



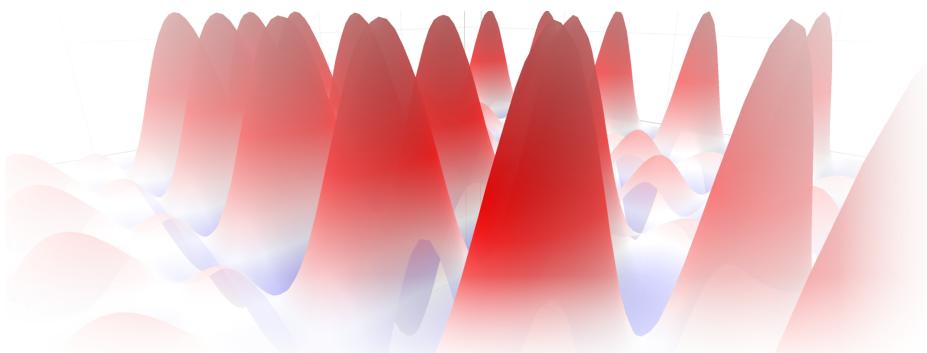
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Variational Quantum Algorithms with Controlled Arbitrary Phase Gates

Nathan Lacroix

Master Thesis

April 17, 2020



Advisors: Prof. Dr. Andreas Wallraff, Dr. Christian K. Andersen
Department of Solid States Physics, ETH Zürich

Abstract

Large-scale, fault-tolerant quantum computing has the potential to impact numerous fields such as medical research, material science and information security. However, near-term quantum computers will only have a limited number of quantum bits and a limited quantum circuit size that can be executed reliably.

Variational quantum algorithms (VQAs) mitigate the consequences of these constraints by outsourcing part of the computation to classical computers and seeking approximate – instead of exact – solutions. However, the gate sequence length that can be executed still remains limited by noise, in particular decoherence. Hence, most implementations so far are restricted to problem instances that can be solved with low depth circuits. Nevertheless, real-world applications involving a higher number of qubits will most likely require deeper circuits to approximate solutions accurately.

In this thesis, we present and implement a two-qubit gate that takes advantage of the structure in the quantum approximative optimization algorithm (QAOA) to reduce the sequence length of the algorithm. Namely, we extend the standard controlled phase gate (CZ gate) to a controlled arbitrary angle phase gate (C-ARB gate) by exploiting off-resonant interaction mechanisms and careful calibration of gate length, interaction strength and dynamic phases. Our implementation of the C-ARB gate achieves an average process fidelity of 97.7%, which we measure with process tomography.

In addition, we demonstrate the advantage of C-ARB gates on a 3-qubit QAOA implementation solving an exact cover problem instance. We reduce the two-qubit gate-count by 50% and the sequence length by a factor of 3 compared to a decomposed implementation with standard CZ gates. Consequently, we achieve a higher success probability with the direct implementation (0.84 versus 0.64) and foresee an increasing advantage for larger-scale problems because the number of layers required for a good approximate solution typically scales with the number of qubits in the problem.

Contents

Abstract	i
Contents	iii
1 Introduction	1
1.1 Quantum computing	2
1.2 Quantum computers with superconducting circuits	3
1.3 The noisy intermediate-scale quantum era	6
1.4 Variational quantum algorithms	7
1.5 Related work	9
2 Calibrating and Characterizing Controlled Arbitrary Phase Gates	11
2.1 Theoretical description	11
2.2 Calibration	15
2.3 Characterization	18
2.3.1 Quantum process tomography	19
2.3.2 Phases errors and leakage	21
2.3.3 Comparison between C-ARB gates and CZ gates . . .	22
2.4 Conclusion	24
3 Solving an exact cover problem instance using the QAOA	27
3.1 Combinatorial optimization	27
3.2 The quantum approximate optimization algorithm	29
3.2.1 Success probability	31
3.2.2 Layer implementation	31
3.2.3 Relation to adiabatic quantum computing	32
3.2.4 Quantum speedup and the importance of depth . . .	34
3.3 Exact cover	34
3.3.1 Definition	35
3.3.2 Example problem instance	36

CONTENTS

3.4	Experiment	37
3.4.1	Cost function landscapes	37
3.4.2	Variational parameter optimization	39
3.4.3	Success probability as function of depth	41
3.5	Conclusion	45
4	Conclusion	47
5	Outlook	49
A	Setup	53
B	High fidelity qutrit single-shot readout and reset	57
B.1	Three-level readout	58
B.2	Readout frequency optimization	59
B.2.1	Analytical derivation	60
B.2.2	Comparison to experimental data	63
B.3	Mode-matched integration weights	64
B.4	Characterization of three-level single-shot readout	64
B.4.1	Single-shot readout characterization	65
B.4.2	Using the three-level readout	68
B.5	Three-level, measurement-based active reset	70
B.5.1	Setup and procedure for low-latency analysis of quantum states	70
B.5.2	Demonstration of three-level active reset	72
	Acknowledgments	77
	Bibliography	79

Acronyms

AWG arbitrary waveform generator. 14, 15, 53, 54

C-ARB gate controlled arbitrary angle phase gate. i, 7, 10, 11, 13, 15–25, 27, 30–32, 39, 41–43, 45, 47–50

CZ gate controlled phase gate. i, 9–11, 13–15, 18, 19, 22–25, 30, 32, 42, 45, 47, 49, 50, 55

dec. impl. decomposed implementation of two qubit interaction terms. 32, 34, 37–39, 42, 44, 45

dir. impl. direct implementation of two qubit interaction terms. 31, 34, 37–39, 42, 44, 45

FIR finite impulse response. 22

GMM Gaussian mixture model. 59, 65, 66, 68, 71

HDAWG high definition arbitrary waveform generator. 53, 70–73

NISQ noisy intermediate-scale quantum. 6, 7, 34

QAOA quantum approximative optimization algorithm. i, 7–10, 25, 27, 29, 30, 32–35, 37, 39–42, 45, 47, 50, 54, 55

QPT quantum process tomography. 19–21

SNR signal to noise ratio. 59, 60, 62, 64, 67, 68, 71

UHFQA ultra-high frequency quantum analyzer. 53, 54, 64, 65, 70, 71, 76

ACRONYMS

VQA variational quantum algorithm. i, 6–10, 25, 29, 47–49

VQE variational quantum eigensolver. 7, 8

Chapter 1

Introduction

Despite the incessant progress over the last seventy years, there are still many problems today's computers cannot solve in a realistic amount of time. While some calmly await the next generation of supercomputers, others might remain intractable for classical computers forever.

Theoretical physicist Richard Feynman popularized this observation in the early eighties. In particular, he argued that classical computers could not simulate quantum mechanics efficiently [1]. Based on the pioneering work of Ed Fredkin and Tomasso Toffoli [2], he proposed an alternative computation model exploiting fundamental properties of quantum mechanics.

The following decade saw a plethora of developments in quantum algorithms and quantum information theory [3–8]. Consequently, potential applications emerged in several other fields such as information security, machine learning and optimization. In this thesis, we focus on optimization and implement an algorithm that finds an approximate solution to a combinatorial problem.

This introduction, which consists of five sections, provides the necessary material to understand the subsequent chapters. The first section introduces fundamental concepts of quantum computing and quantum information. The second provides an overview of how to realize a quantum computing experiment with superconducting circuits, the technology used in our experiments. The third section describes the constraints of near-term quantum computing resulting from imperfect implementation and suggests ways to mitigate their effect: (i) using variational quantum algorithms (VQAs) and (ii) using a more expressive gate set tailored to VQAs, thereby avoiding decomposition in their quantum circuits implementations. The fourth section further details the principles of VQAs and illustrates why the controlled arbitrary phase gate, which we implement in this thesis, is a suitable gate to avoid decomposition. Finally, the last section reports on related work and

introduces the scientific contributions of this work.

1.1 Quantum computing

Both classical and quantum computing aspire to the same goal: solving problems with algorithms. However, they differ fundamentally in the way they represent information. Classical computers encode information in binary variables, called (classical) bits. At any point in time, a bit has a state of either 0 or 1. Quantum computers rely on a generalization of this concept: the quantum bit, referred to as *qubit*. A qubit has a *quantum state*, $|\psi\rangle$, which, like a classical bit, is characterized by two basis states, $|0\rangle$ and $|1\rangle$ ¹. However, unlike its classical counterpart, during a computation, a qubit can adopt any linear combination of these two basis states:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad \{\alpha, \beta\} \in \mathbb{C} \quad (1.1)$$

where α, β are the (complex) amplitudes of the basis states $|0\rangle$ and $|1\rangle$, respectively. We say that the qubit can be in a *superposition* of the two basis states.

Unfortunately, quantum mechanics postulates it is impossible to examine (i.e. measure) the amplitudes of a quantum state directly. Instead, a single measurement projects the states onto only *one* of the basis states with a probability equal to the squared modulus of its amplitude. In the one-qubit case, $|\psi\rangle$ yields $|0\rangle$ with probability $|\alpha|^2$ and $|1\rangle$ with probability $|\beta|^2$ and $|\alpha|^2 + |\beta|^2 = 1$. Nevertheless, it is the leveraging of interference between these amplitudes within the quantum algorithm (i.e. *before* measuring) that makes quantum computing particularly powerful.

Several qubits put together form a quantum state with even more basis states, each of which has its own amplitude. For instance, a two-qubit system encodes the quantum state $|\psi'\rangle = \alpha|00\rangle + \beta|01\rangle + \gamma|10\rangle + \delta|11\rangle$, with 4 basis states and 4 amplitudes. Generally speaking, a N -qubit system has 2^N basis states and can be in a superposition of all of them.

In close analogy to how logical gates (e.g. NOT, AND, OR) act on bits to perform classical computations, quantum algorithms consist of individual operations implemented via quantum gates, acting on single or multiple qubits. These quantum gates take a quantum state as input, shuffle the amplitudes and outputs a modified quantum state. For instance, a single-qubit gate can take $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and exchange the amplitudes α and β such that the output is $|\phi\rangle = \beta|0\rangle + \alpha|1\rangle$. This is analogous to what a NOT gate would do to a classical bit.

¹ $|\cdot\rangle$ is the Dirac notation for state vectors, often used in quantum mechanics.

1.2. Quantum computers with superconducting circuits

There is a convenient mathematical way to visualize how a quantum gate acts on a quantum state. Namely, using vectors and matrices. In the single qubit case, a state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ is encoded into the vector $(\alpha, \beta)^\top$ while the gate G is represented as a unitary² 2×2 matrix,

$$G = \begin{pmatrix} a_0 & a_1 \\ b_0 & b_1 \end{pmatrix} \quad (1.2)$$

in which the coefficients in the first and second column indicate how the gate affects the basis states $|0\rangle$ and $|1\rangle$ respectively. In particular, if the input state is $|0\rangle$, which can be written as $(1, 0)^\top$, then the output state is $a_0|0\rangle + b_0|1\rangle$ and if the input state is $|1\rangle$ then the output state is $a_1|0\rangle + b_1|1\rangle$. In the more general case, if the input state is $|\psi\rangle = (\alpha, \beta)^\top$, then the output state $|\phi\rangle$ is given by the product

$$|\phi\rangle = G|\psi\rangle = \begin{pmatrix} a_0 & a_1 \\ b_0 & b_1 \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \quad (1.3)$$

For the single-qubit gate exchanging α and β , $G = ((0, 1)^\top, (1, 0)^\top)$ such that

$$|\phi\rangle = G|\psi\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \quad (1.4)$$

A similar technique is used for two-qubit gates. They are described by 4×4 matrices multiplying vectors with 4 entries (one for each basis state).

Remarkably, a very small set of single- and two-qubit gates, called *universal gate set*, is sufficient to implement any quantum computation with arbitrarily many qubits because any high level instruction can be decomposed to a list of gates belonging to that universal gate set [9]³.

The existence of universal gate sets is very valuable for building a quantum computer because it implies we can perform quantum computation with few distinct gates implemented in hardware. Nevertheless, building a quantum computer remains an immense challenge. Indeed, it requires to create and manipulate objects displaying quantum mechanical properties, which typically do not materialize at the scale we live in. One option, as we shall see in the next section, consists in mimicking atoms and their quantum mechanical properties with superconducting circuits.

1.2 Quantum computers with superconducting circuits

Since the advent of quantum information theory, many technologies have been explored to build quantum computers: ions trapped in an electro-

²A complex square matrix U is unitary if its conjugate transpose U^\dagger is also its inverse, i.e. if $U^\dagger U = U U^\dagger = I$.

³A similar concept also exists in classical computing, stating that NAND gates alone form a universal set.

1. INTRODUCTION

magnetic trap [10], semiconductor quantum dots [11], photons and linear optics [12], superconducting circuits [13], and many more. While each technology has its pros and cons, in this thesis we focus on superconducting circuits, one of the leading technology platform [14] which we use in our experiments.

In superconducting circuits, qubits are represented by artificial atoms, realized as superconducting electrical circuits with discrete energy levels. Typically, the first two levels are chosen to represent the basis states $|0\rangle$ and $|1\rangle$ while the remaining ones are ignored. The core idea is to create a quantum harmonic oscillator using an LC -circuit (L is an inductor and C is a capacitor). This circuit has a resonance frequency $\omega = 1/\sqrt{LC}$ which corresponds to the frequency of the transition between the energy levels in the system. In a quantum mechanical picture, the resulting harmonic potential has quantized energy levels spaced by the energy quantum $\hbar\omega$ [14], where \hbar is the reduced Planck constant.

Provided the temperature is sufficiently low and there is little dissipation, the energy levels of the oscillator are distinguishable and addressable. However, the equidistant spacing between energy levels prevents from using directly the quantum harmonic oscillator as a qubit. Indeed, while trying to induce a transition from $|0\rangle$ to $|1\rangle$ by injecting an energy quantum, we might induce a transition from $|1\rangle$ to $|2\rangle$ if there is already an excitation in the system. Hence, we cannot restrict the computational space to the first two levels as desired for binary quantum computing.

To circumvent this problem, we create an anharmonic energy potential by using a non-linear inductor called a Josephson tunnel junction [15]. The junction's inductance depends non-linearly on the magnetic flux flowing through it. The introduced anharmonicity enables the separate addressing of each transition with the matching transition frequency.

There are several ways to obtain a Hamiltonian with discretized, anharmonic energy levels using a Josephson junction. One of them consist in creating a superconducting island with one side connected to ground via a Josephson junction, and the other coupled via a capacitor to a voltage source [16, 17].

When the voltage source is unbiased, the system lies in the ground state, $|0\rangle$, also referred to as $|g\rangle$. By increasing the voltage of the source, we polarize the island with the capacitor until a charge quantum⁴ tunnels (without loss) through the Josephson junction onto the island to restore charge balance. The energy stored in the capacitor after the tunneling is called the charging

⁴These charge quanta are called Cooper-pairs and form at very low temperature when two electrons of opposite spin in the metal pair up to form a bosonic state. At sufficiently low temperatures, all valence electrons in the metal form Cooper-pairs and they can all occupy the same ground state.

1.2. Quantum computers with superconducting circuits

energy, E_C , and the energy stored in the junction is called the Josephson energy, E_J . If the gate voltage is subsequently removed, there is *charge offset* on the island and the system is out of equilibrium, i.e., there is one excitation in the system ($|1\rangle$ or $|e\rangle$). Because the state of the system depends on the amount of charge on the island, this implementation is called the *charge qubit*.

In our experiments, we use the *transmon* [18] – a special case of the charge qubit – that operates in the regime $E_J/E_C \approx 50$ to mitigate environmental charge noise [18, 19]. In addition, we use a SQUID to control the Josephson energy with an external magnetic flux, Φ . The total energy required to introduce excitations in the system depends on the energies $E_J(\Phi)$ and E_C . For the first transition energy [18],

$$E_{|1\rangle}/\hbar \approx \omega_{|1\rangle} \approx \sqrt{8E_J(\Phi)E_C - E_C} \quad (1.5)$$

Transition frequencies for frequency tunable transmons are typically located in the micro-wave part of the electromagnetic spectrum (4-8 GHz). Therefore, we use microwave pulses tuned to the appropriate transition frequency to perform the single-qubit gates between the state $|0\rangle$ and $|1\rangle$.

Due to the inherent environmental noise, implementing perfect quantum gates represents an immense challenge in practice. Other experimental considerations further complicate the implementation of quantum computing: for instance, *decoherence*, the process of loosing quantum information encoded in a system over time due to undesirable interactions with the environment⁵, puts a hard bound how many quantum operations can be executed reliably. In addition, information stored in qubits cannot be copied [20] which prevents the use of cloning and majority voting as error correcting scheme. Instead, quantum error correction codes employing many physical qubits to encode the state of few logical qubits [21–23] are required to achieve fault-tolerant quantum computing.

These experimental complications will impose limitations on what (noisy) quantum computers achieve in the near future. Nevertheless, the steady improvement of coherence times and multi-qubit chip control over the last two decades now enable quantum computations which challenge the most powerful classical supercomputers [24]. We are thus at the dawn of a new and exciting quantum computing stage that physicists have named the *noisy intermediate-scale quantum era*.

⁵We typically distinguish between *relaxation*, characterized by the time constant T_1 , which relates to the spontaneous decay of excitation captured in the system after some time, and *dephasing*, characterized by the time constant T_2^* , which relates to the loss of information about the phase of the system.

