{(\mathrm{Q_H})} \left|j\right\rangle_{\mathrm{Q_H}} \left\langle j\right|_{\mathrm{Q_H}}}^{\mathrm{Q_H}} + \overbrace{\hbar\omega_r \hat{a}_r^\dagger \hat{a}_r}^{\text{Resonator}} \\ &+ \sum_i g_{i,i+1}^{(\mathrm{Q_L})} \left(\hat{a}_r^\dagger \left|i\right\rangle_{\mathrm{Q_L}} \left\langle i+1\right|_{\mathrm{Q_L}} + \hat{a}_r \left|i+1\right\rangle_{\mathrm{Q_L}} \left\langle i\right|_{\mathrm{Q_L}}\right) \\ &+ \sum_j g_{j,j+1}^{(\mathrm{Q_H})} \left(\hat{a}_r^\dagger \left|j\right\rangle_{\mathrm{Q_H}} \left\langle j+1\right|_{\mathrm{Q_H}} + \hat{a}_r \left|j+1\right\rangle_{\mathrm{Q_H}} \left\langle j\right|_{\mathrm{Q_H}}\right),\end{aligned} \tag{2.28}$$

where ω_r is the bare resonator frequency, $\omega_{0\to i}^{\mathrm{Q}} = \omega_{0i}(E_C^{\mathrm{Q}}, E_J^{\mathrm{Q}})$ the transmon flux-dependent transition frequencies, and $g_{i,i+1}^{(\mathrm{Q})}$ the level-dependent coupling strengths of the transmons to the resonator. Applying the dispersive approximation, $g_{i,i+1}^{(\mathrm{Q})} \ll \Delta_{i,i+1}^{(\mathrm{Q})} = \hbar\omega_{i\to i+1} - \hbar\omega_r$, truncating at the second excited state of the transmons, and assuming no photons in the resonator, the desired Hamiltonian of two directly coupled transmons emerges[10]:

$$\begin{aligned}\hat{H} \approx &\overbrace{\sum_i \hbar\tilde{\omega}_{0\to i}^{(\mathrm{Q_L})} \left|i\right\rangle_{\mathrm{Q_L}} \left\langle i\right|_{\mathrm{Q_L}}}^{\mathrm{Q_L}} + \overbrace{\sum_j \hbar\tilde{\omega}_{0\to j}^{(\mathrm{Q_H})} \left|j\right\rangle_{\mathrm{Q_H}} \left\langle j\right|_{\mathrm{Q_H}}}^{\mathrm{Q_H}} \\ &+ \overbrace{\sum_{i,j} J_{ij} \left(\left|i\right\rangle_{\mathrm{Q_L}} \left\langle i+1\right|_{\mathrm{Q_L}} \otimes \left|j+1\right\rangle_{\mathrm{Q_H}} \left\langle j\right|_{\mathrm{Q_H}} + \left|i+1\right\rangle_{\mathrm{Q_L}} \left\langle i\right|_{\mathrm{Q_L}} \otimes \left|j\right\rangle_{\mathrm{Q_H}} \left\langle j+1\right|_{\mathrm{Q_H}}\right)}^{\text{Resonator-mediated transmon-transmon coupling}},\end{aligned} \tag{2.29}$$

where we have used the tensor product ($\otimes$) to represent two-transmon states, and effective couplings being given by

$$J_{ij} = \frac{g_{i,i+1}^{(\mathrm{Q_L})} g_{j,j+1}^{(\mathrm{Q_H})} \left(\Delta_{i,i+1}^{(\mathrm{Q_L})} + \Delta_{j,j+1}^{(\mathrm{Q_H})}\right)}{2\Delta_{i,i+1}^{(\mathrm{Q_L})} \Delta_{j,j+1}^{(\mathrm{Q_H})}}. \tag{2.30}$$

2.2.6 Transmon two-qubit gate

There are three classes of two-qubit gates in transmons: flux-pulsing based, all-microwave based and flux-pulse modulated. All of them have been shown to achieve fidelities of $\sim 99\,\%$ [29]. Nevertheless, each class has its own drawbacks. In this subsection we focus on the flux-pulsing scheme explaining how the unitary operator of the CPHASE gate appears naturally when considering two capacitively coupled flux-tunable transmons.

[9] The physical chip in Fig. 2.12 shows eight such buses.
[10] See section 2.2.1 of Ref. 29 for an intermediate step.

The CPHASE unitary can be expressed in the computational space $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ as:

$$U_{\text{CPHASE}}(\phi_{2\text{Q}}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\phi_{2\text{Q}}} \end{bmatrix}, \tag{2.31}$$

which has the effect of applying a relative quantum phase when both qubits are in the excited state $|11\rangle$, hence the name controlled-phase gate. Of central interest for this work is the particular case $\phi_{2\text{Q}} = \pi$ commonly denoted as CZ standing for Controlled-Z gate:

$$\phi_{2\text{Q}} = \pi \;\Rightarrow\; U_{\text{CPHASE}}(\phi_{2\text{Q}}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \equiv U_{\text{CZ}}. \tag{2.32}$$

To elucidate the operating principle behind the CPHASE implementation, on which the CZ is base on, it is enlightening to observe the spectrum of energy levels of two coupled transmons with typical parameters. The Hamiltonian describing this spectrum, including the second excited state, written in the $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle, |02\rangle, |02\rangle, |20\rangle\}$-basis, is approximately given by [42]:

$$\hat{H}_{\text{2-transmons}} = \hbar \begin{bmatrix} \omega_{00} & 0 & 0 & 0 & 0 & 0 \\ 0 & \omega_{01} & J_1 & 0 & 0 & 0 \\ 0 & J_1 & \omega_{10} & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_{11} & J_2 & J_2^{'} \\ 0 & 0 & 0 & J_2 & \omega_{02} & 0 \\ 0 & 0 & 0 & J_2^{'} & 0 & \omega_{20} \end{bmatrix}, \tag{2.33}$$

where $\omega_{ij} = \omega_i^{\text{Q}_\text{L}}(\Phi_{\text{Q}_\text{L}}) + \omega_j^{\text{Q}_\text{H}}(\Phi_{\text{Q}_\text{H}})$ and $\omega_i(\Phi_\text{Q})$ is the i-th flux-dependent energy level of each transmon. To a good approximation the coupling strengths are related by a scaling factor $J_2 \simeq J_2^{'} \simeq \sqrt{2}J_1$. Usually chips are carefully designed such that there is a significant difference between $\omega_1^{\text{Q}_\text{H}}(0)$ and $\omega_1^{\text{Q}_\text{L}}(0)$, hence, we name the higher- (lower-) frequency qubit Q_H (Q_L). Henceforth, we will denote a two-transmons state with the $|\text{Q}_\text{L}\text{Q}_\text{H}\rangle$ convention. Plotting the eigenenergies of Eq. (2.33) $2\pi f_{ij} = \omega_{ij} - \omega_{00}$ as function of the local magnetic flux of Q_H (Fig. 2.9) reveals a rich physics. First, besides the expected quadratic dependence (in first order) (Fig. 2.6), we note that, as we increase the flux, the non computational state $|02\rangle$ decreases in energy much more than $|01\rangle$ as expected. As a consequence, it is expected to "cross" the $|11\rangle$. Furthermore, due to the negative anharmonicity of this type of transmon, the $|02\rangle$ crosses $|11\rangle$ before the crossing of the computational states $|01\rangle$ and $|10\rangle$.

When zooming in at the point where $|11\rangle$ and $|02\rangle$ would become degenerate [Fig. 2.9(b)] we instead find what is know as an *avoided crossing* – the manifestation of the off-diagonal couplings in Eq. (2.33).

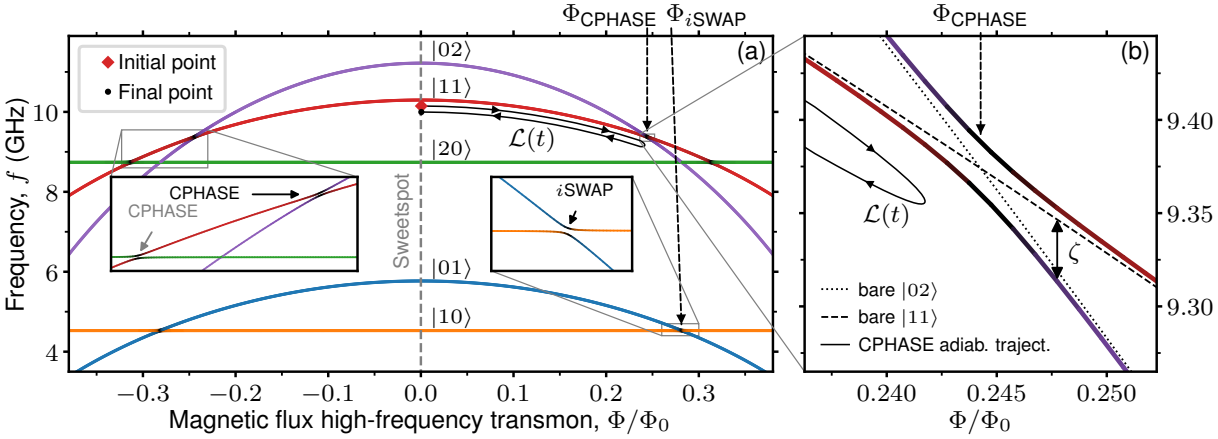


Figure 2.9: (a) Approximate eigenenergies spectrum of two coupled transmons with typical parameters as a function of the (local) magnetic flux in the SQUID loop of the high-frequency transmon, normalized to the flux quantum Φ_0. Only the first five excited levels are included. All branches are symmetrical with respect to the sweetspot. Single qubit excitations $|10\rangle$ and $|01\rangle$ implement an iSWAP gate at $\Phi_{i\text{SWAP}}$ (inset on the right). Interactions of $|11\rangle$ with non-computational levels $|20\rangle$ and $|02\rangle$ give rise to CPHASE-type avoided crossings (inset on the left). Black path between red diamond (initial point) and black dot (final point) indicates typical CPHASE trajectory [$\mathcal{L}(t)$]. The zero point of energy is chosen to be f_{00}. (b) Zoom-in at the avoided crossing between $|11\rangle$ and $|02\rangle$ at Φ_{CPHASE} employed for CZ gating. The "repulsion" of f_{11} from the bare $f_{11} = f_{10} + f_{01}$ is denoted by ζ. The splitting at the center of the avoided crossing corresponds to $2J_2$.

The insets of Fig. 2.9(a) depict 3 such avoided crossings[11] (on each sides of the sweetspot): two of the CPHASE-type and one of the iSWAP-type. The physics involved is the same in all of them, only the interacting levels are distinct. In theory, any can be exploited for the implementation of a quantum gate. However, in practice, only the first avoided crossing is suitable [Fig. 2.9(a), left inset, arrow in black]. One of the main reasons being the increased sensitivity to flux noise at higher flux amplitudes. For completeness, a SWAP gate has the single effect of swapping amplitudes of $|10\rangle$ and $|01\rangle$, while the iSWAP additionally applies an i phase to these states.

To show the accrual of a relative phase we consider exciting both qubits into $|11\rangle$ [Fig. 2.9(a), red diamond] and applying a flux trajectory $\mathcal{L}(\Phi, t)$ towards the avoided crossing with $|02\rangle$ at Φ_{CPHASE} for a total time τ in a slow enough manner compared to the timescale set by J_2 and then returning back following the same excursion [Fig. 2.9(a), black trajectory]. Ideally, this ensures that the moving state is always an eigenstate of the system and $|02\rangle$ is never populated. It is common to label such trajectory as being *adiabatic*.

As a consequence of the dynamic change of the eigenspace along the trajectory, the moving states will experience a time-evolution at different rates compared with the sweetspot time-frame. In the computational space the overall effect is captured by the following unitary:

$$U_{\mathcal{L}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\phi_{10}(\mathcal{L})} & 0 & 0 \\ 0 & 0 & e^{i\phi_{01}(\mathcal{L})} & 0 \\ 0 & 0 & 0 & e^{i\phi_{11}(\mathcal{L})} \end{bmatrix}. \tag{2.34}$$

[11] There is also an avoided crossing between each qubit and the coupling resonator but the by design the interaction with the resonator requires much higher flux. See Ref. 13 for a spectroscopy of an avoided crossing with the resonator.

where

$$\phi_{ij}[\mathcal{L}(t)] = \int_0^\tau \omega_{ij}[\Phi = 0] - \omega_{ij}[\mathcal{L}(t)]\, dt \tag{2.35}$$

is the phase acquired by the state $|ij\rangle$ during the trajectory. Because of the repulsion of $|11\rangle$ due to the $|02\rangle$ near the avoided crossing, the $|11\rangle$ state acquires a phase difference relative to the single excitation states defined as

$$\zeta(t) = \omega_{11}(t) - [\omega_{10}(t) + \omega_{01}(t)] \tag{2.36}$$

Now, if we design a trajectory so that

$$\int_0^\tau \zeta[\mathcal{L}(t)]dt = \phi_{2\mathrm{Q}} = \phi_{11}(\mathcal{L}_\pi) - [\phi_{10}(\mathcal{L}_\pi) + \phi_{01}(\mathcal{L}_\pi)] = \pi, \tag{2.37}$$

where we call $\phi_{2\mathrm{Q}}$ the conditional phase. With this, $U_\mathcal{L}$ becomes

$$U_\mathcal{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & e^{i\phi_{10}(\mathcal{L}_\pi)} & 0 & 0 \\ 0 & 0 & e^{i\phi_{01}(\mathcal{L}_\pi)} & 0 \\ 0 & 0 & 0 & e^{i[\pi+\phi_{10}(\mathcal{L}_\pi)+\phi_{01}(\mathcal{L}_\pi)]} \end{bmatrix}. \tag{2.38}$$

Additionally, $\phi_{10}(\mathcal{L}_\pi)$ and $\phi_{01}(\mathcal{L}_\pi)$ can be nulled with further trajectory design, or by applying weak flux pulses to both qubits; or even by using a virtual-Z gate [55], where virtual means that no physical pulse is applied but the effect is carried over into modifying subsequent single-qubit gates [42]. Finally, we arrive at the CZ unitary:

$$U_\mathcal{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{i\pi} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} = \mathrm{CZ} = \exp\left[-i\frac{\pi}{4}(\hat{Z}\otimes\hat{Z} - \hat{Z}\otimes\hat{I} - \hat{I}\otimes\hat{Z})\right] \tag{2.39}$$

where we emphasize that the avoided crossing with the non-computational state gives rise to an effective ZZ-coupling within the computational state. Of course, the same recipe holds for implementing a CPHASE gate with arbitrary conditional phase $\phi_{2\mathrm{Q}}$. The first CZ gate employing the $|11\rangle$-$|02\rangle$ interaction was experimentally demonstrated a decade ago by DiCarlo et al. [13].

Note that, if the transmons are too close in frequency, the avoided crossing would be much closer to the sweetspot with a non-negligible always-on residual ZZ-coupling that gives rise to detrimental phase accrual [see Appendix A.3 for characterization and impact examples]. Same applies for too strong couplings which motivated reductions from ~ 100 MHz to ~ 10 MHz over the last decade.

Contrasting with the adiabatic approach just presented, we inspect now the ideal extreme of a sudden pulse. After preparing the transmons in $|11\rangle$, we suddenly move the flux amplitude to a point in the vicinity of Φ_{CPHASE}, letting the system evolve for a time τ, then return to the sweetspot in the same sudden manner. In fact, even though real pulses have finite rising time, this is a common experiment used to identify precisely the flux-pulse amplitude (A) of the interaction point. When jumping suddenly into

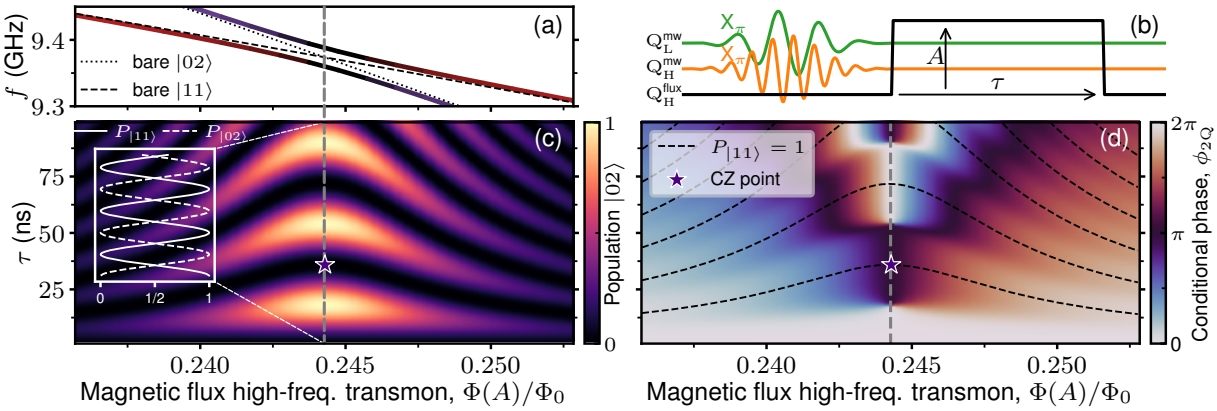


Figure 2.10: Simulation of ideal chevron experiment at the CPHASE avoided crossing. (a) Zoom-in at the avoided crossing at the corresponding flux amplitudes. (b) Experiment waveforms: an X gate is applied to both qubits to prepare $|11\rangle$ followed by a flux pulse of varying amplitude A and duration τ. (c) Chevron pattern: $|02\rangle$ population after the sudden excursion. Complete population swaps occur along the center of the interaction (vertical dashed line) as indicated in the inset. (d) Corresponding conditional phase. Contours of full $|11\rangle$ population recovery from (c) are indicated with black dashed curves. Optimal parameters for a CZ gate are indicated by the star at the center of the chevron crossing the first contour.

the avoided crossing the system performs a Larmor-type rotation [42] – a precession-like phenomenon. Figure 2.10 contains the simulation results of this experiment, namely we plot the probability of finding the system in the $|02\rangle$ state [Fig. 2.10(c)] and the corresponding conditional phase [Fig. 2.10(c)]. The $|02\rangle$ exhibits a typical *chevron* pattern [56, 57]. At the center of the interaction (vertical dashed line) the states swap population at a $2J_2/2\pi$ rate as shown in the inset outlined in white. A CPHASE with arbitrary conditional phase can be implemented along the valleys of full $|11\rangle$ population recovery [black dashed contours extracted from Fig. 2.10(c) and overlapped on Fig. 2.10(d)]. The J_2 rate establishes the fastest CPHASE gates possible, in particular a $t_{\text{lim}} = 2\pi/2J_2 = \pi/J_2$ for the CZ (indicated by the star).

One disadvantage in interacting with states outside the computational space is that, for small offsets from the optimal parameters, the system ends up in superposition with those states being very detrimental in a two-level-based quantum computation model [58]. We name this phenomenon, *leakage* L_1 which represents the rate or probability (per gate) of any computational state leaving the computational space. By extension the inverse process is called seepage L_2, e.g. the energy decay of a leaked or spuriously excited $|02\rangle$ state. We follow the definitions in Ref. 59 for leakage, seepage and average gate fidelity (F).

Being aware of the potential leakage out of the computational state one might consider to employ the iSWAP interaction that only involves computational states and can be faster through its half-duration $\sqrt{i\text{SWAP}}$ variant, however, experimentally there is always a small probability of interaction when trespassing the CPHASE avoided crossing due to finite rise time, which defeats the initial purpose.

Historically, on one side, linear-dynamical distortion in flux-control lines and limited time resolution in arbitrary waveform generators (AWGs) bottlenecked the performance of unipolar square pulses [56]. On the other side, real qubits with limited life-time and a long adiabatic trajectory, as introduced in the beginning of the subsection, is clearly a poor practical combination, and simply speeding up the trajectory translates into leakage. Many efforts went into designing adiabatic trajectories compatible with the quantum error correction threshold ($\sim 1\,\%$). Eventually softened (unipolar) pulses [60, 61]

with fidelities above $99\,\%$ superseded the square counterparts, culminating in a promising fast-adiabatic theory targeted to transmons [34] and only slightly longer than t_{lim}. The research continued in the pursue of higher performance and practical calibrations [62]. However, unipolar pulses have been shown to suffer from long timescale distortions hindering the gate repeatability – essential for quantum circuits of greater depth. As a solution, the Net-Zero (NZ) bipolar pulses based on the fast-adiabatic theory were introduced [39], being the state of the art at the beginning of this work. A further overview of this approach is presented in Section 2.4.4.

2.3 Experimental aspects of quantum computation

2.3.1 Implementations of quantum operations

The physical implementation of quantum operations depends on the type of qubits and chip architecture adopted. In this subsection we briefly look at initial state preparation that is common to most implementations. In Section 2.4 we present an overview of the implementations of quantum gates and measurements (read-out), where we focus on Transmons and the specific chips that were used.

Initial state preparation

Experimentally, some of the difficulties with initial state preparation arise from the thermalization with the environment due to the closeness of the transition energy between the states $|0\rangle$ and $|1\rangle$ and the thermal energy, $k_B T$, typical to many physical implementations.

Two examples of initializing a qubit into a known state are: initialization by measurement and ground state preparation by energy decay. In the first case, we rely on a high fidelity projective measurement that is able to reliably measure the qubit and also project the qubit into the corresponding state of the measurement outcome. Note that, even though the measurement statistics for a specific state may represent the measured state reliably, it is not guaranteed that the post-measurement state is the same, depending on how the measurement is performed. The second example is a common practice that relies on the relaxation of the qubit's energy into the environment. If we just wait enough time the qubit will decay into the ground state, typically represented by $|0\rangle$. In practice, for sub-Kelvin energy scales enough time is on the order of hundreds of microseconds (a few qubit life-times). In the devices used in this work the initialization is based on the energy relaxation.

2.3.2 Faulty physical qubits and quantum gates

In practice, the physical gates applied to the qubits deviate from the theoretical ones. The best error rates achieved in quantum operations, are commonly on the order of 10^{-4}, i.e. one error in 10^4 operations, for single-qubit gates, and 10^{-3} for two-qubit gates [63], while, for comparison, in modern digital circuits gate errors are below 10^{-7} [64].

2.3.3 Qubit connectivity

In order to perform multi-qubit gates, the relevant qubits need to be connected in such a way that it is possible for them to interact. This is achieved in different ways, depending on the implementation. Unlike the quantum circuit abstraction, quantum chips usually do not have all-to-all connectivity between qubits. That is, the qubit connectivity graph is not a complete graph, since such a feature poses great engineering challenges. Furthermore, complete graph connectivity may not even be desirable, since crosstalk[12] is an experimental issue [65]: all-to-all connectivity would facilitate crosstalk, deteriorating the device performance. From a feasibility point of view, the most promising connectivity graph topologies, are the 2D lattices and variants thereof. In such a topology, qubits are connected with the nearest 4 (sometimes 5) qubits. This type of connectivity constraint is called "nearest neighbor" connectivity.

2.3.4 The system stack for a quantum computer

Quantum computers are not entirely quantum, they heavily rely on their classical counterparts plus a lot of specialized, custom made digital and analog electronics. This applies to the current era of Noisy-Intermediate Scale Quantum (NISQ) device and beyond. They will act as an accelerator for the classical processors splitting algorithms optimally between both. An entire technological stack is required for a fully functional general-purpose quantum computer. It might be early to judge if a particular design of the stack is appropriate for a device yet to fully be explored by users, but promising and elaborated designs have already been proposed, such as the one in Ref. 66 developed at TU Delft. It organizes and delineates several interconnected layers of abstraction. In simplified terms, it can be described as follows. A *quantum algorithm* is implemented by a *quantum program* which is interpreted by a (quantum[13]) *compiler*, which outputs instructions of a hardware-aware *microarchitecture* executable by the *quantum-classical* electronics interface to the quantum chip.

2.3.5 Quantum computing long-term prospects

The reader might be wondering how can a useful quantum computer still be built and operated with all the apparent never ending adversities in all the aspects discussed so far. But there is hope! The research has shown that it is possible to achieve useful computations with faulty physical quantum processors [15, 17, 67]. The short and mid term applications [14] might steer the developments into new directions, nevertheless, so far, the long-term goal is to achieve a universal Fault-Tolerant Quantum Computer. Without going deep into the details, the roadmap comprises running a logical quantum circuit encoded redundantly by physical qubits in an architecture such as the Surface Code [17, 67] or variants thereof [68], allowing for Quantum Error Detection and Quantum Error Correction (QEC).

The Surface Code is a quantum error correcting code with attractive properties [42]. It has a fault-tolerance threshold[14] of $\sim 1\,\%$ that has already been achieved in experiments for both single and two-

[12] Crosstalk is loosely used to generally refer to different types of undesired interaction and effects. Ref. [65] is a recent attempt to formalize and systematize the topic.

[13] In the sense that a classical computer is e.g. simplifying and optimally re-arranging the quantum circuit.

[14] Lower bound on the quality of qubit operations that allows for effective quantum error correction.

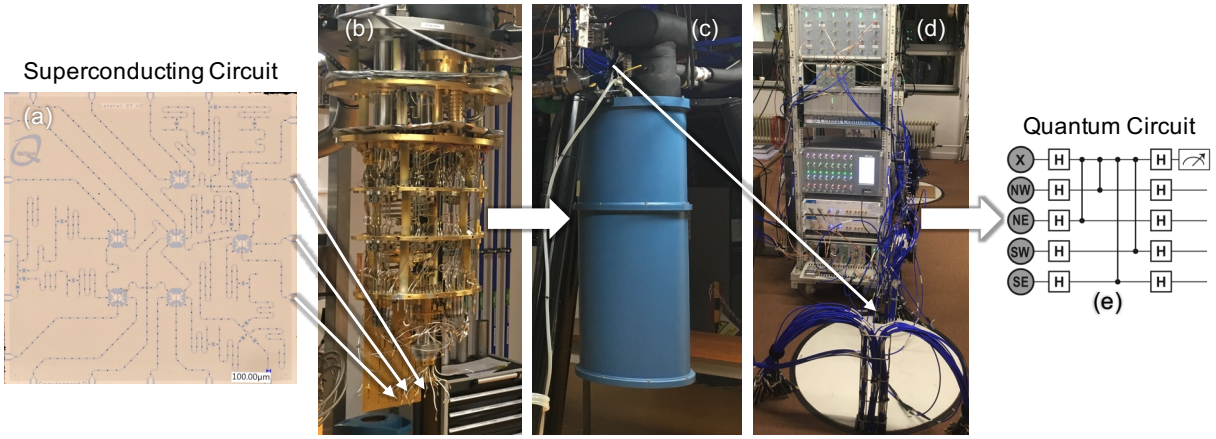


Figure 2.11: Overview of a typical experimental setup employed for transmon research. The quantum chip (a) is placed in a dilution refrigerator (b) maintained in vacuum at mK temperatures (c). By timely outputting electrical pulses and processing output signals, the room-temperature electronics (d) interface with the chip through coaxial cables and implement quantum circuits (e). Credits optical image (a): fabrication team DiCarlo Lab, QuTech. Pictures (b), (c) and (d) were taken in DiCarlo Lab, QuTech. Figure (e) adapted from Ref. 17.

qubit gates, which, in principle, allows the overall error-rate of the logical quantum circuit to be reduced by adding more qubits to the circuit. Combined with the need for only nearest-neighbor coupling, it is one of the most promising quantum error correction codes for medium-to-large scale quantum computing in solid state systems. We are on the corner of the single logical qubit milestone, and many challenges are to be overcome in the field. The work developed for this thesis are aligned with the developments of a Quantum Circuit and Control for a Superconducting Surface Code as described in Ref. 17.

2.4 Transmon experimental overview

An overview of a typical experimental setup used for transmon research is depicted in Fig. 2.11. The superconducting transmon chip is placed inside a dilution refrigerator that maintains it under vacuum at the coldest stage of a few mK. The terminals of the chip are routed up to the room-temperature electronics using coaxial cables (appropriate for each stage of the dilution refrigerator). The electronics are responsible for generating the electrical control pulses and analyzing the output signals. Commonly used instruments include high precision sinusoidal signal generators in the microwave range (in many context referred to as local oscillator, LO), high definition arbitrary waveform generators (HDAWG) with nanosecond resolution and sophisticated analog to digital converters (in our case a Zurich Instrument UHFQA). The instruments responsible for generating the waveforms are orchestrated by a central controller (CC) that is instructed by the classical computer of the experimentalist. The central controller is a critical component that ensures the pulses are timely played according to the intended quantum circuit.

It is important to note that between the room-temperature electronics and the chip there are several other elements (at room-temperature and inside the fridge) that are intended to maximize the signal to noise ratio (SNR) in both directions and isolate the chip environment from the outside noise (and heat). Such components include attenuators, amplifiers and circulators, along with signal splitters, mixer, etc..

Physical property	D1	D2	Z1	X1	Z2	D3	D4
Sweetspot frequency $f_{0\to1,\max}$ (GHz)	6.433	6.252	5.771	5.887	5.785	4.534	4.560
Anharmonicity, α (MHz)	-273.0	-305.0	-318.0	-318.0	-318.0	-317.0	-318.0
T_1 (µs)	39	–	38	53	–	66	–
T_2^* (µs)	55	–	62	93	–	105	–
T_2 (µs)	49	–	58	61	–	40	–
$\|11\rangle$-$\|02\rangle$ coupling strength J_2 (MHz)	D1	D2	Z1	X1	Z2		
Z1	17.34						
X1	17.48	16.66					
Z2		16.74					
D3			13.94	14.57			
D4				13.90	14.59		

Table 2.1: Physical properties of transmons on device C. Some anharmonicities are rough estimates (extrapolated from qubits on the device with same range of frequencies). T_2^* and T_2 are two metrics for the dephasing time. The coherence metrics for D2, Z2 and D4 are omitted due to lack of recent measurements.

2.4.1 The physical chip

During the work developed for this project I had the opportunity to work on two experimental setups, each enclosing in its cryogenic fridge a superconducting device (thin film chip fabricated on silicon wafers), corresponding to two iteration of the same *Lateral Surface-7 (S7)* design. This document refers to them as device C and device S. They were usually operated in an environment at about $10-15$ mK (mixing chamber temperature).

The optical image of the device S is shown in Fig. 2.12 and the corresponding schematic in Fig. 2.14. The chip has 7 qubits designed in a configuration representing a sub-patch of the Surface Code architecture as described in [17]. Qubits color indicate their relative frequencies: purple for high-frequency (targeted at 6.7 GHz), green and blue for mid-frequency (6.0 GHz target) and red for low-frequency (4.9 GHz). For the Surface Code all qubits have connectivity to 4 nearest qubits. In this small patch (Surface 7) only qubit X1 is coupled to other 4 qubits, the rest are coupled only to two other qubits. The inter-qubit bus-resonators are targeted in the $24-29$ GHz frequency range.

All qubits have individual dedicated drive lines (input terminals annotated with brown in Fig. 2.12) that are coupled to microwave sources and used to implement single-qubit gates; and individual flux lines (yellow input terminals in Fig. 2.12) used to implement two-qubit gates between qubits with adjacent frequencies (and also avoid undesired interactions with the non-participating qubits during the two-qubit gates).