1.3 The noisy intermediate-scale quantum era

While Google recently claimed having performed a task on a 53-qubit quantum computer that would take extremely long time on a supercomputer [24], full-scale fault-tolerant quantum computing is still a distant dream. The pivotal period we just entered, called the noisy intermediate-scale quantum (NISQ) era [25], entails two major limitations on near-term quantum computers:

1. The performance of quantum devices is severely limited by noise. There are many possible sources of noise. However, a dominant one is decoherence. For fixed coherence times, quantum computers can only execute a limited number of operations (quantum gates) before quantum information is lost. We say that decoherence limits the *depth* of the quantum circuit.
2. The number of qubits available for computation is limited. Fabricating high quality and multi-qubit quantum processors is a laborious task. Therefore, we expect near-term quantum computers to have fewer than 1000 physical qubits.

Bearing in mind those limitations, how can we make best use of current quantum computers?

Firstly, we shall implement quantum algorithms, such as *variational quantum algorithms* (VQAs) [26], which are believed to cope better with NISQ limitations. VQAs are suited for the NISQ era because they outsource part of the algorithm which does not require quantum properties to classical computers. Therefore, available qubits are used more efficiently and part of the computation is carried out by classical computers which do not suffer from decoherence. In addition, VQAs are intrinsically less sensitive to noise because they seek approximate solutions (instead of exact solutions). Hence, coherent errors may be partially compensated for by the classical optimizer such that they lead to worse approximate solutions but typically do not ruin the whole computation. Nonetheless, errors accumulating over time (such as decoherence) will severely limit the performance of VQAs.

Secondly, we can significantly reduce the time required to execute VQAs on quantum computers by tailoring the available gate set on the quantum computer to match typical operations performed by VQAs. The conceptual underlying principle is presented in Fig. 1.1. When generating a quantum circuit to solve a mathematical problem, the first step consists of formulating the problem we want to solve in a way which the VQA can take it as input. Next, we receive a list of instructions from the algorithm to follow in order to solve the problem. These instructions are passed on to the quantum computer. However, the quantum computer can only perform a restricted set of operations based on the available gates implemented in hardware.

Therefore, it might be necessary to decompose some operations into a longer sequence of available gates.

In particular, decomposition occurs when a VQA is implemented on a quantum computer with a so-called standard gate set. This set comprises of the most commonly implemented two-qubit gates such as the CNOT-, CZ- and SWAP-gate. They receive most attention because a small subset of them, such as the CNOT-gate (or SWAP-gate) combined with arbitrary single-qubit gates, forms a universal gate set which allows to implement any instruction, as discussed in Section 1.1. However, the decomposition of these instructions can be arbitrarily long, which is a major drawback, especially during the NISQ era.

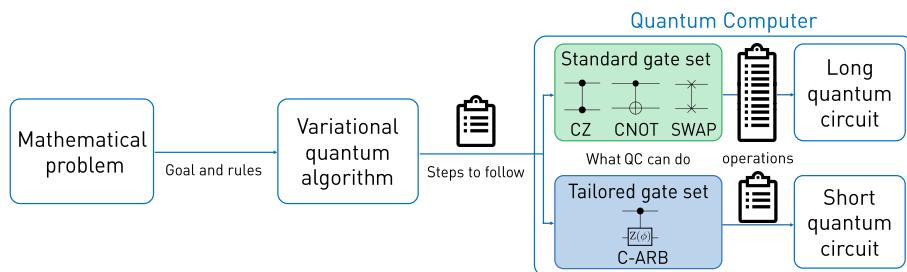


Figure 1.1. Conceptual process flow to generate quantum circuits.

In this thesis, we present and characterize the controlled arbitrary angle phase gate (C-ARB gate), an expressive two-qubit gate which avoids decomposition of instructions for various VQAs. In the next section, we detail the basic working principle of VQAs and explain why C-ARB gates reduces the depth of their quantum circuits implementations.

1.4 Variational quantum algorithms

Variational quantum algorithms (VQAs) encompass hybrid quantum-classical algorithms which leverage a quantum computer to prepare a state in a large state space that would be exponentially hard to prepare classically, while the rest of the computation is executed by a classical computer [26]. The two most prominent VQAs are the variational quantum eigensolver (VQE) [27], which finds approximate solutions to quantum chemistry problems and the quantum approximative optimization algorithm (QAOA) [28], which applies to generic combinatorial optimization problems. Their basic working principle is illustrated in Fig. 1.2.

For both algorithms, the original problem is mapped in polynomial time to a (qubit) cost Hamiltonian, \hat{C} , such that its solution coincides with finding

1. INTRODUCTION

the ground-state energy of that Hamiltonian. Thereafter, the search for the ground-state is performed using both a quantum and a classical computer.

The quantum computer generates a trial state $|\psi(\theta)\rangle$ based on variational parameters θ and the cost Hamiltonian. Repeated single shot measurements allow to estimate the expectation values for each term of the cost Hamiltonian. Based on the expectation value of the total energy, $\langle\psi(\theta)|\hat{C}|\psi(\theta)\rangle$, a classical optimizer suggests new variational parameters θ' to minimize the smooth function of the expectation value of the energy. At convergence, the final state $|\psi(\theta^*)\rangle$ with corresponding final variational parameters θ^* constitutes an approximate solution for the ground-state, from which an approximate solution to the original problem is deduced.

The quality of the approximation strongly depends on the depth of the quantum circuits. Indeed, VQEs and QAOA mimic a *continuous* time evolution of the system with *discrete* steps to find its ground-state [29] (see Section 3.2.3 for a detailed discussion about QAOA's relationship to the time evolution of a quantum system). Each additional layer – or step – in the quantum circuit improves the quality of the approximation but also increases the number of operations and hence the depth of the circuit.

In QAOA, each layer l directly includes the unitary evolution due to the cost Hamiltonian for a time defined by the variational parameter θ_l : $U_{C_l}(\theta_l) = e^{-i\hat{C}\theta_l}$. When \hat{C} is an Ising-like Hamiltonian in the z -basis [30], two-qubit terms in \hat{C} result in unitaries of the form $e^{-i\phi}$ where $\phi = \theta_l\sigma_i^z\sigma_j^z$. Since the parameter θ_l can adopt continuous real values, the operation includes the

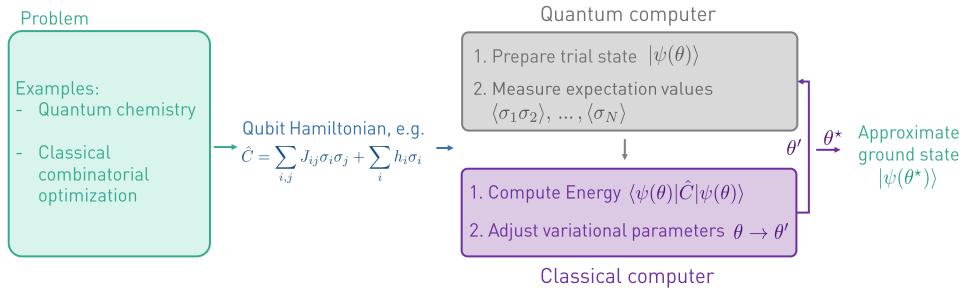


Figure 1.2. Basic principle of a VQA. A problem is mapped onto a qubit (cost) Hamiltonian \hat{C} , where each term consists of tensor-products of individual Pauli operators [9, p. 65]. In this case, we picture an Ising-like Hamiltonian. The quantum computer prepares a parametrized trial state, $|\psi(\theta)\rangle$ and measures the expectation values of each term in \hat{C} . Based on the expectation value of the cost Hamiltonian, $\langle\psi(\theta)|\hat{C}|\psi(\theta)\rangle$ a classical computer adjusts the circuit parameters, θ , to minimize $\langle\hat{C}\rangle$.

addition of an (arbitrary) phase ϕ on the two-qubit state $|11\rangle_{ij}$, which cannot be implemented with a standard controlled phase gate (CZ gate) which adds a π conditional phase. Therefore, early implementations of QAOA decomposed two-qubit terms into two two-qubit gates and additional single qubit gates (see Section 3.2 for more details).

1.5 Related work

The idea of using expressive gate sets for VQAs recently inspired research groups at IBM, Rigetti Computing and Google, see Table 1.1, entries 1-3. In 2019, Ganzhorn et al. implemented exchange-type gates tailored for quantum chemistry simulation [31] with a fidelity of $\sim 95\%$. In the same year, Abrams et al. implemented a more general XY-entangling gate [32] with median fidelity of 97.4%. This gate family directly applies to quantum chemistry simulations, and a combination of two gates of this family also reduces decomposition in quantum circuits for combinatorial optimization. Finally, Foxen et al. [33] implemented both an arbitrary iSWAP-type gate and an arbitrary conditional phase gate which they concatenate to achieve similarly general interactions as with the XY-entangling gate. They report a two-qubit Pauli error of 3.8×10^{-3} .

All three research groups implemented gates for superconducting qubits. Ganzhorn et al. and Foxen et al. implemented tunable couplers-based gates [34, 35] while Abrams et al. realized parametric entangling gates [36] with one fixed and one tunable transmon.

At the time of writing, Abrams et al. is the only group that implemented the QAOA using expressive gates [32]. In combination with standard CZ-gates, their parametrized XY-gates enable an impressive two-qubit gate count reduction of $\sim 30\%$, which they briefly illustrate for a one-layer QAOA implementation of a four-qubit, all-to-all connected MaxCut problem graph.

However, several groups implemented the QAOA with standard gates prior to Abrams et al.'s work, see Table 1.1, entries 4-6. In 2017, Otterbach et al. solved a clustering problem on 19 superconducting qubits [37]. The problem could be solved using a single layer of the QAOA such that the decomposition – which occurs in each layer – did not strongly affect the performance of the algorithm. In 2019, Pagano et al. estimated the ground state energy of a transverse field Ising model with tunable long-range interactions with 20 trapped ions and a two-layer QAOA [38]. Similarly, the best theoretical approximate solution with two layers is only $\sim 1.5\%$ away from the ground state with respect to the full energy scale. Bengtsson et al. were the first to implement a problem requiring a two-layer QAOA on (two) superconducting qubits [39].

1. INTRODUCTION

Table 1.1: Summary and comparison of publications on expressive gate sets (top three entries), QAOA experiments (entries 4 to 6) and this work. Entries are ordered chronologically based on their publication date.

Authors	Main message	qubits	Problem	layers*	gate implementation
Ganzhorn et al. [31]	Exchange-type gates + VQE demo	2	H ₂ -molecular energy	-	parametric, tunable coupler based
Abrams et al. [32]	XY-interaction gate family + QAOA demo	4	MaxCut	1 (?)	parametric
Foxen et al. [33]	Continuous SWAP and phase gates	2	-	-	tunable coupler based (gmon device)
Otterbach et al. [37]	First QAOA implementation	19	MaxCut	1 (1)	parametric CZ gates
Pagano et al. [38]	Analog implementation up to 40	Generic Ising	2 (1)	Analog long-range Ising	
Bengtsson et al. [39]	Solve a simple problem instance with high success probability using QAOA	2	Exact cover	2 (2)	parametric CZ gates with tunable coupler
This work	Demonstrate the advantage of using C-ARB gates for QAOA	3	Exact cover	9 (3)	fixed coupling CZ gates and C-ARB gates

* Implemented layers and between brackets the number of layers required to solve the problem with a success probability larger than 95%.

In this thesis, we present the first in-depth experimental analysis of the advantage of expressive gates, i.e. C-ARB gates, for the implementation of VQAs. In Chapter 2, we describe the calibration and characterization procedure of a C-ARB gate for fixed-coupling, frequency-tunable transmons [40] with an average process fidelity of $\sim 97.7\%$. In Chapter 3, we use C-ARB gates to obtain a 50% two-qubit gate count reduction on a combinatorial problem exploiting the connectivity of our three-qubit device. We also implement for the first time a problem requiring three QAOA layers to be solved with high success probability. In addition, we compare quantitatively on this problem the performance obtained with the direct implementation of the C-ARB gate, and its decomposed alternative. We show that, the QAOA utilizing the direct implementation solves the problem with greater success probability.

Chapter 2

Calibrating and Characterizing Controlled Arbitrary Phase Gates

This chapter details the concepts, calibration and characterization of controlled arbitrary angle phase gates (C-ARB gates). We start by explaining how this gate family can be seen as an extension of standard conditional phase gate (CZ gates). Next, we demonstrate the implementation of a C-ARB gate, allowing us to reach continuous conditional phase in the range $[0, 2\pi]$. We describe the calibration procedure of the gate on qubit 2 and 3 of our quantum processor presented in Appendix A. Next, perform quantum process tomography and evaluate the process fidelity as a function of conditional phase. By exploiting the 3-level readout discussed in Appendix B, we then characterize conditional- and dynamic phase errors and leakage. We compare these phase errors and leakage values to the ones obtained with a CZ gate implemented on the same qubits.

2.1 Theoretical description

In this section, we derive the unitary evolution of the C-ARB gate and explain the physical mechanisms behind its implementation. The goal is to obtain a unitary operator $U_{\text{C-ARB}}$ in a two-qubit subspace which adds a controllable phase ϕ to the $|11\rangle$ state,

$$U_{\text{C-ARB}} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & e^{-i\phi} \end{pmatrix} \quad (2.1)$$

We obtain such unitary by exploiting the same effect as for the CZ gate, namely the collection of geometric phase on the $|11\rangle$ state due to the interaction of the $|11\rangle$ level and the non-computational $|20\rangle$ level [40, 41]. As

2. CALIBRATING AND CHARACTERIZING CONTROLLED ARBITRARY PHASE GATES

pictured in Fig. 2.1(a), the $|11\rangle$ and $|20\rangle$ levels hybridize strongly when they are brought close to resonance. In this regime, we can approximate their interaction as a two-level system where the ground (excited) state is the $|11\rangle$ ($|20\rangle$) state. The system is characterized by the Hamiltonian

$$\hat{H}/\hbar = \omega_{|11\rangle} |11\rangle \langle 11| + \omega_{|20\rangle} |20\rangle \langle 20| + J(|11\rangle \langle 20| + |20\rangle \langle 11|) \quad (2.2)$$

where $\omega_{|11\rangle}$ ($\omega_{|20\rangle}$) is the frequency of the energy level $|11\rangle$ ($|20\rangle$), and J is the fixed coupling strength between the two levels which is a chip design parameter fixed during fabrication. The first and second term in the Hamiltonian correspond to the energy in the $|11\rangle$ and $|20\rangle$ level respectively. The third term represents the coupling energy between the two levels in which the excitation is transferred from one level to the other and vice-versa.

The same Hamiltonian can conveniently be written in matrix form with basis vectors $|11\rangle$ and $|20\rangle$,

$$\hat{H}/\hbar = \begin{pmatrix} \omega_{|11\rangle} & J \\ J & \omega_{|20\rangle} \end{pmatrix} \quad (2.3)$$

We shift the zero energy to $\omega_{|11\rangle}$ and define the frequency detuning between the two levels $\Delta = \omega_{|20\rangle} - \omega_{|11\rangle}$ such that \hat{H} becomes

$$\hat{H}/\hbar = \begin{pmatrix} 0 & J \\ J & \Delta \end{pmatrix} \quad (2.4)$$

From Schrödinger's equation, it follows that the time evolution of an arbitrary state $|\psi\rangle$ in the $|11\rangle$ - $|20\rangle$ subspace is given by $|\psi(t)\rangle = U(t)|\psi\rangle$ where $U(t) = e^{-i\hat{H}/\hbar t}$ is the unitary evolution of the Hamiltonian. Specifically, the unitary $U(t)$ is defined as,

$$U(t) = \begin{pmatrix} e^{-\frac{1}{2}it\Delta} \left(\cos\left(\frac{1}{2}t\tilde{J}\right) + \frac{i\Delta \sin\left(\frac{1}{2}t\tilde{J}\right)}{\tilde{J}} \right) & -\frac{2ie^{-\frac{1}{2}it\Delta} J \sin\left(\frac{1}{2}t\tilde{J}\right)}{\tilde{J}} \\ -\frac{2ie^{-\frac{1}{2}it\Delta} J \sin\left(\frac{1}{2}t\tilde{J}\right)}{\tilde{J}} & e^{-\frac{1}{2}it\Delta} \left(\cos\left(\frac{1}{2}t\tilde{J}\right) - \frac{i\Delta \sin\left(\frac{1}{2}t\tilde{J}\right)}{\tilde{J}} \right) \end{pmatrix} \quad (2.5)$$

where we have defined for clarity $\tilde{J} := \sqrt{4J^2 + \Delta^2}$ which we call the *effective exchange coupling*.

When starting with an excitation in the $|11\rangle$ state (the ground-state of this two-level subsystem), the $|11\rangle$ population oscillates coherently as the excitation is swapped back and forth between the $|11\rangle$ and the $|20\rangle$ state,

$$P_{|11\rangle}(t) = \text{Tr} \left(U(t) |11\rangle \langle 11| U^\dagger(t) \right) = \frac{\Delta^2 + 2J^2(\cos(t\sqrt{4J^2 + \Delta^2}) + 1)}{4J^2 + \Delta^2} \quad (2.6)$$

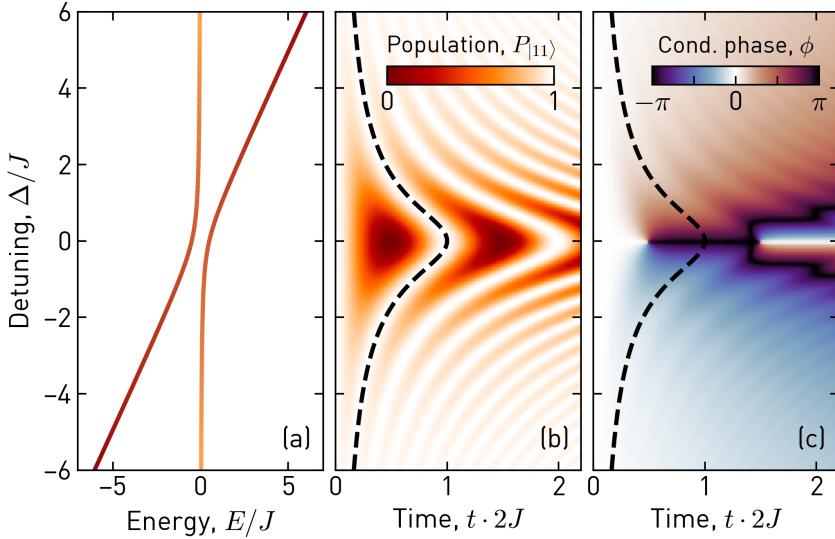


Figure 2.1. Simulation of the C-ARB gate as a generalization of the CZ gate. (a) Avoided crossing between the $|11\rangle$ (brown) and the $|20\rangle$ (orange) energy levels as function of detuning. (b) Population of the $|11\rangle$ state as function of frequency detuning Δ and interaction time t (in units of J) visualizing the coherent population exchange due to hybridization. The dashed line corresponds to the frequency detunings and gate lengths resulting in the first maximum in population recovery in the computational subspace. (c) Conditional phase acquired on the $|11\rangle$ state resulting from the interaction of the two levels. We reach all conditional phases between 0 and 2π by sweeping the detuning and adapting the time accordingly to ensure population recovery (dashed line).

with an oscillation period of $2\pi/\tilde{J}$.