Readout is performed through two feedlines (terminals marked with violet in Fig. 2.12) shared between several qubits. Each feedline has an associated resonator (coupled capacitively). Each qubit is coupled to a feedline through a Purcell filter[15] [see Fig. 2.13(b) for a zoom-in on a section of a Purcell filter]. In some cases, if the coupling of the drive lines to the qubits is too weak (or damaged) it is possible to drive the qubits through the feedline instead, which was the case for qubits X1 and D3 on device C.

[15] See Ref. 41 for details on Purcell effect.

Physical property	D2	X1
Sweetspot frequency, $f_{0\to1,\text{max}}$ (GHz)	7.111	6.259
Anharmonicity, α (MHz)	-258.1	-289.1
T_1 (µs)	∼8-18	∼7-18
T_2^* (µs)	∼8-30	∼7-25
T_2 (µs)	∼10-30	∼6-30
$\|11\rangle$-$\|02\rangle$ coupling strength, J_2 (MHz)	18.18	

Table 2.2: Physical properties of transmons on device S. Some anharmonicities are rough estimates (extrapolated from qubits on the device with same range of frequencies).

Figure 2.13 presents a few zoom-ins on the device in Fig. 2.12. From the initial discovery of the transmon the superconducting device have evolved significantly and grown in complexity. Figure 2.13(a) depicts an exemplary transmon qubit. It is composed of two main islands connected by the SQUID loop [Fig. 2.13(c)]. The geometry of the islands define a parallel-plates capacitor and at the same time couple the qubit capacitively with other qubits (orange terminals), the drive line (brown terminal) and the Purcell filter that connects to the feedline (violet terminal). Figure 2.13(c) show the SQUID loop in detail, the Josephson junctions and the flux line loop that generates the magnetic flux in the SQUID loop. In order to guarantee a uniform electrical ground plane on the device, airbridges are used to connect ground plane islands that are separated by waveguides [Fig. 2.13(b)]. As the number of qubits and control lines on a chip increase, ingenious connectivity is required. A first step on alleviating this is the use of the airbridges to crossover waveguides as shown in Fig. 2.13(d). As a next step vertical input/output connections are being investigated.

Tables 2.1 and 2.2 contain the relevant physical properties for the qubits on device C and device S, respectively. In case of device S we list only two qubits as the rest were not usable in the context of two-qubit gates due to fabrication and technical issues. From these tables it is clear that most qubits deviate significantly from the target (maximum) frequencies, which is still one of the main challenges in the field. Despite being possible to operate qubits below their maximum frequency, in this work all qubits were calibrated and operated at the sweetspot (maximum frequency) to minimize flux-noise impact.

2.4.2 Single-qubit gates

The working principle of a typical setup used to implement the single-qubit gates as described in Section 2.2.4 is depicted in Fig. 2.15. The sinusoidal signal from a local oscillator (LO) at a frequency equal to or close to the qubit frequency ($f_{0\to1}$) is shaped by the Gaussian I and Q envelopes using a mixer whose output controls the qubit state through the drive line capacitively coupled to the qubit. The pulse duration used in our setups and targeted for the Surface Code is 20 ns long. It is often the case that the LO is shared between qubits with distinct frequencies for a more compact setup. Figure 2.15 covers such a case: the I and Q incorporate a modulation frequency such that the driving frequency (ω_d) matches the qubit frequency[16].

In order to farther understand the transmon state evolution subject to a driving pulse, it is necessary

[16]Note that even the most advanced AWGs will not be able to generate frequencies above $\sim$ 1 GHz, therefore a dedicated high-frequency source (LO) is essential.

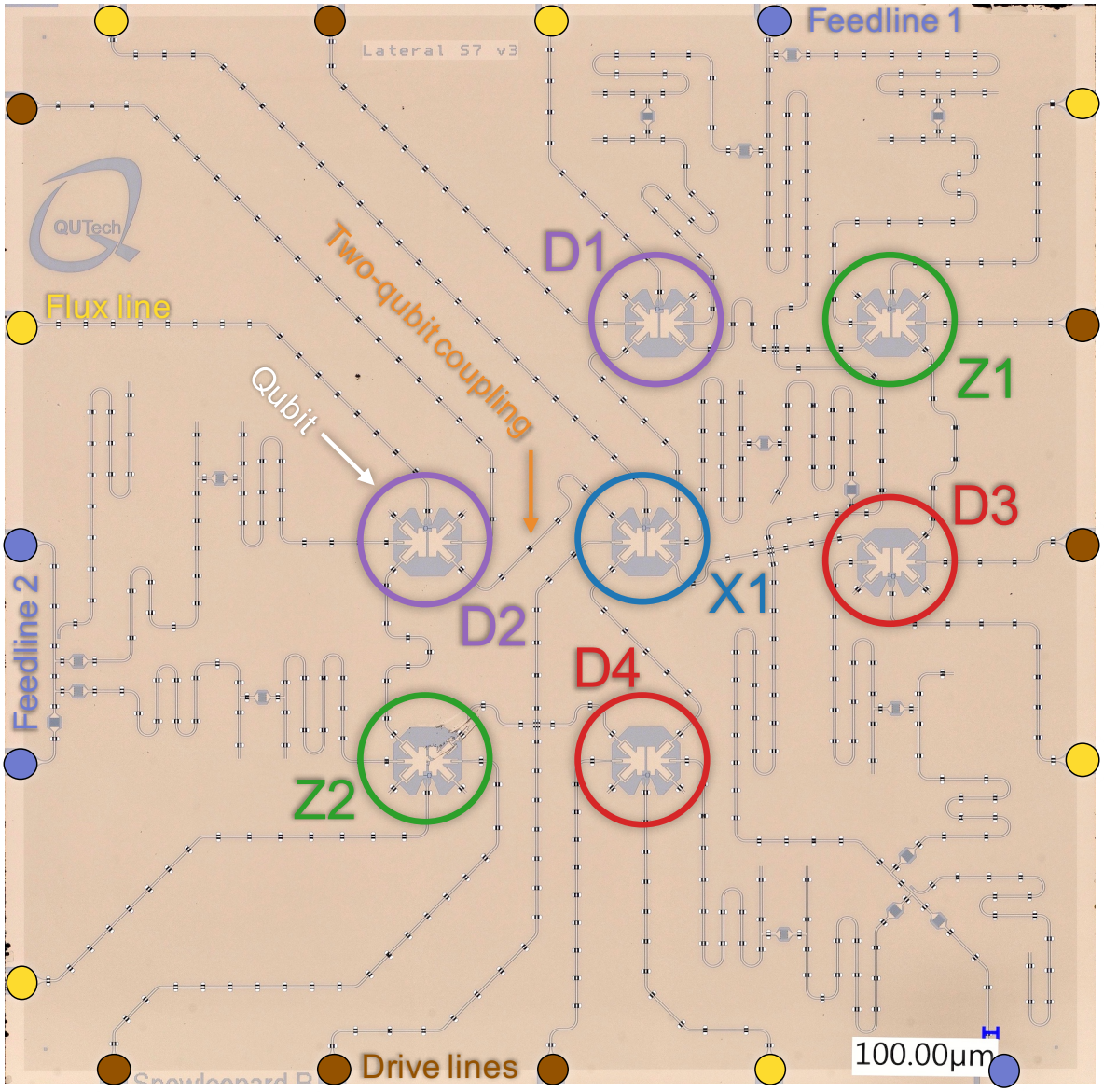


Figure 2.12: Annotated optical image (top view) of device S employing the *Lateral S7* design. Small colored circles on the borders indicate all input/outputs of the chip: flux lines (yellow), drive lines (brown) and feedlines (violet). Corresponding schematics of the chip illustrated in Fig. 2.14. See Fig. 2.13(a) for a zoom-in on an individual qubit. Main text contains further details. Credits original image: fabrication team, DiCarlo Lab, QuTech.

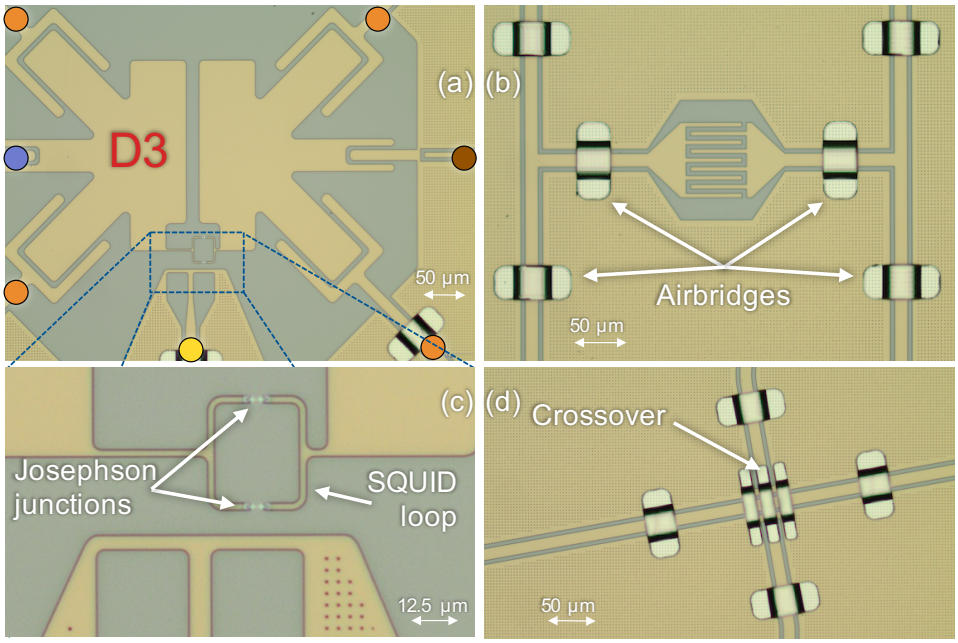


Figure 2.13: Zoom-ins on selected regions of Fig. 2.12. (a) Structure of D3 representative of all qubits. Orange terminals couple to other qubits (only two in this case). Flux line (yellow terminal) is shorted to the ground plane conducting the current that implements the magnetic flux through the SQUID loop. (b) Detailed view of the airbridges and the capacitive couplings of the Purcell filter. (c) Zoom-in on the SQUID loop. (d) Zoom-in on one of the crossovers. Airbridges are used to route a waveguide over another one. Credits original images: fabrication team, DiCarlo Lab, QuTech.

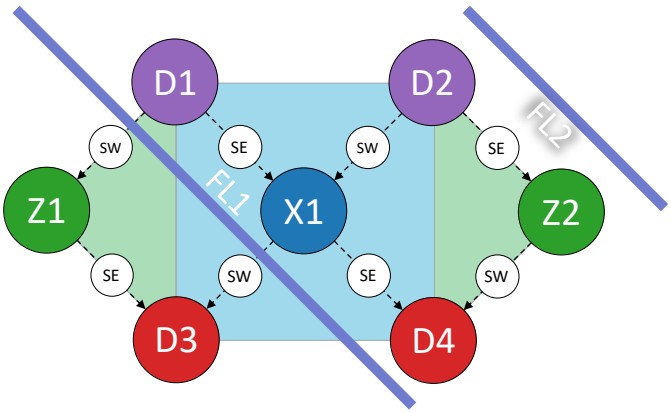


Figure 2.14: Schematic of the Surface-7 devices. Qubits represented by colored circles. Arrows indicate qubit connectivity and their (compass) direction indicates which qubit has its frequency tuned down when performing a two-qubit gate. The two violet diagonal stripes represent the feedlines.

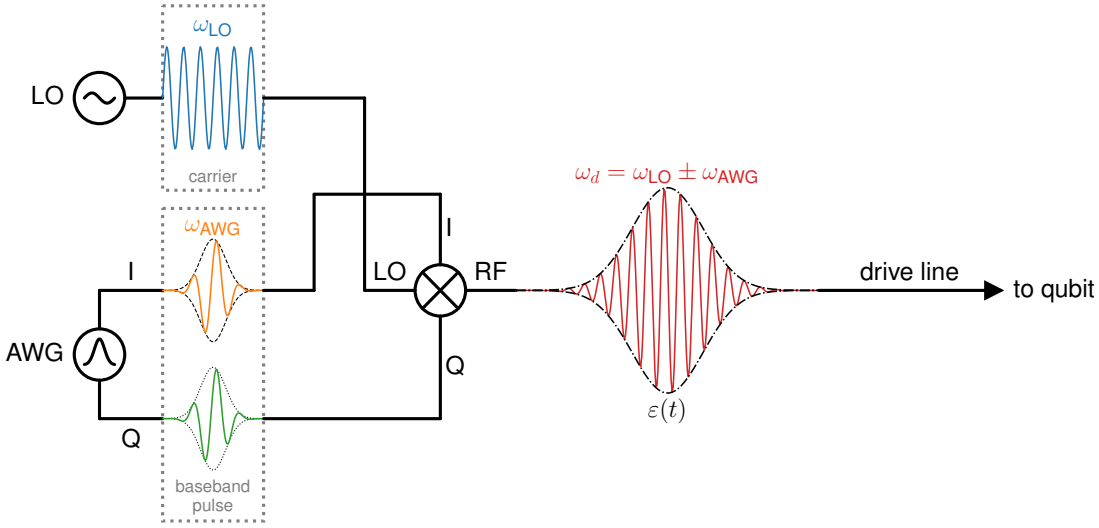


Figure 2.15: Typical experimental setup for implementation of transmon single-qubit gates. The signal from a local oscillator (LO), is combined with the I and Q components generated by an AWG using a mixer. The I and Q components may incorporate a modulation frequency (ω_{AWG}) such that the output signal matches exactly the transmon driving frequency (ω_d).

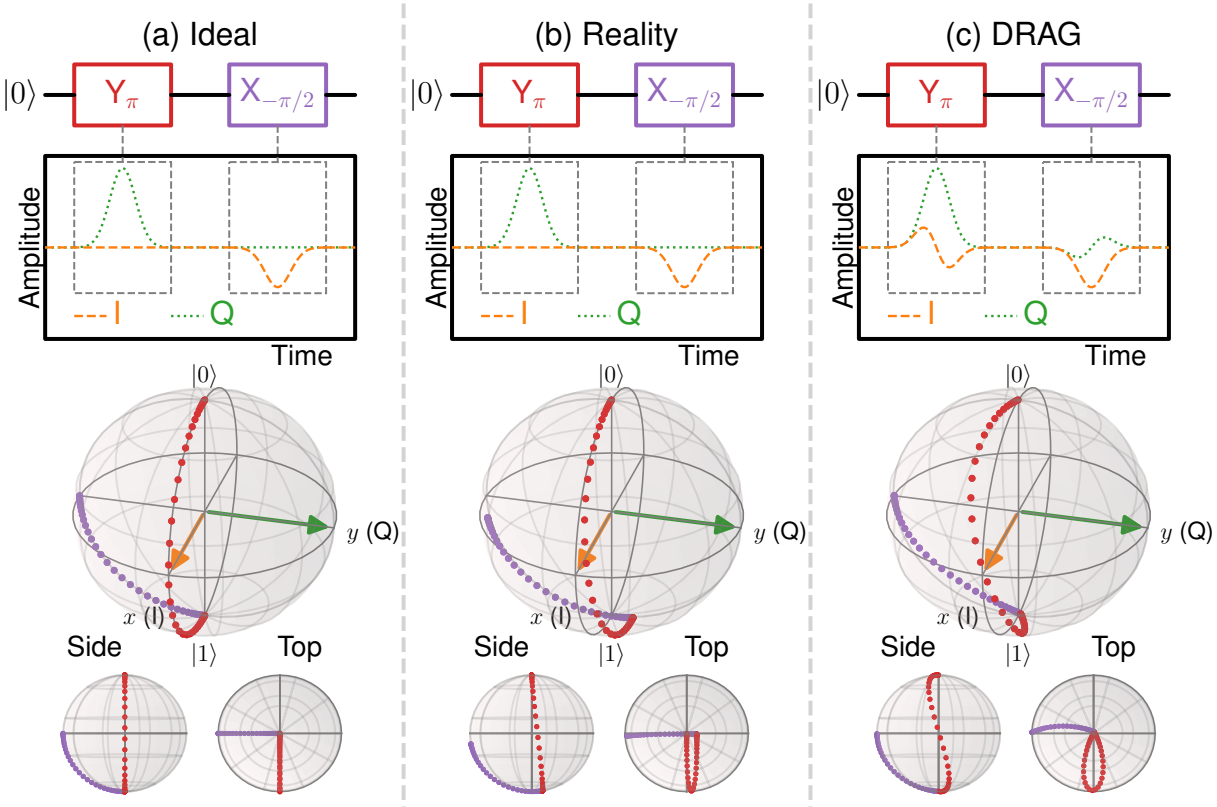


Figure 2.16: Transmon single-qubit gates driving pulse implementation comparison: (a) Ideal qubit, i.e. two-level system. (b) Ideal pulsing for a two-level system applied to a 3-level-transmon. (c) DRAG pulse scheme applied to a 3-level-transmon. Exemplary I and Q driving pulse envelops are depicted inside the black frames as a function of time. Points on the Bloch spheres show the state evolution over time for each gate (colors indicate the corresponding gate). For simplicity only the envelops of the pulses are depicted.

to introduce an intuitive single-qubit gate notation. Throughout this document we write everything in the computational basis $\{|0\rangle, |1\rangle\}$ (eigenstates of $\hat{\sigma}_z = \hat{Z}$) with the corresponding $\{+1, -1\}$ eigenvalues. As the Bloch sphere representation suggests, a single-qubit gate can be seen as a rotation of an angle θ around an axis intersecting the center of the sphere, e.g. rotations around the $\hat{y}$-axis are written as:

$$\mathsf{Y}_\theta = R_\mathsf{Y}(\theta) = e^{-i\frac{\theta}{2}\hat{\sigma}_y} = \cos(\theta/2) - i\sin(\theta/2)\hat{\sigma}_y. \tag{2.40}$$

The shorthand $\mathsf{X} = \mathsf{X}_\pi$ (and similarly for Y and Z) is commonly used and sometimes referred to as a "π-pulse", and similarly "$-\pi/2$-pulse" for $\theta = -\pi/2$.

We can now examine Fig. 2.16 that depicts, for 3 cases, a simple single qubit quantum circuit composed of two single-qubit gates (Y_π and $\mathsf{X}_{-\pi/2}$) applied sequentially in time. Along with it we plot the corresponding exemplary I and Q envelop waveforms. In the same figure, the evolution of the qubit (in the computational basis) is sampled in time and depicted on the Bloch sphere. Fig. 2.16(a) paints the ideal qubit scenario that has only two energy levels and the gates implementation is straightforward with perfect arcs on the Bloch sphere. However, the reality is closer to Fig. 2.16(b), where a transmon with 3 levels is considered, showing how the simple Gaussian envelops are not enough to easily accomplish the desired operations in the computational subspace and without leaving it. The standard technique used to overcome these issues introduced by Motzoi et al. [69] is known as Derivative Reduction by Adiabatic Gate, DRAG for short. It involves modifying the envelops such that one of the envelops is the derivative of the other one up to a factor, λ, to be calibrated:

$$\varepsilon_{DRAG}(t) = \begin{cases} \varepsilon(t) & \text{for I} \\ \lambda\frac{\dot{\varepsilon}(t)}{\alpha} & \text{for Q.} \end{cases} \tag{2.41}$$

Single-qubit gates calibration

The calibration of the derivative amplitude, λ, on its own, presents a trade-off between minimizing leakage out of the computations space and avoiding accumulation of relative phase error between $|0\rangle$ and $|1\rangle$. An extension to the DRAG scheme, that involves changing slightly the driving frequency has been shown to solve both issues [70, 71]. To summarize, the calibration requires a simultaneous optimal choice of the parameters $\{\text{Gaussian envelop amplitude}, \lambda, \omega_d\}$. Note that it is enough to calibrate this parameters for a π-pulse and solve the required parameters for any other rotation. On a daily base, an experimentalist can access qualitatively the calibration state of the single-qubit gates with a simple experiment called AllXY that applies two consecutive single-qubit gates with all the combinations from the $\{\mathsf{I}, \mathsf{X}_\pi, \mathsf{X}_{\pi/2}, \mathsf{Y}_\pi, \mathsf{Y}_{\pi/2}\}$ set. An example is shown in Fig. 2.17. Common patterns of the deviations from the ideal in this figure help troubleshooting for common issues. Each data point on such a plot corresponds to repeating and measuring the same quantum circuit over and over again on the order of 1000-10000 times[17] in order to evaluate the probability of the outcome with low enough statistical error. Besides that there are measurement deviations that only manifest themselves on larger times scales, for

[17] This particular experiment will take a under a minute of wall clock.

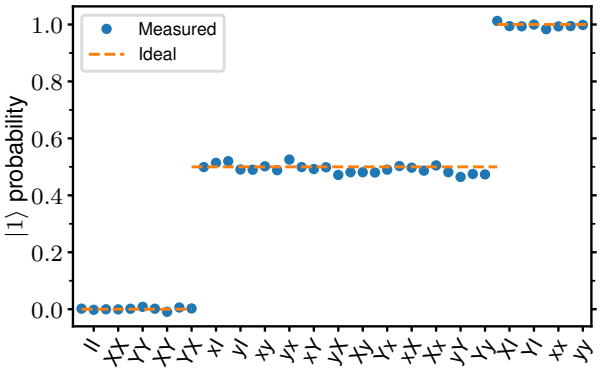


Figure 2.17: AllXY qualitative single-qubit gates metric. Measured on device S, qubit D2. Horizontal axis notation x = $X_{\pi/2}$, y = $Y_{\pi/2}$, X = X_π and Y = Y_π. The two consecutive gates are applied after preparing the qubit in state $|0\rangle$. This example indicates a fairly good calibration. Each combination measured twice (two consecutive data points) for noise evaluation. Dataset generated with routines from Ref. 25.

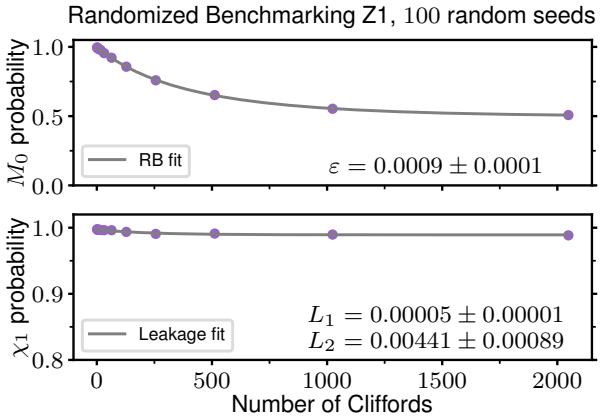


Figure 2.18: Example of an assessment of single-qubit gates using randomized benchmarking with leakage modifications. M_0 is the survival probability of recovering $|0\rangle$. χ_1 is the population in the computation subspace. Horizontal axis corresponds to the length of the random sequences. The RB fit takes into account the leakage but it is low enough to have negligible effect in this case. Dataset generated with routines from Ref. 25.

this reason, in the particular experiment of Fig. 2.17, each data point is evaluated twice (two consecutive blue points) as a mean for qualitatively evaluating the noise (the gates might be well calibrated but the measurement can still be noisy).

For a quantitative assessment the standard technique in our lab is the randomized benchmarking (RB) [72] with modifications for leakage evaluation [73, 74] (and variants thereof). Figure 2.18 presents the analysis of a randomized benchmarking experiment for a good calibration: low error rate $\varepsilon \sim 10^{-3}$ (per gate) and low leakage rate $L_1 \lesssim 10^{-3}$ (per gate). The main plot presents the exponential decay of the probability of recovering the ground state ($|0\rangle$) after applying consecutively longer random sequences of Clifford gates[18] (achieved by appending the single Clifford gate inverting the action of the random sequence). The second plot presents the population of the computation space, χ_1, i.e. the probability of not leaking. L_2 stands for the rate of seepage back into the computational subspace from a leaked or sporadically higher excited state[19].

2.4.3 Qubits readout

As with much of this background overview, there is a lot to be said on the topic of qubits readout (RO) in the context superconducting transmons. This subsection provides a superficial summary based on the more detailed introduction from Ref. 42 and highlights key aspects focused on the implementation in our lab. The quantum measurement is described as entangling a qubit with a variable (observable) of a probe (in our case the superconducting resonator coupling the qubit to a feedline) and performing a classical measurement of the probe – measuring the resonator reaction to a microwave pulse sent into

[18] The Clifford group for single-qubit gates is composed of 24 gates. More than one physical operation per Clifford may be necessary. An extra $\pi_{0\to1}$-pulse needs to be applied at the end of the sequence for L_1 measurement (it has no effect on $|2\rangle$).

[19] L_1 and L_2 are independent rates per gate. L_2 can be higher than L_1 depending on the coherence of computational and non-computational states (here assumed to correspond only to the $|2\rangle$ state).

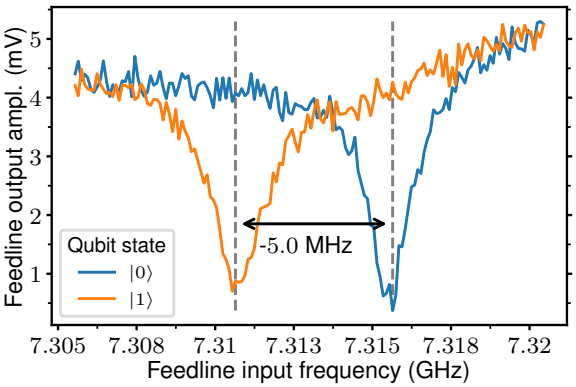


Figure 2.19: Dispersive frequency shift of the readout resonator observed through spectroscopy of the feedline for qubit X1 on device S. Credits datasets: J. F. Marques.

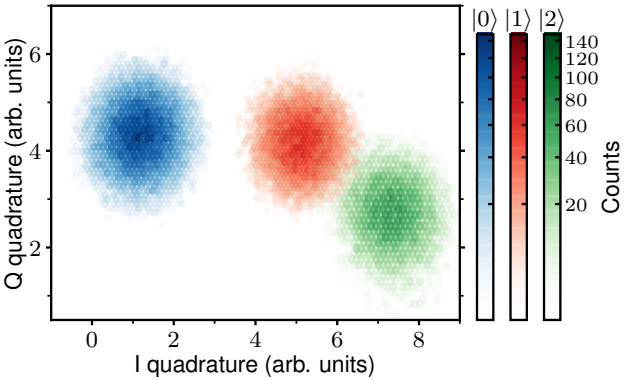


Figure 2.20: Read-out histograms for a transmon, prepared in $|0\rangle, |1\rangle$ and $|2\rangle$, represented on the (I, Q)-plane. For the $|2\rangle$ state the transmon is excited with a driving pulse calibrated for the $|1\rangle \rightarrow |2\rangle$ transition after preparing $|1\rangle$. Dataset generated with routines from Ref. 25.

the feedline. The dispersive (frequency) shift presented earlier allows to map the qubit state onto the response of the resonator. This is clearly seen in Fig. 2.19 where the spectroscopy of the feedline reveals two distinct resonant frequencies depending on the qubit state. There is no energy exchange between the RO resonator and the qubit, instead they affect each other's energy levels. The optimization and calibration of the read-out consists of achieving the best signal-to-noise ration (SNR) of the microwave signal that probes the resonator.

The interaction term, $\chi\hat{\sigma}_z\hat{a}^\dagger\hat{a}$, in Eq. (2.17) commutes with the qubit observable (as an approximation) resulting in a quantum non-demolition (QND) measurement. We aim at QND measurements in our architecture and it is important for running cycles of the Surface Code. Unfortunately, the dispersive regime approximation breaks down for a critical photon number in the resonator which imposes an upper bound on the RO pulse power for a QND measurement. The RO circuit can be implemented sampling the feedline reflection or the transmission. In our architecture transmission is used: each feedline has an input and an output (see Fig. 2.12). Even though Fig. 2.19 might suggest that it would be enough to probe the resonator at the frequency of one of the deeps, it has been shown that the optimal state discrimination happens when probing at the frequency exactly in between the deeps in which case most of the qubit state information is contained in the phase of the demodulated signal. The output RO signal, $s(t)$, can be represented as a static phasor at the known frequency, ω_{RO}, as follows:

$$s(t) = A_{RO}\cos(\omega_{RO}t + \theta_{RO}) = \text{Re}\{\underbrace{A_{RO}\exp(i\theta_{RO})}_{\text{phasor}}\exp(i\omega_{RO}t)\} \tag{2.42}$$

$$A_{RO}\exp(i\theta_{RO}) = A_{RO}\cos(\theta_{RO}) + iA_{RO}\sin(\theta_{RO}) \equiv \text{I} + i\text{Q} \tag{2.43}$$

where A_{RO} and θ_{RO} are the amplitude and the phase of the feedline output, and the exact same information is encoded in an I (in phase) and Q (in quadrature) notation, often referred to as the (I, Q)-plane. Figure 2.20 exemplifies this concept for the (demodulated) read-out voltages of the first 3 levels of an exemplary transmon. The position of the centroids of the histograms can vary significantly even between

similar transmons, namely, with the $|2\rangle$ state at different locations relative two the other two states. The best qubit state discrimination is achieved maximizing the separation between the computational states on the (I, Q)-plane – an optimization that starts at the chip-design level with optimal choice of qubit-resonator frequency detuning.

Conceptually, in order to measure the I and Q components the experimental setup required corresponds (putting it in very simple terms) to the inverse of the one employed for single-qubit gates (Fig. 2.15), i.e. the feedline output is demodulated and the I and Q signals are sampled with a DAC (digital to analog converter). In a single qubit world the most simple setup that could be used, in principle, is known an as *homodyne* measurement that involves an LO at frequency ω_{RO} and simply averaging the digitized signals for the I and Q components. Beside several technical disadvantages, this scheme does not support FDM (frequency division multiplexing), i.e. it does not allow to have a combination of signals with distinct frequencies, each frequency probing a specific qubit resonator simultaneously. The solution scheme is called *heterodyne* measurement that involves a two-step demodulation through an intermediate frequency, ω_{IF}. In a first step, an analog demodulation to unique IF frequencies is used, followed by an extra digital demodulation step that extracts the desired I and Q voltages (or equivalently the amplitude and phase, A_{RO} and θ_{RO}). Reasonable spacing between the frequencies is necessary to avoid interference during the demodulation process. We use heterodyne demodulation in our setups. A feedline is shared between several qubits allowing for more compact hardware and improved scalability.

Readout calibration

The RO calibration comprises the optimization of 3 parameters of the feedline input RO signal: pulse amplitude, duration and frequency. Choosing the frequency does not comprise a big challenge. However, the pulse amplitude and duration are interdependent. Their calibration is time consuming. A calibrated RO involves the integration[20] of the demodulated I(t) and Q(t) signals (usually over about 400 ns) which are weighted by a time series, $w(t) \in \mathbb{C}$, called optimal weights, that is obtained from the difference between the feedline response when probing the transmon in the $|0\rangle$ and $|1\rangle$ states. For each combination of RO amplitude and duration that is tested, it is necessary to determine the optimal weights from new measurements which make the RO optimization take on the order of $0.5 - 1$ h per transmon. In principle, it could be parallelized to calibrate the multiplexed RO of all qubits sharing the same feedline.