The population as function of frequency detuning Δ and interaction time t results in a Chevron pattern shown in Fig. 2.1(b). For the operating point of the CZ gate, $\Delta = 0$, the population is given by $P_{|11\rangle} = \frac{1}{2} + \frac{1}{2} \cos 2Jt$. This corresponds to a complete population exchange between the $|11\rangle$ and the $|20\rangle$ state and with an oscillation period of π/J . By contrast, $\Delta \neq 0$ results in a partial population exchange between the $|11\rangle$ and the $|20\rangle$ state. The oscillation period is also reduced due to the detuning Δ in the denominator.

The gate length must be an integer multiple of the oscillation period to ensure a unitary operation within the computational subspace (i.e. full population recovery). In Fig. 2.1(b), we highlight (dashed line) the first oscillation period which corresponds to the shortest possible gate length as a function of the frequency detuning,

$$t_{\text{C-ARB}} = \frac{2\pi}{\sqrt{4J^2 + \Delta^2}} \quad (2.7)$$

In Fig. 2.1(c), we illustrate the acquired conditional phase, ϕ , as a function of detuning and gate length. For the gate lengths $t_{\text{C-ARB}}$, the corresponding phase on the $|11\rangle$ state acquired by the closed loop state evolution is

$$\phi = \text{Arg}\left(\langle 11 | U(t_{\text{C-ARB}}) | 11 \rangle\right) = \pi \cdot \left(1 + \frac{\Delta}{\sqrt{\Delta^2 + 4J^2}}\right) \quad (2.8)$$

which spans the interval $[0, 2\pi[$ for a sufficiently large detuning sweep, see dashed line in Fig. 2.1(c). As expected, a zero detuning yields a conditional phase of π and therefore corresponds to the standard CZ gate.

In practice, we achieve these timed interactions with flux pulses¹ which shift the transition frequency of the individual qubits. For a gate in the off-state, we ensure $|\Delta| \gg J$ such that the acquired conditional phase is small. Ideally, the acquired phase should be zero but in practice it is not due to the residual ZZ-coupling between the $|11\rangle$ and the $|20\rangle$ state. To turn on the gate, we apply a square flux pulse with amplitude a and length l to one of the qubits via its flux line. The pulse shifts the $|11\rangle$ state non-adiabatically close to the $|20\rangle$ state.

As mentioned in Section 1.2, the flux affects the Josephson energy [18, Eq. 2.18],

$$E_J(a) = E_{J,\max} \sqrt{\cos^2\left(\frac{\pi}{\Phi_0} \frac{\partial\Phi}{\partial V} a\right) + d^2 \sin^2\left(\frac{\pi}{\Phi_0} \frac{\partial\Phi}{\partial V} a\right)} \quad (2.9)$$

where d is the asymmetry between the junctions in the SQUID loop of the transmon, $\partial\Phi/\partial V$ is the flux sensitivity (voltage to flux conversion parameter), $\Phi_0 = h/2e$ is the superconducting flux quantum and $E_{J,\max}$ is the maximal junction energy.

In turn, the Josephson energy shifts the qubit transition frequency as defined by Eq. (1.5) such that

$$\omega_{|01\rangle}(a) \simeq \sqrt{8E_{J,\max} \sqrt{\cos^2\left(\frac{\pi}{\Phi_0} \frac{\partial\Phi}{\partial V} a\right) + d^2 \sin^2\left(\frac{\pi}{\Phi_0} \frac{\partial\Phi}{\partial V} a\right)}} E_C - E_C \quad (2.10)$$

where we approximate $|E_c|$ to be equal to the anharmonicity $|\alpha_2|$ of the fluxed transmon [18, Eq. 2.12]. The frequency detuning, Δ , as function of the flux pulse amplitude is then simply

$$\begin{aligned} \Delta(a) &= \omega_{|20\rangle} - \omega_{|11\rangle} \\ &= 2\omega_{|10\rangle} - \alpha_1 - (\omega_{|10\rangle} + \omega_{|01\rangle}(a)) \\ &= \omega_{|10\rangle} - \alpha_1 - \omega_{|01\rangle}(a) \end{aligned} \quad (2.11)$$

¹In fact, we apply a *voltage* pulse at the output of the arbitrary waveform generator (AWG) and the latter creates a current in the flux line which results in magnetic flux through the SQUID loop

where α_1 is the anharmonicity of the transmon which is not fluxed.

It follows that varying the amplitude of the flux pulse indeed controls the frequency detuning between the $|11\rangle$ and the $|20\rangle$ state, which in turn controls the collected phase on the $|11\rangle$ state. To ensure full population recovery in the computational subspace, the length of the flux pulse must be $t_{\text{C-ARB}}$ which can be related back to the amplitude of the flux pulse by combining Eq. (2.7), (2.10) and (2.11),

$$t_{\text{C-ARB}}(a) \simeq \frac{2\pi}{\sqrt{4J^2 + \left(\omega_{|10\rangle} - \alpha_1 - \sqrt{8E_{J,\max}\sqrt{\cos^2\left(\frac{\pi}{\Phi_0}\frac{\partial\Phi}{\partial V}a\right) + d^2\sin^2\left(\frac{\pi}{\Phi_0}\frac{\partial\Phi}{\partial V}a\right)}}\alpha_2 - \alpha_2\right)^2}} \quad (2.12)$$

In Section 2.2, we fit Eq. (2.12) to experimental data to estimate J , $\partial\Phi/\partial V$ and d and thereby also the expected conditional phase using Eq. (2.8).

In addition to the conditional phase, the flux pulse also results in a dynamic phase ϕ_D on the fluxed qubit, because it takes the qubit out of its rotating frame [40],

$$\phi_D = \int_0^l (\omega(t) - \omega_{\text{park}}) dt \quad (2.13)$$

where $\omega(t)$ and ω_{park} denote the instantaneous qubit frequency during the pulse and the frequency at parking position², respectively. We compensate for this single-qubit phase shift with a virtual Z gate after the C-ARB gate.

While the CZ gate is composed of from a single detuning and gate length, the C-ARB gate requires a careful interpolation of distinct detunings and corresponding gate lengths to span all possible conditional phases with high $|11\rangle$ population recovery. In addition, since the qubit frequency excursion depends on Δ , dynamic phases have to be calibrated for all possible detuning/gate-length pairs. We detail the calibration procedure of these three continuous parameters in the next section.

2.2 Calibration

This section describes the calibration procedure of a C-ARB gate using a square flux pulse on qubit 2 and 3 of the device presented in Appendix A. To account for the finite rise-time and ensure a stable voltage at the output of the AWG, we filter the square pulse with a 1 ns Gaussian kernel. The

²The parking position refers to the frequency of the qubit when the gate is off. We typically choose it to be where the derivative of the frequency with respect to flux is 0 to minimize the charge noise, a point we call the "sweet spot" [15]

calibration of the C-ARB gate consists of three measurements, illustrated in Fig. 2.2.

First, we calibrate the pulse lengths to ensure maximum population recovery. We prepare the $|11\rangle$ state and perform a two-dimensional sweep of the flux pulse amplitude and length. For each amplitude and length, we measure the $|1\rangle$ population of qubit 2 after the flux pulse using 3-level readout, yielding the Chevron pattern shown in Fig. 2.2(a). Brighter areas correspond to high level population of qubit 2, indicating the population is fully swapped back from the $|2\rangle$ level into the computational subspace. To find the pulse lengths corresponding to the first period of the population exchange, we fit the $|1\rangle$ population oscillation to a cosine for each pulse amplitude of the two-dimensional sweep. We then retrieve the pulse length corresponding to the first maximum of the population recovery.

We fit the pulse lengths corresponding the maximum population recovery to Eq. (2.12) to extract the flux sensitivity $\frac{\partial\Phi}{\partial V}$, the asymmetry between the junctions in the SQUID loop d and the coupling strength J between the $|11\rangle$ and $|20\rangle$ states. The fit is shown as a dashed line in Fig. 2.2(a). We obtain a flux sensitivity of $\frac{\partial\Phi}{\partial V}/\Phi_0 = 0.34 \pm 0.01 \frac{1}{V}$, an asymmetry of 0.7884 ± 0.0003 and a coupling of $J/2\pi = 4.72 \pm 0.05$ MHz, all of which are consistent with previous characterization of this device [42, 43].

Next, we measure the conditional phase as function of flux pulse amplitudes (see Fig. 2.2(b)) while adapting the pulse length to maximize population recovery (see Fig. 2.2(a), black circles). If necessary, pulse length is interpolated linearly between the calibration points of the pulse length calibration. The colored flux pulses shown in the pulse scheme in Fig. 2.2(e) correspond to the diamond-shaped data points of respective color in 2.2(b). Using the parameter values obtained from the fit in Fig. 2.2(a), we compute the expected conditional phase from Eq. (2.8) and plot it in dashed line in Fig. 2.2(b). The measured conditional phase saturates at a value of different than 0 due the phase acquired via the a residual ZZ-coupling of $\alpha_{ZZ} = 400$ kHz over the period the pulse length and buffers (25ns) that separates the two $R_x^{\pi/2}$ -pulses.

Finally, we measure the dynamic phase the fluxed qubit acquires during its frequency excursion (see Fig. 2.2(c)) with the pulse scheme presented in Fig. 2.2(f)). With the approximation that its frequency is constant over the entire pulse at the value given by Eq. (2.10), we compute the expected dynamic phase from Eq. (2.13) and show it with a dashed line. This measurement requires a high number of calibration points, as the dynamic phase grows rapidly as a function of detuning yet can only be measured in the $[0, 2\pi[$ interval (a minimum of 3 points per $[0, 2\pi[$ interval are required to ensure correct unwrapping). In Fig. 2.2(c), only one out of three points is shown for clarity.

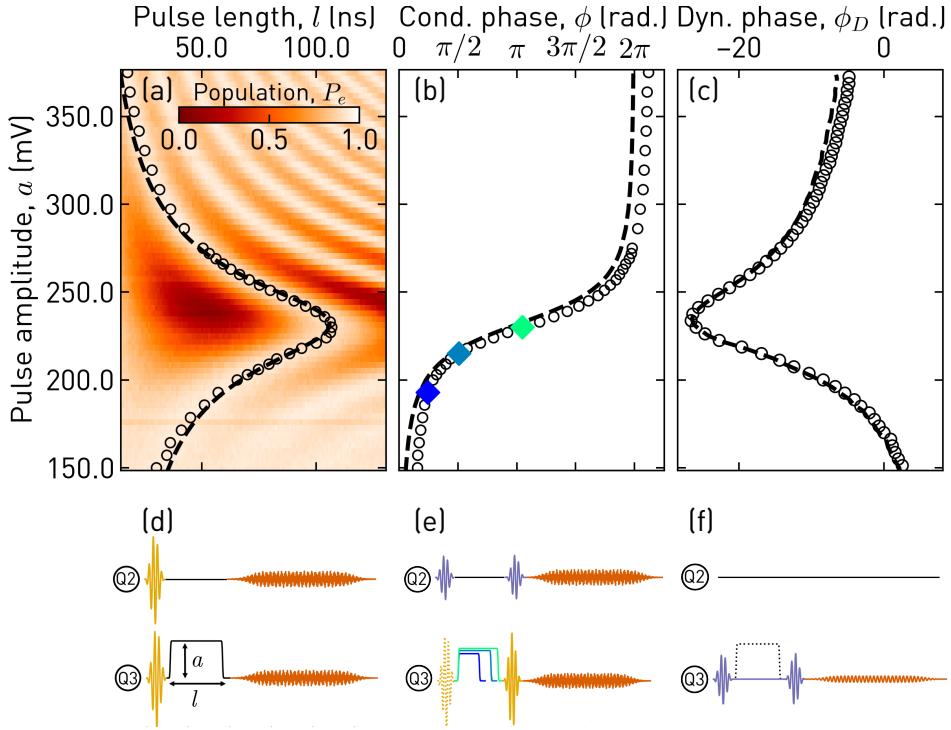


Figure 2.2. Three calibration measurements (a, b, c) of a C-ARB gate and their corresponding pulse schemes (d, e, f). (a) Qubit 2 $|e\rangle$ level population as function of the flux pulse amplitude and length. The inset details the pulse scheme: the $|11\rangle$ state is prepared, the flux pulse (black) is applied to qubit 3, the two qubits are read out with a multiplexed readout pulse. The scattered points correspond to the amplitudes and lengths used to calibrate the conditional phase (see b). The dashed line corresponds to the fit to the theoretical model. (b) Conditional phase as function of flux pulse amplitude. The amplitude sampling is not uniform to account for the non-linearity of the conditional phase. The dashed line indicates the phase predicted of the model (see main text for details). (c) Unwrapped dynamic phase of qubit 3 as function of pulse amplitude. Only one out of three data point is shown for clarity. The dashed line corresponds to the dynamic phase predicted by theory (see main text for details). (d) Pulse scheme corresponding to the two-dimensional parameter sweep shown in (a). The $|11\rangle$ state is prepared with two single-qubit R_x^π (yellow), followed by the flux pulse (black) of amplitude a and length l , and finally a readout pulse (orange). (e) Pulse scheme to measure conditional phase (see b). The phase is measured on qubit 2 with a Ramsey experiment (two $R_x^{\pi/2}$ pulses (purple)), while qubit 3 is fluxed once with and once without preceding single-qubit R_x^π (dashed yellow). The difference between the two measurements yields the conditional phase. The different flux pulse colors result in the conditional phases indicated with diamonds of corresponding color in (b). (f) The dynamic phase is measured by comparing the phase of the qubit via a Ramsey experiment with and without flux pulse (dashed black).

Due to flux cross-talk between the flux lines, the qubit staying at its parking position also acquires a non-zero dynamic phase. However, the latter is not calibrated, as flux cross-talk between the flux line of qubit 1 (3) and qubit 2 is smaller than 2% (3%) for the two two-qubit gates considered in this report [44].

The calibration time of the C-ARB gate on qubits 2 and 3 amounts to approximately 2 hours and 45 minutes i.e. about 5 to 8 times slower than a full calibration of a CZ gate. The high-resolution, two-dimensional amplitude/gate length sweep last for 1 hour, whilst for the conditional phase measurement (45 calibration points) requires 15 minutes, and the dynamic phase, 1 hour and 30 minutes. The following actions would reduce the calibration time:

- Use the active reset scheme discussed in Section B.5. This scheme allows an increase of the repetition rate by a factor of 5-10 leading to a total measurement time of 20-40 minutes.
- Use a non-linear amplitude sampling for the dynamic phase based on the theoretical model. Namely, have an adaptable spacing between amplitudes at which the dynamic phase is calibrated based on whether the model predicts a large or small increase in dynamic phase. This would reduce the number of measurement point for the dynamic phase calibration by a factor of ~ 2 , compared to a dense linear pulse amplitude spacing.
- Rewrite the dynamic phase measurement such that all waveforms are uploaded on the AWGs at once, instead of one at the time. This would reduce communication overhead.

Implementing these changes could in principle reduce the calibration time to under 20min per gate. Parallel tune up of two-qubit gates has the potential to decrease the calibration time of the device even further.

In this section, we demonstrated the ability to calibrate a C-ARB gate. The next section details the characterization of the gate.

2.3 Characterization

Randomized benchmarking – the most common method to characterize two-qubit gates – cannot be performed on the C-ARB gate because it is not part of the Clifford group [45, 46]. Hence, we use another common method called quantum process tomography to characterize the C-ARB gate. Next, we perform a fine target phase sweep and characterize conditional and dynamic phase errors as well as leakage out of the computational subspace. Finally, we compare the performance of the C-ARB gate to the CZ gate and assess their stability over a time (up to 15 hours after calibration).

2.3.1 Quantum process tomography

Quantum process tomography (QPT) is a method providing full description of a quantum process [47, 48]. In brief, it consists of preparing an ensemble of input states $\{\rho_1^{\text{in}}, \dots, \rho_k^{\text{in}}\}$ spanning the Hilbert space of interest, passing them through the quantum process \mathcal{E} and finally identifying the resultant states $\{\rho_1^{\text{out}}, \dots, \rho_k^{\text{out}}\}$ using quantum state tomography [49]. The output states are then

$$\rho_k^{\text{out}} = \mathcal{E}(\rho_k^{\text{in}}) = \sum_{mn} \chi_{mn} P_n \rho_k^{\text{in}} P_m^\dagger \quad (2.14)$$

where χ is the process matrix and P is the Pauli operator basis for 2 qubits: $P_n \in \{I, X, -iY, Z\}^{\otimes 2}$ [50]. The inversion of the above equation yields the desired process matrix χ , which describes how the process affects an arbitrary input density matrix.