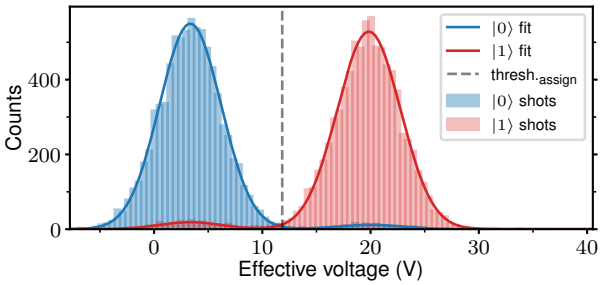


Figure 2.21: Single Shot Read-Out (SSRO) histogram. Gaussian on the left (right) corresponds to the counts when preparing the $|0\rangle$ ($|1\rangle$) state and projecting onto the line crossing the two state histograms in Fig. 2.20. Generated with routines from Ref. 25.

After calibrating the optimal weights, we can rotate the (I, Q)-plane, based on the experiment of Fig. 2.20, in such a way that the states $|0\rangle$ and $|1\rangle$ end up, for example, on the I axis. This is done directly in the physical instrument responsible for the RO (UHFQA) with the advantage that a single

[20] Discrete integration at the sampling rate of the DAC.

quadrature contains the information of the qubit state. Such a RO mode (see Fig. 2.21) is known as Single Shot Read-Out (SSRO) due to the possibility to determine, ideally, the qubit state from a single RO pulse. Furthermore, as indicated by the vertical line in the same figure, an optimal threshold can be used (configured in the UHFQA) to discriminate both states for each individual measurement with a binary output, i.e. a digitized RO mode. This mode is essential for future applications, namely classical feedback, essential for quantum error correction, that requires the ability to apply quantum gates conditional on the output of previous measurements, in real time, within tens of nanoseconds (fraction of qubit lifetime).

2.4.4 Two-qubit gate: Net-Zero

The CZ implementation makes it inherently vulnerable to flux-noise due to increased sensitivity away from the sweetspot, and to leakage of information from the computational subspace due to the interaction with the $|02\rangle$ state. The latter issue was significantly alleviated by the fast-adiabatic pulse shaping [34]. However, unipolar pulses (adiabatic or not), if uncompensated, suffer from distortions in the flux-control lines originating from limited waveform-generator bandwidth, high-pass bias tees[21], low-pass filters, impedance mismatches, on-chip response, etc., which may result in a history-depend flux pulse impacting the gate repeatability. For experiments with quantum operations predetermined before execution, calculating predistortions has been the common practice, unfortunately, as we move towards a programmable quantum computer supporting classical control flow and feedback, this becomes unfeasible at run-time [39].

At the beginning of this work, the Net-Zero (NZ) bipolar flux pulse for the implementation of the CZ gate was the state of the art after its recent discovery in the same laboratory by Rol et al. [39] and quickly adopted in several experiments [18, 75, 76]. It solves the repeatability problem with great performance, both in terms of fidelity ($> 99\,\%$) and leakage ($> 0.1\,\%$). This is achieved with a zero-average bipolar flux-pulse [Fig. 2.22(a)] made possible by the transmon flux-symmetry [Fig. 2.22(b)] that allows to use the $|11\rangle$-$|02\rangle$ avoided crossing twice (one time on each side of the sweetspot). The resulting gate is as fast as the unipolar counterpart. Additionally, when interacting with $|02\rangle$ for the second time, an interference effect contributes to reducing leakage (partial analogy to a Mach-Zehnder interferometer). Most importantly, the zero-average property of the NZ provides resilience against long-timescales distortions (remaining in the flux lines after pre-compensation with real-time filters), hence the dependence on the flux-line history is diminished and the CZ becomes repeatable [29, 77]. Following the main NZ pulse, due to current hardware limitations, the single-qubit phases, ϕ_{01} and ϕ_{10} (see Section 2.2.6), are calibrated to zero

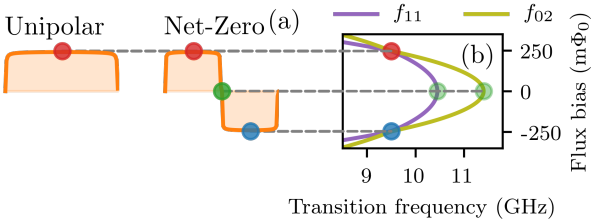


Figure 2.22: (a) Schematic comparison between an unipolar adiabatic pulse and the Net-Zero bipolar pulse implementing CZ gates. (b) Corresponding transmon energy levels, $|11\rangle$ and $|02\rangle$, as a function of the flux (linearly dependent on pulse amplitude). Adapted from Ref. 39.

[21] In general, a transmon requires a static magnetic flux to compensate sweetspot offsets. A bias tee can be used to combine a DC current and the output of an AWG for flux pulsing. Repeated unipolar pulsing may result in a drifting bias tee.

through two simultaneous weaker and shorter bipolar flux pulses on each qubit [inset in Fig. 2.24(a)].

Cryoscope technique overview

Tapping into the high performance NZ-based gates requires a good characterization and compensation of the distortions in the flux lines that prevent a target waveform $A(t)$ generated by an AWG to retain its shape when reaching the transmons. Modeling the flux line as a linear time-invariant system, the issue is solved by first characterizing its impulse response $h(t)$ and reverting it with a filter $\tilde{h}^{-1}$ [29]. The Cryoscope technique has been developed to achieve this [31]. As a brief overview, the Cryoscope essence consists of an experiment in which we move the Bloch vector of the qubit to the equator, by applying e.g. an $X_{\pi/2}$ gate, followed by a square flux pulse A_τ of fixed amplitude. When the qubit is moved away from its operating point (sweetspot) its Bloch vector will experience a precession on the equator plane (when measured in the sweetspot frame). The accumulated quantum phase of the qubit, at the end of the pulse, corresponds to an integration of its frequency detuning due to the (distorted) flux trajectory that it experiences. In subsequent experiments, we increase the pulse duration τ in steps at the maximum resolution allowed by the AWG. To an excellent extent, the accumulated phase as a function of τ can be discretely derived to recover the frequency trajectory, and therefore also the flux experienced by the transmon at each τ, hence $h(t)$ is determined. Most of the time, $\tilde{h}^{-1}$ is well approximated by a combination of one finite impulse response (FIR) and 3 or 4 infinite impulse response (IIR/exponential) filters, both of which are readily available at hardware level in the Zurich Instruments HDAWG waveform generator with the predistortions calculated in real-time.

NZ calibration landscapes

After we are confident that the flux-line response has been precompensated, the calibration of the NZ waveform begins. Usually, the total pulse duration is fixed above the minimum imposed by the coupling strength (J_2). Two main parameters θ_f and λ_2 determine the shape of a half-pulse through a non-linear parameterization inspired from Ref. 34. Loosely speaking, θ_f defines the overall amplitude of the pulse with the center of the avoided crossing at $\theta_f = \pi/2$, and λ_2 the sharpness of the pulse rise/fall. Figure 2.23 displays an example of the measured landscapes of the conditional phase ϕ_{2Q} and corresponding leakage estimate $\tilde{L}_1$ in this parameter space for the transmon-pair used by Rol et al. [39]. Along with it, the authors have demonstrated well matching simulations when modeling the system realistically based on experimental characterization of the device [Figs. 2.23(c) and 2.23(d)]. It is easy to see that an optimal region where the 180° conditional phase contour coincides with the low leakage fringe at $\theta_f \approx 63^\circ$ is available for a CZ calibration.

With further optimization it was possible to push the performance of this gate above $99\,\%$ as characterized by the fidelity (F) to the CZ unitary extracted from the standard interleaved two-qubit randomized benchmarking (2QIRB) with modifications to account for leakage [Fig. 2.24]. Unfortunately, part of the optimization is somewhat laborious because, besides feeding the simulations with accurate characterization of the device, it requires to first heuristically optimization the pulse duration in simulation. This

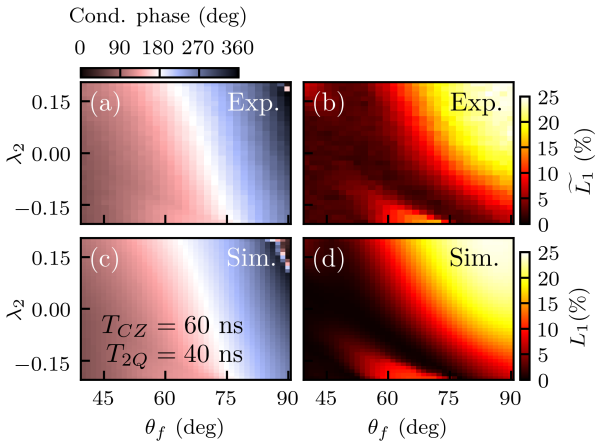


Figure 2.23: Conventional NZ calibration landscapes example and matching simulation. For rough intuition, θ_f controls (non-linearly) the overall pulse amplitude and λ_2 the sharpness of the rise and fall of the pulse. (a, b) Experimental landscapes of the leakage estimate $\tilde{L}_1$ and corresponding conditional phase. Optimal calibration conditions were met at the overlap of the 180 ° contour with the low-leakage fringe. (c, d) Closely matching simulations. Adapted from Ref. 39.

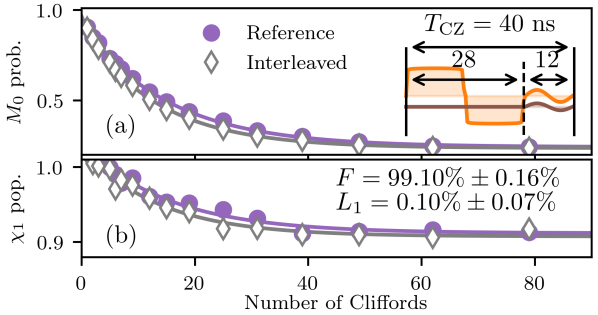


Figure 2.24: NZ CZ performance assessment characterized by two-qubit interleaved randomized benchmarking (2QIRB) with leakage modification. Horizontal axis corresponds to the number of two-qubit cliffords in the randomized sequence in the reference curve. The interleaved curve corresponds to interleaving the two-qubit cliffords of the reference curve with the CZ gates for the same randomized sequence. (a) $|00\rangle$ survival probability, M_0. (b) Computation subspace population, χ_1. Adapted from Ref. 39.

scales up the complexity of the problem even in simulation. Note that in experiment it is almost unfeasible owing to the $\sim 30\,\mathrm{s}$ necessary to evaluate a single point on the landscape shown in Fig. 2.24. Furthermore, despite its current adoption, the complex structure and interdependence of the calibration landscapes limits the use of the NZ as quantum processors grow in qubit number. It is worth emphasizing that the qualitative features of the landscapes of Figure 2.23 vary significantly between pulse durations and among transmon-pairs as testified during this work. The overlap of the lowest leakage with $\phi_{2Q} = 180°$ contour is a condition rather rare and hard to meet. This remaining challenge is successfully addressed in Chapter 5 superseding the conventional NZ discussed so far. For this reason, some aspects specific to the NZ only will not be explored thoroughly.

Conditional oscillation experiment: CZ calibration

We shall now elaborate on how the conditional phase (ϕ_{2Q}), the leakage (L_1) and the single-qubit phases (ϕ_{10} and ϕ_{01}) are estimated experimentally and how these experiments are used to calibrate a CZ gate (or an arbitrary CPHASE gate). To this end, several iteration of the conditional oscillation experiment [29] are carried out. The procedure is illustrated in Fig. 2.25 and was extensively used throughout this work. The left-side plots correspond always to the high-frequency qubit. Due to potential drift of the system parameters during the optimization process, RO calibration points for all 4 computational states are measured and used as normalization reference (purple triangles). The uncertainty of the quantities of interest measured is omitted for figure clarity.

Each Ramsey-like experiment consists of two variants of the same quantum circuit. In the first variant, the control qubit ($\mathrm{Q_{contr.}}$) is prepared in $|0\rangle$ and considered to be *off* [Fig. 2.25(a)]. In the *on* variant, it is

prepared in $|1\rangle$ with an X_π gate [Fig. 2.25(b)]. The two variants allow to estimate ϕ_{2Q}, L_1 and the single-qubit phase of the qubit taking the role of the target ($Q_{targ.}$). In both cases, $Q_{targ.}$ is prepared in a maximal superposition by moving its Bloch vector to the equator through an $X_{\pi/2}$ gate. Next, follows the flux pulse intended to implement the CZ. At the end, a $\varphi_{\pi/2}$ operation is applied to $Q_{targ.}$ that corresponds to a single-qubit rotation of $\pi/2$ around an axis on the equator that makes an angle φ with the $\hat{x}$ axis. In parallel, in the *on* sequence a second X_π pulse is applied to $Q_{contr.}$. The blue solid (silver dashed) curves indicate the experimental results from the *on* (*off*) variant. Each data-point on a curve is obtained by averaging (typically in hardware) over 1000 repetitions of the quantum circuit for each φ.

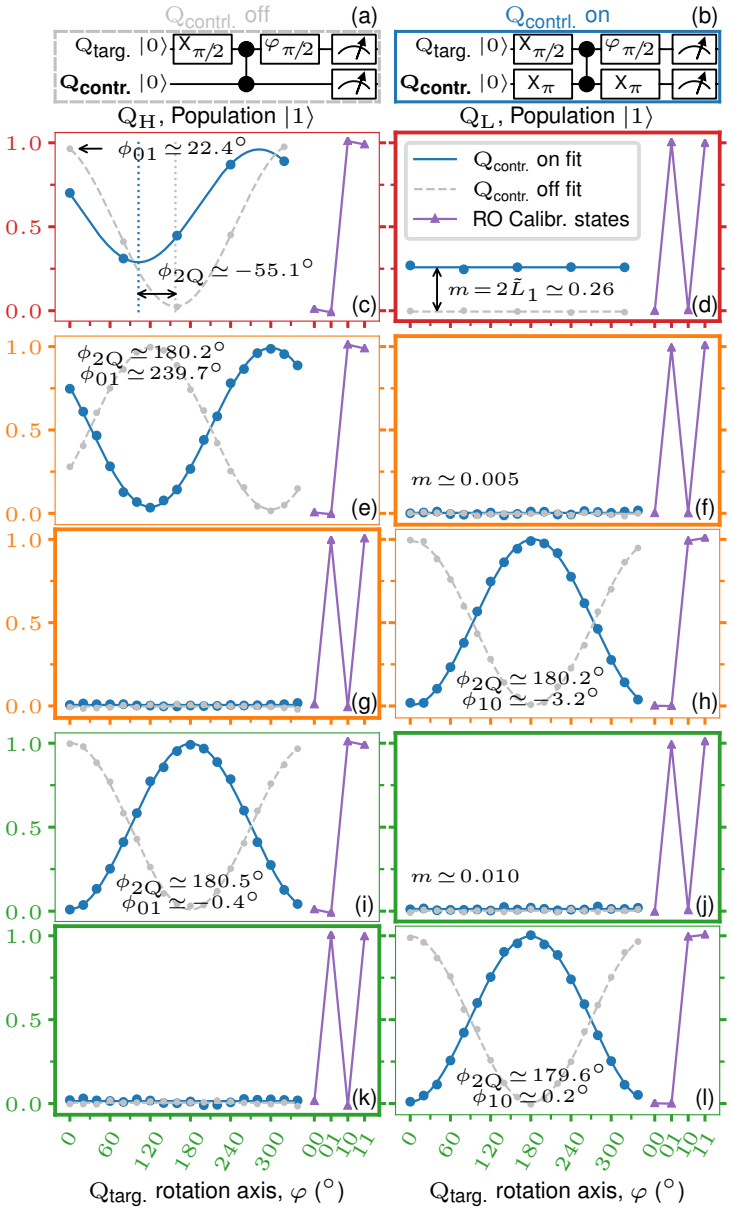


Figure 2.25: CZ calibration steps via conditional oscillation experiments. See main text for details. Datasets generated with routines from Ref. 25.

The single-qubit phase is measured directly from $Q_{targ.}$ oscillation when $Q_{contr.}$ is *off* [see Fig. 2.25(c)]. The other single-qubit phase requires a measurement of the *off* sequence in which the qubits swap the control and target roles [Figs. 2.25(h) and 2.25(l), the thick frame indicates the qubit assuming $Q_{contr.}$ role]. The conditional phase is calculated as the difference between the phase acquired by the $Q_{targ.}$ in the *on* and *off* sequences: $\phi_{2Q} = \phi(Q_{targ.})^{off} - \phi(Q_{targ.})^{on}$. When the role of the qubits is swapped we might find some small discrepancies in its value due to systematic coherent errors (e.g. RO crosstalk, residual ZZ-coupling).

To understand how leakage is estimated, consider the case when $Q_L = Q_{contr.}$ and the system is left in the $|02\rangle$ state after the flux pulse. The second X gate on $Q_{contr.}$ will produce an excitation ($|02\rangle \rightarrow |12\rangle$), instead of removing it as it would be with no leakage ($|11\rangle \rightarrow |01\rangle$). Under no noise and no leakage to other levels, defining the *missing fraction* $m = P_{|1\rangle}(Q_{contr.})^{on} - P_{|1\rangle}(Q_{contr.})^{off}$ we would have $L_1 = m/2$. Realistic simulations have shown that this holds well enough [39], and therefore we estimate leakage as $\tilde{L}_1 = m/2$. Even though the transmons suffers from energy relaxation which makes $\tilde{L}_1$

overestimate L_1 slightly, for calibration purposes the locations of the lowest leakage will be accurately determined. Note that it is not possible to estimate leakage with this procedure when $\mathrm{Q_H} = \mathrm{Q_{contr.}}$. As an alternative the $|2\rangle$ population could be measured directly (see Fig. 2.20) at the cost of added complexity to the RO scheme, data processing, etc. and crucially time overhead. Similar considerations discard the randomized benchmarking for practical calibrations of this two-qubit gate.

With rough knowledge of the flux-pulse parameters that bring $|11\rangle$ and $|02\rangle$ close to interaction a non-zero conditional phase will be observed likely with significant leakage [Figs. 2.25(c) and 2.25(d)]. This is an uncalibrated gate (denoted by the red frames). The calibration is composed of two sequential stages (the final state of each is denoted by the orange and green frames). In the first stage, the parameter space of the main flux pulse is explored in order to find the target $\phi_{2\mathrm{Q}}$ (180° for CZ) and minimize $\tilde{L}_1$. In general, these quantities are tightly interdependent in non-trivial ways in the NZ case, hence the search space is at least two-dimensional. This stage is the most challenging and time-consuming.

Upon success [Figs. 2.25(e) to 2.25(h)] we move to the next stage: single-qubit phases calibration. After the main CZ pulse, the fluxed transmon will acquire an arbitrary single-qubit phase determined by the excursion it was subject to, while the static transmon will typically acquire a very small phase, yet both must be compensated. This is done through a weak flux pulse on each qubit. Virtual-Z gates would be preferred instead, if precise real-time control of the phase of the microwave sources for single-qubit gates was available [22]. For weak enough pulses the effect on leakage and conditional phase is negligible, hence they can be calibrated independently (and ideally also independently from the other weak pulse). Figures 2.25(e) and 2.25(h) quantify the single-qubit phases to be canceled. From the known dependence of the qubit frequency on the flux, the amplitude of a (bipolar) flux pulse with fixed duration ($\sim 10\,\mathrm{ns}$) can be estimated for each qubit such that ϕ_{10} and ϕ_{01} are nulled. A few ($\sim 5-10$) repetitions of the conditional oscillation measurement, for each transmon, as a function of that amplitude, followed by a fit complete the CZ calibration.

Two final conditional oscillation experiments confirm a well calibrated gate (green panels, Figs. 2.25(i) to 2.25(l)) with phases targeted under 1° and $\tilde{L}_1$ below $1\,\%$. The more robust metrics are used afterwards for a final performance assessment, e.g. 2QIRB [Fig. 2.24].

Two-qutrit flux-pulsing simulations

The developments of the NZ CZ gate was extensively backed by simulations [29, 39]. The core code-base of these simulations (available in [25]) has been used throughout this work to emulate the physical system. This allowed for faster testing and development of analysis and calibrations tools. First, it removes the need for a live connection to the experimental setup, which is an expensive and shared resource in the lab. And second, it is a one order of magnitude faster to simulate a data point (no noise) than to measure a conditional oscillation experiment ($\sim 3\,\mathrm{s}$ vs $\sim 30\,\mathrm{s}$). In this subsection we give a brief overview on these simulations (see supplementary materials in Ref. 39 for thorough details).

Among other quantities, the simulations allow to evaluate the conditional phase ($\phi_{2\mathrm{Q}}$), leakage (L_1)

[22] For experiments with no classical logic involved, the single-qubit phases could also be absorbed by modifying the microwave waveforms following a CPHASE gate. However, this capability is poorly supported in the codeword-based architecture of the lab in which a limited number of waveforms can be triggered during the execution of the quantum circuit.

and fidelity F to a target unitary for a single application of an arbitrarily shaped flux pulse. These can be obtained for configurable system parameters and noise models. The noise model can include relaxation and dephasing effects, as well as (remaining) flux line distortions. Including noise demands more computational resources, however, the location of optimal points is not affected significantly (except for flux-line distortions). Therefore, in this work, the noise modeling was employed only for exploring performance boundaries based on experimental characterization of the system.

The flux pulse is modeled as a trajectory in a two-qutrit[23] Hamiltonian. Internally, only the fluxing of the higher-frequency transmon is considered. For each waveform the solution to the Lindblad equation gives the propagator or time-evolution superoperator from which the quantities of interest are extracted. The time evolution is performed in short steps ($\sim$ 0.1 ns) in which the pulse amplitude is considered constant. That amplitude is determined by linear interpolation of the pulse waveform sampled at the AWG sampling rate ($2.4\,\mathrm{ns}^{-1}$). The resonator coupling the two transmons is excluded from simulation by assuming that it is never physically excited. The coupling strengths J_1 and J_2 take into account the weak flux dependence.

2.4.5 Timings calibration

A critical aspect in the field of modern multi-qubit quantum experiments (and sometimes overlooked) are the demanding requirements on the synchronicity between the different analog channels, distributed among several room-temperature instruments. These requirements become more strict as the number of channels grows along the qubit count. The issue is that we want the signals to arrive in sync at the ports of the qubits at the chip level, however the delays intrinsic to the instruments and the latency of the lines, amplification stages, etc. are generally not known to high precision.

For simple experiments the practice has been to separate the operation with generous time windows such that potential mismatches are accounted for [13]. Of course this strategy has a detrimental effect on the depth of the quantum circuits that can be executed within a qubit lifetime.

For precise synchronization two challenges must be addressed. First, the instruments must support sub-nanosecond configurable delays per individual channel. Modern instruments, like the HDAWG used for flux control, usually have good support for this feature. Second, in order to make use of that feature, it is crucial that the relative latencies between all channels are well characterized. Often, this is far from trivial, given that in general there is no direct access to the signal in fridge. In fact, most of the signal lines are unidirectional and designed to minimize crosstalk.

In the architecture of the central controller the triggering of instruments obeys a $20\,\mathrm{ns}$ timing grid, i.e. an instrument can only be triggered to play a waveform every $20\,\mathrm{ns}$ and pulses that do not match the grid exactly will be penalized with extra idling time. This timing grid also provides a coarse mechanism for controlling the relative delays between instruments. It can be used for a coarse timings calibration for single-qubit control as illustrated in Fig. 2.26. The latency (with respect to the measurement) of instruments responsible for generating the microwave and flux pulses are varied in the steps of the timing grid for an experiment that, when calibrated, corresponds to a square flux-pulse sandwiched

[23] Qutrits: each transmons is considered to have 3 energy levels: $|0\rangle$, $|1\rangle$ and $|2\rangle$

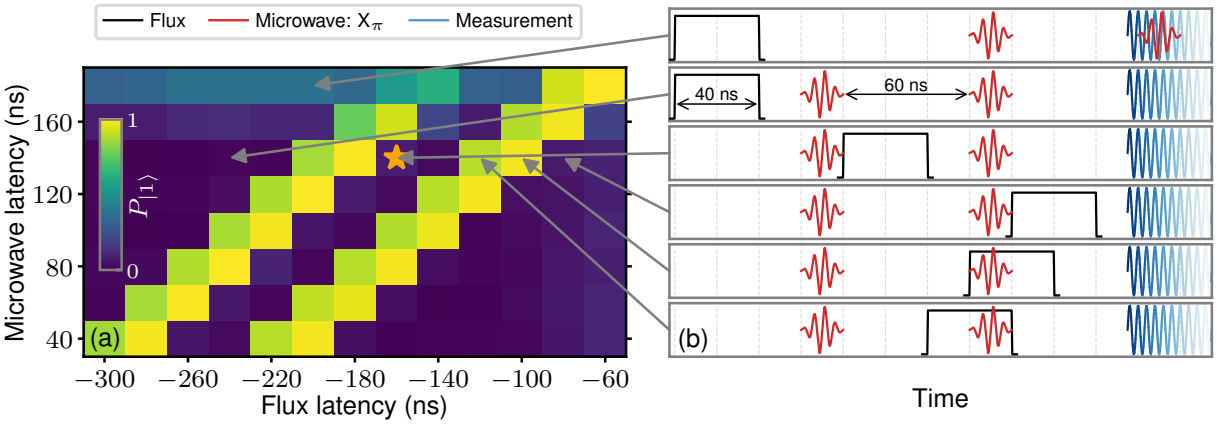


Figure 2.26: Exemplary coarse timing diagram for single-qubit control and measurement. The relative latency of the microwave and flux lines are varied with respect to the measurement in coarse steps of 20 ns. The star indicates a possible optimal choice. (a) Probability of measuring the qubit in the excited state indicating an overlap in time between pulses. Dataset generated with routines from Ref. 25. (b) Illustrative waveform time-lines as experienced by the qubit for cases of special interest.

between two equal microwave gates, followed by a final measurement of the qubit. In this example, we make use of a microwave X gate with a full π rotation such that it prepares $|1\rangle$ and the second microwave gate (ideally) resets it to $|0\rangle$. The result is a timing diagram [Fig. 2.26(a)] with useful features as a consequence of two key cases.

If the flux pulse overlaps either microwave pulse, the transmon will be away from the sweetspot and therefore the microwave frequency will not be resonant with the transmon transition anymore. This leads to measuring the qubit in the $|1\rangle$ state. The two bright diagonals each spanning 40 ns on the horizontal axis (determined by the flux pulse duration) correspond to the overlap with each microwave pulse. The second important case corresponds to the microwave pulse running into the measurement [first panel in Fig. 2.26(b)]. Though not very easy to know what are the resulting dynamics of the system when overlapping the measurement, in most cases the contrast in the timing diagram will change [the first row of Fig. 2.26(a)]. From the several latencies that lead to sandwiched combinations with non-overlapping pulses, we pick the one (indicated by the yellow star) that plays the flux right after the microwave and simultaneously reserves a conservative window before the measurement. Finer relative delays between the microwave and the flux line can be achieved with the sub-nanosecond delay of the HDAWG.

However, it is worth stressing that in multi-qubits experiments all ports of all qubits must be synchronized with a global reference imposed by the synchronicity-constraints in two-qubit operations. Because the timing diagram presented does not allow to accurately characterize the latency between microwave/flux and the RO pulse, it remains a challenge to design a procedure that would synchronize all channels with respect to a global reference. This is already a potential issue for qubits across different feedlines pulsed by separate instruments (two UHFQA for a S7 chip). Finally, we must note that using $X_{\pi/2}$ pulses can be advantageous because the quantum phase of a Bloch vector on the equator is more sensitive to disturbances, nonetheless it might be harder to discern the relevant features on the diagram, hence, it was not used for this illustration.

Chapter 3

Automatic flux bias offset calibration and analysis

After determining the coupling strength between two (connected) transmons through a Chevron-type scans (Fig. 2.10), along with the rough flux-pulse amplitude of the $|11\rangle - |02\rangle$ interaction, and before we can proceed to sampling the CZ calibration landscapes, the essential step is to ensure the higher-frequency transmon is sweetspot-centered to high precision through the static flux-bias. This corresponds to a flux bias offset calibration, which in the lab is also known as the chevron(s) alignment because we employ a chevron-like experiment to this end.

In this chapter we present how this calibration was automated achieving a six-fold speedup of this procedure from $\sim$30 min to $\sim$5 min. The necessary steps were systematized within an adaptive measurement routine followed by a specialized analysis. The measurement routine focuses on yielding a high-quality dataset within the least amount of time. The data acquisition is managed by an algorithm (`Learner1D_Minimizer`) developed to optimally sample a 1D domain, nevertheless providing capabilities beyond its initial purpose and serving as template for a generalization to multi-dimensional domains (`LearnerND_Minimizer`, Section 4.3). Moreover, due to peculiar technical details and requirements of the flux bias offset calibration the capabilities of the measurement control of PycQED (experimental framework of the lab [25]) was extended, to support multiple adaptive sampler instances per dimension (flux pulse amplitude) with outer loops along other dimensions (flux bias current).

Chapter outline The chapter begins by outlining the problem while at the same time presenting the chosen approaches to tackle the sub-problems. It describes the required steps and the developed calibration routine together with the illustration of the associated analysis. The second part of the chapter elaborates on the inner workings of the developed adaptive sampling algorithm responsible for the major speed up. First, we gain some intuition about the framework used to build the custom sampler. Second, we dive into the technical description of the sampling strategy used in the calibration routine. Finally, we overview other interesting sampling modes and features.

3.1 The problem and solution

We discuss the problem while presenting the developed solution at the same time, as this makes it easier to illustrate the former. The deeper technical details of the solution follow in the next sections.

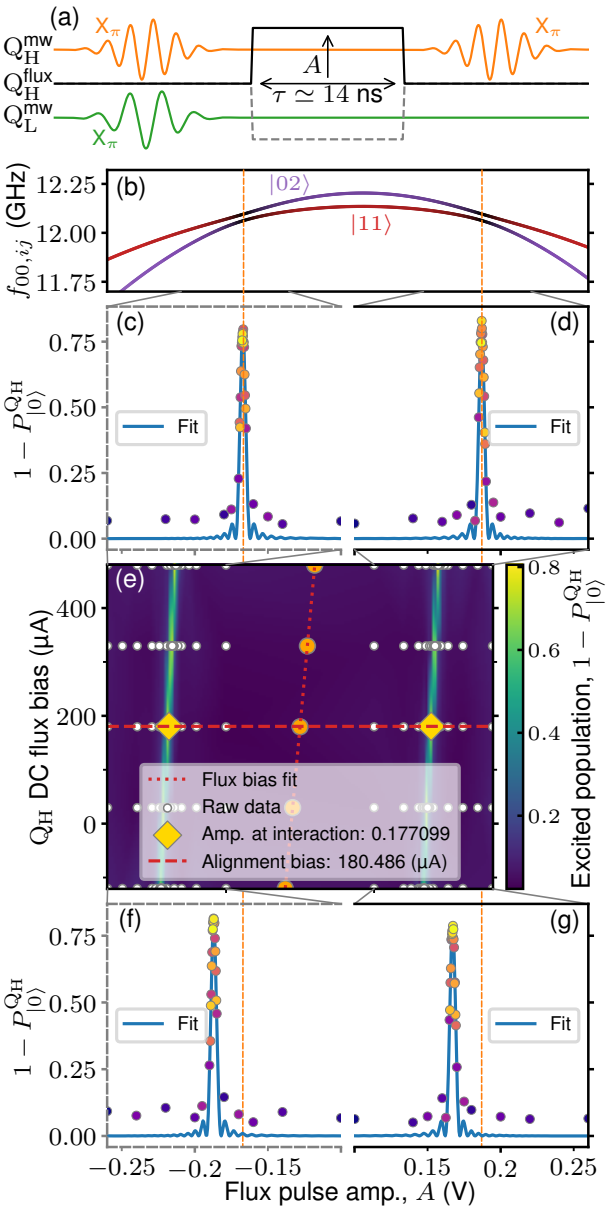


Figure 3.1: Flux bias offset calibration concept and exemplary experimental results analysis. Brightness of data points in (c), (d), (f) and (g) indicates dynamically-determined measurement order, from first (dark) to last (bright). Sampling each 1D cut (fixed bias) takes ∼1.5 min. In (e) only the white dots are measured points, the rest is interpolated to aid the visualization. See main text for details.