The process fidelity \mathcal{F} quantifies the closeness between the measured process matrix χ_{exp} and the target matrix χ_{targ} [51]:

$$\mathcal{F} = \text{Tr}(\chi_{\text{exp}} \chi_{\text{targ}}) \quad (2.15)$$

This metric quantifies how well the measured gate would act on half of a fully entangled system and if that action is identical to the target operation. In the presence of leakage, the tomographic reconstruction of the computational subspace is not a trace-preserving map [52]. Therefore, we use the 3-level readout scheme discussed in Appendix B to detect and exclude leakage events from the tomography measurement. We characterize leakage occurring during the gate separately in Section 2.3.2. Readout noise can also lead to non-physical³ density matrix reconstructions. We use a maximum likelihood algorithm to find physical density matrices which are most likely to have been measured under the assumption of Gaussian noise [53]. Finally, we correct for readout errors by multiplying the measured probabilities with the inverse of the readout probability matrix (see Section B.4.2 and Supplementary Material [54, suppl. mat.]).

We perform QPT for various target conditional phases of the C-ARB gate. Two resulting process matrices, χ^π and $\chi^{7\pi/4}$, for target phases of π and $7\pi/4$ respectively, are shown in Fig. 2.3. The black frame represents the target process matrices while the filled bars correspond to the reconstructed process matrices. We observe a good agreement between them. In the chosen operator basis, a process matrix which can be described by a real unitary operator will also be real [9]. This is the case for χ^π , which is described by the real diagonal CZ gate unitary (see Eq. (2.1) with $\phi = \pi$). By contrast, $\chi^{7\pi/4}$ is not described by a real unitary operator and therefore contains an imaginary part. On the diagonal of the real part, the same terms appear as

³A physical density matrix is positive semidefinite, Hermitian and trace preserving.

2. CALIBRATING AND CHARACTERIZING CONTROLLED ARBITRARY PHASE GATES

in χ^π , but with a different relative contribution. In particular, the II contribution is higher, as adding a phase of $7\pi/4$ (nearly 2π) on the $|11\rangle$ state is closer to an identity operation than adding a phase of π .

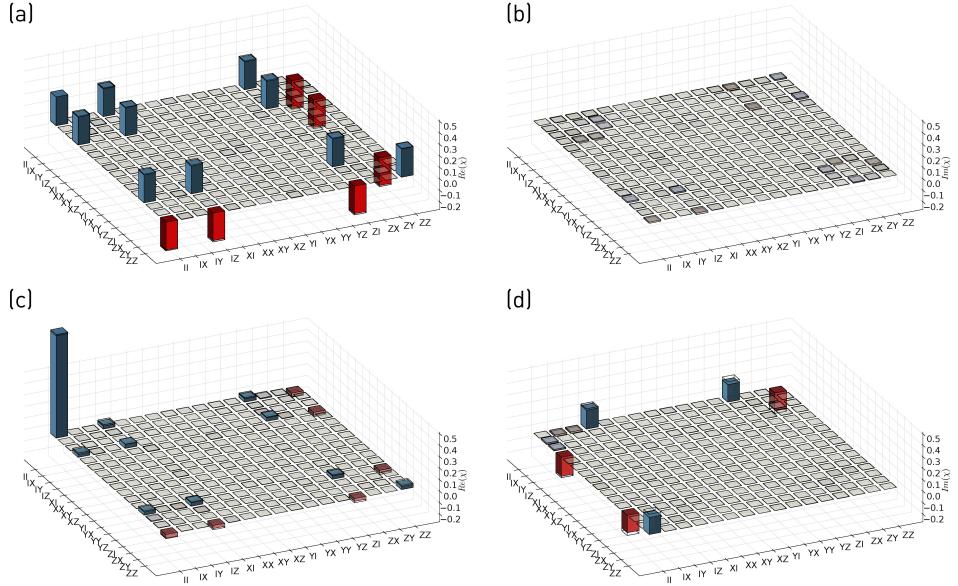


Figure 2.3. Real (a, c) and imaginary (b, d) elements of the process matrices χ^π and $\chi^{7\pi/4}$ respectively. The black wire frame corresponds to the target process matrix, while the blue (positive values) and red (negative values) filled bars correspond to the reconstructed process matrix from measurements. The process fidelity for the phase π ($7\pi/4$) is 97.3% (98.3%).

In Fig. 2.4, we show the process fidelity for 7 phase angles of the C-ARB gate. The maximum (minimum) fidelity is 98.3% (97.3%), for a phase of $7\pi/4$ (π). The expected fidelity from a master equation simulation including effects of decoherence during the gate is shown as dotted line. It explains about 1 %, and is largest near phases of π . Thermal population (0.4% (1.3%) at the time of the measurement on qubit 2 (3)) also affects the fidelity of the gate. Nevertheless, evaluating its effect precisely is not trivial: during QPT the thermal population is mixed into other states with the tomography preparation pulses. Therefore, we scale the fidelity obtained by simulation by the probability that both qubits are in the ground state before starting the measurement. We suggest that this constitutes an approximate lower bound for the fidelity. This lower bound corresponds to the dashed line in Fig. 2.4. While this lower bound closely matches the measured fidelities near phases of π , it underestimates the fidelity for small and large phases. We expect this underestimation arises because both low and large conditional phase gates are closer to an identity operation and in the limit of a fully mixed (thermal) state, any gate behaves the identity operation. A detailed

theoretical analysis is beyond the scope of this thesis (see e.g. [55]).

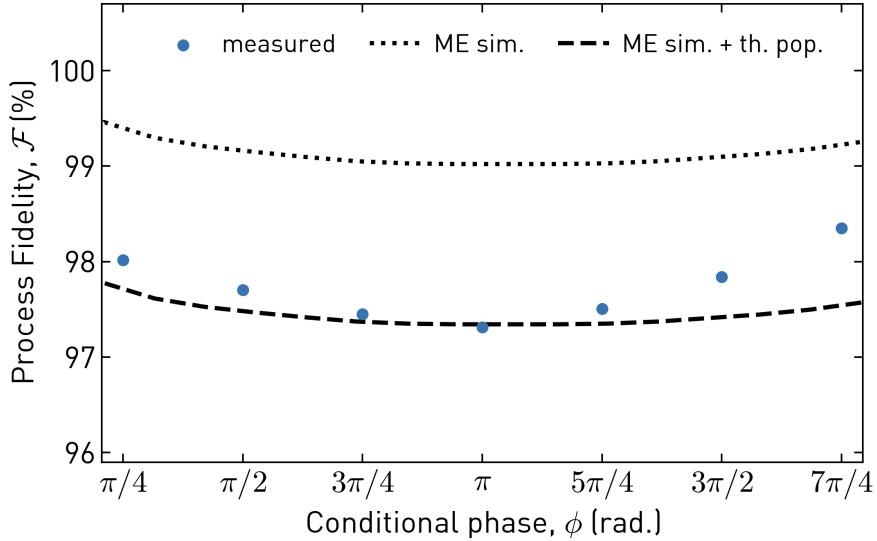


Figure 2.4. Process fidelity as function of conditional phase angle of the C-ARB gate. Measurements are shown in scattered points. Master equation simulation including effects of decoherence in dotted line. The dashed line corresponds to the fidelity of the simulation scaled by the probability of starting the measurement in the ground state.

Although QPT yields full information of the quantum process, it also has several downsides. It is a time-consuming measurement and does not provide a direct and intuitive explanation about the different origin of errors. In addition, the maximum likelihood estimation of the process matrix relies on the assumption that state preparation and measurement errors are negligible, which is not necessarily the case when the length of the two-qubit pulse is on the same scale as the single-qubit tomography pulses (i.e. for small and large target conditional phases). Finally, it does not include information about leakage. We therefore perform additional characterization measurements in the next section which portray leakage and phase errors directly.

2.3.2 Phases errors and leakage

A natural way to test the gate before its usage in an algorithm is to target a conditional phase and assess how close the measured conditional phase is from the target. When using 3-level readout, this measurement also characterizes, for each target phase, the $|f\rangle$ level population of the qubit leaving the computational subspace (in this case, qubit 2). A similar approach allows to

assess how well we are able to predict the dynamic phase the fluxed-qubit acquires.

We perform such a sweep over a target conditional phase in the range $[0^\circ, 360^\circ]$ with a spacing of 8° and present the results in Fig. 2.5(a). The conditional (dynamic) phase error shows a distribution with mean and standard deviation of $0.23 \pm 0.97^\circ$ ($-0.28 \pm 1.38^\circ$) respectively. To reduce the phase error even further, we could investigate different interpolation strategies. However, we argue that these phase errors will not be the limiting factor when used in algorithms. Indeed, the fidelity⁴ of a gate subjected to an over-rotation $\delta\phi$ degrades with $\cos^2 \delta\phi$ [57]. Therefore, for $\delta\phi = 2^\circ \approx 2\sigma_{\text{err}}$, the expected infidelity is on the order of 0.1% (0.2% when adding both conditional and dynamic phase errors). By comparison, the effect of decoherence leads to a 5 times larger infidelity, and the 400 kHz residual ZZ-coupling for a duration of 100 ns induces a phase error of 14.4° on the $|11\rangle$ state⁵.

The leakage population shows an angular dependency. It is minimal ($P_f \approx 10^{-3}$) for conditional phases far from 180° , namely when the gate is short and the $|11\rangle$ and $|20\rangle$ states are off-resonant. By contrast, we observe leakage of up to 2% when the $|11\rangle$ and $|20\rangle$ states are near resonance. At resonance, the population transfer is maximal, leading to a maximum in leakage population when assuming a fixed probability excitations being trapped in the $|20\rangle$ state. The dominant source of leakage are the short timescale distortions of the flux pulse [58] which are not perfectly compensated by the finite impulse response (FIR) filters used for pre-distortion [59]. Leakage of qubit 3, which does not go into the $|f\rangle$ level during the pulse, is on the order of 10^{-4} for all target conditional phases.

2.3.3 Comparison between C-ARB gates and CZ gates

As final characterization procedure, we compare the C-ARB gate and CZ gate (implemented on the same qubits) in terms of phase errors, leakage and long term stability. We calibrate both gates and repeat the measurement described in Section 2.3.2. For the CZ gate, we target a conditional phase of 180° , repeated N times. For the C-ARB gate, we sweep the range of target phases from 0 to 360° with N points. These measurements are repeated for up to 15 hours after calibration. The full distributions of deviations from

⁴We use the definition of fidelity between two pure states ψ_ρ and ψ_σ as the squared overlap between the amplitudes: $\mathcal{F} = |\langle \psi_\rho | \psi_\sigma \rangle|^2$ [56]. We define the gate fidelity as by the fidelity between the state resulting for the ideal gate and the state resulting from the faulty gate implementation, minimized over all input states [57].

⁵Note that errors resulting from residual ZZ-coupling do not accumulate during the two-qubit gate but rather during single-qubit pulses and idle time because the calibration of the conditional phase includes the phase resulting from residual ZZ-coupling

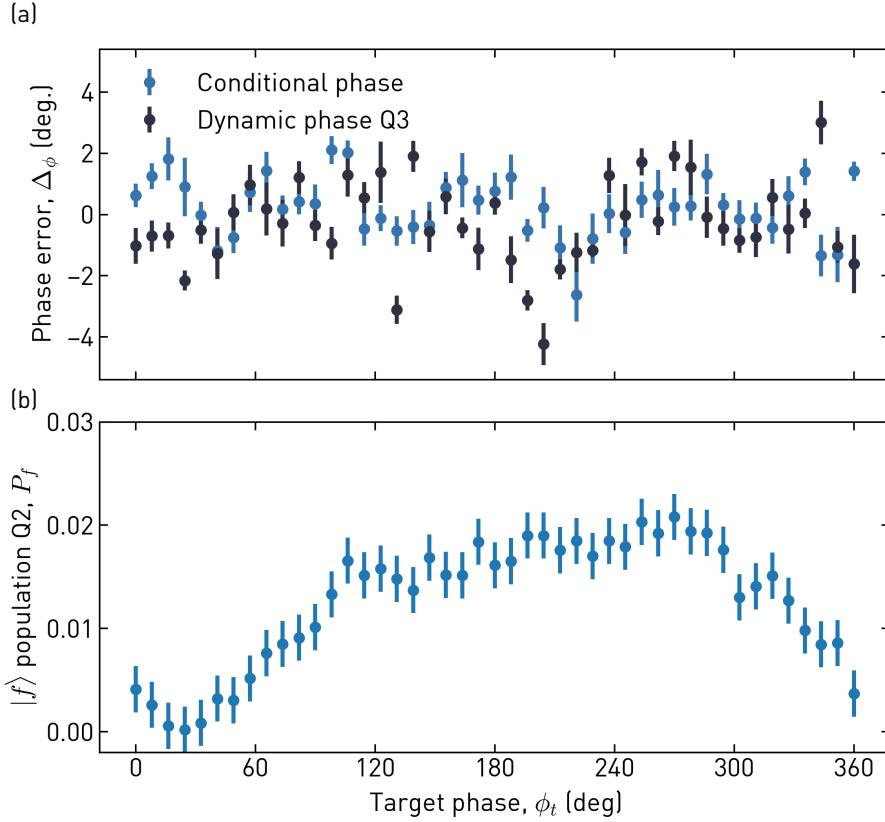


Figure 2.5. Characterization of phase errors and leakage as function of target conditional phase with a granularity of 8° . Each point is averaged over 40000 single shots. (a) Conditional (light blue) and dynamic phase error (dark blue) show distributions with mean and standard distributions of $0.23 \pm 0.97^\circ$ and $-0.28 \pm 1.38^\circ$ respectively. (b) We use 3-level readout to characterize leakage and measure up to 2% leakage.

measured phases to target phases and leakages are presented as individual histograms, visualized as shaded filling in Fig. 2.6.

Both gates exhibit a similar range of conditional phase errors (standard deviation of 0.95° for the C-ARB gate versus 0.80° for the CZ gate). We conclude that extending the CZ gate to a C-ARB gate does not strongly impact the ability of preparing a target conditional phase. Note that the CZ gate shows a consistent offset of approximately 1° with respect to its target conditional phase. This originates from the fact that the gate is calibrated with a single point, which in this case was on the lower tail of the distribution.

The C-ARB gate displays a wider distribution of dynamic phase errors, with an average standard deviation of 1.26° versus 0.83° for the CZ gate. This

suggests it is harder to calibrate the dynamic phase for many flux pulse amplitudes and lengths simultaneously compared to calibrating only one. Preliminary analysis also suggest that the interpolation method has an influence on the standard deviation of the dynamic phase error. Using a cubic spline interpolation could help reduce the dynamic phase interpolation errors, compared to the implemented linear interpolation.

The qubit 2 leakage trends are consistent with the discussion in Section 2.3.2: the CZ gate experiences an average leakage of approximately 2% since the $|11\rangle$ and $|20\rangle$ levels are on resonance. By contrast, the C-ARB gate experiences an average leakage of 1.4% as several conditional phases result in low leakage. Its worse case leakage, however, is comparable to the one of the CZ gate.

The phase error distributions stay relatively stable in time for both gates which suggests they will provide stable performance in algorithms for up to 15 hours after calibration.

The gates also differ in their average lengths. This effect is best visualized in Fig. 2.2(a). While the CZ gate has a fixed flux pulse length of 107 ns, the C-ARB gate's flux pulse length is target phase dependent. Averaged over all conditional phases, the flux pulse length is 84 ns long or $\sim 20\%$ shorter than for a conditional phase of 180 °. For both implementations, we add a 10 ns (15 ns) buffer at the start (end) to avoid overlap of other gates with the rising and falling edges of the flux pulse.

2.4 Conclusion

In this chapter we have shown how to achieve a two-qubit unitary operation $U_{\text{C-ARB}}$ which adds an arbitrary phase ϕ on the $|11\rangle$ state. We implement this unitary by varying the frequency detuning between the $|11\rangle$ and the non-computational $|20\rangle$ state with a square, Gaussian-filtered flux pulse. We adjust the length of the flux pulse to ensure high population recovery in the computational subspace. In practice, we calibrate the gate for a finite number of conditional phases (45) and interpolate linearly all parameters to reach phases between the calibration points.

We have characterized the gate using quantum process tomography for different conditional phases yielding a fidelity between 97.3% and 98.3%. The infidelity is dominated by decoherence and thermal population. In addition, we have shown that phase errors are small compared to other error mechanisms such as decoherence and residual ZZ-coupling. The average leakage amounts to 1.4% and originates from the distortions of the flux pulse which are not corrected by the finite impulse response filter. The C-ARB gate shows similar time stability and phase error distributions as a CZ gate implemented on the same qubits while being 20% shorter on average.

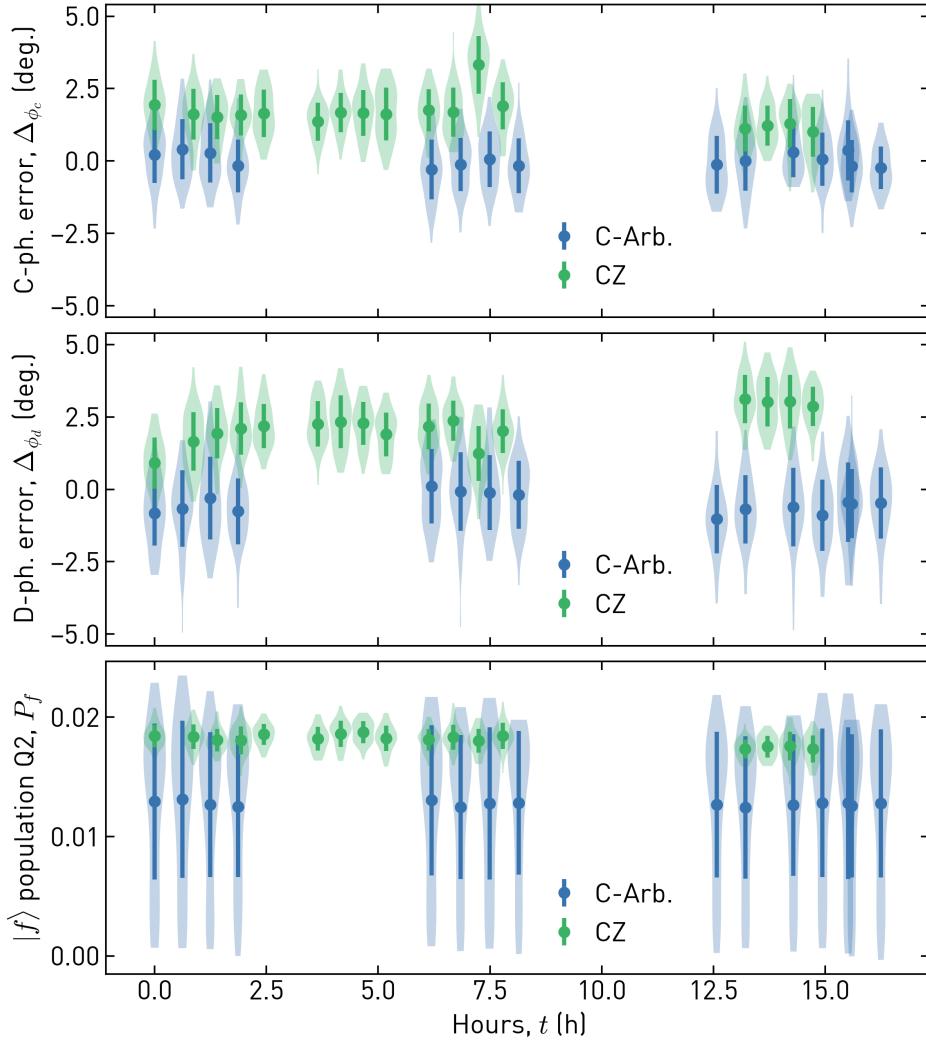


Figure 2.6. Comparison between the CZ gate (green) and C-ARB gate (blue) in terms of phase errors and leakage as a function of time elapsed since calibration. The error bars correspond to the mean and standard deviation while the full distribution ($N = 45$ points) is shown as shaded filling.

Although its calibration procedure is more elaborate, the C-ARB gate enables a significant gate count reduction for VQA circuits. In the next chapter, we implement a three-qubit problem instance which we solve using the QAOA to demonstrate this advantage.

Chapter 3

Solving an exact cover problem instance using the QAOA

In this chapter, we first introduce combinatorial optimization. Next, we explain the quantum approximative optimization algorithm (QAOA), a quantum heuristic algorithm providing approximate solutions to combinatorial optimization problems. We show that using C-ARB gates can reduce the length of a QAOA’s gate sequence implemented on the quantum computer. Next, we demonstrate this advantage experimentally by solving an exact cover problem instance on a three-qubit quantum processor.

3.1 Combinatorial optimization

An optimization problem for which the aim is to find the optimal (or close to optimal) solution amongst a finite set of possibilities is called a *combinatorial optimization problem*. Such problems are pervasive and appear in applications such as logistics, hardware verification, telecommunication network design, task scheduling and many more [60–62].

In its general form, a combinatorial optimization problem is specified by a set of rules, called clauses, acting on strings of N bits, in which each bit represents a binary decision variable [28]. Each clause is satisfied for certain assignments of the bits and unsatisfied for other assignments. Solving the problem consists in finding a combination of bits (forming together the bit string, $b := b_1 \dots b_N$) satisfying the largest (weighted) amount of clauses. The objective function can be written as

$$C(b) = \sum_{m=1}^M w_m C_m(b) \quad (3.1)$$

where $C_m(b)$ is the m -th clause and w_m is its corresponding (real and non-negative) weight. If $C_m(b)$ is satisfied by the bit string b then its value is 1

3. SOLVING AN EXACT COVER PROBLEM INSTANCE USING THE QAOA

and otherwise 0. In these terms, solving the problem is expressed as finding the bit string b_{opt} maximizing C . An approximate solution provides a bit string b_{approx} for which $C(b_{\text{approx}})$ is close to the maximum of C .

We illustrate this abstract concept with a simple example. Imagine you have to choose an outfit – consisting of one t-shirt and one pair of pants – from your closet. You have 2 shirts to choose from, a white and blue one. Similarly, you have a white and a blue pair of pants. You believe you look better when your whole outfit is of the same color. Additionally, you think your blue shirt suits you better than the white one. Which outfit should you choose for the party?

While the answer is trivial in this example, we formalize the problem with the terminology introduced above. The outfit, $b = b_1 b_2$, can be formulated as a 2-bit string, one bit for the shirt (b_1), and one for the pair of pants (b_2). Colors are encoded by the values of the bits: 0 for white and 1 for blue. The four possible outfits are: 00 (white-white), 01 (white-blue), 10 (blue-white) and 11 (blue-blue). The problem has two clauses. $C_1(b) = \neg(b_1 \oplus b_2)$, leading to a value of 1 if you choose a shirt and pair of pants of the same color and 0 otherwise¹. $C_2(b) = b_1$ has a value of 1 only if you decide on wearing your blue shirt. The objective function is then:

$$C(b) = C_1(b) + C_2(b) = \neg(b_1 \oplus b_2) + b_1 \quad (3.2)$$

where we assigned equal weights $w_1 = w_2 = 1$ to both clauses. The score of each output is then:

$$C(00) = 1 + 0 = 1 \quad (3.3a)$$

$$C(01) = 0 + 0 = 0 \quad (3.3b)$$

$$C(10) = 0 + 1 = 1 \quad (3.3c)$$

$$C(11) = 1 + 1 = 2 \quad (3.3d)$$

The bit string maximizing $C(b)$ is indeed 11 (blue shirt, blue pants). An algorithm capable of providing this bit string is a combinatorial optimization algorithm. In this case, we used the exhaustive search algorithm, meaning that we enumerated all possible answers and picked the best one.

In many real applications, exhaustive search is not tractable. In fact, even a slight modification of the presented problem makes it unrealistic to use exhaustive search. Instead of 2, imagine having to choose 25 items, each available in 8 different colors (each item has now 3 bits to encode the colors). The number of possible outfits becomes $2^{3 \cdot 25} \approx 4 \cdot 10^{22}$. It would take about

¹ \oplus denotes the "XOR" operation and \neg is the logical not.

3.2. The quantum approximate optimization algorithm

a million years to search sequentially through all possibilities for a computer requiring 1 ns to evaluate the cost of each outfit².

Several (classical) heuristic algorithms exist providing an approximate solution for this type of problems. In the next section, we explain the QAOA, a quantum heuristic algorithm capable of finding approximate solutions to this type of problems.

3.2 The quantum approximate optimization algorithm

The QAOA [28] (depicted in Fig. 3.1(a)) is a VQA designed to find approximate solutions to combinatorial optimization problems. It is a hybrid algorithm executed partially on a quantum computer and partially on a classical computer.

First, the quantum computer prepares a quantum state $|\vec{\gamma}, \vec{\beta}\rangle$ based on the objective function³ C characterizing the clauses of the problem (see Eq. (3.1)), and a set of variational parameters $\theta = (\vec{\gamma}, \vec{\beta})$. The state is prepared with p layers of two unitaries corresponding to the time-evolution of two non-commuting Hamiltonians, \hat{C} and \hat{B} , applied to a uniform superposition state of N qubits,

$$|\vec{\gamma}, \vec{\beta}\rangle = \underbrace{e^{-i\beta_p \hat{B}} e^{-i\gamma_p \hat{C}}}_{\text{layer } p} \cdots \underbrace{e^{-i\beta_1 \hat{B}} e^{-i\gamma_1 \hat{C}}}_{\text{layer 1}} |+\rangle^{\otimes N} \quad (3.4)$$

where $|+\rangle^{\otimes N} = \left(\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)\right)^{\otimes N}$ is the initialization state, and $\vec{\gamma} = (\gamma_1, \dots, \gamma_p)^T$ and $\vec{\beta} = (\beta_1, \dots, \beta_p)^T$ are the real variational parameters for the p layers. Referring to the number of layers, we say that the implementation is of depth p . \hat{C} is the cost Hamiltonian, which for all objective functions of NP-complete problems can be mapped to an Ising Hamiltonian in polynomial time [30],

$$\hat{C} = \sum_{n < m} J_{nm} \sigma_n^z \sigma_m^z + \sum_n h_n \sigma_n^z \quad (3.5)$$

and \hat{B} is the mixing Hamiltonian

$$\hat{B} = \sum_n \sigma_n^x \quad (3.6)$$

where σ_n^z (σ_n^x) are the Pauli Z (X) operators applied to the n -th qubit, and h_n and J_{nm} are real coefficients. In the combinatorial optimization picture, these coefficients represent the relative weights of each clause, see Eq. (3.1).

²Note that you are confronted to a similar situation every morning when dressing up. But you do not use the exhaustive search algorithm to decide what you wear, otherwise you would never get out of your house.

³Also often named cost function in the context of a minimization problem.

3. SOLVING AN EXACT COVER PROBLEM INSTANCE USING THE QAOA

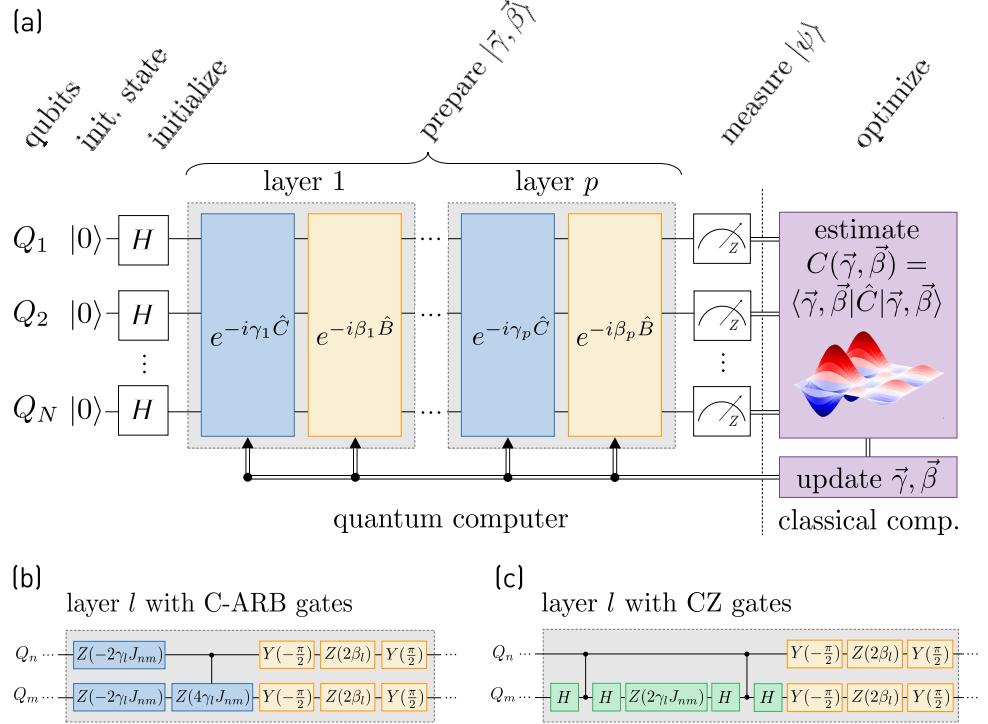


Figure 3.1. The QAOA scheme. $Z(\cdot)$ and $Y(\cdot)$ correspond to single-qubit z- and y-rotations while H is the Hadamard gate. (a) The scheme consists of two steps. First, the quantum computer applies p alternating sequences – called layers – of two unitaries corresponding to the time evolution of two Hamiltonians \hat{C} (blue), \hat{B} (yellow) scaled by layer-dependent variational parameters to an equal superposition state. The resulting state, $|\vec{\gamma}, \vec{\beta}\rangle$, is prepared and measured several times to obtain an estimate for the expectation value of a cost function $C(\vec{\gamma}, \vec{\beta})$, based on the expectation values of single and two-qubit terms. Based on this estimate, a classical optimizer alters the variational parameters to minimize $C(\vec{\gamma}, \vec{\beta})$. The output of the algorithm is the projection in the z-basis of state $|\psi_{\text{opt}}\rangle$ yielding a bit-string with a low cost. (b) Implementation of a layer of the QAOA for a cost Hamiltonian consisting only of two-qubit interaction terms. Only two generic qubits, Q_n and Q_m , are represented. The gate’s background color indicates which part in (a) it implements. The single-qubit Z gates are virtual gates implemented by adding phase on the subsequent pulses. The two-qubit interaction is mediated by a C-ARB gate. The arbitrary single-qubit x-rotations are not calibrated on the device and are thus decomposed into two y-rotations and a z-rotation. (c) Equivalent gate sequence diagram as in (b) but featuring a decomposition of the C-ARB gate into CZ gates and single-qubit gates.

The resulting state $|\vec{\gamma}, \vec{\beta}\rangle$ is measured in the Z basis. A classical computer uses the measurement outcome (i.e. bit strings corresponding to the eigen-

states of $|\vec{\gamma}, \vec{\beta}\rangle$) to compute the expectation value of \hat{C} and iteratively updates the variational angles to optimize the objective function. When the objective function is mapped to the Ising Hamiltonian, the resulting problem consists in finding the lowest energy state of the system, i.e. minimizing the expectation value of the cost Hamiltonian. Once the optimizer has converged, the state $|\psi_{\text{opt}}\rangle$ constitutes an approximate solution for the optimization problem.

3.2.1 Success probability

Measuring successive preparations of $|\psi_{\text{opt}}\rangle$ yields a distribution on the eigenstates $|\psi_i\rangle$. We define the success probability, P_s , as the probability of measuring an eigenstate that is a solution to the combinatorial problem.

Note that the optimization is performed on the smooth landscape of the objective function $C(\vec{\gamma}, \vec{\beta})$ and not on P_s directly. But for problems with a discrete domain of real values, a low expectation value of the cost function generally guarantees a high concentration of probability on optimal solutions upon measurement [63, 64].

3.2.2 Layer implementation

The algorithm is subdivided into layers, in which each layer l includes the unitary $e^{-i\gamma_l \hat{C}}$. Since \hat{C} is diagonal in the computational basis, all its terms commute. In addition, in this thesis we consider the case where all $h_n = 0$. Therefore, we look at the implementation of a single generic term $J_{nm} \sigma_n^z \sigma_m^z$ involving qubit Q_n and qubit Q_m . Any other term in \hat{C} can be applied before or after in the exact same way.

The corresponding unitary to implement is $e^{-iJ_{nm}\gamma_l \sigma_n^z \sigma_m^z}$, yielding in the nm -two-qubit subspace,

$$U_{nm}^l = \begin{pmatrix} e^{-i\gamma_l J_{nm}} & 0 & 0 & 0 \\ 0 & e^{i\gamma_l J_{nm}} & 0 & 0 \\ 0 & 0 & e^{i\gamma_l J_{nm}} & 0 \\ 0 & 0 & 0 & e^{-i\gamma_l J_{nm}} \end{pmatrix} \quad (3.7)$$

We present two approaches to implement this unitary. The first one, which we call the *direct implementation of two qubit interaction terms* (*dir. impl.*), implements Eq. (3.7) with a single C-ARB gate and two single-qubit Z-gates (see Fig. 3.1(b)). The C-ARB gate introduces a phase of $4\gamma_l J_{nm}$ on the $|11\rangle$ state which is then redistributed to the other diagonal terms by single-qubit Z-gates rotating their respective qubit subspace in the opposite direction with half the angle, i.e. $2\gamma_l J_{nm}$. Up to a global phase, this implements Eq. (3.7). Note that, due to the presence of the variational parameter γ_l , the angle

on the $|11\rangle$ state may have any real value and therefore indeed requires a C-ARB gate to rotate the $|11\rangle$ state with a single gate.

In case we choose to limit the gate set to the use of CZ gates only (i.e. rotation of π on the $|11\rangle$ state), we use a *decomposed implementation of two qubit interaction terms (dec. impl.)* consisting of two CZ gates and additional single-qubit gates (see green colored elements in Fig. 3.1(c)).

In Section 3.4, we compare experimentally the performance of both implementations.

3.2.3 Relation to adiabatic quantum computing

The depth p of a QAOA implementation is defined as the number of layers used for the state preparation. This parameter plays an important role both in theory and in experiments.

For $p \rightarrow \infty$, the QAOA finds the global optimum of C [28]. Intuitively, this is because the QAOA can be seen as a Trotterized approximation [65] of the quantum adiabatic algorithm [66]. Similarly to the QAOA, the quantum adiabatic algorithm starts in the lowest energy state of the mixing Hamiltonian \hat{B} , i.e. the equal superposition over all qubits in the Z basis, and then evolves adiabatically to the lowest eigenstate of the cost Hamiltonian \hat{C} . The time-dependent Hamiltonian during this evolution is

$$\hat{H}(t) = \hat{H}_1(t) + \hat{H}_2(t) = (1 - t/T)\hat{B} + (t/T)\hat{C} \quad (3.8)$$

where T is a real and positive parameter called run time. As long as the evolution is sufficiently slow i.e. T is sufficiently large, this process forces the state of the system to remain in the ground state of the slowly varying Hamiltonian $\hat{H}(t)$, ultimately resulting to the ground state of \hat{C} [28, 66].