So far we have illustrated the flux dependence of the transmon energy levels always with the sweetspot aligned at zero flux, or, equivalently, pulse amplitude [see e.g. Fig. 2.9]. However, experimentally this is never the case. When cooling down a device the sweetspot of the transmons will be found at a non-zero amplitude [see schematic in Fig. 3.1(b)]. Even though a transmon can be operated away from the sweetspot, this is highly undesired due to the increased sensitivity to the flux noise that impacts the qubit coherence. The offset can be compensated with a DC current injected into the flux line through a bias tee[1]. For the higher frequency transmons[2], the bias amplitude that nulls the offset has to be very well calibrated in order to efficiently implement high-fidelity CZ gates of the NZ type which are based on symmetric bipolar flux pulses.

One of the most accurate and precise ways to evaluate if this condition is met is to perform a Chevron-type experiment [see Fig. 2.10]. The pulsing sequence[3] is shown in Fig. 3.1(a) (measurement implied at the end of the sequence). We fix the flux pulse duration and vary the amplitudes in the range of the $|11\rangle - |02\rangle$ interaction. A series of such experiments is performed for negative amplitudes [Fig. 3.1(c)] and another set for positive amplitudes [Fig. 3.1(d)]. The pulse duration can be chosen roughly such that a peak corresponding to the maximum leakage of the high frequency transmon shows up (and an inverted peak on the lower frequency transmon, not shown here). As these figures suggest, if the absolute amplitudes

[1]The other port of the bias tee is AC-coupled to the AWG responsible for dynamic flux control, namely two-qubit gating.

[2]The offset calibration for the lowest frequency transmons is less demanding as they are never fluxed close to an avoided crossing. For these transmons more coarse calibrations can be used.

[3]The final X_π gate allows to confirm the measurement of the $|2\rangle$ state, on which the X gates have no effect.

at the center of the peaks do not match, we conclude that the flux arc has an offset to be compensated for (i.e. the transmon is not exactly at the sweetspot). When changing the DC flux bias both peaks will move by an amount (in units of the AC flux pulse) not known a priori and not straightforward to determine without keeping track of several details regarding the flux line properties. It is worth noting that when running these probing experiments the changes to the flux bias should be as small as possible. This is mainly because the frequency of the previously calibrated single-qubit gates must be very close to the transmon frequency in order to excite it and perform the chevron sequence. In fact, for the same reason, after calibrating the flux bias offset, it is often necessary to re-calibrate the single-qubit gates to recover their high fidelities.

Prior to this work, the bias offset calibration was performed heuristically by a human operator taking on the order of $30\,\mathrm{min}$ to complete. It involved: manually running 1D cuts of the chevron experiment (with evenly spaced points), one for each side of the sweetspot; taking note of the approximate center; manually change the flux bias; repeat the two experiments and determine the centers; iterate a few times until the centers on both sides matched "well enough".

A six-fold speedup was achieved with an automated measurement routine and analysis reducing the procedure to $3-5\,\mathrm{min}$. The entire procedure is contained within a single calibration routine generating a single dataset. The calibration experiment and its results are summarized in Fig. 3.1(e). For illustration purposes, the data corresponds to a transmon that was already aligned at the sweetspot using this procedure. The measurement routine takes the flux pulse amplitude boundaries as input (beside other optional configurations with sensible defaults). The boundaries should span a wide enough range such that the peaks are still within the sampling range on both sides when the DC bias is shifted.

Several chevron 1D cuts with the same fixed flux pulse duration are measured sequentially for at least two distinct DC flux bias values. Here we have measured 5 values to validate the procedure, however in practice 3 values are more than enough: the current bias and shifting the current bias by $\pm(30-100\,\mu\mathrm{A})$. In order to identify the centers of the peaks, for each DC bias, a custom adaptive sampling strategy was developed that finds the peak and optimally samples it. This approach has two main advantages. First, it greatly reduces the measurement time by sampling mostly the regions of interest (the peak). This is specially beneficial considering that a wide enough range of pulse amplitudes is to be sampled when changing the bias that would results in a large amount data for a evenly-spaced grid. Second, by increasing the amount of samples on the peak itself, it facilitates the peak identification in the analysis. This sampler is launched twice, once on each side of the sweetspot. The same is performed for the other bias values. Figures 3.1(c), 3.1(d), 3.1(f) and 3.1(g) show the excellent performance of the sampler despite the noisy measurement.

To integrate these procedures in the experimental framework of the lab (PycQED [25]) while confining all the information within a single dataset, the measurement control[4] was substantial extended to accommodate for the peculiar requirements of this measurement routine. Namely, the generalization to support mixing adaptive sampling (in this case along the dimension of the flux pulse amplitude) with a grid of predetermined data points (here in the DC flux bias dimension). Previously the measurement

[4]See Ref. 29 for more details on the measurement control concept within PycQED.

control supported only one of these modes, either a grid of points or an adaptive sampler in charge of all dimensions. Furthermore, support for running several adaptive samplers within a dimensions was built in to allow sampling the peaks on both sides.

As a final note on the calibration procedure, due to potential static flux cross talk on more crowded chips that lead to flux interdependence among transmons, when targeting high performance it might be necessary to perform chip-level iterations among several transmons in order to converge on DC flux bias values that are optimal for all transmons.

Once the experiment is concluded, the dataset is processed with a specialized analysis developed for this end. The processing involves two steps. First, we fit the analytical model of the chevron (assuming only two coupled levels), simultaneously for both sides for each measured DC bias values. The results of the fit yield the center of each peak very accurately. Second, we take the average of the centers for each bias [denoted by orange circles in Fig. 3.1(e)], and evaluate a simple linear fit (dotted red line) from which the desired alignment bias is determined (dashed red line). For convenience, the flux pulse amplitude at the center of the interaction is also estimated (yellow diamonds) such that it can be used to normalize the pulse amplitude for the SNZ CZ gate.

3.2 Overview of the `adaptive` package

To elucidate how the optimal sampling of the peaks in Fig. 3.1 was achieved we first overview the `adaptive` Python package [78] from which the custom sampler was derived. Quoting the package homepage, *With adaptive you just supply a function with its bounds, and it will be evaluated at the "best" points in parameter space*. This is the feature that we are most interested in as it can reduce experiments duration while enhancing the quality of the sampled data. The package provides the `Learner1D`, `Learner2D` and `LearnerND` algorithms for sampling 1D, 2D and ND functions, respectively, among other algorithms. For the calibration presented in Section 3.1 we build upon the `Learner1D`.

Figure 3.2 showcases the advantage of the `Learner1D` over an homogeneous sampling (evenly spaced grid) under the same budget of sampling points. From Figs. 3.2(a) and 3.2(d) it is easy to see the advantage: the adaptive sampler distributes the points in real-time in such a way that it maximizes the resolution in the regions with interesting features, namely the peak, while in the homogeneous case we got only 4 points on the peak and would require another pass at lower resolution to achieve comparable results.

In Fig. 3.2 we have used the default strategy of the `Learner1D` to determine what are the next "best" segments to sample. Internally, the algorithm assigns a score to each segment which then determines the sampling priority of these segments. The segments with highest score are sampled next. Figure 3.3 illustrates this process. The score is calculated through a function (called `loss` function) that takes as inputs the coordinates of the segment and the corresponding values (of the function being sampled). The package provides an interface for user-defined `loss` functions. The default one returns the area of the trapezium defined by the extremities of a segment. In most cases, this simple scoring strategy leads to very good results when we are interested in discriminating the features of our target function with the

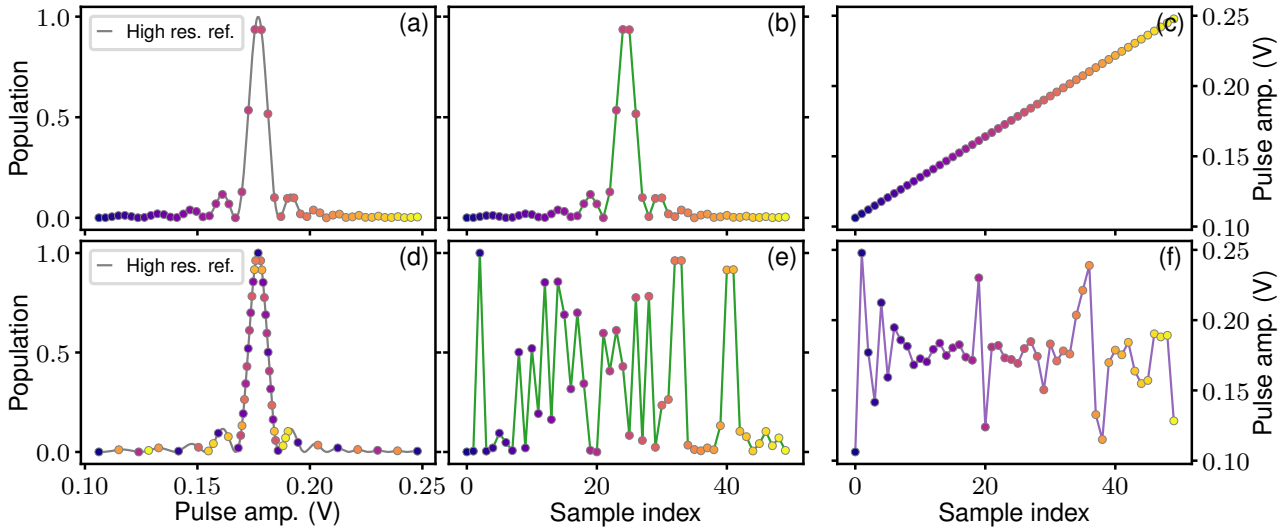


Figure 3.2: Comparison between a homogeneous gridded sampling (a) and the default strategy of the `Learner1D` from the `adaptive` Python package [78] (d) for an equal budget of sampling points. The same chevron model (with constant pulse duration and no noise) is used in both cases. The lightness of the data points indicates the sampling order (sample index). In the simple case of an evenly spaced grid (a), points are predetermined beforehand (c) resulting in a non-optimal resolution. The adaptive strategy (d) determines the next point to be sampled (f) dynamically at run-time based on the accumulated knowledge of the sampled function (e) resulting in an optimized resolution adequate for each part of the domain.

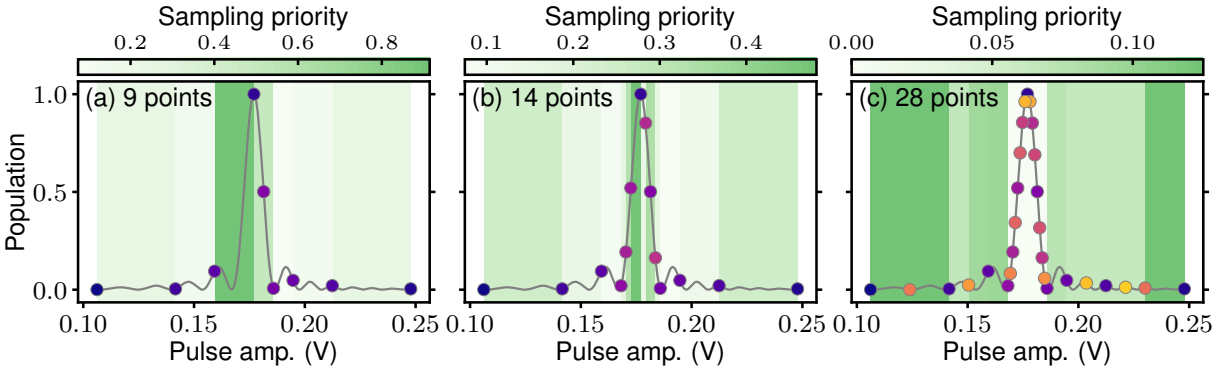


Figure 3.3: Working principle of an adaptive sampler (default strategy of `Learner1D`). The lightness of the data points indicates the sampling order. The next domain segment to be sampled is determined by an internal scoring mechanism which assigns the sampling priority to each segment after receiving new data. The default scoring function uses the area of the trapezium defined by the data points of each segment. As soon as an "interesting" segment is found [(a), dark green] it gets split into smaller segments (b) until their sampling priority becomes comparable to the larger segments (c). The sampling continues until a defined goal is achieved, e.g. maximum number of points.

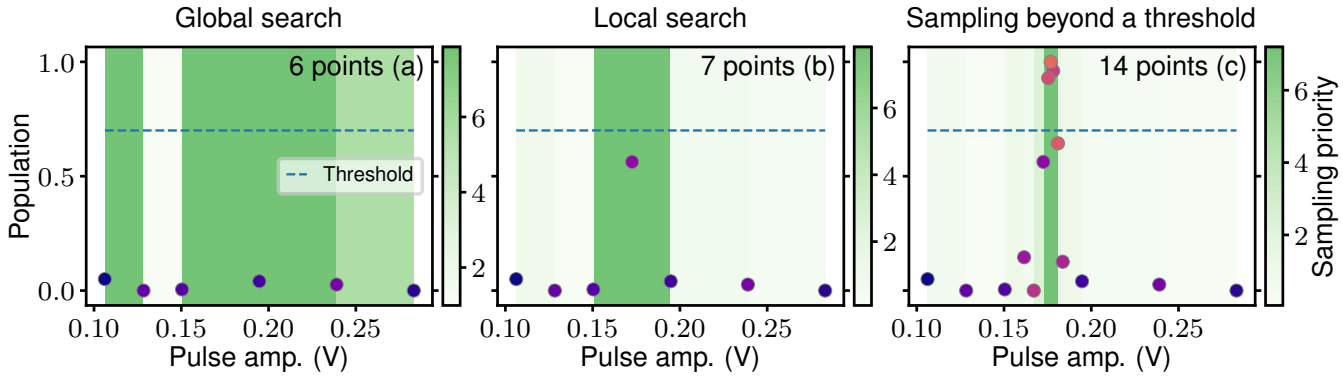


Figure 3.4: `Learner1D_Minimizer` evolution during a flux bias offset calibration. The algorithm has the goal of finding and optimally sampling a peak higher than the threshold. The lightness of the data points indicates the sampling order. (a) In the beginning the sampling priority is mostly guided by the segments size. (b) A local search begins when a new sampled point exhibits a value beyond the range of values known up to that moment. (c) If a value beyond the threshold is detected, the peak will be optimally sampled on both sides (above the threshold) which later leads to accurate estimation of the peak in the analysis. The final result along with all the iterations are shown in Fig. 3.5.

least amount of evaluations.

3.3 `Learner1D_Minimizer`: custom 1D adaptive sampler/optimizer

With a basic understanding of the default sampler, we now describe in more detail how the optimal sampling for the flux bias offset calibration was achieved. Even thought the default strategy gives fairly good results, it is not agile enough to address the needs of our calibration. Hence, the custom `Learner1D_Minimizer`[5] algorithm was developed along with a dedicated scoring (`loss` function) toolbox[6].

The `Learner1D_Minimizer` inherits and preserves the core functionalities of the `Learner1D`. One of the main differences is that `Learner1D_Minimizer` recomputes the score of each segment as soon as a data point with a value outside the previously known range of the sampling function is found. The default `Learner1D` has a latency in this mechanism which is adequate for mathematical functions but not for measurements where function evaluations are significantly more expensive, compared to the overhead of re-evaluating all the scores. The second difference consists of additional logic to support more complex internal states and the control flow of these states beyond the simple scoring function. In particular, we expose the full learner knowledge in the `loss` function (a not only the local information as in the original `Learner1D`).

In the context of the DC bias offset calibration, we can distinguish three states of the algorithm: *global search*, *local search* and *sampling beyond a threshold*. The key difference is how segments are scored.

Global search This is the initial and main state of the learner. The search starts after the boundary points of the domain are sampled. Next, segments are sampled according to their sampling priority

[5]As the name indicates, the resulting algorithm can also be used as an optimizer in the typical case of e.g. minimizing a function. See Section 3.3.1 for other modes of operation.

[6]The resulting code library is available in `pycqed.utilities.learner1D_minimizer.py`.

determined by evaluating a custom scoring function (`loss` function). This stage is depicted in Fig. 3.4(a). Apart from normalizations and some technical details, in simplified terms, the score assigned to each segment is the product of (A) the length of the segment in the domain of the sampling function (x axis here); and (B) the absolute difference (on the y axis) of the corresponding measured values (at the segment extremities) with respect to the most extreme value known[7]. In very loose terms: we combine "how large is a segment is" and "how optimal are the measured values of that segment". The contribution of these two terms to the priority score, is controlled by a parameter, `interval_weight`, specified at the beginning of the sampling. The `interval_weight` takes a value in the [0; 1000] range, where 0 means "always sample the global optimal segment" (known so far), 500 means roughly "the biggest segments have similar priority to the global optimal segment", and 1000 translates into uniform sampling. Using such a definition makes it intuitive for the users and keeps this tool general enough for other tasks beyond this calibration. A glance at the 2D generalization of this mode in Figs. 4.7 and 4.8 might bring further insight.

Local search Every time the learner samples a point whose value lies outside the previously known boundaries (on the y axis), it enters a local search [Fig. 3.4(b)]. During the local search maximum priority is given to the segments that share that new point. The expectation it that it is likely to find even a better point locally. Therefore, this scoring strategy is maintained for a few (configurable) maximum iterations even if new best points are not found. However, the counter keeping track of this state is reset again (allowing for more local sampling) if new best optimal points are encountered. The counter itself is the mechanism that allows to escape from the local optima back into the *global search*.

Optimal sampling beyond a threshold This is the desired final state of the learner for the offset bias calibration [Fig. 3.4(c)]. When launching the sampling, a `threshold` parameter is specified (dashed line in Fig. 3.4), which is an educated guess between the noise background and the expected peak maximum. This state is activated a soon as a data point falls beyond the threshold. We assume this point is part of the peak. The challenge now is to sample both sides such that we acquire enough information for a fit. There many technical nuances in defining a global `loss` function that locally achieves this "bilateral alternating sampling" that we shall skip. In short, we add a new term to the score that is the product between the segment size and the "distance" (on the y axis) from the `threshold`. Effectively, in loose terms, it translates into "sample all the segments beyond the `threshold` with a priority for the segments closer to the maximum of the peak". Finally, the sampling stops after a certain amount of points beyond the `threshold` is accumulated (10 points in out case.). Figure 3.1 summarizes all the iterations from the beginning till this target is met.

[7] This part might sound confusing, and it is! The idea is to translate having a value close to the optimal in a high score. Note, also, that internally the algorithm is always looking for a minimum as its name indicates, to maximize we simply take the negative of the sampled values.

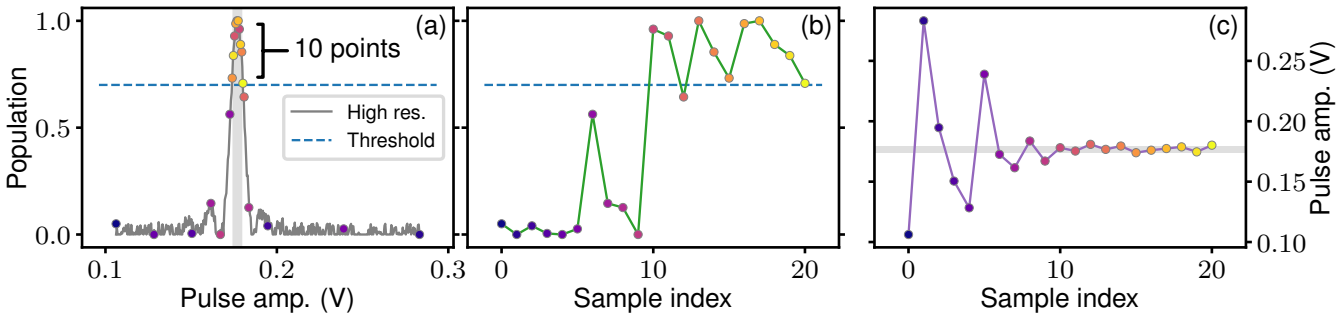


Figure 3.5: Sampling strategy developed for the chevron alignment calibration illustrated with a simulated noisy chevron model (a). The lightness of the data points indicates the sampling order. The domain is sampled with a bias towards high population values. The peak is detected when a value above a predetermined threshold is found. After that, only domain segments with at least one value above the threshold will be sampled [(b, c), sample index > 9]. From this moment onwards, the scoring strategy ensures both sides of the peak are sampled, until a target number of points beyond the threshold is reached (10 points).

3.3.1 Other features and modes of operation

In the previous section we have covered the main use case of the developed algorithm. Nevertheless, as a byproduct of the many development attempts and iterations several other useful features are now part of the algorithm which we briefly overview here[8]. The following operation modes and combinations thereof are possible:

- Limit the maximum resolution allowed. This is useful to enforce physical constraints into the algorithm, e.g. resolution of an instrument output.
- Global optimization of arbitrary continuous potentially noisy functions. The algorithm is robust under moderate noise.
- Converge in a local optimal. This is useful when we already know we are in a local optimal (e.g. from previous measurements) and just want to to converge to the best point.
- Converge into a local optimal beyond a threshold. This is similar to the previous section with the distinction that we just want to converge into the peak/deep.
- Adaptive sampling with a priority bias for optimal regions. This corresponds to always staying in the *global search*. The reason this type of scan is useful is because, as an experimentalist, even though we are usually searching for optimal parameters it is important to understand the landscape at the same time. In general, typical black-box optimizers do not provide this flexibility.
- Launching multiple samplers (sequentially) within a single experiment. This was a necessary feature for the calibration of the bias offset, but is general enough to run any number of adaptive samplers, each with its own configuration. An exotic example could be running a minimizer and a maximizer within the same domain.

[8]The interested readers can find a detailed Jupyter notebook tutorial under `/examples/MeasurementControl-adaptivesamplingusecases.ipynb` within PycQED [25].

- Sampling a predefined set of points before engaging a sampling strategy. This is specially useful when the location of the optimal regions can be roughly guessed beforehand which helps the learner build the knowledge of the sampling function more quickly.

Chapter 4

Analysis and sampling of CZ calibration landscapes

A crucial element towards the automation of CZ gates calibration is the sampling and analysis of the calibration landscapes (such as the ones of Fig. 2.23). Even thought the experimental code-base of the lab (PycQED framework [25]) provides a bare-bone analysis for 2D landscapes, at the beginning of this work, a dedicated analysis for CZ landscapes was missing. Furthermore, most experiments were sampled as evenly-spaced grids of predetermined data points scaling poorly for a 2D landscape. This chapter showcases the capabilities of the analysis and the adaptive sampler that were developed to bridge these gaps.

Following the structure in PycQED, the analysis is a Python class that inherits from the a base analysis class which provides common interfaces for extracting data, fitting and saving the generated results including figures. The class goes under the name `Conditional_Oscillation_Heatmap_Analysis`[1]. As its the name indicates, each measured data point corresponds to a conditional oscillation experiment (Fig. 2.25) from which several quantities are estimated, among others the conditional phase ϕ_{2Q} and the missing fraction m (serving as leakage estimate). The analysis was build with very few assumptions due to the lack of structure in the NZ calibration landscapes, as opposed to the SNZ discovered later (Chapter 5). Along the chapter the many functionalities are illustrated by running the analysis on exemplary datasets (for both NZ and SNZ). The figures presented closely match the version produced by the analysis as part of PycQED. The plotting is based on `matplotlib` [79], while the real-time plotting that was improved as well is based on `pyqtgraph` [80] (for performance reasons) through QCoDeS [26].

During the measurements, adaptive sampling (`LearnerND_Minimizer`[2]) is employed to optimally discriminate the features of the landscapes or to optimize the CZ parameters. For this, an empirically defined cost function is fed into the sampling algorithm and is also saved in the dataset as a "measured quantity". The adaptive sampler has evolved throughout this work targeting the CZ calibration landscapes, nonetheless, the final algorithm is not limited in anyway to this particular case. It provides

[1] The source code is located in `pycqed.analysis_v2.fluxing_analysis.py`.
[2] The source code library is available in `pycqed.utilities.learnerND_minimizer.py`.

similar functionality to the `Learner1D_Minimizer` (Section 3.3) with multiple modes of operation generalized to multi-dimensional domains.

Chapter outline The chapter starts by presenting the goals and requirements for the analysis followed by technical descriptions and illustrations of the implemented capabilities and features: flexibility, interpolation, colormaps, contours, real-time plotting support and extraction of optimal points. The chapter ends with an overview of the `LearnerND_Minimizer` adaptive sampler and its inner workings for the most relevant operating mode.

4.1 Main goals and requirements

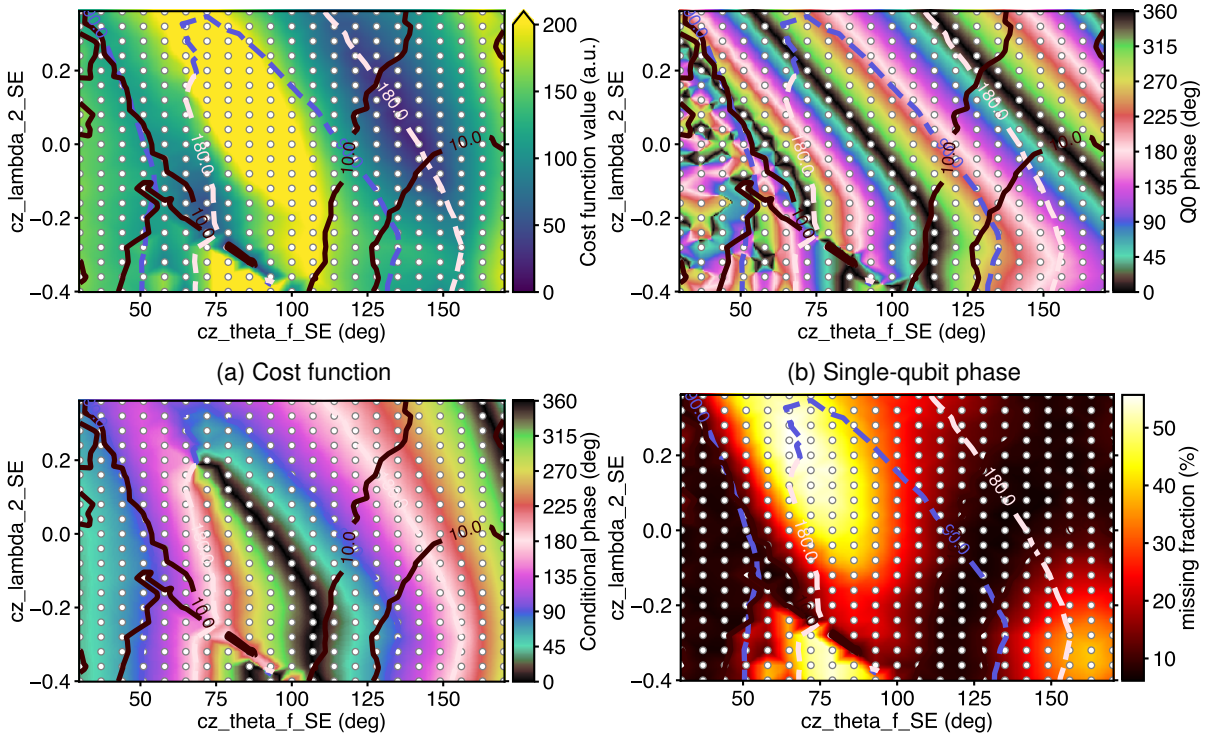
We now briefly discuss and list the key goals and requirements of the analysis. Despite not being very obvious in the beginning, from the experience in the lab it became clear that having flexible tools was very important to an experimentalist. Very rigid tools are cumbersome or useless when dealing with unexpected results and needs that were not predicted when defining initial requirements. A second driver for flexibility was the unclear calibration path yet to be established. This analysis was developed taking those facts into account and proved to be of great value many times. Beside flexibility, the following requirements were targeted:

- Integrated in PycQED such that later it could be part of a calibration routine;
- Automated as much as possible with little to no input from the user in most cases;
- Support for both experimental and simulated datasets;
- Intuitive and easy to understand visualization of post-measurement results (and real-time plotting);
- Extract and save quantities of interest;
- Detection and extraction of multiple local optima per landscape.

4.2 Analysis capabilities and features

Flexibility

Several features were introduced in the design of the analysis in order to match the flexibility requirements. Because the NZ calibration is not straightforward, initially the calibration process was intended to tightly integrate simulations with the experiment in order to fasten the process, at the cost of the additional complexity. Consequently, the developed routine accommodates for the analysis of both experimental and simulated datasets (Figs. 4.1 and 4.3 respectively). Additionally, under the consideration that an experimentalist might be interested to monitor any other parameter(s) of the system (which generate additional landscapes in the dataset), the analysis ensures that these are processed and plots are generated.



(a) Cost function

(b) Single-qubit phase

(c) Conditional phase, ϕ_{2Q}

(d) Missing fraction, m ($\simeq 2\tilde{L}_1$)

Figure 4.1: Analysis example of experimental NZ CZ calibration landscapes for a qubit pair on device C. Each data point corresponds to a conditional oscillation experiment [see Fig. 2.25]. In total, 420 (21 x 20) points were grid-sampled over $\sim$ 3.5 h [$\sim$ 30 s per point]. The white dots indicate the location of the sampling points. The rest of the landscape is linearly interpolated. The cost function is an arbitrarily weighted sum of the missing fraction and the conditional phase deviation from a target phase (in this case 180° for the CZ gate). The single-qubit phase is usually calibrated later through an independent knob, hence not included in the cost function. Dashed contours are extracted from the conditional phase landscape, while the solid contours are extracted from the missing fraction landscape.

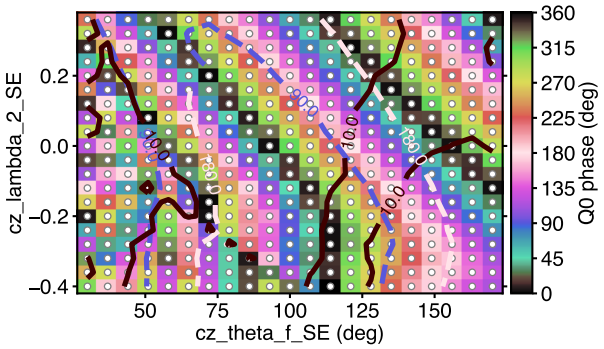


Figure 4.2: Non-interpolated single-qubit phase landscape corresponding to Fig. 4.1(b). Generated with the same analysis when setting the interpolation off.

As further described, the analysis is able to generate/extract many quantities of interest. Care was taken to ensure these were automatically saved in a persistent manner along with the dataset and made accessible (under the `Analysis/quantities_of_interest` group inside the dataset). Beside that, the analysis is robust to being re-executed against the same dataset with distinct settings. This might seem obvious and expected, nevertheless, needs to be proactively designed to avoid side-effects (e.g. inconsistent overriding of the quantities of interest).

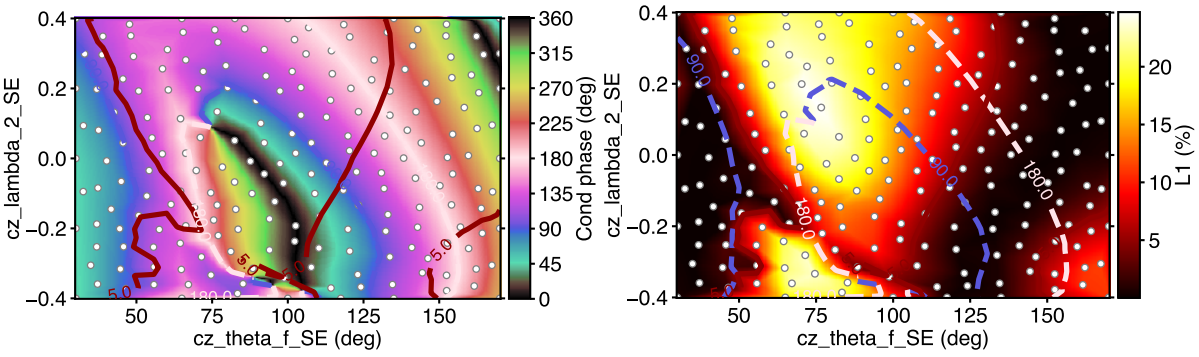
Using color to map the third dimension of these landscapes is the natural approach in order to easily generate a static human-interpretable format. However, it is often the case that after a dataset was generated we are interested only in a specific range of values for a particular measured quantity. For example, we are only interested in the minima of the cost function. To accommodate for this, the analysis settings (configurable at each execution) permit to specify the limits of the third dimension for each measured quantity. This effectively translates into a "saturated" color scale on one or both extremities, as exemplified in Fig. 4.1 where we have opted to focus on regions of the cost function below 200 units (note the arrow of the color bar).

Interpolation

As the reader might have noticed, the landscapes in Fig. 4.1 are presented with an overlaid grid of white dots. These denote the exact coordinates of the measured experimental data points, i.e. these are the pixels of our detector in this particular experiment. The rest of the landscape, for each measured value, is interpolated linearly to aid the visualization.

It might be deemed redundant to do so, considering that there is no new physical information obtained. The relevance of this feature is better understood observing the the landscape of the single-qubit phase [Fig. 4.1(b)]. If we would not interpolate and instead show only single-colored pixels with the size of the step in each dimension [Fig. 4.2], then it is arguably more difficult to discern the monotonicity of the single qubit phase along the diagonal. Of course, this is a double-edge sward and it is up to the human operator to decide if interpolation is appropriate or farther sampling instead is necessary to discriminate the relevant features. This would be the case of the bottom left corner of this landscape (if we were interested in that region).

Nonetheless, the primary use-case for interpolation are non-gridded datasets originated, for exam-



(a) Conditional phase, ϕ_{2Q}

(b) Leakage, L_1

Figure 4.3: Analysis example of simulated NZ CZ calibration landscapes for the same qubits pair as in Fig. 4.1. In total, 200 points were sampled adaptively on a non-regular grid. The white dots indicate the location of the sampling points. The rest of the landscape is linearly interpolated. Dashed contours are extracted from the conditional phase landscape, while the solid ones are extracted from the leakage landscape.

ple, from an adaptive sampler or an optimizer. In such a case linear interpolation is one of the simplest procedures that allows to visualize the underlying data within our framework. As an example, the analysis of a roughly uniform but not gridded dataset is depicted in Fig. 4.3 from a simulation (matching qualitatively the landscapes of Fig. 4.1). These figures were generated with the same settings as in Fig. 4.1.

A special attention is needed when interpolating the conditional and single-qubit phases. Due to the circular nature of these quantities, a direct interpolation would result in non-physical artifacts at the $0° \leftrightarrow 360°$ wraparound. This issue is solved by first taking the sine and cosine of the phase, and then interpolating each separately, after which we recover the angle from the complex landscape defined by these components.

Custom circular colormap

Given the importance of phases as the main parameters of the native gate-set of our transmons, it became important for the daily operation to be able to quickly evaluate if experiments were yielding the targeted phases. Which is usually monitored through real-time plotting of the measured quantities. Driven by this, a custom circular colormap was developed [see e.g. the color bar of Fig. 4.3(a)].

The goal of this color mapping is to always associate a specific color with each multiple of $45°$ as these are often of higher interest in certain quantum circuits. Additionally, smooth transitions between these colors are desired to map to the continuum of measured values. As there are not enough "rainbow colors" as need, the colormap adds white-pink in the middle and black for the wraparound point.

Implementation-wise, the color mapping was built using the HLS color model (Hue, Saturation, Lightness), an alternative of the RGB model. The colors for each $45°$-multiple was hand picked following linear interpolation along each dimension of the HLS model to fill the rest of the color space. Unfortunately, given the initial intent, the resulting colormap cannot be perceptually uniform[3], i.e. it does not result in a monotonic grayscale when converted, namely by the eye of color-blind people. If needed the

[3]See the phase landscapes in Chapter 5 for perceptually uniform colormaps.

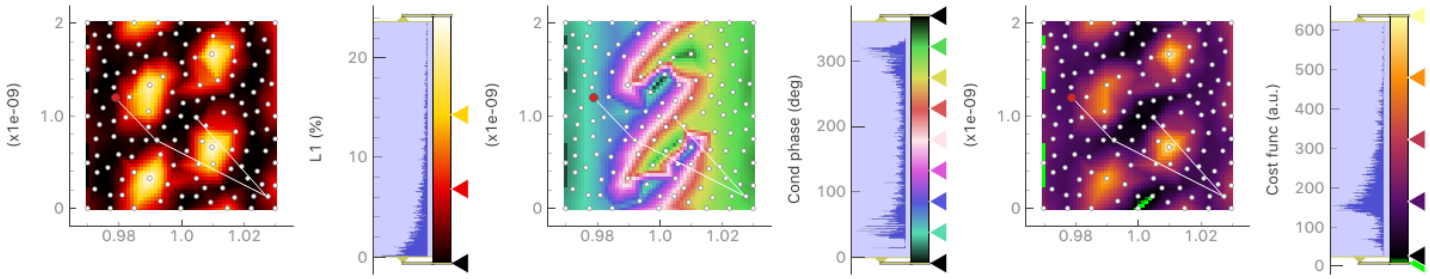


Figure 4.4: An instantaneous view of the real-time plotting monitor of simulated ideal SNZ CZ landscapes (adaptively sampled). The x and y axes of each landscape correspond to the A and t_ϕ SNZ parameters (see Fig. 5.1). The circular custom colormap for the conditional phase matches the one used in analysis [Figs. 4.1 and 4.3] and it is selected and mapped automatically. The cost function colormap, is also selected automatically, highlighting in green the optimal regions. The white dots denote the coordinates of the points sampled so far. The rest of the landscape is interpolated linearly. The red point corresponds to the last sampled point and the white lines connect a few of the points sampled immediately before, which allows to follow the behavior of the sampling algorithm.

colormap can be easily changed according to the purpose.

Contours overlay

A very important feature of the analysis is the extraction and plotting of the conditional phase and leakage contours on the the landscapes of the rest of the measured quantities. This guides the eyes to easily evaluate if the desired phase crosses the valley(s) of minimum leakage. Some technical details had to be accounted for to achieve this because, for a circular quantity as the phase, the contours-estimation routines suffer from the same artifacts as the interpolation at the wraparound boundary.

By juxtaposing both contour types on the cost function [Fig. 4.1(a)] we are able to better understand how each is contributing to the deviation from our target gate. For both quantities contour levels can be specified by the user. Here we have chosen to plot the $90°$ and $180°$ levels of conditional phase and the $5\,\%$ leakage estimate (roughly half of the missing fraction). In addition, a set of conditional phase contours can be specified to be saved as a quantity of interest. These can later be used, for example, to sample the landscape along the contour [see Fig. 5.6].

Real-time plotting monitor improvements

Though not strictly part of the (post-experiment) analysis, an essential tool for the operation of the experimental setup is the real-time plotting monitor of the measurement control module in PycQED [25]. While the measurement points are being acquired, it allows to monitor an experiment in real-time through interactive plots [Fig. 4.4]. Thus, an experimentalist can quickly judge weather the experiment is running as expected or not, and take decisions instantly.

Since the beginning of this work, the functionality of the plotting monitor was extended in order to greatly improve the support of the CZ calibrations driven by the recurrence of such measurements.

By default the 2D plotting monitor used the same colormap for all measured quantities. Besides that, the range of measured values, know at each moment in time during the experiment, was mapped to

the color spectrum dynamically with the incoming data. This is indeed the desired effect in most cases. However, for a circular domain, e.g. the conditional phase, it is required to have a fixed mapping to [0°; 360°]. In the absence of such a functionality, for each experiment, the user had to manually define the colors and change the boundaries of the colormap.

That frustration was solved by implementing a programmatically controlled interface that allows to specify a colormap as well as the boundaries of the mapping at the begging of each experiment (retained until explicitly changed or reset). For convenience, this process was automated based on the units and the label of each measured quantity, while allowing the user (or other calibration routines) to override these settings. In Fig. 4.4 (middle) the conditional phase uses the same colormap as in the analysis (Fig. 4.1) and its boundaries are automatically clipped to the [0°; 360°] range. Additionally to meeting the requirements of circular domains, a colormap with highlighted (gree) optimal extreme is assigned to cost functions (Fig. 4.4, right).

Similarly to the analysis, the coordinates of the measured points are marked in white to aid the interpretation of the interpolated landscapes. Motivated by the development of adaptive sampling algorithms within this work, the real-time behavior of the sampling algorithm is denoted by a red dot marking the last point connected by white lines to the points sampled immediately before. This work has inspired some of the features of Quantify [81] – the recent successor of PycQED.

Extraction of optimal points

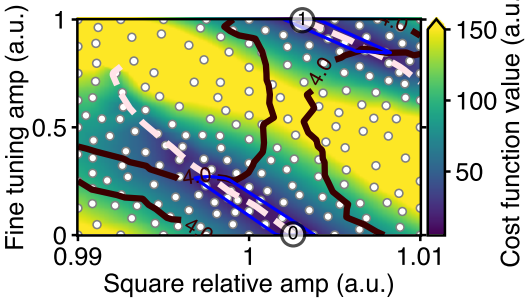
After sampling a calibration landscape, we usually proceed to extract the point (or region) that is expected to yield the best CZ gate and configure the necessary parameters and instruments. Following which, either a narrower region is sampled or we evaluate the gate performance through more elaborated metrics (e.g. randomized benchmarking).

These steps are relatively straight forward when the calibration landscape contains only a single and well defined global optimum (minimum in our case). Unfortunately, or rather fortunately, sometimes[4] the CZ calibration landscapes for NZ pulses contained more than one optimal region, i.e. a contour of the desired conditional phase crossing a valley of low leakage. Furthermore, with the new NZ variant as proposed in Chapter 5, multiple optima per landscape became expected (Fig. 4.4 already illustrates this). A need for a more advanced analysis becomes clear when combined with the fact that a sampled point rarely falls on the contour of the desired conditional phase within high precision. For example, one local optimum might correspond to a lower leakage but its cost function might not reflect that because the sampled point fell further from the desired phase contour.

Figure 4.5 depicts the additional capabilities of the analysis that solve these issues. First, it provides flexible settings with sensible defaults that allow to split apart the local minima. Second, a visualization toolbox provides easy interpretation of the results.

First we discard all points that have too high leakage and are too far from the target phase, with thresholds that can be configured by the user if needed. After this step we are left with (potentially)

[4] Depending on transmon pair and pulse duration. However, it is not straightforward to predict or to design beforehand for (multiple) optima in the NZ case.

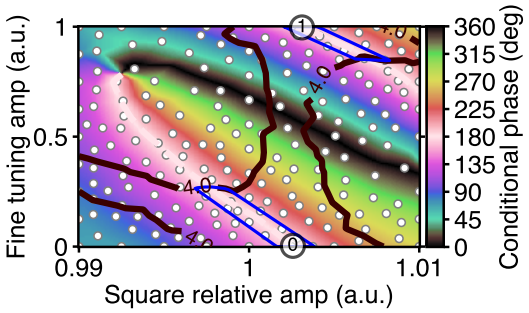


==========
Optimal #0
==========
Square relative amp = 1.003 a.u.
Fine tuning amp = 0.003344 a.u.

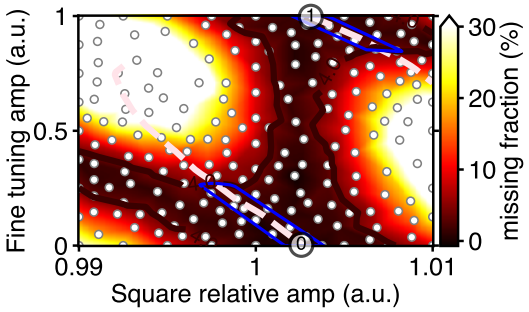
Cost function value = 7.63 a.u.
Conditional phase = 179.5 deg
missing fraction = 0.768 %
Q0 phase = 133.2 deg
==========
Optimal #1
==========
Square relative amp = 1.003 a.u.
Fine tuning amp = 0.9967 a.u.

Cost function value = 15.84 a.u.
Conditional phase = 179.2 deg
missing fraction = 1.489 %
Q0 phase = 136.4 deg

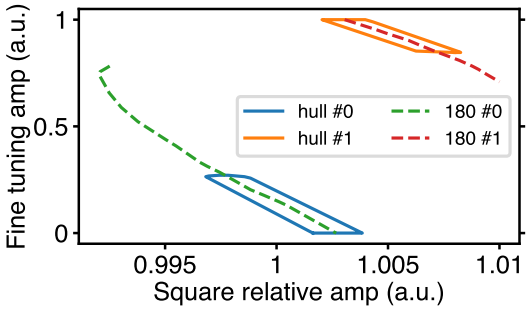
(a) Cost function and optimal points summary



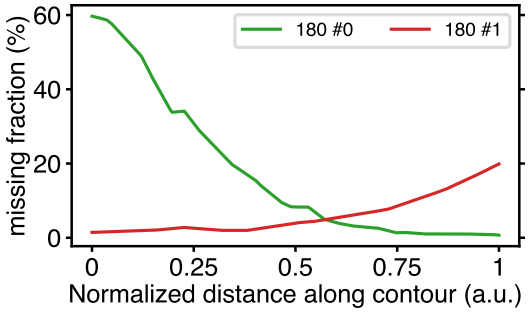
(b) Conditional phase, ϕ_{2Q}



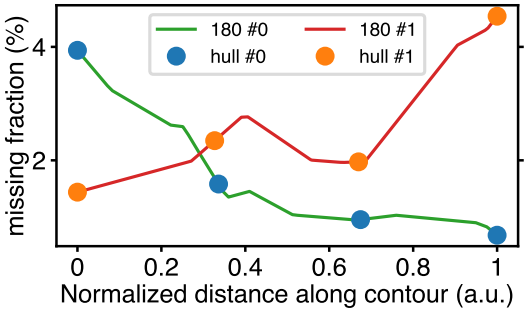
(c) Missing fraction, m ($\simeq 2\tilde{L}_1$)



(d) Optimal hulls and contours

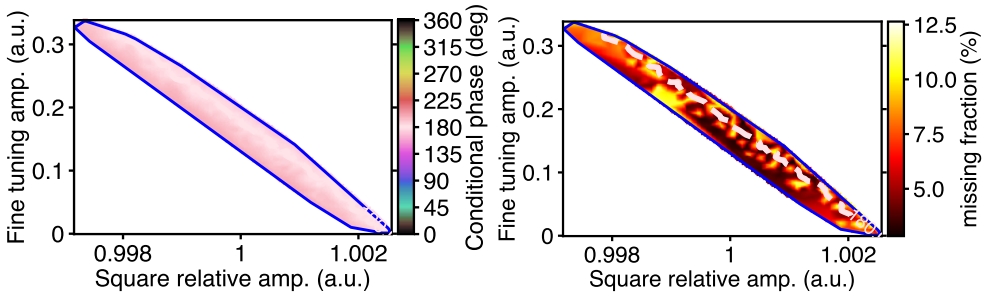


(e) m along $\phi_{2Q} = 180°$



(f) m along the hull-restricted $\phi_{2Q} = 180°$

Figure 4.5: Analysis of SNZ CZ calibration landscapes containing multiple optima. (a, b, c) Dashed curves corresponds to $\phi_{2Q} = 180°$ contours, while solid thick curves correspond to missing fraction. Blue convex hulls denote local optimal regions for CZ calibration. Numbered white circles correspond to the estimated points yielding best CZ gates inside each hull based on a strategy that takes into account the local average leakage (missing fraction). (d, e, f) Auxiliary plots that aid the inspection of the leakage along the contours. (d) Contours and optimal hulls are colored and annotated as a reference for figures (e) and (f). (e) Missing fraction along the entire contours. (f) Missing fraction along the portion of the contours inside the corresponding hull. Colored dots are only indicative of the hull [in (d)] to which the curve belongs to.



(a) Conditional phase, ϕ_{2Q}

(b) Missing fraction, m

Figure 4.6: Analysis of SNZ CZ landscapes sampled only inside an optimal region previously obtained from the analysis of a larger landscapes. The landscape is interpolated, however, the location of the measured points is omitted for clarity. A narrow region like this can be used to recalibrate a gate when its performance deteriorates.

several "groups" of points that would be classified by a human as being disjoint. In our analysis, this classification was achieved by using a distance-based clustering algorithm from the `scipy` Python module. As sometimes the number of measured points might be very limited, by default, the clustering uses the interpolated landscape to obtain more reliable results without the intervention of the user.

The resulting clusters are denoted in Figs. 4.5(a) to 4.5(c) by blue convex hulls that delimit the boundaries of each identified cluster. For clarity and reference, these hulls are plotted again in Fig. 4.5(d), now with distinct colors to aid their association with the corresponding conditional phase contours. In addition, the points defining the convex hulls are saved along with the quantities of interest and can be retrieved later in order to perform calibrations (or re-calibrations) only within the optimal region – a functionality provided by the Python `adaptive` module. The analysis of such a measurement is presented in Fig. 4.6.

Next, we identify the best point within each cluster [annotated with numbered circles in Figs. 4.5(a) to 4.5(c)]. As a last step, when several local minima are present, these best points can be score either based on the cost function value, or the average surrounding leakage (missing fraction). Higher scoring points are assigned the lowest indices, here `#0`. For ease of inspection and future reference, a summary of each optima is printed together with the cost function landscape [Fig. 4.5(a)].

Finally, it is often of interest to inspect how each measured quantity varies along conditional phase contours, with special interest in the missing fraction. The analysis automatically creates such plots, both for the full extension of the contours [Fig. 4.5(e)] and for the contours within the corresponding hull [Fig. 4.5(f)]. The later confirms the lower leakage and correct scoring of the optima.

Future improvements of the analysis

Here we discuss a few potential improvements of the analysis. The extraction of the optimal point(s) could benefit from recalculating the cost function in the analysis based on the interpolated landscapes of the measured quantities. This has the potential to better estimate the best optima because interpolating the cost function linearly (based on the measured points) cannot yield a lower minimum than its value at the best measured point, while re-evaluating the cost function at each interpolated point might pinpoint the optima more accurately. A similar strategy could be extended to an optimizer that is aware of the

contributions that are usually lumped together in the cost function.

4.3 Optimal sampling of landscapes: `LearnerND_Minimizer`

Most of the adaptively sampled landscapes presented in this work have been acquired using an algorithm, `LearnerND_Minimizer`, developed for this purpose to speed up the calibration experiments as well as simulations. It corresponds to the multidimensional generalization of the `Learner1D_Minimizer` presented in Section 3.3 while the core is inherited form the `LearnerND` of the `adaptive` package [78].

The `LearnerND_Minimizer` provides all the functionalities of the `Learner1D_Minimizer` with some minor differences. We focus here on the differences and point the reader to Section 3.3 for an extensive description of the sampling strategies and modes of operation.

Because the domain is now multidimensional, instead of segments we have simplexes which in 2D corresponds to triangles. Naturally, the key parameter `interval_weight` (`Learner1D_Minimizer`) now becomes `volume_weight` (`LearnerND_Minimizer`). Furthermore, as there are more data points per simplex, instead of just the extremities of a segment in the 1D case, the learner takes advantage of this to incorporate resilience to noise and outliers, namely we discard the worst value in a simplex[5] when calculating its sampling priority.

The main use of the `LearnerND_Minimizer` within this work has been to sample CZ calibration landscapes both for NZ gates (calibrated in the scope of a variational quantum algorithm [19]) and SNZ gates (Chapter 5). As opposed to other algorithms it allowed to acquire an overall understanding of the landscape while finding the local optima, all within the same experiment. For this we use the *global search* mode (see Section 3.3). We exemplify such a case in Fig. 4.7 where the internal state of the learner gives a better understanding of its operation. By the end of the sampling the non-optimal, high-cost, regions are barely sampled as intended, while the sampling resolution is greater in the regions of interest (low-cost values). Usually, for a practical calibration, much less points are enough ($\sim 50 - 60$). Further insight into the capabilities of the algorithm can be drawn from Fig. 4.8. There we showcase how the `volume_weight` parameter can be used to balance the sampling between optimal regions and high-volume simplexes (large triangles in this 2D case).

Combining the sampling algorithm and the landscape analysis toolbox with the simple calibration landscapes of the gate introduced in Chapter 5 CZ calibrations and re-calibrations were streamlined. Routine calibrations, required roughly on a biweekly basis, take now on the order of 30 min, while calibrations from "scratch" can be performed overnight. In either case, further speedups are plausible at chip-level calibration with simultaneous landscapes for distinct qubits pairs being measured within the same pulse sequence, which will minimizes overheads in communication with hardware, qubit initialization and measurement time.

[5] We still take into account the volume of the simplex as defined by the vertices. We "ignore" the value of "worst" data point to avoid unlucky noisy samples or outliers (not so rare for the CZ calibration landscapes).

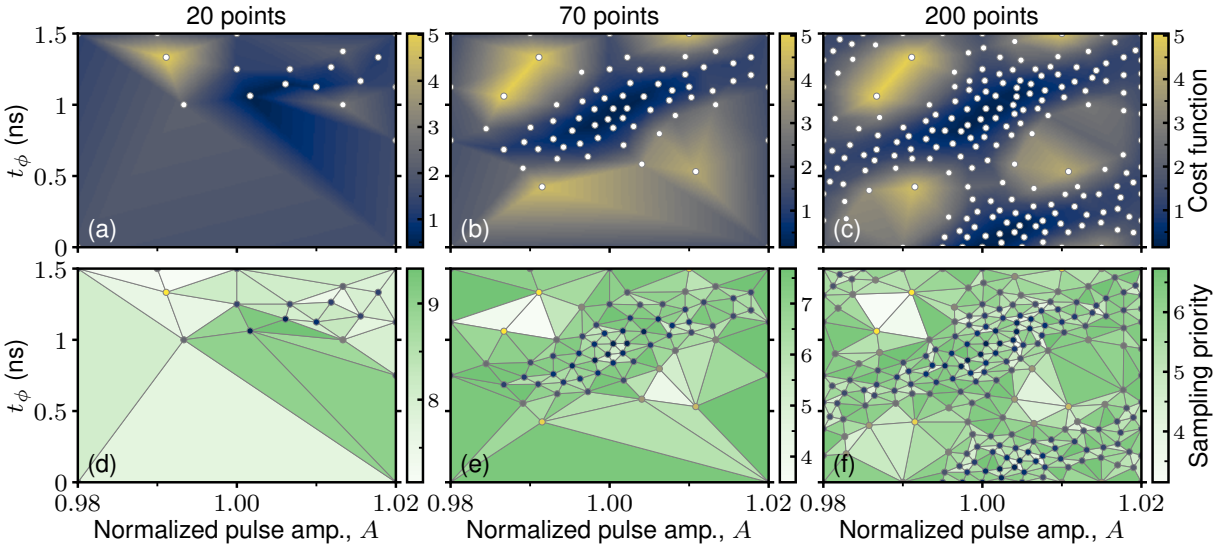


Figure 4.7: `LearnerND_Minimizer` sampling evolution example applied to the cost function of a simulated ideal SNZ calibration landscape (the cost function components shape are plotted in Fig. 5.1). The algorithm settings target an overall understanding of the landscape with focus on areas of low cost. The `volume_weight` was configured to `1/1000` implying a significant bias towards areas with low cost function. The three columns correspond to the interpolated cost function landscapes (top row) and the corresponding internal state of the algorithm (bottom row) in terms of the sampling priority, at three moments in time during a continuous run. Sampled points are annotated with small circles. In the bottom row, these are colored according to their cost value from the top row. The darkest simplexes (triangles) indicate the next areas to be sampled. The biggest triangles with highest-cost vertices are assigned the lowest sampling priority (white-colored).

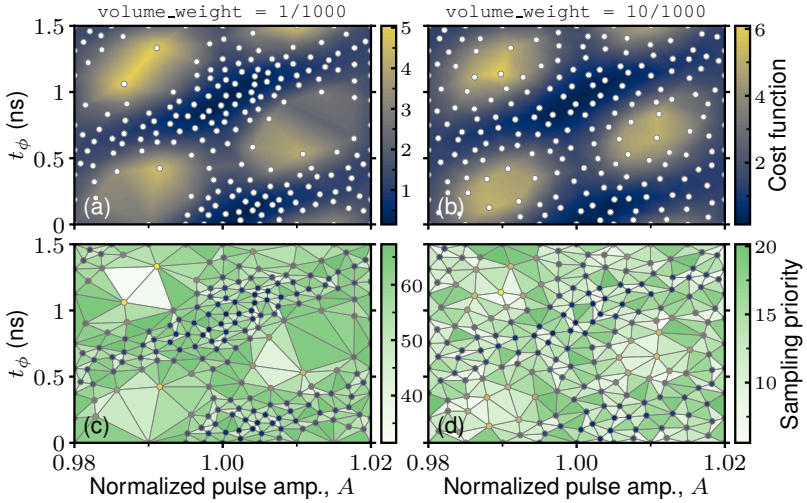


Figure 4.8: Effect of the `volume_weight` configuration parameter for a fixed budget of sampling points. When increasing its value from (a, c) to (b, d), the volume (area) of a triangle has a bigger weight in the calculation of its sampling priority score relative to how low is the cost value at its vertices. This results in a more uniform grid (second column) while losing resolution in the areas of the lowest cost function.

Chapter 5

High-fidelity CZ gate with maximal intermediate leakage operating at the speed limit

Simple tuneup of two-qubit gates is essential for the scaling of quantum processors. In this chapter, we introduce the sudden variant (SNZ) of the Net-Zero scheme realizing controlled-Z (CZ) gates by baseband flux control of the transmon frequency. The SNZ CZ gates operate at the speed limit of transverse coupling (J_2) between computational and non-computational states by maximizing intermediate leakage. The key advantage of the SNZ is tuneup simplicity, owing to the regular structure of conditional phase and leakage as a function of two control parameters. We realize SNZ CZ gates in a four-transmon patch, achieving $99.93\,\%$ [$99.54(27)\,\%$ average] fidelity and $0.10\,\%$ [$0.18(4)\,\%$ average] leakage on the best pair. The SNZ is compatible with scalable schemes for quantum error correction and adaptable to generalized conditional-phase (CPHASE) gates useful in intermediate-scale applications.

Chapter outline The work presented in this chapter is an extended version of the preprint Ref. 20. The chapter starts with an introduction that includes an overview of the field of two-qubit gates, previous approaches for flux-tunable qubits and their limitations, and a summary of the obtained results with our new approach. Next we introduce the SNZ concept starting with an idealized scenario and an analogy to a Mach-Zehnder interferometer. For completeness, we contrast the SNZ pulses with the previous state-of-art conventional NZ pulses (CNZ). Next, we proceed to experimental realization of the gate, where we first present the device and the employed qubits, followed by the calibration procedure and examination of hardware limitations and their impact. In the following subsection, the experimental performance of four SNZ CZ gates is evaluated and discussed. Finally, dominant gate error sources are estimated resorting to numerical simulations and compared between CNZ and SNZ pulses.

5.1 Introduction

Superconducting quantum processors have recently reached important milestones for quantum computing [82], notably the demonstration of quantum supremacy on a 53-transmon processor [5]. On the path to quantum error correction (QEC) and fault tolerance [67], recent experiments have used repetitive parity measurements to stabilize two-qubit entanglement [75, 83] and to perform surface-code quantum error detection in a 7-transmon processor [18]. These developments have relied on two-qubit controlled-phase (CPHASE) gates realized by dynamical flux control of transmon frequency, harnessing the transverse coupling J_2 between a computational state $|11\rangle$ and a non-computational state such as $|02\rangle$ [13, 84]. Compared to other implementations, e.g., cross-resonance using microwave-frequency pulses [85] and parametric radio-frequency pulsing [86], baseband flux pulses achieve the fastest controlled-Z (CZ) gates (a special case of the CPHASE), operating near the speed limit $t_{\rm lim} = \pi/J_2$ [87].

Over the last decade, baseband flux pulsing for two-qubit gating has evolved in a continuous effort to increase gate fidelity and to reduce leakage and residual ZZ coupling. In particular, leakage has become a main focus for its negative impact on QEC, adding complexity to error-decoder design [58] and requiring hardware and operational overhead to seep back into the computation space [88–92]. To reduce leakage from linear-dynamical distortion in flux-control lines and limited time resolution in arbitrary waveform generators (AWGs), unipolar square pulses [13, 56] have been superseded by softened unipolar pulses [60, 61] based on fast-adiabatic theory [34]. In parallel, coupling strengths have reduced roughly fourfold (to $J_2/2\pi \sim 10{-}20$ MHz) in the effort to reduce residual ZZ coupling, which affects single-qubit gates and idling at bias points, and produces crosstalk from spectator qubits [93]. Many groups are actively developing tunable coupling schemes to suppress residual coupling without incurring the slowdown of the gates [94–98].

A main limitation to the fidelity of flux-based CPHASE gates is dephasing from flux noise, as one qubit is displaced $0.5 - 1$ GHz below its flux-symmetry point (i.e., sweetspot [99]) to reach the $|11\rangle - |02\rangle$ resonance. To address this limitation, Ref. 39 recently introduced a bipolar variant [termed Net Zero (NZ)] of the fast-adiabatic scheme, which provides a built-in echo reducing the impact of low-frequency flux noise. The double use of the transverse interaction also reduces leakage by destructive interference, as understood by analogy with a Mach-Zehnder interferometer (MZI). Finally, the zero-average characteristic avoids the buildup of long-timescale distortions remaining in flux-control lines after compensation, significantly improving gate repeatability. NZ pulsing has been successfully used in several recent experiments [18, 75, 76], elevating the state of the art for CZ gate fidelity in a transmon processor to $99.72(35)\,\%$ [82]. However, NZ suffers from complicated tuneup, owing to the complex dependence of conditional phase and leakage on fast-adiabatic pulse parameters. This complication limits the use of NZ for two-qubit gating as quantum processors grow in qubit count.