To implement this on a quantum computer, we break down the total time evolution $U(0, T)$ from 0 to T into K smaller intervals $U_k((k-1)\Delta t, k\Delta t)$ of length Δt such that

$$|\psi_{\text{opt}}\rangle = U(0, T)|+\rangle^{\otimes N} = U_K((K-1)\Delta t, K\Delta t) \dots U_0(0, \Delta t)|+\rangle^{\otimes N} \quad (3.9)$$

with the time evolution in each interval given by

$$U_k(k\Delta t, (k+1)\Delta t) = \mathcal{T} \exp \left(-i \int_{k\Delta t}^{(k+1)\Delta t} \hat{H}(t) dt \right) \quad (3.10)$$

and the time ordering operator \mathcal{T} [67, p. 143]. It is shown that removing the time ordering operator introduces an error $\varepsilon \propto \Delta t^2$ [68]. Intuitively, this suggests that as long as the time intervals are small, the order of operations within the step leads to negligible errors. It is also more practical to approximate the integral by a sum, for instance using Monte-Carlo integration [68].

3.2. The quantum approximate optimization algorithm

The latter replaces the integral by the average of M evaluations of the function at random times τ_m^k within the interval Δt which introduces an error $\varepsilon_{\text{MC}} \propto \Delta t / \sqrt{M}$. The time evolution step becomes

$$U_k(k\Delta t, (k+1)\Delta t) = \exp\left(-\frac{i}{M} \sum_{m=1}^M \hat{H}(\tau_m^k)\right) = \prod_{m=1}^M e^{-\frac{i}{M} \hat{H}(\tau_m^k)} \quad (3.11)$$

Within the product, the Hamiltonian is not time dependent anymore, such that we can apply the Trotter formula to decompose it into the two components \hat{B} and \hat{C} . The Trotter product formula stipulates for two non-commuting time-independent Hamiltonians \hat{H}_1 and \hat{H}_2 [65]:

$$e^{-i(\hat{H}_1 + \hat{H}_2)t} = \lim_{N \rightarrow \infty} \left(e^{-i\hat{H}_1 t/N} e^{-i\hat{H}_2 t/N} \right)^N \quad (3.12)$$

Truncating this equation to a finite value for N yields an error with lowest order term in $\mathcal{O}\left(\frac{t^2}{2N} [\hat{H}_1, \hat{H}_2]\right)$ [29, 69].

In this case, we make an approximation of order $N = 1$ with $\hat{H}_1 = (1 - \frac{\tau_m^k}{T})\hat{B}$ and $\hat{H}_2 = \frac{\tau_m^k}{T}\hat{C}$. The full adiabatic evolution of Eq. (3.9) becomes the alternating sequence between the two Hamiltonians \hat{B} and \hat{C} ,

$$U(0, T) = \prod_{k=0}^{K-1} U_k(k\Delta t, (k+1)\Delta t) = \prod_{k=0}^{K-1} \prod_{m=1}^M e^{-i \underbrace{\frac{1 - \tau_m^k/T}{M} \hat{B}}_{\beta_{km}}} e^{-i \underbrace{\frac{\tau_m^k/T}{M} \hat{C}}_{\gamma_{km}}} \quad (3.13)$$

Note that applying this discretized adiabatic time evolution on $|+\rangle^{\otimes N}$ as presented in Eq. (3.9) ends in the global extremum of \hat{C} *only* for $K \rightarrow \infty$ and large T . In that case, the number of layers is given by $p = K \cdot M$ and corresponds intuitively to approximating the time evolution from 0 to T by infinitely many time steps. The error in the approximation of the integral intrinsically tends to zero if $\Delta t \rightarrow 0$. In such a scenario, the set of variational parameters $(\vec{\gamma}, \vec{\beta}) = \{(\gamma_0, \beta_0), \dots (\gamma_p, \beta_p)\}$ highlighted in Eq. (3.13) suffices to find the optimum.

However, in practice the number of layers we can implement on a quantum computer is limited. Therefore, we truncate Eq. (3.13) to a finite value for both K and M . This truncation introduces approximation errors and finding the global optimum is no longer guaranteed.

To mitigate this approximation, we relax the adiabatic constraint $(\beta_l = 1 - \gamma_l)$ at each layer to allow for a potentially more direct path to the ground state than the adiabatic one. By optimizing these parameters at each layer variationally, the QAOA can reach the ground state with high accuracy for some problem instances even with a relatively small number of layers [26].

3.2.4 Quantum speedup and the importance of depth

At the time of its first publication, the QAOA achieved a higher approximation ratio on MAX-3-LIN-2 – a specific instance of the NP-complete MaxCut problem [70] – than any other classical algorithm. This is no longer the case [71, 72]. In fact, several research articles have shown that classical algorithms could outperform the QAOA on a variety of problem instances [71–73]. On the other hand, several articles claim to achieve Grover speedup for unstructured search [74] and state transfer [75] using the QAOA. Whether or not the QAOA will provide a provable or practical speedup compared to classical algorithms remains unanswered.

Understanding the number of layers required to open a path to the ground state as function of (i) the number of qubits in the Hamiltonian, and (ii) the chosen problem instance are both open research questions. In addition, it remains unclear whether the path can be found efficiently as the number of variational parameters grows with $2p$. Early results suggests that, for QAOA to provide any quantum speedup, the scaling of p must be faster than $\mathcal{O}(1)$ otherwise the circuit is likely simulated efficiently on a classical computer [73]. On the other hand, to ensure a quantum speedup the scaling of p should also not grow faster than $\mathcal{O}(\log N)$ because faster scaling could imply that finding good variational parameters becomes exponentially hard [76].

While the above considerations provide insights into the theoretical performance of the QAOA as function of depth, they do not provide insights into the experimental performance on NISQ devices. In this work, we highlight a crucial experimental trade-off related to the number of layers. Namely, in theory the performance of QAOA circuits can only improve with increasing number of layers [77]. However, the performance on near-term quantum computers will not increase monotonically with depth due to decoherence and other noise sources. There will be an optimal number of layers beyond which additional layers only decrease the performance of the algorithm. We illustrate this trade-off in the next section and show that the dir. impl. increases the number of layers which can be executed reliably compared to the dec. impl. This in turn allows to solve more complex problem instances using the dir. impl.

3.3 Exact cover

We choose to solve an instance of the exact cover problem [78] to compare the direct and decomposed implementation of QAOA. The exact cover is NP-complete [70], meaning that there is no known algorithm to find answers in

polynomial time⁴, but answers can be checked in polynomial time. In other words, the problem is "hard to solve" but it is "easy" to check the validity of a proposed solution. In addition, the NP-completeness guarantees that any problem in NP can be reduced to the exact cover in polynomial time, such that studying the exact cover is closely related to studying other NP problems. Finally, the exact cover can be easily construct via a problem graph that is not complete, which allows to create instances respecting the hardware connectivity of the quantum device (see Chapter 5 for a discussion on hardware connectivity).

In this section, we provide the formal definition of the exact cover and introduce a problem instance which we solve in Section 3.4 using QAOA.

3.3.1 Definition

The exact cover is an NP-complete decision problem formulated as follows: *given a set \mathcal{N} , and a collection \mathcal{L} of subsets of \mathcal{N} , is there a subcollection \mathcal{L}' of \mathcal{L} for which each element in \mathcal{N} is included exactly once?* In other words, elements of the subcollection \mathcal{L}' must be disjoint and their union must be \mathcal{N} .

As mentioned in Section 3.2, any NP-complete problem can be mapped onto an Ising Hamiltonian in polynomial time. For the exact cover, this is most conveniently done using an incidence matrix M [79], where each row corresponds to an element of \mathcal{L} and each column to an element of \mathcal{N} . The matrix element M_{ln} is equal to 1 if the l -th element of \mathcal{L} contains the n -th element of \mathcal{N} and 0 otherwise. The Ising parameters J_{nl} and h_l are obtained from M [30, 80]:

$$J_{ln} = \frac{1}{2} \sum_{k=1}^N M_{lk} M_{nk} \quad (3.14a)$$

$$h_l = \frac{1}{2} \sum_{n=1}^N M_{ln} \left(\sum_{k=1}^L M_{kn} - 2 \right) \quad (3.14b)$$

This formulation requires $L = |\mathcal{L}|$ binary variables $\{b_1, \dots, b_L\}$ with $b_l \in \{0, 1\}$. To yield an Ising Hamiltonian formulation, see Eq. (3.5), each binary variable b_l is mapped to a spin variable $\sigma_l \in \{-1, 1\}$ with the mapping,

$$\sigma_l = 1 - 2b_l \quad (3.15)$$

Finally, each spin in the problem is modelled by a qubit on the quantum processor.

⁴An algorithm runs in polynomial time if its execution time, in the worst case scenario, is upper-bounded by a polynomial function of the input N , i.e. its time complexity scales as $\mathcal{O}(N^k)$ with k a positive and real constant.

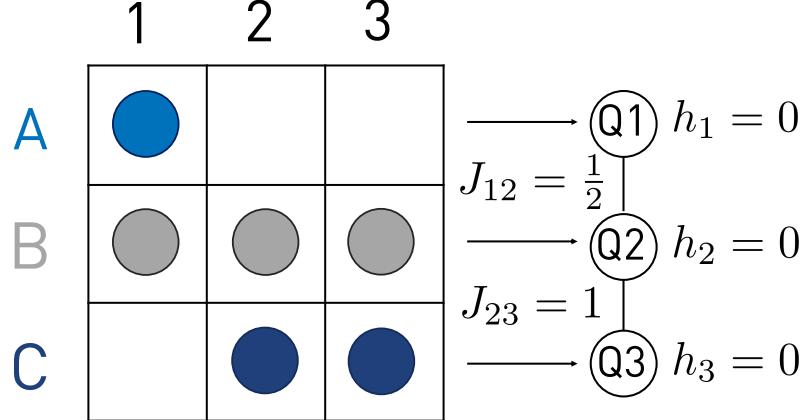


Figure 3.2. Incidence matrix representation of the exact cover problem instance implemented in this work. A dot indicates a 1 and an empty square indicates a 0. The mapping to the 3 qubit chain and the corresponding Ising Hamiltonian parameters are indicated on the right hand side.

3.3.2 Example problem instance

In this work, we consider the 3-qubit problem instance depicted in Fig. 3.2. $\mathcal{N} := \{1, 2, 3\}$ and the elements of \mathcal{L} are $A := \{1, 2\}$, $B := \{1, 2, 3\}$ and $C := \{2, 3\}$, encoded in qubit 1, 2 and 3 respectively. In this picture, the decision problem becomes: *is there a subcollection \mathcal{L}' of letters for which each number in \mathcal{N} is included exactly once?* The answer is yes, and the corresponding subcollections are $\mathcal{L}'_1 = \{A, C\}$ and $\mathcal{L}'_2 = \{B\}$. These subcollections are encoded in the states $|\psi\rangle = |101\rangle$ and $|\psi\rangle = |010\rangle$, where a 1 in position l indicates that the l -th element of \mathcal{L} is included in the subcollection \mathcal{L}' .

“Solving” this problem instance corresponds to measuring $|\psi\rangle = |010\rangle$ and $|\psi\rangle = |101\rangle$ with high probability as measuring and checking whether this state is part of the solution set would allow to answer the yes/no-decision problem.

The cost Hamiltonian of the problem is

$$\hat{C} = \frac{1}{2} \hat{\sigma}_1^z \hat{\sigma}_2^z + \hat{\sigma}_2^z \hat{\sigma}_3^z \quad (3.16)$$

and the corresponding cost function $C(\vec{\gamma}, \vec{\beta}) = \langle \vec{\gamma}, \vec{\beta} | \hat{C} | \vec{\gamma}, \vec{\beta} \rangle$ reaches its minimum value of $C(\vec{\gamma}, \vec{\beta}) = -1.5$ when $|\vec{\gamma}, \vec{\beta}\rangle = |101\rangle$ or $|010\rangle$.

3.4 Experiment

In this section, we implement the QAOA on a quantum processor to find an approximate solution to a combinatorial optimization problem with 3 qubits. In addition, we compare the performance of the dir. impl. and dec. impl. for different depths. We use qubits 1, 2 and 3 of the device presented in Appendix A. Information relative to the single- and two-qubit gates are reported in Appendix A, Table A.2. All measurements are performed with 3-level single shot readout, see Appendix B. In the context of the QAOA, leakage events can be seen as measuring bit strings which do not belong to the space of possible solutions. We therefore condition the measurements outcome on no leakage event. Hence, leakage events reduce the effective number of shots on which we compute estimates. They do not, however, influence the algorithm as long as the excitation stays in the $|f\rangle$ level until the state is measured. We make this assumption, as sequences of interest are shorter than $4\ \mu s$, and the shortest $|f\rangle$ level T_1 is approximately $10\ \mu s$. All measurements are done without ground state heralding or readout error correction.

3.4.1 Cost function landscapes

Although $p = 1$ does not always yield a good approximate solution to the optimization problem, it is a useful intermediate benchmark towards implementing multi-layer QAOA circuits [38, 39]. A single layer implementation has two variational parameters, γ and β , such that the cost function landscape can be mapped out and visualized with a grid search over γ and β . We make several observations to restrict the search space for each parameter.

First, if \hat{C} has only integer eigenvalues⁵, then γ is at least 2π -periodic because $e^{-i(\gamma+2\pi)\hat{C}} = e^{-i2\pi\hat{C}}e^{-i\gamma\hat{C}} = e^{-i\gamma\hat{C}}$ since $2\pi\hat{C}$ yields integer multiples of 2π for all elements on the diagonal. If in addition the eigenvalues are either all odd or all even, then γ shows a periodicity of π , as $e^{i\pi\hat{C}}$ yields in that case $\pm I$ depending on whether the eigenvalues are all odd or all even. By construction, this is the case when all J_{nm} in \hat{C} are integers.

Second, the parameter β yields single-qubit x -rotations of angle 2β on all qubits and is therefore at least π -periodic.

We thus restrict our search space for this problem instance⁶ to the domain $(\gamma, \beta) \in [0, \pi] \times [0, \pi]$. For each parameter pair, we repeat the state prepa-

⁵ \hat{C} is diagonal and therefore elements on its diagonal are the eigenvalues by definition.

⁶In the problem instance we study, the eigenvalues are odd integers multiples of $1/2$ and not odd integers. Therefore, the periodicity in γ is only 2π and not π . We make use of the fact that the landscape is point symmetric around $(\gamma, \beta) = (\pi, \pi/2)$ due to the linear connectivity to restrict the search space of γ between 0 and π . Another option would have been to multiply \hat{C} by a constant factor such that all eigenvalues are integers.

3. SOLVING AN EXACT COVER PROBLEM INSTANCE USING THE QAOA

ration and measurement 20000 times to estimate the expectations values $\langle \sigma_1^z \sigma_2^z \rangle$ and $\langle \sigma_2^z \sigma_3^z \rangle$ from which the cost $C(\vec{\gamma}, \vec{\beta})$ is computed.

We execute this measurement with a resolution of 45 points per axis both for the dir. impl. and the dec. impl., and present the result in Fig. 3.3 along with the landscape obtained with a noise-free unitary evolution (which does not take into account decay and dephasing rates) simulated in QuTip [81].

The simulation reveals two global minima for $C(\vec{\gamma}, \vec{\beta}) \approx -1.05$ at $(\gamma, \beta) \approx (\frac{2\pi}{9}, \frac{\pi}{7})$ and $(\frac{2\pi}{9}, \pi/2 + \frac{\pi}{7})$. The landscape also exhibits two pairs of local minima with slightly different depths; one at $\gamma \approx \frac{5\pi}{8}$ and one at $\gamma \approx \frac{7\pi}{8}$ also separated by $\pi/2$ in the β direction. The $\pi/2$ periodicity in β can be understood intuitively from the Bloch sphere perspective: for $p = 1$, we start on the equator and apply parametrized entangling Z-rotation between qubits. Each qubit therefore stays on the equator of its Bloch sphere. Thereafter, we apply parametrized X-rotations on all qubits which rotate the vectors back towards the z-axis⁷. For a X-rotation angle larger than π (i.e. $\beta > \pi/2$), we simply inverse the polarity of all states (all $|1\rangle$ are now mapped to $|0\rangle$ and vice-versa). The correlations between qubits remain unaffected by this polarity inversion and therefore the landscape is periodic.

The locations of all extrema in the measurements are in good agreement with the simulations, suggesting a low amount of total coherent errors. Incoherent errors due to decoherence affect the contrast of the landscapes, resulting in a minimum measured cost of ~ -0.95 (~ -0.85) for the direct (decomposed) implementation. The decomposed implementation does not reach the same cost value as its gate sequence is longer and thereby more affected by decoherence. In addition, it suffers more from coherent errors than the direct implementation. The latter are caused predominantly by residual ZZ-coupling, and therefore accumulate over the longer sequence. The combination of these effects could explain the diagonal feature observed in its landscape, which does not appear as strongly for the direct implementation. The average total leakage (any qubit measured in the $|f\rangle$ -level) amounts to 4.5% for the direct implementation and 7.8% for the decomposed implementation. Most of the leakage occurs due to the two-qubit gate between qubit 1 and qubit 2, which suffers from higher leakage (see Appendix A, Table A.2 for a detailed discussion).

Note that even in the noise-free simulation, $C(\vec{\gamma}, \vec{\beta})$ never reaches its minimum value of -1.5 thereby indicating that the performance of the algorithm is limited by the chosen depth of $p = 1$. Nevertheless, additional layers introduce new variational parameters, making grid search quickly intractable. Therefore, we use a classical optimizer to find the optimal variational parameters.

⁷Except if the vectors are degenerate with the $\pm x$ -axis.