In this chapter, we introduce the sudden variant (SNZ) of the Net Zero scheme implementing CZ gates using baseband flux pulsing. SNZ offers two key advantages while preserving the built-in echo, destructive leakage interference, and repeatability characteristic of conventional Net Zero (CNZ). First, SNZ operates at the speed limit of transverse coupling by maximizing intermediate leakage to the non-

computational state. The second and main advantage is greatly simplified tuneup: the landscapes of conditional phase and leakage as a function of two pulse parameters have very regular structure and interrelation, easily understood by exact analogy to the MZI. We realize SNZ CZ gates among four pairs of nearest neighbors in a seven-transmon processor and characterize their performance using two-qubit interleaved randomized benchmarking (2QIRB) with modifications to quantify leakage [39, 59, 100]. The highest performance achieved has $99.93 \pm 0.24\%$ fidelity with $0.10 \pm 0.02\%$ corresponding leakage. Using numerical simulation with experimental input parameters, we dissect an error budget finding SNZ to slightly outperform CNZ. SNZ CZ gates are fully compatible with scalable approaches to QEC [17]. The generalization of SNZ to arbitrary CPHASE gates is straightforward and useful for optimization [101], quantum simulation [102], and other noisy intermediate-scale quantum (NISQ) applications [14].

5.2 SNZ concept

A flux pulse harnessing the $|11\rangle$-$|02\rangle$ interaction implements the unitary

$$U_{\text{CPHASE}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & e^{i\phi_{01}} & 0 & 0 & 0 \\ 0 & 0 & e^{i\phi_{10}} & 0 & 0 \\ 0 & 0 & 0 & \sqrt{1-4L_1}e^{i\phi_{11}} & \sqrt{4L_1}e^{i\phi_{02,11}} \\ 0 & 0 & 0 & \sqrt{4L_1}e^{i\phi_{11,02}} & \sqrt{1-4L_1}e^{i\phi_{02}} \end{bmatrix}$$

in the $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle, |02\rangle\}$ subspace, neglecting decoherence and residual interaction between far off-resonant levels. Here, ϕ_{01} and ϕ_{10} are the single-qubit phases, and $\phi_{11} = \phi_{01} + \phi_{10} + \phi_{2Q}$, where ϕ_{2Q} is the conditional phase. Finally, L_1 is the leakage parameter[1].

In a rotating frame that absorbs the single-qubit phases, the Hamiltonian is

$$H = \Delta_{02}(t)\,|02\rangle\langle 02| + J_2\,(|02\rangle\langle 11| + |11\rangle\langle 02|)\,,$$

where $\Delta_{02}(t)$ is the dynamical detuning between $|02\rangle$ and $|11\rangle$. Each half of the bipolar

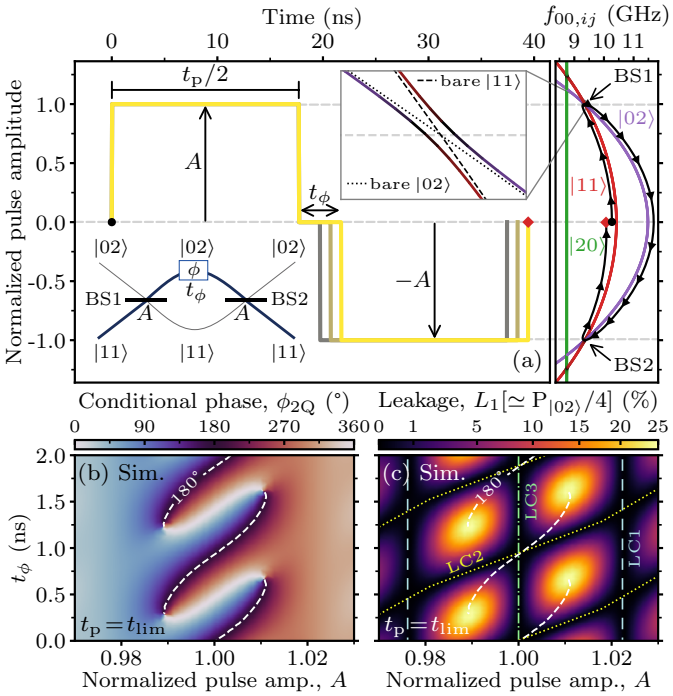


Figure 5.1: Numerical simulation of an SNZ pulse with infinite time resolution for the pair Q_L-Q_{M2}, with $\Delta_{02}^{max}/2\pi = 1.063$ GHz (at bias point) and $t_{lim} = 35.40$ ns for the $|11\rangle$-$|02\rangle$ interaction. (a) Schematic of the SNZ flux pulse, with $t_p = t_{lim}$ and variable A and t_ϕ. Amplitude A is normalized to the $|11\rangle$-$|02\rangle$ resonance. (b, c) Landscapes of conditional phase ϕ_{2Q} (b) and leakage L_1 (c) as a function of A and t_ϕ. Dashed white curves are contours of $\phi_{2Q} = 180°$. In (c), the vertical valley at $A = 1$ is due to LC3: each half pulse fully transmits $|11\rangle$ to $|02\rangle$, and vice versa. The other two vertical valleys are due to LC1: each half pulse implements a complete (off-resonant) oscillation and thus perfectly reflects. The diagonal valleys are due to LC2. Note the simultaneous overlap of $\phi_{2Q} = 180°$ contours with the crossing of LC3 and LC2 valleys. These crossing points occur at $A = 1$ and t_ϕ satisfying $\Delta_{02}^{max} t_\phi = 0 \pmod{2\pi}$.

[1] Even though we follow the leakage definition where it accounts for any computational state to leak into any other non-computation state [59], for this illustration we simplify and attribute all leakage to the $|11\rangle \to |02\rangle$ channel, hence a 4 factor.

NZ pulse implements the unitary

$$U_{\mathrm{A}} = U_{-\mathrm{A}} = \begin{bmatrix} \gamma e^{i\phi_{\mathrm{a}}} & \beta e^{i\phi_{\mathrm{b}}} \\ \beta e^{i\phi_{\mathrm{c}}} & \gamma e^{i\phi_{\mathrm{d}}} \end{bmatrix},$$

in the $\{|11\rangle, |02\rangle\}$ subspace, where $\gamma, \beta \in [0,1]$ satisfy $\gamma^2 + \beta^2 = 1$ and $\phi_{\mathrm{a}} + \phi_{\mathrm{d}} = \phi_{\mathrm{b}} + \phi_{\mathrm{c}} + \pi \pmod{2\pi}$. In the MZI analogy, this unitary is the action of each beamsplitter, ideally identical $\mathrm{BS1} = \mathrm{BS2}$.

In SNZ [Fig. 5.1(a)], each half pulse is a square pulse with amplitude $\pm A$ and duration $t_{\mathrm{p}}/2 = t_{\mathrm{lim}}/2$. SNZ intentionally adds an idling period t_ϕ between the half pulses to perfect the analogy to the MZI (Fig. 5.1 inset), allowing accrual of relative phase ϕ in between the beamsplitters $\mathrm{BS1}$ and $\mathrm{BS2}$. The unitary action of this idling is

$$U_\phi = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{bmatrix}.$$

An ideal CZ gate, our target here, achieves $\phi_{01} = \phi_{10} = 0 \pmod{2\pi}$, $\phi_{2\mathrm{Q}} = \pi \pmod{2\pi}$ (phase condition PC), and $L_1 = 0$ (leakage condition LC), with arbitrary ϕ_{02}. Accomplishing both conditions with $U_{-\mathrm{A}} U_\phi U_{\mathrm{A}}$ requires

$$\gamma^2 e^{i2\phi_{\mathrm{a}}} + \beta^2 e^{i(\phi_{\mathrm{b}} + \phi_{\mathrm{c}} + \phi)} = -1 \ (\mathrm{PC})$$

simultaneously with either one of three conditions:

$$\beta = 0 \ [\mathrm{LC1}]; \qquad \phi_{\mathrm{a}} - \phi_{\mathrm{d}} - \phi = \pi \pmod{2\pi} \ [\mathrm{LC2}]; \qquad \text{or } \gamma = 0 \ [\mathrm{LC3}].$$

$\mathrm{LC1}$ corresponds to perfect reflection at each beamsplitter, while $\mathrm{LC3}$ results in perfect transmission. $\mathrm{LC2}$ corresponds to destructive interference at the second beamsplitter of the $|02\rangle$ leakage produced by the first.

The key advantage of SNZ over CNZ is the very straightforward procedure to simultaneously meet the PC and low leakage. To appreciate this, consider first the ideal scenario where the pulses can have infinite time resolution. For, $t_{\mathrm{p}} = t_{\mathrm{lim}}$, $t_\phi = 0$ ($\phi = 0$), and $A = 1$ (the $|\Delta_{02}|$ minimum) each half pulse implements an iSWAP gate between $|11\rangle$ and $|02\rangle$. Thus, $\gamma = 0$ (meeting $\mathrm{LC3}$) and $\phi_{\mathrm{b}} = \phi_{\mathrm{c}} = -\pi/2$ (meeting PC). In the MZI analogy, the first beamsplitter fully transmits $|11\rangle$ to $-i|02\rangle$ (producing maximal intermediate leakage), and the second fully transmits $-i|02\rangle$ to $-|11\rangle$.

Consider now the effect of non-zero t_ϕ. The idealized two-qutrit[2] numerical simulation with infinite time resolution and no decoherence shows that the landscapes of $\phi_{2\mathrm{Q}}$ and L_1 as a function of A and t_ϕ [Figs. 5.1(b) and 5.1(c)] have a clear structure and link to each other. Evidently, $U_{-\mathrm{A}} U_\phi U_{\mathrm{A}}$ is 2π-periodic in ϕ, so both landscapes are vertically periodic. The $L_1(A, t_\phi)$ landscape shows a vertical leakage valley at $A = 1$ where $\mathrm{LC3}$ is met. $\mathrm{LC2}$ gives rise to additional, diagonally running valleys. Juxtaposing the contour of $\phi_{2\mathrm{Q}} = 180°$ shows that PC is met at the crossing points between these valleys. This regular leakage landscape provides on its own useful crosshairs for simultaneously meeting PC. We note that along the $\mathrm{LC3}$ vertical valley, $\phi_{2\mathrm{Q}}$ changes monotonically as a function of t_ϕ, allowing the realization of

[2] A qutrit is the extension of a qubit to three energy levels: $|0\rangle$, $|1\rangle$ and $|2\rangle$.

CPHASE gates with any desired ϕ_{2Q}. We leave this useful generalization for future work, focusing here on CZ gates.

While in this idealized scenario the idling period is not needed, there are practical reasons to include t_ϕ in experiment: any flux-pulse distortion remaining from the first half pulse during the second (e.g., due to finite pulse rise time) will break the symmetry $U_{-A} = U_A$. Due to the fixed time resolution t_s of the AWG used for flux control, ϕ can only increment in steps of $-\Delta_{02}^{max} t_s$, where Δ_{02}^{max} is the detuning at the bias point. As typically $\Delta_{02}^{max}/2\pi = 0.5 - 1$ GHz and $t_s \sim 1$ ns, one may only use the number of intermediate sampling points in t_ϕ for very coarse control. For fine control, we propose the (simultaneous) tunning of the amplitude $\pm B$ of two sampling points: the first and last sampling points during t_ϕ [see Fig. 5.2(b)].

5.2.1 Comparison between CNZ and SNZ pulses

The conventional NZ (CNZ) strong pulse [Fig. 5.2(a)] consists of two back-to-back half pulses of duration $t_p/2$ each, applied on the higher-frequency transmon. Typically, $t_p/t_{lim} \sim 1.1{-}1.6$. The strong half pulses are formally parametrized as in [34]. For the purposes of illustration, here we can loosely lump this parametrization as affecting the amplitude ($\pm A$) and curvature (A') of the strong half pulses. Immediately following the strong pulse, weak bipolar pulses of duration t_{1Q} are applied on both the higher- and lower-frequency transmons with amplitudes $\pm C$ and $\pm D$, respectively, in order to null the single-qubit phases acquired by each. Typically, $t_{1Q} = 10$ ns. In conventional NZ there is no intermediate idling period between the strong half pulses, so the analogy to the MZI is not exact [Fig. 5.2(c)]. During tuneup, one searches the (A, A') space to achieve $U_{-A} U_A = U_{\text{CPHASE}}(\phi_{2Q} = \pi)$ by only affecting the unitary action of the two beamsplitters. Because for typical t_p conventional NZ produces significant leakage at the first strong pulse, achieving minimal leakage relies on meeting LC2. The structure of the $\phi_{2Q}(A, A')$ and $L_1(A, A')$ landscapes and especially their interrelation are not straightforward, so the search for an (A, A') setting satisfying both PC and LC2 is not easily guided [see Fig. 2.23 and Ref. 39].

The SNZ pulses introduced here [Fig. 5.2(b)] differ in two key ways. First, the strong half pulses are replaced by square half pulses each with duration $t_p/2$ as close as possible to $t_{lim}/2$ (as allowed by the AWG sampling period) but not shorter. Second, an intermediate idling period t_ϕ is added to accrue relative phase ϕ between $|02\rangle$ and $|11\rangle$, perfecting the analogy to the MZI [Fig. 5.2(d)]. We use the amplitude $\pm B$ of the first and last sampling points in t_ϕ and the number of intermediate zero-amplitude points to achieve fine and coarse control of ϕ, respectively. As in CNZ, we use weak bipolar pulses on both transmons (also with $t_{1Q} = 10$ ns) to null the single-qubit phases. During tuneup, we search the (A, B) space to achieve $U_{-A} U_\phi U_A = U_{\text{CPHASE}}(\phi_{2Q} = \pi)$. In Section 5.3.2 we will show that the SNZ pulse design gives a very simple structure to the $\phi_{2Q}(A, B)$ and $L_1(A, B)$ landscapes. Crucially, the crossing point of leakage valleys satisfying LC2 and LC3 matches $\phi_{2Q} = 180°$. This simplicity of tuneup is the key advantage of the SNZ over the conventional NZ.

Another advantage of SNZ over conventional CZ is the reduced total time $t_{total} = t_p + t_\phi + t_{1Q}$ required for a CZ gate. However, due to the 20 ns timing grid of the control electronics and the transverse coupling strengths in this device, this speedup is insufficient to reduce the total time allocated per CZ

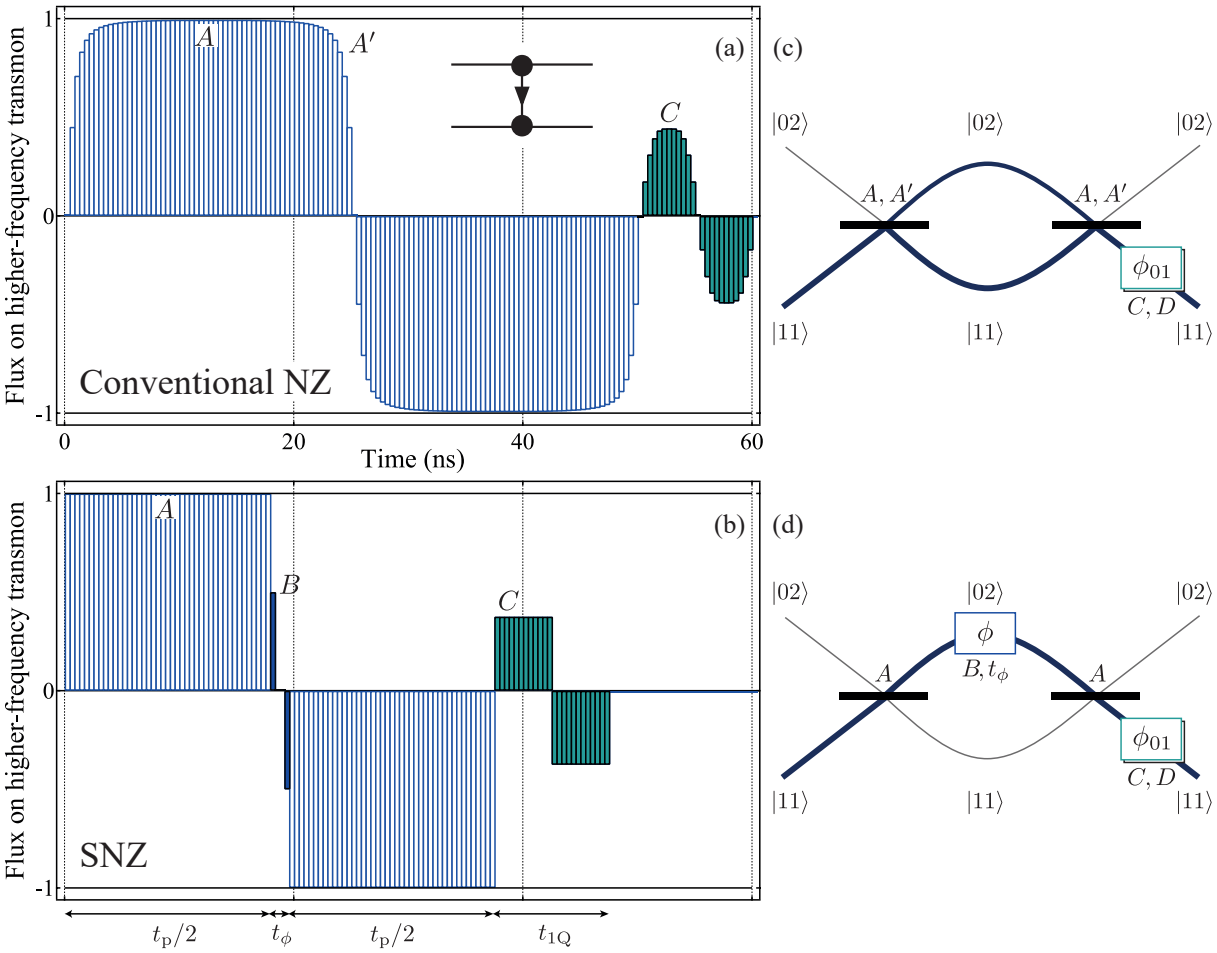


Figure 5.2: Comparison of conventional NZ and SNZ pulses for CZ gates. (a) Conventional NZ CZ pulses consist of two back-to-back strong half pulses of duration $t_p/2$ each, followed by two weak back-to-back half pulses of duration $t_{1Q}/2$ each on the higher-frequency qubit. The amplitude ($\pm A$) and curvature (A') of the strong pulses are jointly tuned to set the conditional phase ϕ_{2Q} at minimal leakage L_1, while the amplitude $\pm C$ of the weak pulses is used to null the single-qubit phase on the higher-frequency transmon. Weak pulses (amplitude $\pm D$) on the lower-frequency qubit (not shown here) are also used to null its single-qubit phase. (b) In SNZ, the strong pulses are replaced by square pulses with t_p as close as possible to t_{lim} but not shorter. Also, an intermediate idling period t_ϕ is added to accrue relative phase ϕ between $|02\rangle$ and $|11\rangle$. The amplitude $\pm B$ of the first and last sampling points in t_ϕ and the number of intermediate zero-amplitude points provide fine and coarse control of this relative phase, respectively. SNZ CZ gates also use weak bipolar pulses (now square) of total duration t_{1Q} to null single-qubit phases. (c) The MZI analogy for CNZ pulses is incomplete. Each strong half pulse implements a beamsplitter (ideally identical) with scattering parameters affected by A and A'. However, there is no possibility to independently control the relative phase in the two arms between the beamsplitters. (d) The MZI analogy is exact for SNZ pulse. The scattering at the beamsplitters is controlled by A and the relative phase ϕ is controlled finely using B and coarsely using t_ϕ. Credits: L. DiCarlo.

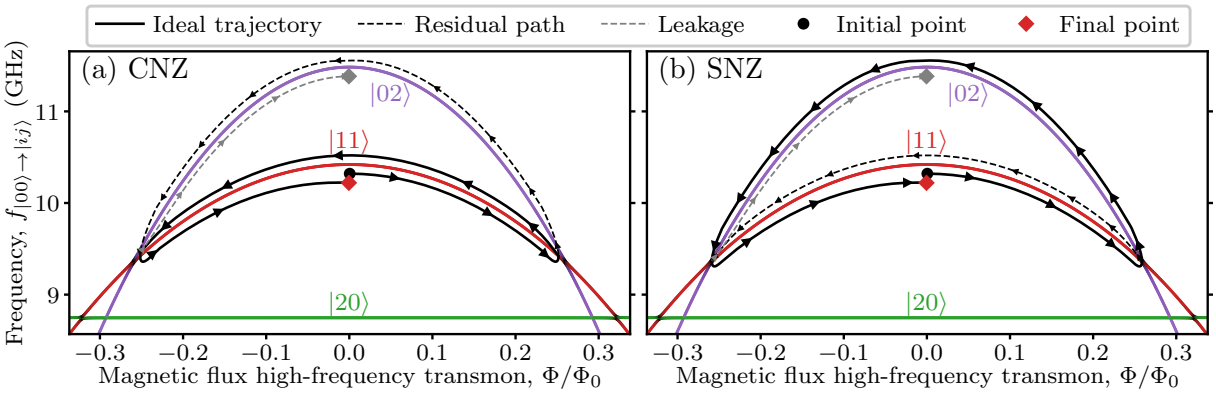


Figure 5.3: Schematic comparison between realistic trajectories for CNZ and SNZ pulses when applied to the double-excited state $|11\rangle$. The (lower) single-qubit energy levels are depicted in Fig. 2.9. The trajectory for single-qubit phase corrections are omitted. Note that in both cases, the most of the trajectory is spent close to (CNZ) or at the center (SNZ) of the interaction point. (a) Trajectory for a CNZ pulse. The system remains in the $|11\rangle$ state during the full trajectory for a sufficiently long pulse such that the adiabatic condition holds. Imperfections and/or too short pulses can lead to part of the population leaking into (or seeping back from) the $|02\rangle$ state. (b) Trajectory for an SNZ pulse. Ideally, maximal leakage is produced during the first half-pulse and the full population of the $|11\rangle$ state is recovered during the second half-pulse.

gate from 60 to 40 ns. Nonetheless, in SNZ, the fluxed transmon spends more time at its sweetspot, which reduces the dephasing due to flux noise.

Additionally, in Fig. 5.3 we compare schematically the trajectories implemented by the strong CNZ and SNZ pulses on a system initialized in the $|11\rangle$ state. A CNZ pulse attempts to adiabatically approach twice the interaction point while remaining in the $|11\rangle$ state [Fig. 5.3(a)]. Because it is hard to meet such condition, during the positive branch of the pulse, some population is lost through the residual path of $|02\rangle$. Imperfect destructive interference of the $|02\rangle$ state during the negative branch of the pulse leads to leakage at the end of the gate. In case of an SNZ pulse [Fig. 5.3(b)], after suddenly moving into the avoided crossing, during the first half-pulse, we maximize the transfer of population to the $|02\rangle$ state by fixing the amplitude $A = 1$ for a duration of $t_\mathrm{p}/2 \sim t_\mathrm{lim}/2$. As we switch to the second half-pulse approaching the avoided crossing for the second time, the single-sample amplitudes $\pm B$, together with t_ϕ, provide robust tunning knobs to maximize the recovery of the $|11\rangle$ population with the desired conditional phase.

5.3 Experimental realization of SNZ CZ gates

5.3.1 Device and transmon parameters

With these considerations, we turn to the experimental realization of SNZ CZ gates between the nearest-neighbor pairs among four transmons (Fig. 5.4) in the 7-transmon processor (device C, similar design to Fig. 2.12). High- and low-frequency transmons ($\mathrm{Q_H}$ and $\mathrm{Q_L}$, respectively) connect to two mid-frequency transmons ($\mathrm{Q_{M1}}$ and $\mathrm{Q_{M2}}$) using bus resonators dedicated to each pair. Each transmon has a flux-control line for two-qubit gating, a microwave-drive line for single-qubit gating, and a dispersively coupled

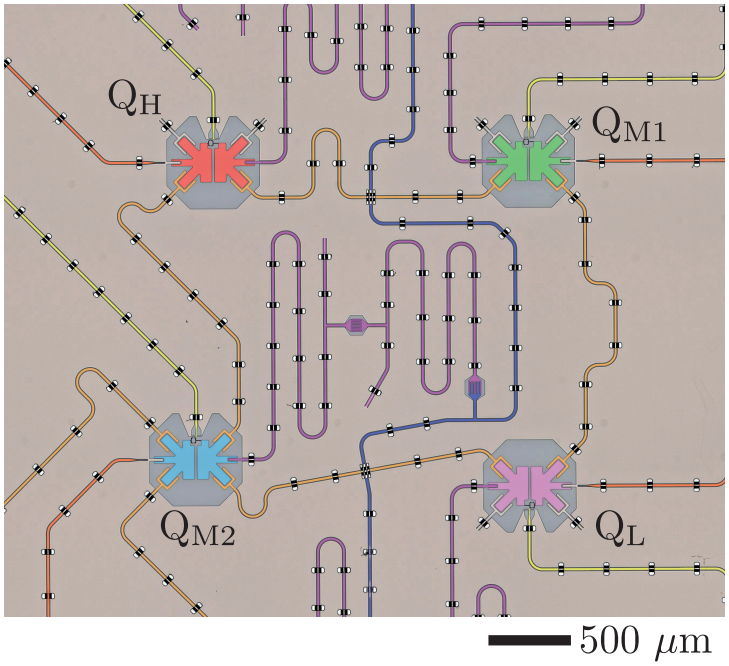


Figure 5.4: False colored optical image of the device, zoomed in to the four transmons used in this study. Transmons Q_H (red) and Q_L (pink) each connect to Q_{M1} (green) and Q_{M2} (cyan) using dedicated coupling bus resonators(light orange). Each transmon has a flux-control line (yellow), a microwave-drive line (dark orange), and dispersively-coupled resonator with Purcell filter for readout (purple) [75, 103]. The readout-resonator/Purcell-filter pair for Q_{M2} is visible at the center. The vertically running common feedline (blue) connects to all Purcell filters, enabling simultaneous readout of the four transmons (frequency multiplexing). Credits optical image: fabrication team, DiCarlo Lab, QuTech. Credits false coloring: A. Bruno, DiCarlo Lab, QuTech.

resonator with Purcell filter for readout [75, 103]. All transmons can be measured simultaneously by frequency multiplexing using a common feedline. Table 5.1 provides a summary of measured parameters for the four transmons. See Appendix A.1 for single-qubit gates details and performance. Each transmon is biased at its sweetspot using static flux bias to counter residual offsets. Flux pulsing is performed using a Zurich Instruments HDAWG-8 ($t_s = 1/2.4$ ns). Following prior work [31, 39], the linear-dynamical distortions in the flux-control lines are measured using the Cryoscope technique and corrected using real-time filters built into the AWG.

5.3.2 Calibration procedure

We exemplify the tuneup of SNZ pulses using pair Q_L-Q_{M2} (Fig. 5.5). We first identify t_{lim} for the $|11\rangle$-$|02\rangle$ interaction and the amplitude A bringing the two levels on resonance. The rightmost index indicates the excitation level of the fluxed transmon, here Q_{M2}. These parameters are extracted from the characteristic chevron pattern of $|2\rangle$-population in Q_{M2} as a function of the amplitude and duration of a unipolar square flux pulse acting on $|11\rangle$ [Fig. 5.5(a)]. The chevron symmetry axis corresponds to $A = 1$ and the oscillation period along this axis gives t_{lim}. We set $t_p = t^+_{lim} \equiv 2nt_s$, where n is the smallest integer satisfying $2nt_s \geq t_{lim}$. Next, we measure the landscapes of ϕ_{2Q} and leakage estimate $\tilde{L}_1$ in the range $A \in [0.9, 1.1]$, $B \in [0, A]$. These quantities are extracted from the conditional-oscillation experiments described in [39]. As expected, the landscape of $\tilde{L}_1$ [Fig. 5.5(c)] reveals a vertical valley

	Q_H	Q_{M1}	Q_{M2}	Q_L
Sweetspot transition frequency, $f_{0\to1}$ (GHz)	6.4329	5.7707	5.8864	4.5338
Transmon anharmonicity, α (MHz)	−280	−290	−285	−320
Readout frequency, $\omega_r/2\pi$ (GHz)	7.4925	7.2248	7.0584	6.9132
Relaxation time, T_1 (µs)	37(1)	40(1)	47(1)	66(1)
Ramsey dephasing time, T_2^* (µs)	38(1)	49(1)	47(1)	64(1)
Echo dephasing time, T_2 (µs)	54(2)	68(1)	77(1)	94(2)
Residual qubit excitation, (%)	1.4	1.2	4.3	1.7
Best readout fidelity, F_{RO} (%)	99.1	98.5	99.4	97.8

Table 5.1: Summary of frequency, coherence, residual excitation, and readout parameters of the four transmons. The statistics of coherence times for each transmon are obtained from 30 repetitions of standard time-domain measurements [42] taken over ~ 4 h. The residual excitation is extracted from double-Gaussian fits of single-shot readout histograms with the qubit nominally prepared in $|0\rangle$. The readout fidelity quoted is the average assignment fidelity [104], extracted from single-shot readout histograms (Fig. 2.21) after mitigating residual excitation by post-selection on a pre-measurement. Credits data collection and analysis: H. Ali, DiCarlo Lab, QuTech.

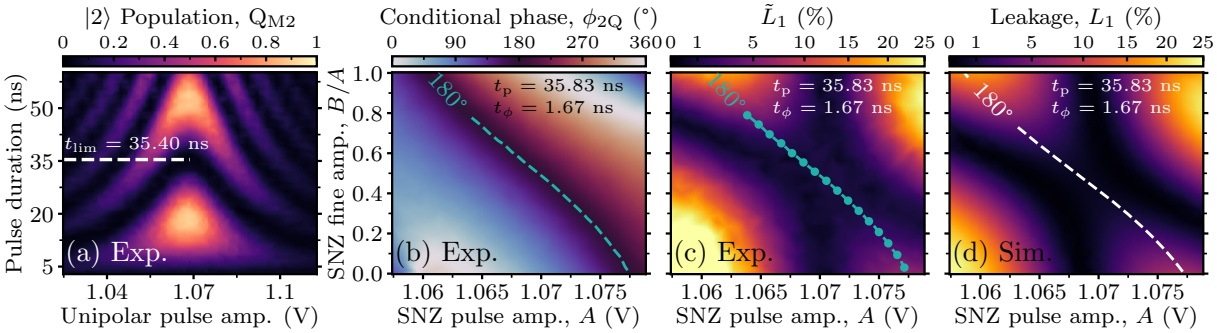


Figure 5.5: Calibration of the SNZ pulse for pair Q_L-Q_{M2} and comparison to simulation. (a) $|2\rangle$-state population of Q_{M2} as a function of the amplitude and duration of a unipolar square pulse making $|11\rangle$ interact with $|02\rangle$. The characteristic chevron pattern is used to identify t_{lim} and the amplitude bringing $|11\rangle$ and $|02\rangle$ on resonance. (b,c) Landscapes of conditional phase ϕ_{2Q} and leakage estimate $\tilde{L}_1$ as a function of SNZ pulse amplitudes A and B, with $t_p = t_{lim}^+$ and $t_\phi = 1.67$ ns. The juxtaposed $\phi_{2Q} = 180°$ contour runs along the opposite diagonal compared to Figs. 5.1(b) and 5.1(c) because, contrasting with t_ϕ, increasing B (which decreases Δ_{02}) changes ϕ in the opposite direction. Data points marked with dots are measured with extra averaging for detailed examination in Fig. 5.6. (d) Numerical simulation of leakage L_1 landscape and $\phi_{2Q} = 180°$ contour. The simulation uses pulse and transmon parameters from experiment, and includes the measured flux-pulse distortion. All landscapes are sampled using the adaptive algorithm developed in this work (`LearnerND_minimizer`, Section 4.3) based on [78].

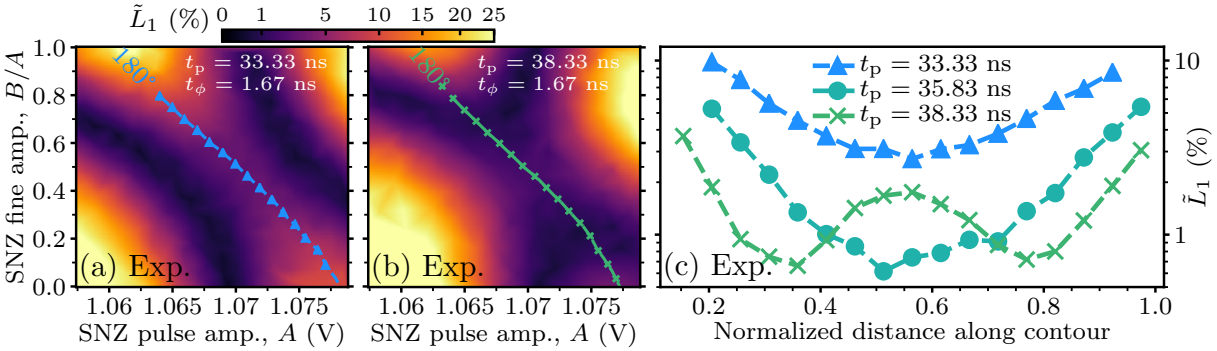


Figure 5.6: Landscapes of the leakage estimate $\tilde{L}_1$ for (a) intentionally short ($t_p = t_{lim}^+ - 6t_s$) and (b) intentionally long ($t_p = t_{lim}^+ + 6t_s$) flux pulses on Q_{M2}. These landscapes are sampled using the adaptive algorithm developed in this work (`LearnerND_minimizer`, Section 4.3) based on [78]. (c) Extracted $\tilde{L}_1$ along the $\phi_{2Q} = 180°$ contours from (a), (b), and Fig. 5.5(c). For $t_p < t_{lim}$, a single minimum is observed, but at a higher value than the minimum for $t_p = t_{lim}^+$. For $t_p > t_{lim}$, two leakage minima are found, matching the single minimum for $t_p = t_{lim}^+$.

at $A = 1$ and a diagonal valley. Juxtaposing the $\phi_{2Q} = 180°$ contour extracted from the ϕ_{2Q} landscape [Fig. 5.5(b)], we observe the matching of PC at the crossing of these valleys. These experimental observations are in excellent agreement with a numerical two-qutrit simulation [Fig. 5.5(d)].

SNZ robustness to hardware limitations

Experimentally, it is nearly impossible to precisely match $t_p = t_{lim}$ due to the discreteness of t_s. To understand the consequences of t_p mismatch, we examine the ϕ_{2Q} and $\tilde{L}_1$ landscapes for SNZ pulses with intentionally set $t_p = t_{lim}^+ \pm 6t_s$ (Fig. 5.6). We find that the PC contour remains roughly unchanged in both cases. However, there are significant effects on $\tilde{L}_1$. In both cases, we observe that $\tilde{L}_1$ lifts at the prior crossing of LC2 and LC3 valleys where $\phi_{2Q} = 180°$. For too-short pulses [Fig. 5.6(a)], there remain two valleys of minimal $\tilde{L}_1$, but these are now curved and do not cross the $\phi_{2Q} = 180°$ contour. For too-long pulses [Fig. 5.6(b)], there are also two curved valleys. Crucially, these cross the $\phi_{2Q} = 180°$ contour, and it remains possible to achieve PC and minimize leakage at two (A, B) settings. Extracting $\tilde{L}_1$ along the $\phi_{2Q} = 180°$ contours [Fig. 5.6(c)] confirms that the minimal leakage obtainable for $t_p = t_{lim}^+ + 6t_s$ matches that for $t_p = t_{lim}^+$. The observed impossibility to achieve minimal leakage at $\phi_{2Q} = 180°$ for $t_p < t_{lim}$ is a clear manifestation of the speed limit set by J_2. In turn, the demonstrated possibility to do so for $t_p > t_{lim}$ (even when overshooting the duration significantly) is an important proof of the viability of the SNZ pulse in practice.

5.3.3 Experimental performance of SNZ CZ gates

With these insights, we follow similar procedures to tune SNZ CZ gates for the four transmon pairs. Namely, we use final weak bipolar pulses of total duration $t_{1Q} = 10$ ns to null the single-qubit phases in the frame of microwave drives. Also, since the codeword-based control electronics has a 20 ns timing grid, and $40 \text{ ns} < t_{total} = t_p + t_\phi + t_{1Q} < 60$ ns for all pairs, we allocate 60 ns to every CZ gate. Some pair-specific details must be noted. Owing to the overlap of qubit frequencies between mid-frequency qubits, implementing CZ between Q_H and Q_{M1} (Q_{M2}) requires parking of Q_{M2} (Q_{M1}) during the SNZ pulse on

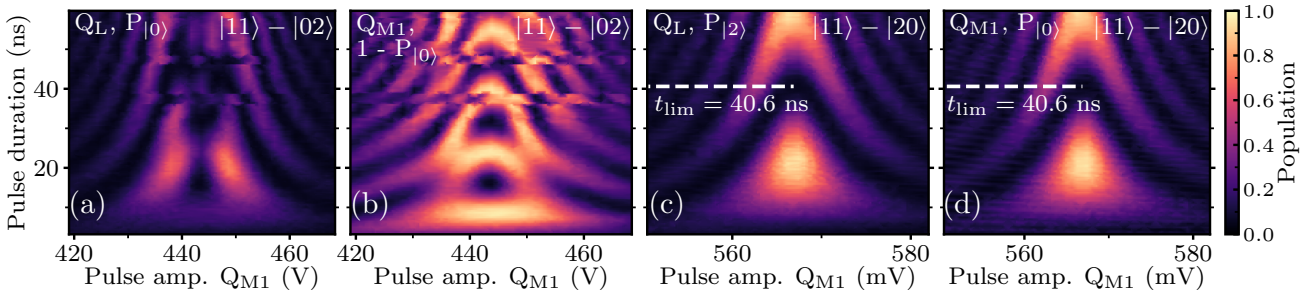


Figure 5.7: Time-domain characterization of the $|11\rangle$-$|02\rangle$ and (c,d) $|11\rangle$-$|20\rangle$ interactions for pair Q_L-Q_{M1}. (a,b) Landscapes of (a) ground-state population $P_{|0\rangle}$ of Q_L and (b) total excited-state population $1 - P_{|0\rangle}$ of Q_{M1} as a function of the amplitude and duration of a unipolar square pulse near the $|11\rangle$-$|02\rangle$ resonance. The absence of the expected chevron pattern in these landscapes reflects a flickering TLS resonant with the qubit transition of Q_{M1} at this pulse amplitude. Horizontally shifting fringes in (a) and (b) are due to flickering of the TLS on the scale of a few minutes. These observations preclude the use of the $|11\rangle$-$|02\rangle$ interaction to realize the CZ gate. In contrast, the landscapes of (c) two-state population $P_{|2\rangle}$ of Q_L and (d) $P_{|0\rangle}$ of Q_{M1} and as a function of unipolar square pulse parameters near the $|11\rangle$-$|20\rangle$ resonance reveal a standard, stable chevron pattern. All landscapes are sampled using the adaptive algorithm developed in this work (`LearnerND_minimizer`, Section 4.3) based on [78].

Q_H [17, 18]. The parking flux pulse is also bipolar, with each half a square pulse lasting $(t_p + t_\phi)/2$. Its amplitude is chosen to downshift the parked qubit by ~ 300 MHz, and fine tuned to null its single-qubit phase.

For most pairs, we employ the $|11\rangle$-$|02\rangle$ interaction, which requires the smallest flux amplitude (reducing the impact of dephasing from flux noise) and does not require crossing any other interaction on the way to and from it. However, for pair Q_L-Q_{M1}, we cannot reliably use this interaction as there is a two-level system (TLS) interacting intermittently with Q_{M1} at the flux amplitude placing $|11\rangle$ and $|02\rangle$ on resonance. Figures 5.7(a) and 5.7(b) show the negative impact of this TLS when attempting to characterize the $|11\rangle$-$|02\rangle$ interaction by the standard time-domain chevron measurement. While experience shows that it is probable that such a TLS could be displaced or eliminated by thermal cycling at least above the critical temperature of aluminum, we chose instead to use the more flux distant $|11\rangle$-$|20\rangle$ interaction to realize the SNZ CZ gate for this pair. For this interaction, a standard, stable chevron pattern is observed [Figs. 5.7(c) and 5.7(d)]. Using square pulses is a side benefit of SNZ in this case: it minimizes exchange between $|01\rangle$ and the TLS, $|11\rangle$ and $|20\rangle$, and $|01\rangle$ and $|10\rangle$ as their resonances are crossed as suddenly as possible.

Table 5.2 summarizes the timing parameters and performance attained for the four SNZ CZ gates. The CZ gate fidelity F and leakage L_1 are extracted using a 2QIRB protocol modified to quantify leakage [39, 59]. For each pair, we report the best, average and standard deviation of both values based on at least 10 repetitions of the protocol spanning more than 8 h [see Appendix A.2 for details]. Several observations can be drawn. First, CZ gates involving Q_H perform better on average than those involving Q_L. This is likely due to the shorter t_{lim} and correspondingly longer time 60 ns $- t_p$ spent near the sweetspot. Another possible reason is that the frequency downshifting required of Q_H to interact with Q_{M1} and Q_{M2} is roughly half that required of the latter to interact with Q_L. This reduces the impact of dephasing from flux noise during the pulse. Not surprisingly, performance is worst for the pair Q_L-Q_{M1}. Here, the pulse must downshift Q_{M1} the most to reach the distant $|11\rangle$-$|20\rangle$ interaction, increasing

Parameter	$Q_{M1} - Q_H$	$Q_{M2} - Q_H$	$Q_L - Q_{M1}$	$Q_L - Q_{M2}$
t_p [t_{lim}] (ns)	32.50 [32.20]	29.10 [29.00]	40.83 [40.60]	35.83 [35.40]
t_ϕ (ns)	2.92	3.75	1.25	1.67
$(t_p + t_\phi)/t_{lim}$	1.10	1.13	1.04	1.06
t_{total} (ns)	45.42	42.91	52.08	47.50
Interaction	$\|11\rangle - \|02\rangle$	$\|11\rangle - \|02\rangle$	$\|11\rangle - \|20\rangle$	$\|11\rangle - \|02\rangle$
Parked qubit	Q_{M2}	Q_{M1}	–	–
Avg. F (%)	98.89(35)	99.54(27)	93.72(210)	97.14(72)
Max. F (%)	99.77(23)	99.93(24)	99.15(120)	98.56(70)
Avg. L_1 (%)	0.13(2)	0.18(4)	0.78(32)	0.63(11)
Min. L_1 (%)	0.07(4)	0.10(2)	0.04(8)	0.41(10)

Table 5.2: Summary of SNZ CZ pulse parameters and achieved performance for the four transmon pairs. All SNZ CZ gates null single-qubit phases with weak bipolar square pulse of duration $t_{1Q} = 10$ ns immediately following the strong pulse. We allocate 60 ns to every CZ gate to conform to the 20 ns timing grid of the control electronics. Gate fidelities and leakage are obtained from 2QIRB keeping the other two qubits in $|0\rangle$. Statistics (average and standard deviation) are taken from repeated 2QIRB runs [for details see Appendix A.2]. The maximum F and minimum L_1 quoted are not necessarily from the same run.

dephasing from flux noise. Also, there may be residual exchange with the identified TLS and as the $|11\rangle$-$|02\rangle$ and $|01\rangle$-$|10\rangle$ resonances are crossed. Overall, there is significant temporal variation in the performance as gleaned by repeated 2QIRB characterizations. This is likely the reflection of the underlying variability of qubit relaxation and dephasing times, which, however, were not tracked simultaneously. In addition to having the best average performance, pair Q_{M2}-Q_H also displays the hero performance based on a single 2QIRB run (Fig. 5.8). Peaking at $F = 99.93\,\%$, to the best of our knowledge, is the highest CZ fidelity extracted from 2QIRB characterization in a transmon processor.

5.3.4 Error sources: SNZ vs CNZ

In an effort to identify the dominant sources of infidelity $\varepsilon = 1 - F$ and leakage for SNZ CZ gates, we perform a two-qutrit numerical simulation for pair Q_{M2}-Q_H with an error model taking parameters from experiment [Fig. 5.9]. As in the previous work on CNZ [39], the simulation cumulatively adds: (A) no noise; (B) energy relaxation; (C) Markovian dephasing; (D) dephasing from low-frequency flux noise; and (E) the remaining measured flux-pulse distortion. The experimental inputs for models B, C and D combine measured qubit relaxation time T_1 at the bias point, and measured echo and Ramsey dephasing times (T_2 and T_2^*) as a function of qubit frequency. The input to E consists of a final Cryoscope measurement of the flux step response using all real-time filters.

The simulations suggest that the main source of ε is Markovian dephasing (as in [39]), while the dominant contribution to L_1 is low-frequency flux noise. The latter contrasts with Ref. 39, where simulation identified flux-pulse distortion as the dominant leakage source. We identify two possible reasons for this difference: in the current experiment, the $1/f$ low-frequency flux noise is ~ 4 times larger (in units of $\Phi_0/\sqrt{\text{Hz}}$) and the achieved flux step response is noticeably sharper. Finally, we use the simulation to compare performance of SNZ to CNZ CZ. For the latter, we fix $t_\phi = 0$, $t_{1Q} = 60$ ns $- t_p$, and use the fast-adiabatic pulse shape and $t_p = 45.83$ ns optimized by simulation. Overall, the error sources

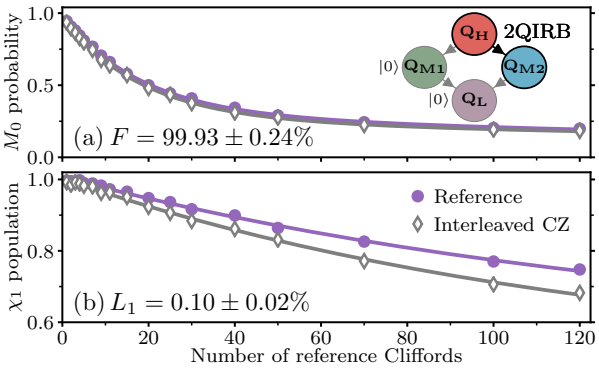


Figure 5.8: Best SNZ CZ gate performance achieved from a single run of 2QIRB. (a) Reference and CZ-interleaved return probability M_0 to $|00\rangle$ and (b) population in the computational space χ_1 as a function of the number of two-qubit Cliffords in the reference curve. Errors bars in F and L_1 are obtained from the uncertainty of exponential-decay fits.

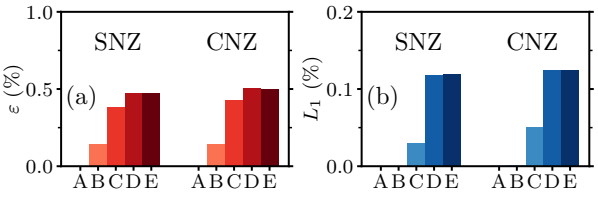


Figure 5.9: Error budgets for infidelity ε (a) and leakage L_1 (b) obtained by a numerical simulation (as in [39]) of the Q_{M2}-Q_H SNZ CZ gate with parameters in Fig. 5.8 and for a CNZ gate with optimized parameters (see text for details). The simulation incrementally adds errors using experimental input parameters for this pair: (A) no noise; (B) relaxation; (C) Markovian dephasing; (D) dephasing from quasistatic flux noise; and (E) flux-pulse distortion. Credits simulations and analysis: F. Battistel, Terhal group, QuTech.

contribute very similarly to the error budget for both cases. The marginally higher overall performance found for SNZ is likely due to the increased time spent at the sweetspot during the 60 ns allocated for each CZ.

Chapter 6

Conclusions and outlook

This chapter provides a summary of the work developed within this thesis reflecting back on important conclusions from the hands-on experience of operating a transmons quantum processor in the quest to automatic and efficient calibration of two-qubit gates. In addition we outlook the next steps towards a fully automated calibration of an entire quantum chip and some implication for large scale devices.

6.1 Conclusions

In Chapter 2 we gained some intuition for the physical principles behind the transmons qubits and glanced over the experimental interaction with a superconducting transmon quantum processor.

The automatic flux bias offset calibration and analysis in Chapter 3 solved the key challenge of precisely flux-biasing a transmon to its maximum-frequency sweetspot, as required for high-fidelity two-qubit CZ gates. A six-fold speedup over a human operator was achieved based on a tailor-made 1D adaptive sampling algorithm (`Learner1D_Minimizer`) yielding robust and easy to analyze measurement datasets. Furthermore, the algorithm was noted to be suitable for other sampling and/or optimization problems owing to its configurable mode of operation and served as a template for an N-dimensional generalization. In addition, as a result of the peculiarities of this calibration, the control of data acquisition within PycQED has been extended to support more exotic acquisition procedures.

In the subsequent Chapter 4 we detailed the technicalities of the multi-featured sampling and analysis developed for the CZ 2D calibration landscapes. These tools found immediate application when calibrating CZ gates in the context of a Quantum Approximate Optimization Algorithm (QAOA) NISQ application [19]. Though harder to quantify quantitatively, the provided features of the analysis significantly simplify the work required from an experimentalist to troubleshoot and calibrate a CZ. It provides user-friendly visualization of results, high flexibility and, crucially, automatic identification and extraction of (multiple) optimal points, besides other quantities of interest ready to be leveraged for re-calibrations. At end of the chapter, in Section 4.3, we introduce the generic `LearnerND_Minimizer` which has been routinely employed, both within experiments and simulations, to optimally discriminate the features of those landscapes with a configurable bias towards regions of interest or to optimize the CZ parameters.

In Chapter 5, the sudden variant (SNZ) of the NZ scheme [39] realizing flux-based CZ gates by exploiting transverse coupling between computational and non-computation states was proposed and implemented. SNZ operates at the speed limit of the transverse coupling by maximizing intermediate leakage to the non-computational state. The key advantage of SNZ over CNZ is ease of tuneup, owing to the simple structure of error landscapes as a function of pulse parameters. High-fidelity CZ gates were demonstrated between four transmon pairs in a patch of a transmon processor. Control architectures without a timing grid will further benefit from the increased speed of SNZ over the conventional NZ by reducing the total gate time and thus the impact of decoherence. Taking advantage of the tuning simplicity, SNZ CZ gates are already employed in the Starmon-5 quantum processor publicly available via the QuTech Quantum Inspire platform [21]. Moving forwards, the full compatibility of SNZ with the previously proposed [17] scalable scheme for surface coding makes SNZ the two-qubit gating choice for quantum error correction. Finally, the noted straightforward extension of SNZ to arbitrary conditional-phase gates can find immediate use in NISQ applications.

From personal experience in operating and troubleshooting the experimental setup during this project, it is worth sharing a few practical learnings. It became clear to me how valuable it is for an experimentalist to have a flexible (software) toolbox that can be reused and adapted to unexpected results without developing new tools from scratch. Considering that currently that toolbox is developed by the same experimentalist, it often not easy to balance the research deadlines of yesterday with the investment in flexible and reusable software. Nonetheless, it is usually a compounding interest that pays off in the future. If you need a script twice, it is likely it will be needed a hundred times, make a generic routine (without over-engineering it), you and your peers will thank you later. Solve a problem well the first time, shortcuts are a highway to compounding technical debt.

6.2 Outlook

It is not clear at this point which quantum chip architecture and two-qubit gating choice will yield the holy grail of a useful quantum supremacy. Hence, it is worthwhile to further research and develop the different technologies with such potential.

For our platform, automation already brings significant benefits on a 7-transmon device and it will become imperative as we advance towards tens and hundreds of qubits. Therefore, it is essential to continue laying down strong foundations on this front. The next steps comprise the full automation of all the dependencies of the CZ calibrations and the integration of all nodes together in an autonomous chip-level calibration and state maintenance. Given that there are several well-defined nodes, these routines can be developed in parallel through multiple project like this one.

When scaling up in qubit count, the need for massive parallelization should be equated on the measurement and software level, if the calibration of a quantum chip as whole is to be finished on a practical time-scale, before its individual constituents drift away. Fortunately, it is plausible to perform many calibrations and optimization in parallel (e.g. Multiplexed readout, Cryoscope, appropriate Chevron/CZ combinations). With the increasing complexity of the calibration and data processing software, one must

also shift towards parallel computation inside the classical processor. Even though the benefits might be a few years ahead, I cannot stress enough how critical it is to take these considerations into account today!

Farther into the future, mastering kilo-qubit devices are expected to require fundamental changes on how the calibration problem is approached as there is only so many cables one can route into a fridge. This will likely require bringing the classical computation as closes as possible to the quantum chip itself and reducing to the minimum the communications overhead (between the user computer and the controlling hardware).

Bibliography

[1] E. Gibney. Hello quantum world! Google publishes landmark quantum supremacy claim, 10 2019. URL http://www.nature.com/articles/d41586-019-03213-z.

[2] Quantum Computing Participants | Quantum Computing Report, . URL https://quantumcomputingreport.com/players/.

[3] E. Gibney. Quantum gold rush: the private funding pouring into quantum start-ups. *Nature*, 574(7776):22–24, 10 2019. ISSN 0028-0836. doi: 10.1038/d41586-019-02935-4. URL http://www.nature.com/articles/d41586-019-02935-4.

[4] E. SPERLING. Quantum Effects At 7/5nm And Beyond. URL https://semiengineering.com/quantum-effects-at-7-5nm/.

[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis. Quantum supremacy using a programmable superconducting processor. *Nature*, 574(7779):505–510, oct 2019. ISSN 0028-0836. doi: 10.1038/s41586-019-1666-5. URL http://www.nature.com/articles/s41586-019-1666-5.

[6] M. A. Rol. Quantum Computational Supremacy – Bits of Quantum, 2019. URL https://blog.qutech.nl/index.php/2019/10/31/quantum-computational-supremacy/.

[7] N. Davis. Exploring Implications of Exponential Technologies - SU Blogs. URL https://su.org/blog/implications-of-exponential-technologies/.

[8] M. Jackson. 6 Things Quantum Computers Will Be Incredibly Useful For, 2017. URL

https://singularityhub.com/2017/06/25/6-things-quantum-computers-will-be-incredibly-useful-for/.

[9] Google AI Quantum. URL https://research.google/teams/applied-science/quantum/.

[10] Accenture. Consortium Applies Quantum Computing to Drug Discovery for Neurological Diseases - Drug Discovery and Development. URL https://www.drugdiscoverytrends.com/consortium-applies-quantum-computing-to-drug-discovery-for-neurological-diseases/.

[11] R. Stevens. Bitcoin's race to outrun the quantum computer - Decrypt. URL https://decrypt.co/8498/bitcoins-race-to-outrun-the-quantum-computer.

[12] A. Purcell. Exploring quantum computing for high-energy physics | CERN. URL https://home.cern/news/news/computing/exploring-quantum-computing-high-energy-physics.

[13] L. DiCarlo, J. M. Chow, J. M. Gambetta, L. S. Bishop, B. R. Johnson, D. I. Schuster, J. Majer, A. Blais, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf. Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature*, 460:240, 2009. URL http://www.nature.com/nature/journal/v460/n7252/abs/nature08121.html.

[14] J. Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, Aug. 2018. ISSN 2521-327X. doi: 10.22331/q-2018-08-06-79. URL https://doi.org/10.22331/q-2018-08-06-79.

[15] P. W. Shor. Fault-tolerant quantum computation. *Annual Symposium on Foundations of Computer Science - Proceedings*, pages 56–65, 1996. ISSN 02725428. doi: 10.1142/9789812385253_0008.

[16] Fault-tolerant Quantum Computing Archives - QuTech. URL https://qutech.nl/roadmap/fault-tolerant-quantum-computing/.

[17] R. Versluis, S. Poletto, N. Khammassi, B. M. Tarasinski, N. Haider, D. J. Michalak, A. Bruno, K. Bertels, and L. DiCarlo. Scalable quantum circuit and control for a superconducting surface code. *Phys. Rev. Appl.*, 8:034021, Sep 2017. doi: 10.1103/PhysRevApplied.8.034021. URL https://link.aps.org/doi/10.1103/PhysRevApplied.8.034021.

[18] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, N. Lacroix, G. J. Norris, M. Gabureac, C. Eichler, and A. Wallraff. Repeated quantum error detection in a surface code. *Nat. Phys.*, 16(8):875–880. ISSN 1745-2481. URL https://doi.org/10.1038/s41567-020-0920-y.

[19] R. Sagastizabal, S. P. Premaratne, B. A. Klaver, M. A. Rol, V. Negîrneac, M. Moreira, X. Zou, S. Johri, N. Muthusubramanian, M. Beekman, C. Zachariadis, V. P. Ostroukh, N. Haider, A. Bruno, A. Y. Matsuura, and L. DiCarlo. Variational preparation of finite-temperature states on a quantum computer. 2020. URL https://arxiv.org/abs/2012.03895.

[20] V. Negîrneac, H. Ali, N. Muthusubramanian, F. Battistel, R. Sagastizabal, M. S. Moreira, J. F. Marques, W. Vlothuizen, M. Beekman, N. Haider, A. Bruno, and L. DiCarlo. High-fidelity controlled-Z gate with maximal intermediate leakage operating at the speed limit in a superconducting quantum processor. 2020. URL https://arxiv.org/abs/2008.07411.

[21] Qutech Quantum Inspire, . URL https://www.quantum-inspire.com/. https://www.quantum-inspire.com/.

[22] C. Zachariadis. Fabricating a superconducting quantum processor in a cleanroom. URL https://blog.qutech.nl/2020/04/24/fabricating-a-superconducting-quantum-processor-in-a-cleanroom/.

[23] J. Martinis. Quantum supremacy using a programmable superconducting processor - J. Martinis - 11/1/2019 - YouTube. URL https://www.youtube.com/watch?v=FklMpRiTeTA&t=1760s.

[24] C. Dickel. A Cloud Quantum Computer Business Plan. URL https://blog.qutech.nl/2018/07/18/a-cloud-quantum-computer-business-plan/.

[25] M. A. Rol, C. Dickel, S. Asaad, C. C. Bultink, R. Sagastizabal, N. K. L. Langford, G. de Lange, B. C. S. Dikken, X. Fu, S. R. de Jong, and F. Luthi. PycQED, 2016. URL https://github.com/DiCarloLab-Delft/PycQED_py3.

[26] A. Johnson, G. Ungaretti, et al. QCoDeS, 2016. URL https://github.com/QCoDeS/Qcodes.

[27] J. Kelly, P. O'Malley, M. Neeley, H. Neven, and J. M. Martinis. Physical qubit calibration on a directed acyclic graph. 2018. URL http://arxiv.org/abs/1803.03226.

[28] M. A. Rol, L. Ciorciaro, and P. Eendebak. AutoDepGraph, 2017. URL https://github.com/AdriaanRol/AutoDepGraph.

[29] M. A. Rol. *Control for Programmable Superconducting Quantum Systems*. PhD thesis, Delft University of Technology, Oct. 2020. URL https://doi.org/10.4233/uuid:0a2ba212-f6bf-4c64-8f3d-b707f1e44953.

[30] M. A. Rol, C. C. Bultink, T. E. O'Brien, S. R. de Jong, L. S. Theis, X. Fu, F. Luthi, R. F. L. Vermeulen, J. C. de Sterke, A. Bruno, D. Deurloo, R. N. Schouten, F. K. Wilhelm, and L. DiCarlo. Restless tuneup of high-fidelity qubit gates. *Phys. Rev. Applied*, 7:041001, Apr 2017. doi: 10.1103/PhysRevApplied.7.041001. URL https://link.aps.org/doi/10.1103/PhysRevApplied.7.041001.

[31] M. A. Rol, L. Ciorciaro, F. K. Malinowski, B. M. Tarasinski, R. E. Sagastizabal, C. C. Bultink, Y. Salathe, N. Haandbaek, J. Sedivy, and L. DiCarlo. Time-domain characterization and correction of on-chip distortion of control pulses in a quantum processor. *Applied Physics Letters*, 116(5): 054001, 2020. doi: 10.1063/1.5133894. URL https://doi.org/10.1063/1.5133894.

[32] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7 (4):308–313, 1965.

[33] N. Hansen. The CMA Evolution Strategy: A Tutorial. 2016. URL http://arxiv.org/abs/1604.00772.

[34] J. M. Martinis and M. R. Geller. Fast adiabatic qubit gates using only σ_z control. *Phys. Rev. A*, 90: 022307, Aug 2014. doi: 10.1103/PhysRevA.90.022307.

[35] M. Reagor, C. B. Osborn, N. Tezak, A. Staley, G. Prawiroatmodjo, M. Scheer, N. Alidoust, E. A. Sete, N. Didier, M. P. Da Silva, E. Acala, J. Angeles, A. Bestwick, M. Block, B. Bloom, A. Bradley, C. Bui, S. Caldwell, L. Capelluto, R. Chilcott, J. Cordova, G. Crossman, M. Curtis, S. Deshpande, T. El Bouayadi, D. Girshovich, S. Hong, A. Hudson, P. Karalekas, K. Kuang, M. Lenihan, R. Manenti, T. Manning, J. Marshall, Y. Mohan, W. O'Brien, J. Otterbach, A. Papageorge, J. P. Paquette, M. Pelstring, A. Polloreno, V. Rawat, C. A. Ryan, R. Renzas, N. Rubin, D. Russel, M. Rust, D. Scarabelli, M. Selvanayagam, R. Sinclair, R. Smith, M. Suska, T. W. To, M. Vahidpour, N. Vodrahalli, T. Whyland, K. Yadav, W. Zeng, and C. T. Rigetti. Demonstration of universal parametric entangling gates on a multi-qubit lattice. *Science Advances*, 4(2), 2018. ISSN 23752548. doi: 10.1126/sciadv.aao3603.

[36] Y. Wu, L. P. Yang, M. Gong, Y. Zheng, H. Deng, Z. Yan, Y. Zhao, K. Huang, A. D. Castellano, W. J. Munro, K. Nemoto, D. N. Zheng, C. P. Sun, Y. x. Liu, X. Zhu, and L. Lu. An efficient and compact switch for quantum circuits. *npj Quantum Information*, 4(1), 2018. ISSN 20566387. doi: 10.1038/s41534-018-0099-6.

[37] S. Sheldon, L. S. Bishop, E. Magesan, S. Filipp, J. M. Chow, and J. M. Gambetta. Characterizing errors on qubit operations via iterative randomized benchmarking. *Phys. Rev. A*, 93:012301, 2016. doi: 10.1103/PhysRevA.93.012301.

[38] H. Paik, A. Mezzacapo, M. Sandberg, D. T. McClure, B. Abdo, A. D. Córcoles, O. Dial, D. F. Bogorin, B. L. T. Plourde, M. Steffen, A. W. Cross, J. M. Gambetta, and J. M. Chow. Experimental demonstration of a resonator-induced phase gate in a multiqubit circuit-qed system. *Phys. Rev. Lett.*, 117:250502, 2016. URL https://link.aps.org/doi/10.1103/PhysRevLett.117.250502.