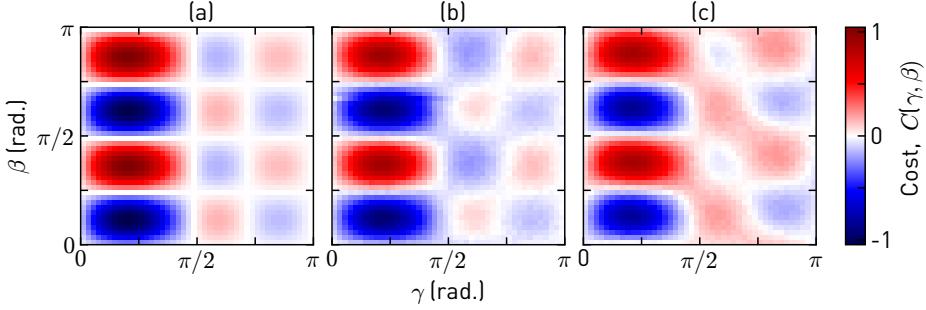


Figure 3.3. Optimization landscapes for a QAOA implementation with one layer. The two variational parameters, γ and β are swept over a 45×45 grid, and the cost function $C(\vec{\gamma}, \vec{\beta})$ is evaluated for each parameter pair. (a) Landscape obtained with noise-free simulation. The minimal cost function value is approximately -1.05. (b) Landscape obtained with the direct implementation of the C-ARB gate. Taken with 20000 single shot measurements per parameter pair. (c) Landscape obtained with the decomposed implementation of the C-ARB gate. Taken with 20000 single shot measurements per parameter pair.

3.4.2 Variational parameter optimization

When no strategy is known to find the best variational angles and $p > 1$, classical optimizers constitute a way to navigate through the parameter space. Desirable properties to ensure rapid and consistent convergence to the optimal parameters are:

- **Low cost function sampling.** Each query to the quantum computer is costly in time. Therefore, optimizers minimizing the number of queries sent to the quantum computer are better suited for this task. Algorithms relying on Hessian matrix estimation are not suited, as the number of elements in the Hessian matrix grows quadratically with p .
- **Robustness to noise.** Measurements from quantum computers are subject to noise. Especially when the gradient of the cost function is small, it is crucial for the algorithm to be robust to fluctuations.
- **Capability of avoiding local minima.** At low depths, the cost functions show local minima. Algorithms capable of finding the global optimum are better suited.

We leave the task of finding the optimal classical optimizer for future work, and instead focus on comparing the performance difference of a simple optimizer between the dir. impl. and dec. impl. For the sake of completeness, we briefly report on optimizers which have been applied to QAOA in recent literature.

To find variational parameters for a single layer QAOA implementation, Ref. [37] applied Bayesian optimization (BO) [82], a gradient-free, global op-

3. SOLVING AN EXACT COVER PROBLEM INSTANCE USING THE QAOA

timization method relying on Gaussian processes. Although BO performs excellently in low-dimensional parameter spaces, the number of evaluations grows exponentially with the number of parameters, rendering it impractical when p is large. Ref. [39] compared BO on a two-layer QAOA implementation with the Nelder-Mead (NM) algorithm [83] and the covariance matrix adaptation evolution strategy (CMA-ES) [84]. On their problem instance, NM has the lowest cost function sampling but is also the most sensitive to local minima due to its locality. By contrast, the CMA-ES requires more function evaluations but is stochastic and is believed to scale favorably with the number of parameters [39]. For $p = 2$, BO yielded a good trade-off between low cost function sampling and sensitivity to local minima. Ref. [38] used problem-specific heuristics combined with a bootstrapping algorithm. Other heuristics [77] and deep-learning based approaches [85] have been proposed but not tested beyond simulations. Finally, the simultaneous perturbation stochastic approximation (SPSA) [86] is a promising stochastic global optimization approach not yet applied to QAOA. The major advantage of SPSA is that it requires only two cost function evaluations per optimization step, independently of the dimensions of the parameter space.

In this work, we use the NM algorithm, due to its simplicity and low cost function sampling. To mitigate the effect of local minima, we increase the size of the initial simplex to the same order of magnitude as the distance between minima observed in Fig. 3.3.

For each optimization trajectory, we initialize parameters randomly and then let the NM algorithm run 40 cost function evaluations. One function evaluation takes approximately 20-25 seconds such that an optimization trajectory takes approximately 12 minutes. In Fig. 3.4, we show the results of 15 optimization trajectories of both implementations for up to 4 QAOA layers. Each data point of the cost function is estimated from 20000 shots.

The motivation for a higher number of layers becomes apparent. For $p = 1$, we observe for either implementation three different convergence values of the cost function, two corresponding to the local minima and one to the global minimum, as expected from the three pairs of minima visible in Fig. 3.3. The local minima significantly impact the average cost at convergence. In addition, the best case performance is limited by the minimal achievable cost for $p = 1$.

For $p = 2$, the best trajectories of the direct implementation achieve cost below what is possible with only one layer and local minima are less apparent, although still present.

Three layers are required to reach the minimum cost of -1.5 in noise-free simulation. But adding layers does not always yield better performance. Indeed, it also results in longer gate sequences, aggravating the effect of

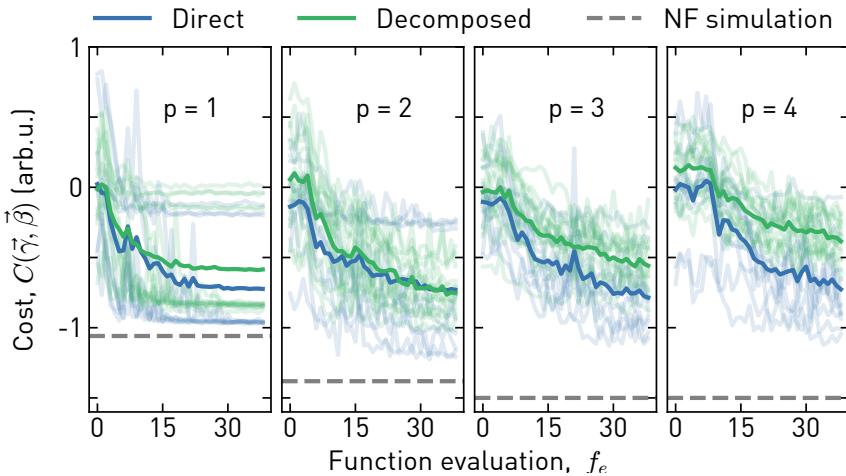


Figure 3.4. Cost optimization trajectories for depths of $p = 1, 2, 3$ and 4 as a function of the number of function evaluations, f . The variational parameters are initialized randomly and then optimized iteratively with the Nelder-Mead algorithm. Each cost value is computed from 20000 single shot measurements. The trajectories obtained with the direct (decomposed) implementation of the C-ARB gate are shown in blue (green). The faded lines correspond to the 15 individual trajectories while the darker color correspond to their average. The grey dashed line indicates the minimal achievable cost for each depth as calculated with a noise-free simulation.

decoherence. Moreover, additional layers increase the dimension of the parameter space, resulting in slower convergence, which is clearly visible by comparing the average trajectories for e.g. 1 and 4 layers. In fact, the average trajectories suggest 40 cost function evaluations do not allow the algorithm to fully converge for $p > 1$.

For $p = 1, 3, 4$, the direct implementation reaches lower cost on average (for $p = 2$, only its median is lower), indicating that the gate sequence length plays a major role in the performance of the algorithm. In the next section, we analyze this effect in more details.

3.4.3 Success probability as function of depth

The previous section suggests there is an inherent trade-off: additional layers increase the performance of the algorithm in theory but also aggravate the effect of sequence-length-dependent errors in experiments. Similar conclusions were found by Alam et al. [87]. The goal of this section is to analyze how this trade-off affects the two implementations of the QAOA.

Note that the (approximate) solution of a problem found with QAOA is the measured bit-string after optimization of the variational parameters. Therefore, a crucial metric to assess the performance is the success probability, P_s

3. SOLVING AN EXACT COVER PROBLEM INSTANCE USING THE QAOA

(see Section 3.2.1) and not the cost function used for classical optimization of the variational parameters.

To investigate the effect of the depth p on P_s accurately, we seek to mitigate the influence of the classical optimizer on P_s (namely, the influence of local minima and the convergence speed, which are expected to be equal for both implementations). To this extend, we provide an educated initial guess for the variational parameters. Namely, we first find optimal parameters for each depth in a noise-free simulation and provide these theoretically optimal parameters for the initialization of the quantum computer. Under the assumption of equal depolarization, bit-flip and dephasing channels for all qubits, the optimal variational parameters remain nearly the same in a noise-free and noisy environment, according to a recent theoretical investigation [88]. Because coherence times and coherent errors are not exactly equal for all qubits on our device, we additionally optimize locally with up to 20 function evaluations on the quantum computer.

We show the result of the above-mentioned protocol in Fig. 3.5. The success probability for various depths is plotted versus sequence length, L , to facilitate the comparison of sequences of similar lengths independently of the implementation. The dashed lines correspond to the expected success probability in a noise-free environment.

The dir. impl. exhibits a clear advantage: in about $1\ \mu\text{s}$, it is able to execute a 3 layers of QAOA while the decomposed implementation only carries out one. For a fixed depth, it consequently suffers less from errors scaling with the sequence length.

The difference in sequence length arises from the combination of two factors (see inset of Fig. 3.5 for the full pulse sequence for $p = 1$):

1. The decomposition of each C-ARB gate requires two individual CZ gate and 4 single-qubit gates.
2. The C-ARB gate is shorter than the CZ gate for any conditional phase different from π .

Moreover, the number of C-ARB gates grows with $\mathcal{O}(m \cdot p)$, where m indicates the number of two-qubit terms in \hat{C} which cannot be executed simultaneously. Therefore, the total decomposed sequence length rapidly becomes much longer.

As mentioned in Section 3.4.2, noise-free simulations require in the ideal case 3 layers to reach $P_s \approx 1$, which for the dec. impl. is realized with $L \approx 3\ \mu\text{s}$. For that sequence length, the dir. impl. can implement up to nine layers. However, in Fig. 3.5 we observe that both implementations show similar performance for $L \approx 3\ \mu\text{s}$, which directly indicates that performance is limited by sequence length rather than the number of layers. In partic-

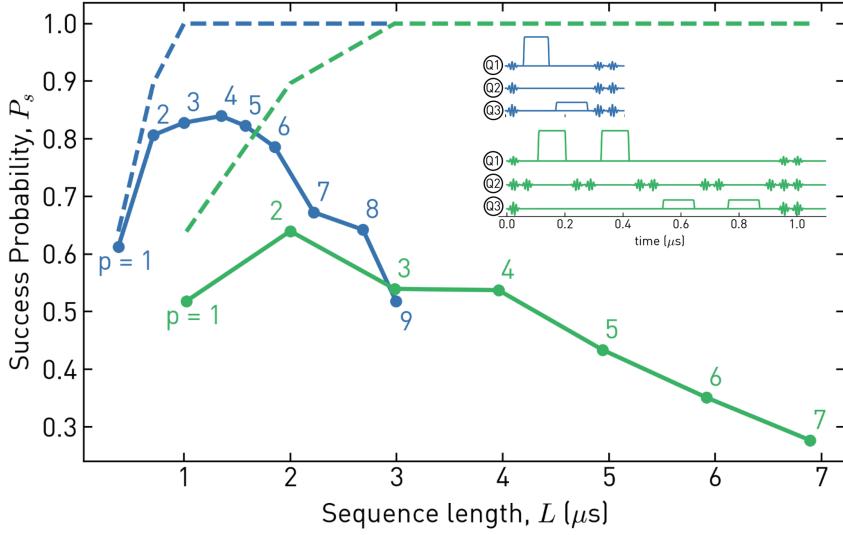


Figure 3.5. Success probability as a function of sequence length for the direct (blue) and decomposed (green) implementations of the C-ARB gate. The dashed line of corresponding color indicates the success probability as calculated by a noise-free simulation. The depth is annotated next to the data points. For each depth, the quantum computer is initialized with variational parameters obtained from the noise-free simulation. The corresponding pulse schemes for one layer are shown in the inset.

ular, in our case we expect both conditional phase errors due to residual ZZ-coupling and decoherence to have a major impact which scales with sequence length.

To gain further insights into the performance of the algorithm, we present in Fig. 3.6 the full distribution of measured bit strings for each of the first 4 points ($p = 1, 2, 3, 4$ for each implementation) of Fig. 3.5. The colored dots in the right panel indicate the respective success probabilities, i.e. the sum of the probabilities for states $|010\rangle$ and $|101\rangle$ (colored bars). The right panel also shows the success probability distribution (faded filling) of the optimization trajectories discussed in Section 3.4.2. If the trajectories constitute a representative sampling of the parameter space, the dot should lie at the right edge of the distribution. A data point outside the distribution suggests the trajectories did not converge to the lowest possible value, or did not find the global cost optimum, for instance because of a local minimum. Conversely, a data point inside the distribution suggests the initialization from theoretically optimal variational parameters is not perfect, or fluctuating performance compared to when the trajectories were recorded.

In the state distributions, the most likely states are $|101\rangle$ and $|010\rangle$, corre-

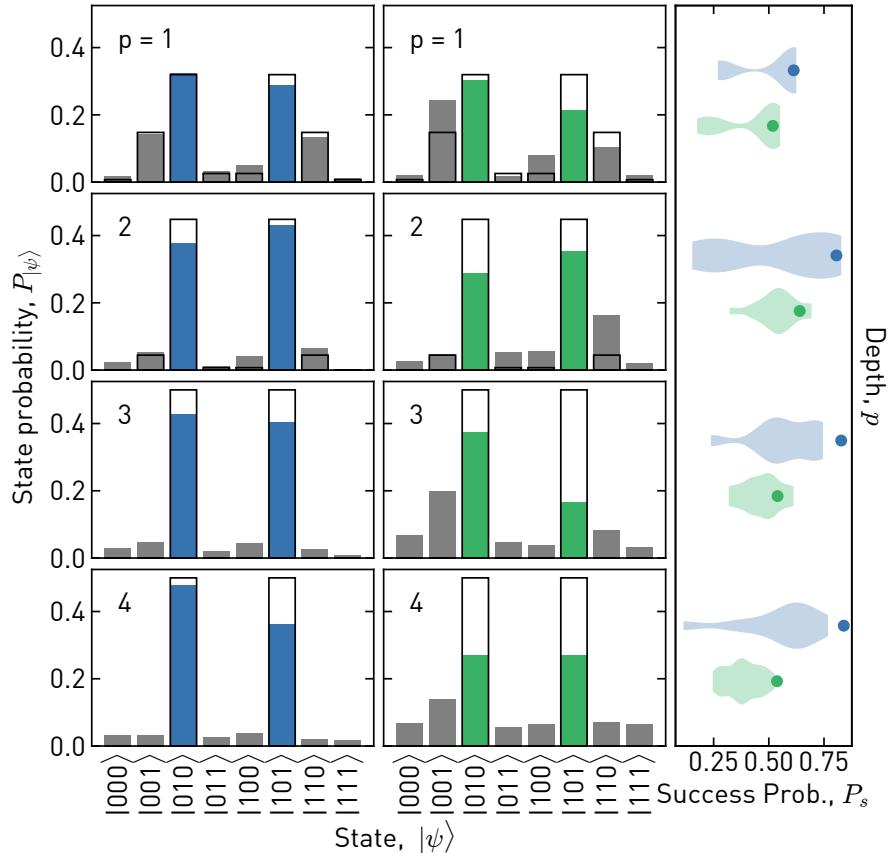


Figure 3.6. Distribution of measured states for the dir. impl. (left) and dec. impl. (middle) for $p = 1, 2, 3, 4$. The black wire-frame corresponds to the noise-free simulation distribution of states and is equal for both implementations, at fixed depth. For each distribution, the highlighted bars correspond to solutions of the exact cover problem. Their sum amounts to P_s , shown on the right panel with dots of corresponding color (also corresponding to the first 4 dots for each implementation in Fig. 3.5). The faded distributions in the right panel correspond to the success probability of the optimization trajectories presented in Fig. 3.4.

sponding to the desired sets $\mathcal{L}'_1 = \{A, C\}$ and $\mathcal{L}'_2 = \{B\}$ defined in Section 3.3. The dir. impl. is in excellent agreement with the noise-free simulation (black wire-frame) for all depths, and achieves a success probability of 0.84 at $p = 4$ ⁸. By contrast, the dec. impl. consistently shows more de-

⁸In principle three layers are sufficient to reach $P_s \approx 1$ and additional layers are expected to suffer from additional errors, thereby showing a decrease in success probability. However, the classical optimizer also allows to compensate for some errors and more parameters

viations from the simulation and achieves a maximal success probability of 0.64⁹ at $p = 2$.

We conclude that the trade-off of additional layers affects the two implementations of the C-ARB gate differently. The success probability of the dir. impl. increases for up to four layers before being decreasing due to decoherence and accumulated phase errors, thereby achieving a success probability of 0.84. On the other hand, the dec. impl. can only execute two layers before being affected by sequence length dependent errors, which is not sufficient to reach the regime where the success probability is near unity even in the noise-free simulation.

3.5 Conclusion

In this chapter, we have found approximate solutions for a three-qubit exact cover problem instance using the quantum approximative optimization algorithm (QAOA). The problem instance requires at least three layers of the QAOA in the ideal case to reach a success probability approaching unity, as determined by noise-free simulations. To the best of our knowledge, this constitutes the deepest problem considered in the experimental QAOA literature.