[39] M. A. Rol, F. Battistel, F. K. Malinowski, C. C. Bultink, B. M. Tarasinski, R. Vollmer, N. Haider, N. Muthusubramanian, A. Bruno, B. M. Terhal, and L. DiCarlo. Fast, high-fidelity conditional-phase gate exploiting leakage interference in weakly anharmonic superconducting qubits. *Phys. Rev. Lett.*, 123:120502, 2019. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.123.120502.

[40] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, 2000.

[41] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A*, 76:042319, 2007. URL http://journals.aps.org/pra/abstract/10.1103/PhysRevA.76.042319.

[42] P. Krantz, M. Kjaergaard, F. Yan, T. P. Orlando, S. Gustavsson, and W. D. Oliver. A quantum engineer's guide to superconducting qubits. *Applied Physics Reviews*, 6(2):021318, jun 2019. ISSN 1931-9401. doi: 10.1063/1.5089550. URL https://aip.scitation.org/doi/10.1063/1.5089550.

[43] C. Dickel. *Scalability and modularity for transmon-based quantum processors*. PhD Dissertation, Delft University of Technology, 2018. URL https://doi.org/10.4233/uuid:78155c28-3204-4130-a645-a47e89c46bc5.

[44] L. DiCarlo. An introduction to circuit QED - Superconducting Quantum Bits Quantum Science and Quantum Information Minor - Lecture Slides, 2019.

[45] J. Johansson, P. Nation, and F. Nori. QuTiP 2: A Python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 184:1234 – 1240, 2013. doi: https://doi.org/10.1016/j.cpc.2012.11.019. URL https://doi.org/10.1016/j.cpc.2012.11.019.

[46] C. D. Hill, E. Peretz, S. J. Hile, M. G. House, M. Fuechsle, S. Rogge, M. Y. Simmons, and L. C. L. Hollenberg. A surface code quantum computer in silicon. *Sci. Adv.*, 1(9), 2015. doi: 10.1126/sciadv.1500707. URL http://advances.sciencemag.org/content/1/9/e1500707.

[47] J. Lisenfeld, A. Bilmes, A. Megrant, R. Barends, J. Kelly, P. Klimov, G. Weiss, J. M. Martinis, and A. V. Ustinov. Electric field spectroscopy of material defects in transmon qubits. *npj Quantum Information*, 5(1), Nov 2019. ISSN 2056-6387. doi: 10.1038/s41534-019-0224-1. URL http://dx.doi.org/10.1038/s41534-019-0224-1.

[48] D. P. DiVincenzo. The physical implementation of quantum computation. *Fortschritte der Physik*, 48:771, 2000. ISSN 00158208. doi: 10.1002/1521-3978(200009)48:9/11<771::AID-PROP771>3.0.CO;2-E. URL https://onlinelibrary.wiley.com/doi/10.1002/1521-3978(200009)48:9/11%3C771::AID-PROP771%3E3.0.CO;2-E.

[49] G. Wendin. Quantum information processing with superconducting circuits: A review. *Reports on Progress in Physics*, 80(10), 2017. ISSN 00344885. doi: 10.1088/1361-6633/aa7e1a.

[50] J. Benhelm, G. Kirchmair, C. F. Roos, and R. Blatt. Towards fault-tolerant quantum computing with trapped ions. *Nature Physics*, 4(6):463–466, 2008. ISSN 17452481. doi: 10.1038/nphys961.

[51] X. Rong, J. Geng, F. Shi, Y. Liu, K. Xu, W. Ma, F. Kong, Z. Jiang, Y. Wu, and J. Du. Experimental fault-tolerant universal quantum gates with solid-state spins under ambient conditions. *Nature Communications*, 6, 2015. ISSN 20411723. doi: 10.1038/ncomms9748.

[52] M. H. Devoret. Houches_fluctuations.pdf, 1995.

[53] N. K. Langford. Circuit QED - Lecture Notes, 2013. URL http://arxiv.org/abs/1310.1897.

[54] N. K. Langford. Errors in quantum tomography: diagnosing systematic versus statistical errors. *New J. Phys.*, 15(3):035003, 2013. URL http://iopscience.iop.org/article/10.1088/1367-2630/15/3/035003.

[55] D. C. McKay, C. J. Wood, S. Sheldon, J. M. Chow, and J. M. Gambetta. Efficient z gates for quantum computing. *Phys. Rev. A*, 96:022330, Aug 2017. doi: 10.1103/PhysRevA.96.022330. URL https://link.aps.org/doi/10.1103/PhysRevA.96.022330.

[56] L. DiCarlo, M. D. Reed, L. Sun, B. R. Johnson, J. M. Chow, J. M. Gambetta, L. Frunzio, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf. Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature*, 467:574, 2010. URL https://doi.org/10.1038/nature09416.

[57] T. Yamamoto, M. Neeley, E. Lucero, R. C. Bialczak, J. Kelly, M. Lenander, M. Mariantoni, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner, Y. Yin, A. N. Cleland, and J. M. Martinis. Quantum process tomography of two-qubit controlled-z and controlled-not gates using superconducting phase qubits. *Phys. Rev. B*, 82:184515, 2010.

[58] B. Varbanov, F. Battistel, B. M. Tarasinski, V. P. Ostroukh, T. E. O'Brien, B. M. Terhal, and L. DiCarlo. Characterization and detection of leakage in simulations of a transmon-qubit surface-17 architecture. *arXiv:2002.07119*, in prepartion (2020). URL https://arxiv.org/abs/2002.07119.

[59] C. J. Wood and J. M. Gambetta. Quantification and characterization of leakage errors. *Phys. Rev. A*, 97:032306, Mar 2018. doi: 10.1103/PhysRevA.97.032306. URL https://link.aps.org/doi/10.1103/PhysRevA.97.032306.

[60] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis. Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508(7497):500, Appl. Phys. Rev. 2014. ISSN 1476-4687. URL http://www.nature.com/nature/journal/v508/n7497/abs/nature13171.html.

[61] J. Kelly, R. Barends, A. G. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, B. Campbell, Y. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, A. N. Cleland, J. Wenner, and J. M. Martinis. State preservation by repetitive error detection in a superconducting quantum circuit. *Nature*, 519(7541):66–69, 2015. URL https://www.nature.com/nature/journal/v519/n7541/full/nature14270.html.

[62] Realisation of high-fidelity nonadiabatic CZ gates with superconducting qubits. *npj Quantum Information*, 5(1):84, 2019. ISSN 2056-6387. doi: 10.1038/s41534-019-0202-7. URL https://doi.org/10.1038/s41534-019-0202-7.

[63] Quantum Computing Report. Scorecards: Qubit Quality. URL https://quantumcomputingreport.com/scorecards/qubit-quality/.

[64] E. Taylor and J. Fortes. Device variability impact on logic gate failure rates. *16th IEEE Int. Conf. Application-Specific Syst., Archit. & Processors (ASAP 2005)*, pages 247–253, 2005.

[65] M. Sarovar, T. Proctor, K. Rudinger, K. Young, E. Nielsen, and R. Blume-Kohout. Detecting crosstalk errors in quantum information processors. *ArXiv:1908.09855*, 2019. URL https://arxiv.org/abs/1908.09855.

[66] X. Fu, M. A. Rol, C. C. Bultink, J. van Someren, N. Khammassi, I. Ashraf, R. F. L. Vermeulen, J. C. de Sterke, W. J. Vlothuizen, R. N. Schouten, C. G. Almudever, L. DiCarlo, and K. Bertels. An experimental microarchitecture for a superconducting quantum processor. In *Proceedings of the 50th Annual IEEE/ACM International Symposium on Microarchitecture*, MICRO-50 '17, pages 813–825, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4952-9. doi: 10.1145/3123939.3123952. URL http://doi.acm.org/10.1145/3123939.3123952.

[67] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Phys. Rev. A*, 86:032324, Sep 2012. URL https://link.aps.org/doi/10.1103/PhysRevA.86.032324.

[68] C. Chamberland, G. Zhu, T. J. Yoder, J. B. Hertzberg, and A. W. Cross. Topological and Subsystem Codes on Low-Degree Graphs with Flag Qubits. *Physical Review X*, 10(1):1–21, 2020. ISSN 2160-3308. doi: 10.1103/physrevx.10.011022.

[69] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Phys. Rev. Lett.*, 103:110501, 2009. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.103.110501.

[70] J. M. Chow, L. DiCarlo, J. M. Gambetta, F. Motzoi, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf. Optimized driving of superconducting artificial atoms for improved single-qubit gates. *Phys. Rev. A*, 82:040305, 2010. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.82.040305.

[71] E. Lucero, J. Kelly, R. C. Bialczak, M. Lenander, M. Mariantoni, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner, T. Yamamoto, A. N. Cleland, and J. M. Martinis. Reduced phase error through optimized control of a superconducting qubit. *Phys. Rev. A*, 82:042339, 2010.

[72] E. Knill, D. Leibfried, R. Reichle, J. Britton, R. B. Blakestad, J. D. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. J. Wineland. Randomized benchmarking of quantum gates. *Phys. Rev. A*, 77: 012307, 2008.

[73] J. M. Epstein, A. W. Cross, E. Magesan, and J. M. Gambetta. Investigating the limits of randomized benchmarking protocols. *Phys. Rev. A*, 89:062321, Jun 2014. doi: 10.1103/PhysRevA.89.062321. URL https://link.aps.org/doi/10.1103/PhysRevA.89.062321.

[74] S. Asaad, C. Dickel, S. Poletto, A. Bruno, N. K. Langford, M. A. Rol, D. Deurloo, and L. DiCarlo. Independent, extensible control of same-frequency superconducting qubits by selective broadcasting. *npj Quantum Inf.*, 2:16029, 2016. URL https://www.nature.com/articles/npjqi201629.

[75] C. C. Bultink, T. E. O'Brien, R. Vollmer, N. Muthusubramanian, M. W. Beekman, M. A. Rol, X. Fu, B. Tarasinski, V. Ostroukh, B. Varbanov, A. Bruno, and L. DiCarlo. Protecting quantum entanglement from leakage and qubit errors via repetitive parity measurements. *Science Advances*, 6 (12), 2020. doi: 10.1126/sciadv.aay3050. URL https://advances.sciencemag.org/content/6/12/eaay3050.

[76] M. Kjaergaard, M. E. Schwartz, A. Greene, G. O. Samach, M. O. A. Bengtsson, C. M. McNally, J. Braumüller, D. K. Kim, P. Krantz, M. Marvian, A. Melville, B. M. Niedzielski, Y. Sung, R. Winik, J. Yoder, D. Rosenberg, S. L. K. Obenland, . T. P. Orlando, I. Marvian, S. Gustavsson, and W. D. Oliver. A quantum instruction set implemented on a superconducting quantum processor. *ArXiv:2001.08838*, 2020. URL https://arxiv.org/pdf/2001.08838.

[77] B. R. Johnson. *Controlling Photons in Superconducting Electrical Circuits*. PhD Dissertation, Yale University, 2011. URL https://rsl.yale.edu/sites/default/files/files/RSL_Theses/johnson-thesis.pdf.

[78] B. Nijholt, J. Weston, J. Hoofwijk, and A. Akhmerov. *Adaptive*: parallel active learning of mathematical functions, 2019. URL https://doi.org/10.5281/zenodo.1182437.

[79] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3): 90–95, 2007. doi: 10.1109/MCSE.2007.55.

[80] L. Campagnola. Pyqtgraph: Scientific graphics and gui library for python, 2012. URL http://www.pyqtgraph.org.

[81] M. A. Rol, C. Attryde, J. van Oven, and V. Negîrneac. Quantify. URL https://pypi.org/project/quantify-core/.

[82] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver. Superconducting qubits: Current state of play. *Annual Review of Condensed Matter Physics*, 11(1):369–395, 2020. doi: 10.1146/annurev-conmatphys-031119-050605. URL https://doi.org/10.1146/annurev-conmatphys-031119-050605.

[83] C. K. Andersen, A. Remm, S. Lazar, S. Krinner, J. Heinsoo, J.-C. Besse, M. Gabureac, A. Wallraff, and C. Eichler. Entanglement stabilization using ancilla-based parity detection and real-time feedback in superconducting circuits. *npj Quantum Information*, 5(1):1–7, 2019. URL https://doi.org/10.1038/s41534-019-0185-4.

[84] F. W. Strauch, P. R. Johnson, A. J. Dragt, C. J. Lobb, J. R. Anderson, and F. C. Wellstood. Quantum logic gates for coupled superconducting phase qubits. *Phys. Rev. Lett.*, 91:167005, 2003. URL https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.91.167005.

[85] S. Sheldon, E. Magesan, J. M. Chow, and J. M. Gambetta. Procedure for systematically tuning up cross-talk in the cross-resonance gate. *Physical Review A*, 93:060302, Jun 2016.

doi: 10.1103/PhysRevA.93.060302. URL https://journals.aps.org/pra/abstract/10.1103/PhysRevA.93.060302.

[86] S. S. Hong, A. T. Papageorge, P. Sivarajah, G. Crossman, N. Didier, A. M. Polloreno, E. A. Sete, S. W. Turkowski, M. P. da Silva, and B. R. Johnson. Demonstration of a parametrically activated entangling gate protected from flux noise. *Phys. Rev. A*, 101:012302, Jan 2020. doi: 10.1103/PhysRevA.101.012302. URL https://link.aps.org/doi/10.1103/PhysRevA.101.012302.

[87] R. Barends, C. M. Quintana, A. G. Petukhov, Y. Chen, D. Kafri, K. Kechedzhi, R. Collins, O. Naaman, S. Boixo, F. Arute, K. Arya, D. Buell, B. Burkett, Z. Chen, B. Chiaro, A. Dunsworth, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, T. Huang, E. Jeffrey, J. Kelly, P. V. Klimov, F. Kostritsa, D. Landhuis, E. Lucero, M. McEwen, A. Megrant, X. Mi, J. Mutus, M. Neeley, C. Neill, E. Ostby, P. Roushan, D. Sank, K. J. Satzinger, A. Vainsencher, T. White, J. Yao, P. Yeh, A. Zalcman, H. Neven, V. N. Smelyanskiy, and J. M. Martinis. Diabatic gates for frequency-tunable superconducting qubits. *Phys. Rev. Lett.*, 123:210501, Nov 2019. doi: 10.1103/PhysRevLett.123.210501. URL https://link.aps.org/doi/10.1103/PhysRevLett.123.210501.

[88] P. Aliferis and B. M. Terhal. Fault-tolerant quantum computation for local leakage faults. *Quantum Info. Comput.*, 7(1):139–156, Jan. 2007. ISSN 1533-7146. URL http://dl.acm.org/citation.cfm?id=2011706.2011715.

[89] J. Ghosh, A. G. Fowler, J. M. Martinis, and M. R. Geller. Understanding the effects of leakage in superconducting quantum-error-detection circuits. *Phys. Rev. A*, 88:062329, Dec 2013. doi: 10.1103/PhysRevA.88.062329. URL https://link.aps.org/doi/10.1103/PhysRevA.88.062329.

[90] A. G. Fowler. Coping with qubit leakage in topological codes. *Phys. Rev. A*, 88:042308, 2013. URL https://link.aps.org/doi/10.1103/PhysRevA.88.042308.

[91] M. Suchara, A. W. Cross, and J. M. Gambetta. Leakage suppression in the toric code. *Quantum Info. Comput.*, 15(11-12):997–1016, Sept. 2015. ISSN 1533-7146. URL http://dl.acm.org/citation.cfm?id=2871350.2871358.

[92] J. Ghosh and A. G. Fowler. Leakage-resilient approach to fault-tolerant quantum computing with superconducting elements. *Phys. Rev. A*, 91:020302, Feb 2015. doi: 10.1103/PhysRevA.91.020302. URL https://link.aps.org/doi/10.1103/PhysRevA.91.020302.

[93] S. Krinner, S. Lazar, A. Remm, C. Andersen, N. Lacroix, G. Norris, C. Hellings, M. Gabureac, C. Eichler, and A. Wallraff. Benchmarking coherent errors in controlled-phase gates due to spectator qubits. *arXiv preprint arXiv:2005.05914*, 2020. URL https://arxiv.org/abs/2005.05914.

[94] Y. Chen, C. Neill, P. Roushan, N. Leung, M. Fang, R. Barends, J. Kelly, B. Campbell, Z. Chen, B. Chiaro, A. Dunsworth, E. Jeffrey, A. Megrant, J. Y. Mutus, P. J. J. O'Malley, C. M. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, M. R. Geller, A. N. Cleland, and J. M. Martinis. Qubit architecture with high coherence and fast tunable coupling. *Phys. Rev. Lett.*, 113:220502,

Nov 2014. doi: 10.1103/PhysRevLett.113.220502. URL https://link.aps.org/doi/10.1103/PhysRevLett.113.220502.

[95] F. Yan, P. Krantz, Y. Sung, M. Kjaergaard, D. L. Campbell, T. P. Orlando, S. Gustavsson, and W. D. Oliver. Tunable coupling scheme for implementing high-fidelity two-qubit gates. *Physical Review Applied*, 10(5):054062, 2018. URL https://link.aps.org/doi/10.1103/PhysRevApplied.10.054062.

[96] P. Mundada, G. Zhang, T. Hazard, and A. Houck. Suppression of qubit crosstalk in a tunable coupling superconducting circuit. *Physical Review Applied*, 12(5):054023, 2019. URL https://link.aps.org/doi/10.1103/PhysRevApplied.12.054023.

[97] M. C. Collodo, J. Herrmann, N. Lacroix, C. K. Andersen, A. Remm, S. Lazar, J.-C. Besse, T. Walter, A. Wallraff, and C. Eichler. Implementation of conditional-phase gates based on tunable zz-interactions. *arXiv preprint arXiv:2005.08863*, 2020. URL https://arxiv.org/abs/2005.08863.

[98] Y. Xu, J. Chu, J. Yuan, J. Qiu, Y. Zhou, L. Zhang, X. Tan, Y. Yu, S. Liu, J. Li, F. Yan, and D. Yu. High-fidelity, high-scalability two-qubit gate scheme for superconducting qubits. *arXiv preprint arXiv:2006.11860*, 2020. URL https://arxiv.org/pdf/2006.11860.

[99] J. A. Schreier, A. A. Houck, J. Koch, D. I. Schuster, B. R. Johnson, J. M. Chow, J. M. Gambetta, J. Majer, L. Frunzio, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Suppressing charge noise decoherence in superconducting charge qubits. *Phys. Rev. B*, 77:180502, 2008. URL https://link.aps.org/doi/10.1103/PhysRevB.77.180502.

[100] E. Magesan, J. M. Gambetta, and J. Emerson. Characterizing quantum gates via randomized benchmarking. *Phys. Rev. A*, 85:042311, 2012.

[101] N. Lacroix, C. Hellings, C. Kraglund Andersen, A. Di Paolo, A. Remm, S. Lazar, S. Krinner, G. J. Norris, M. G. Gabureac, A. B. Blais, C. Eichler, and A. Wallraff. Improving the performance of deep quantum optimization algorithms with continuous gate sets. *arXiv preprint arXiv:2005.05275*, 2020. URL https://arxiv.org/abs/2005.05275.

[102] R. Barends, L. Lamata, J. Kelly, L. García-Álvarez, A. Fowler, A. Megrant, E. Jeffrey, T. White, D. Sank, J. Mutus, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, I.-C. Hoi, C. Neill, P. O'Malley, C. Quintana, P. Roushan, A. Vainsencher, J. Wenner, E. Solano, and J. Martinis. Digital quantum simulation of fermionic models with a superconducting circuit. *Nat. Commun.*, 6: 7654, 2015. URL http://www.nature.com/articles/ncomms8654.

[103] J. Heinsoo, C. K. Andersen, A. Remm, S. Krinner, T. Walter, Y. Salathé, S. Gasparinetti, J.-C. Besse, A. Potočnik, A. Wallraff, and C. Eichler. Rapid high-fidelity multiplexed readout of superconducting qubits. *Phys. Rev. Appl.*, 10:034040, Sep 2018. doi: 10.1103/PhysRevApplied.10.034040. URL https://link.aps.org/doi/10.1103/PhysRevApplied.10.034040.

[104] C. C. Bultink, B. Tarasinski, N. Haandbaek, S. Poletto, N. Haider, D. J. Michalak, A. Bruno, and L. DiCarlo. General method for extracting the quantum efficiency of dispersive qubit readout in circuit qed. *Appl. Phys. Lett.*, 112:092601, 2018. URL https://doi.org/10.1063/1.5015954.

Appendix A

Supplementary material

A.1 Single-qubit gates performance (four-transmon patch)

All single-qubit gates are implemented as DRAG-type [69, 70] microwave pulses with a total duration of $4\sigma = 20$ ns, where σ is the Gaussian width of the main-quadrature Gaussian pulse envelope. Two sets of experiments are performed to jointly quantify the infidelity ε and leakage L_1 of these gates. First, individual performances are evaluated with single-qubit randomized benchmarking (1QRB) keeping the other three qubits in $|0\rangle$. Second, the simultaneous single-qubit randomized benchmarking (S1QRB) on pairs of qubits is evaluated, keeping the other two qubits in $|0\rangle$. The results obtained from both types of experiment are reported as diagonal and off-diagonal elements in the matrices presented in Fig. A.1.

A.2 Technical details on 2QIRB

Table A.1 details technical aspects of the characterization of CZ gates by repeated 2QIRB runs.

Parameter	$\mathrm{Q_{M1}} - \mathrm{Q_H}$	$\mathrm{Q_{M2}} - \mathrm{Q_H}$	$\mathrm{Q_L} - \mathrm{Q_{M1}}$	$\mathrm{Q_L} - \mathrm{Q_{M2}}$
Number of 2QIRB runs	39	10	88	35
Number of randomization seeds	100	300	100	100
Same randomization seeds	No	No	Yes	No
Avg. time per 2QIRB run (min)	17	50	9	17
Total wall-clock time (h)	28.8	16.9	16.7	14.8

Table A.1: Technical details of the characterization of CZ gates by repeated 2QIRB. The average time per 2QIRB run is the time required to perform back-to-back measurements of the reference and the CZ-interleaved curves. The total wall-clock time includes the overhead from compilation of RB sequences and other measurements performed in between the CZ 2QIRB runs, e.g., idling 2QIRB (Fig. A.3).

A.3 Residual ZZ coupling at bias point

Coupling between nearest-neighbor transmons in the device used for SNZ realization (Chapter 5) is achieved using dedicated coupling bus resonators. The non-tuneability of said couplers leads to residual

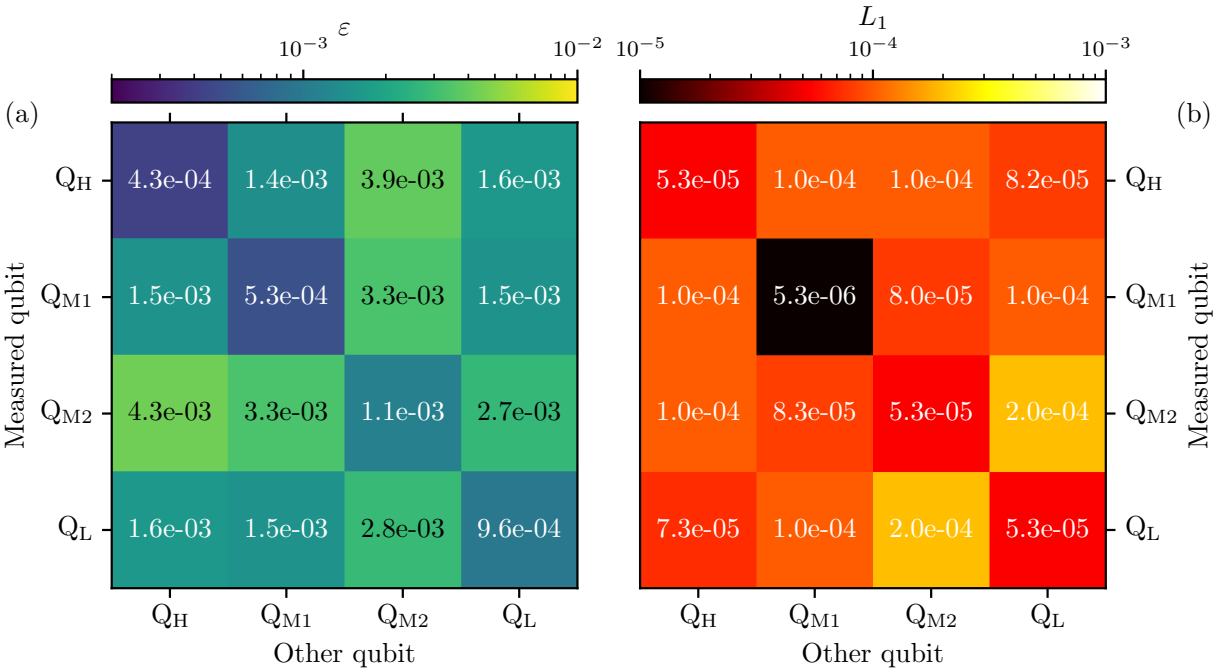


Figure A.1: Characterization of single-qubit gate infidelity ε (a) and leakage L_1 (b) using randomized benchmarking (100 randomization seeds). Diagonal elements are extracted from individual single-qubit randomized benchmarking keeping the other 3 qubits in $|0\rangle$. Off-diagonal elements are extracted from simultaneous one-qubit randomized benchmarking on pairs of qubits, keeping the other two qubits in $|0\rangle$. Credits: H. Ali, DiCarlo Lab, QuTech.

ZZ coupling between the transmons at the bias point. We quantify the residual ZZ coupling between every pair of qubits as the shift in frequency of one qubit when the state of the other changes from $|0\rangle$ to $|1\rangle$. We extract this frequency shift using a simple time-domain measurement [42]: we perform a standard echo experiment on one qubit (the echo qubit), but add a π pulse on the other qubit (control qubit) halfway through the free-evolution period simultaneous with the refocusing π pulse on the echo qubit. The results are presented as a matrix in Fig. A.2. We observe that the residual ZZ coupling is highest between $\mathrm{Q_H}$ and the mid-frequency qubits $\mathrm{Q_{M1}}$ and $\mathrm{Q_{M2}}$. This is consistent with the higher (lower) absolute detuning between $\mathrm{Q_H}$ ($\mathrm{Q_L}$) and the mid-frequency transmons, and the higher (lower) transverse coupling $J_2 = \pi/t_{\mathrm{lim}}$ for the upper (lower) pairs.

An alternative way to evidence this residual ZZ coupling is to extract the fidelity of idling using 2QIRB and to compare this fidelity to that of CZ. To this end, we perform 2QIRB of idling (for 60 ns) on pairs $\mathrm{Q_{M2}}$-$\mathrm{Q_H}$ and $\mathrm{Q_L}$-$\mathrm{Q_{M2}}$. The results (Fig. A.3), show striking differences for the two pairs. For $\mathrm{Q_{M2}}$-$\mathrm{Q_H}$, the pair with strongest residual coupling, the idling fidelity is significantly lower than the CZ fidelity. This is because the residual ZZ coupling is a source of error during idling but is absorbed into the tuneup of SNZ. For $\mathrm{Q_L}$-$\mathrm{Q_{M2}}$, for which the residual coupling is one order of magnitude lower, this trend is not observed.

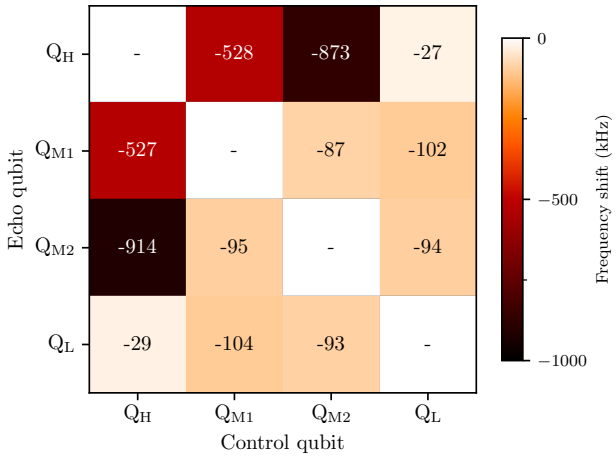


Figure A.2: Extracted residual ZZ coupling between all pairs of qubits at their bias points. We report the frequency shift in one qubit (named echo qubit) when the computational state of another qubit (named control qubit) is shifted from $|0\rangle$ to $|1\rangle$. Credits: H. Ali, DiCarlo Lab, QuTech.

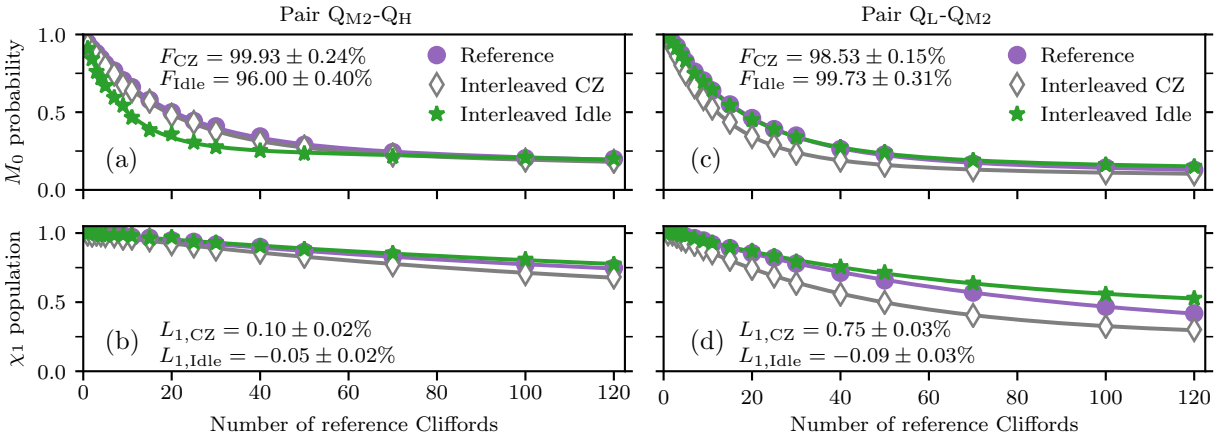


Figure A.3: Comparison by 2QIRB of the fidelity and leakage of SNZ CZ versus idling (for an equivalent 60 ns) for pairs Q_{M2}-Q_H and Q_L-Q_{M2}. SNZ gate parameters are provided in Table 5.2 of the main text. (a,c) Return probability to $|00\rangle$ as a function of the number N_{2Q} of two-qubit Clifford operations in the reference curve. For Q_{M2}-Q_H, the extracted idling fidelity is significantly lower than the SNZ CZ fidelity. This is due to the high residual ZZ coupling between these two qubits as reported in Fig. A.2, which is not refocused during idling but absorbed into the tuneup of the SNZ CZ gate. For Q_L-Q_{M2}, idling fidelity exceeds SNZ CZ fidelity as the residual coupling is one order of magnitude weaker. (b,d) Population in the computational subspace as a function of N_{2Q}. Leakage as a function of N_{2Q} is weakest when interleaving idling steps, leading to negative $L_{1,Idle}$. This is due to seepage (during idling) of the leakage produced by the reference two-qubit Cliffords. Credits: H. Ali, DiCarlo Lab, QuTech.