We have shown that for this problem instance, the direct implementation of two-qubit interaction terms in each layer of the QAOA reduces the two-qubit gate-count by 50% and the executed gate-sequence length by a factor of ~ 3 compared to a decomposition into 2 CZ gates and 4 single-qubit gates. Exploiting this advantage, we demonstrate a success probability of 0.84 using 4 layers of QAOA. In comparison, the decomposed implementation reached a maximal success probability of 0.64 with 2 layers as it is more affected by decoherence and residual ZZ-coupling. Experimental observations also strongly suggest that decoherence and residual ZZ-coupling are the dominant error sources in both implementations.

The sequence length reduction factor enabled by C-ARB gates scales approximately linearly with the number of two-qubit terms in the cost Hamiltonian which cannot be executed simultaneously. Typically, the number of two-qubit terms grows with the number of qubits in the problem instance. We therefore expect a considerable performance gain on near-term quantum processors with 10-1000 qubits, as a linear reduction in sequence length avoids exponentially increasing incoherent errors. In summary, this work provides tangible evidence that C-ARB gates open the door to solving more

provide additional degree of freedom to correct for errors.

⁹Note that one of the optimization trajectory starting from random initialization reaches a success probability of 0.69 at $p = 2$. We attribute this difference to fluctuations in coherence times and gate performance

3. SOLVING AN EXACT COVER PROBLEM INSTANCE USING THE QAOA

complex problem instances on quantum computers for as long as the algorithm's performance are limited by decoherence.

Chapter 4

Conclusion

In this thesis, we demonstrated a way to enhance the performance of variational quantum algorithms (VQAs) by reducing the sequence length required to implement their circuit on quantum hardware. In particular, we enlarged the gate-set available on the quantum computer with a controlled arbitrary angle phase gate (C-ARB gate), which enables us to avoid decomposition of higher order operations into multiple gates.

The C-ARB gate is a generalized version of the controlled π -phase two-qubit gate (CZ gate). It is able to reach any phase on the $|11\rangle$ state between 0 and 2π . Our implementation achieves an average process fidelity of 97.7%, whereby the remaining errors are dominated by decoherence and effects of thermal population. The average leakage per gate amounts to 1.4%, which is slightly lower than the leakage we obtain for a CZ gate implemented on the same physical qubits.

We demonstrated the advantage of C-ARB gate on a three qubit exact cover problem instance which we solved using the quantum approximative optimization algorithm (QAOA), a VQA that finds approximate solutions to combinatorial problems. The instance we considered is the first experimental implementation requiring three QAOA layers to be solved with high success probability.

We obtain a 50% two-qubit gate count reduction and a gate sequence length reduction factor of 3 using the direct implementation of the C-ARB gate compared to a decomposition into controlled phase gates (CZ gates) and single qubit gates. Consequently, we achieve a higher success probability with the direct implementation (0.84 versus 0.64) because it executes more layers for a fixed sequence length.

We foresee an even more pronounced advantage for larger-scale experiments because the number of layers required to solve problems typically scales with the number of qubits involved in the experiment. Therefore, for as long

4. CONCLUSION

as quantum devices are limited by decoherence, C-ARB gates will open the door to solving more complex problem instances with VQA.

Chapter 5

Outlook

While C-ARB gates will provide a considerable performance advantage for larger quantum computers in the NISQ era, their up-scaling to larger systems also entail new challenges and open questions.

A first challenge relates to the elaborate C-ARB gate calibration procedure, although a few straight-forward improvements are expected to increase the calibration speed by an order of magnitude (see Section 2.2). While we implemented the gate for fixed-coupling, frequency-tunable transmons using unipolar flux pulses, other gate architectures might provide a higher level of convenience. One promising approach is a gate based on a tunable ZZ-coupling achieved with a frequency tunable coupler qubit [89]. Independently of the flux pulse parametrization, this gate acquires conditional phase linearly as function of interaction time which eases the calibration and potentially reduces interpolation errors. In addition, the acquired dynamic phase changes slowly as a function of other gate parameters, which reduces the number of calibration points and therefore the overall tune-up time of the gate.

Developing rapid and adequate characterization techniques for continuously parametrized gates designed for VQAs is a second aspect arising from this thesis, and is not clearly established in literature. Although quantum process tomography provides full information about the quantum process, it also suffers from leakage, preparation errors and does not offer an intuitive interpretation for the different error sources.

Inspired from Ref. [32], we propose an alternative characterization method for benchmark pairs of subsequent gates, with conditional phases of $\pi - \alpha$ and α respectively. Since the conditional phase of these two gates sums up to π , the combination of the two gates corresponds to a CZ gate which is in the Clifford group. Randomized benchmarking therefore allows to put a bound on the error of the combination of two gates (but not on the individual gates).

5. OUTLOOK

Cross-entropy benchmarking [90] is another approach that can assess the performance of non-Clifford gates and has been used by Ref. [33] for their continuously parametrized gates.

However, both randomized and cross-entropy benchmarking are insensitive to small coherent errors which can accumulate quickly and cause deleterious effects in algorithms with repetitive structures (such as QAOA) [91]. A recent calibration method based on quantum process tomography of long strings of CZ gates alleviate these effects [91] and could be extended to C-ARB gates.

As we progress towards larger experiments, we will have the opportunity to explore more complex problem instances. In Fig. 5.1(b) and (c), we present two exact cover problem instances designed for the device described in Ref. [92], whose hardware connectivity graph is reproduced in Fig. 5.1(a). The former instance respects the connectivity of the device and has two solutions, $\mathcal{L}_1 = \{A, D, G\}$ and $\mathcal{L}_2 = \{B, C, E, F\}$. By contrast, the latter instance represents a dense problem graph that require interactions between qubits which are not physically connected and has a single answer, $\mathcal{L} = \{C, E, F\}$. The implementation thereof requires SWAP-gates which lengthen the gate sequence for each QAOA layer. While the SWAP-gates themselves can be decomposed into CZ gates and single qubit gates, the sequence would be much shorter if the SWAP operations are part of the available hardware gate set. Due to the higher problem graph connectivity and the uniqueness of the solution, the number of layers required to find a good approximate solution is higher for the second problem instance.

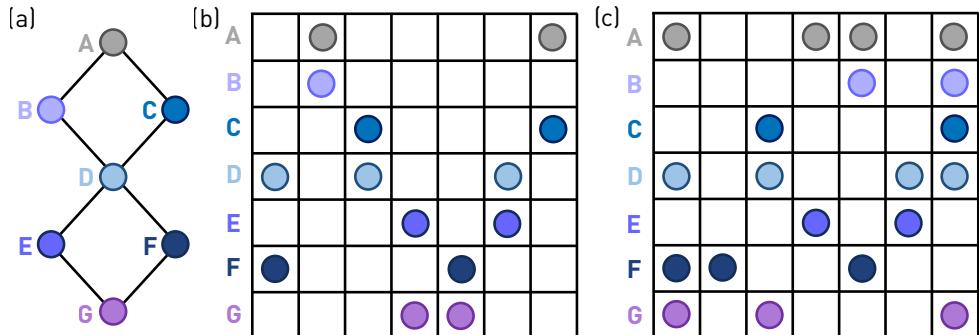


Figure 5.1. Example of future experiments designed for a 7-qubit quantum processor. (a) Hardware graph connectivity of the 7-qubit quantum processor described in Ref. [92]. (b) Incidence matrix of an exact cover problem instance respecting the connectivity of (a). (c) Incidence matrix of an exact cover problem instance requiring full “logical” connectivity and therefore SWAP operations between the physical qubits.

Finally, the use of the C-ARB gate is foreseen to extend beyond QAOA,

with potential applications in quantum chemistry [93] and quantum neural networks [94].

Appendix A

Setup

In this Appendix, we provide additional information with respect to the quantum processor and the experimental setup used for this thesis. In Fig. A.1, we present an optical micrograph of the four-qubit quantum processor, from which we use the first three in this thesis.

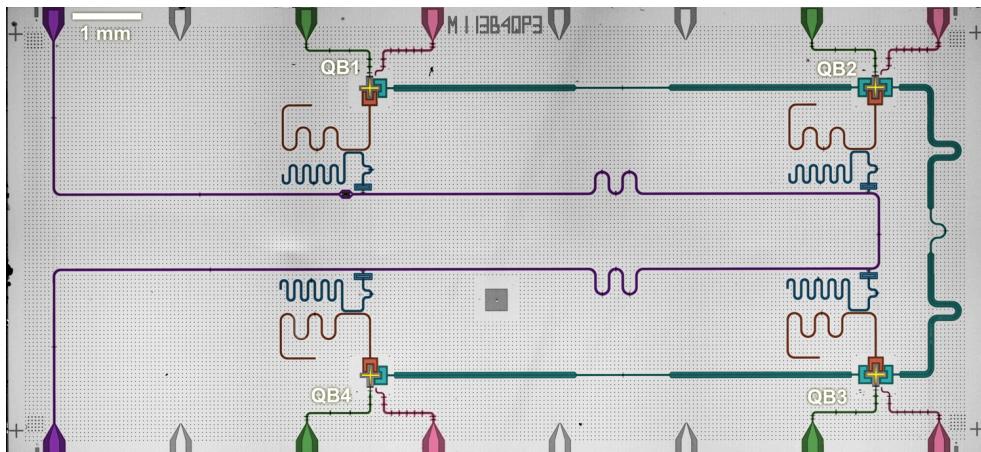


Figure A.1. False-colored optical micrograph of the four-qubit quantum processor. The transmon qubits are colored in yellow, coupling resonators in cyan, flux lines for single-qubit tuning in green, charge lines for single-qubit manipulation in pink, a common feedline for readout in purple, and transmission line resonators for readout and for Purcell filtering in red and blue, respectively. Figure and caption text adapted and reproduced from Ref. [43].

In Fig. A.2, we show the detailed schematics of the the experimental setup with all relevant instruments. We use a ultra-high frequency quantum analyzer (UHFQA) for readout pulse generation and data acquisition, a high definition arbitrary waveform generator (HDAWG) for single qubit pulse generation, the flux arbitrary waveform generator (AWG) for two-qubit gate

A. SETUP

flux pulses generation and a trigger AWG to synchronize all instruments. The readout line consists of a traveling-wave parametric amplifier (TWPA), a high-electron-mobility transistor (HEMT) amplifier, and the warm amplifier board (WAMP). The filters and attenuators are indicated at each temperature stage of the cryogenic system.

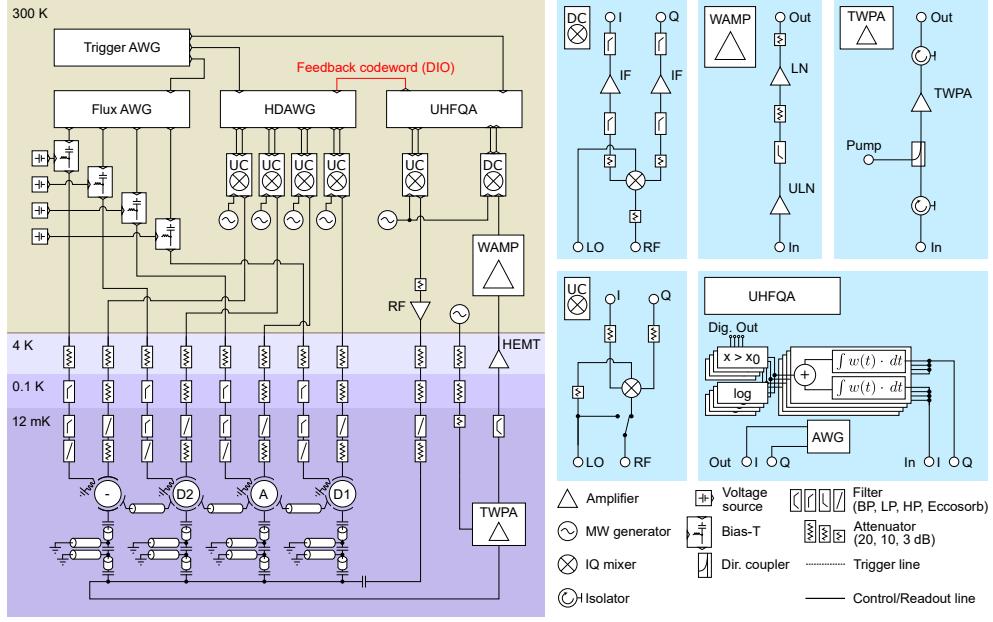


Figure A.2. Detailed schematics of the experimental setup. The filtering of the input and output signals including low-pass (LP), band-pass (BP) and infrared-blocking Eccosorb filters and attenuators are shown at the indicated temperature stages. Extended schematics of the components of the readout chain are shown in the light blue panels. The readout signal is amplified using a traveling-wave parametric amplifier (TWPA), a high-electron-mobility transistor (HEMT) amplifier, and the warm amplifier board (WAMP) consisting of an ultra-low-noise (ULN) amplifier, a low-noise (LN) amplifier and a high-pass (HP) filter. The UHFQA supports weighted integration, data logging (log) and thresholding ($x > x_0$) of the readout signal. Figure and caption text reproduced and adapted from Ref. [42]

In Table A.1, we report the qubit parameters including coherence times, thermal populations and 3-level readout correct assignment probability extracted as detailed in Appendix B.

In Table A.2, we report the two-qubit gate parameters used for the QAOA experiments.

Table A.1: Coherence, frequency, coupling, and readout properties of the three qubit used in this thesis. The T_1 and T_2 times are measured at the maximum qubit frequency for qubit 1, and at the minimum qubit frequency for qubits 3. Qubit 2 is not tunable. RO stands for readout. Parameters followed by an asterisk (*) are reproduced from Ref. [43]. See Ref. [42] for multiplexed readout.

Parameter	Q1	Q2	Q3
Maximum qubit frequency* (GHz)	5.721	5.210	5.530
Minimum qubit frequency* (GHz)	5.083	4.880	4.386
Qubit lifetime (μ s)	27.8	16.3	22.9
Qubit Ramsey coherence time (μ s)	25.4	20.0	12.5
Qubit echo coherence time (μ s)	32.4	22.8	22.4
Readout resonator frequency (GHz)	6.890	7.085	6.686
Purcell resonator – RO resonator detuning* (MHz)	29.5	27.5	19.4
Purcell resonator – RO resonator coupling* (MHz)	10.9	8.2	9.5
Purcell resonator to half-feedline coupling* (MHz)	27.2	34.7	10.7
Effective readout resonator linewidth* (MHz)	2.4	2.0	1.5
Dispersive shift* (MHz)	-3.9	-1.6	-1.8
Thermal population (%)	0.9	0.4	1.4
3-level RO assignment prob. (no preselection) (%)	97.42	96.57	95.71

Table A.2: CZ gate parameters as used for QAOA.

Parameter	Q1-Q2	Q2-Q3
Pulse length (ns)	94	107
Pulse amplitude (V)	0.785	0.229
Gaussian filter standard deviation (ns)	1	1
Buffer before pulse (ns)	15	10
Buffer after pulse (ns)	15	15
Maximum leakage (%)	9	2.5

Appendix B

High fidelity qutrit single-shot readout and reset

As mentioned in Section 1.2, information stored in qubits cannot be copied which prevents redundancy and majority voting as an error correcting scheme. Instead, fault-tolerant quantum computing requires quantum error correction codes [21, 22] such as the surface code [23].

However, standard decoders for the surface code typically assume there is no population leakage outside of the computational subspace spanned by the $|0\rangle$ and $|1\rangle$ states [95]. Unfortunately, most physical implementation of qubits are multi-level systems (see Section 1.2). Consequently, qubit manipulation occasionally results in *leakage*, the phenomenon of populating levels beyond $|0\rangle$ or $|1\rangle$ which are not part of the computational basis.

Detecting and calibrating low leakage gates requires accurate measurement of additional states of the qubit outside its computational basis. This appendix describes how we perform high-fidelity single-shot readout of the first 3 eigenstates of a transmon, respectively named $|g\rangle$ (ground state) or $|0\rangle, |e\rangle$ (first excited state) or $|1\rangle$, and $|f\rangle$ (second excited state) or $|2\rangle$. Higher levels are not considered the scope of this work but are further discussed in [96, 97].

The first section provides a theoretical description of the readout scheme and outlines the steps required to achieve high fidelity readout. In particular, we describe how to find the optimal readout frequency and derive the statistically optimal state discrimination procedure. The next section details the implementation of the scheme within our measurement framework. Thereafter, we provide experimental data achieving a correct state assignment probability of 98.5% and illustrate the scheme on a T_1 measurement of the $|f\rangle$ level. Finally, the last section demonstrates that this high fidelity readout scheme can enable high quality three-level active reset of the transmon to the ground state.

B.1 Three-level readout

The state of a qutrit (the three-level extension of a qubit), $|\psi\rangle$, reads,

$$|\psi\rangle = \alpha_g |g\rangle + \alpha_e |e\rangle + \alpha_f |f\rangle \quad (\text{B.1})$$

and is measured with probability $P_i = |\alpha_i|^2$ in $|i\rangle$ for $i \in \{g, e, f\}$. The probabilities P_g, P_e, P_f are also referred as the ground state, first-excited and second-excited state populations respectively.

To readout the state of a qutrit, we make use of a dispersive interaction between the transmon and a resonator [13, 98] which we probe with a voltage pulse. The output signal, $S_{\text{out}}(t)$, consists of an in-phase real component, $I(t)$, and an imaginary quadrature, $Q(t)$,

$$S_{\text{out}}(t) = I(t) + iQ(t) \quad (\text{B.2})$$

After several stages of amplification, we use the state-dependent time-response of the resonator $S_{\text{out}}(t)$ to discriminate between the ground state, the first excited state and the second excited state.

In average readout, the measurement is repeated many times and averaged to reduce noise. By contrast, the objective of single-shot readout is to construct a function acting on a single time trace $S_{\text{out}}(t)$ to determine the projected state of the qutrit on one of its three eigenstates with high probability.

With the approximation that the qutrit is perfectly prepared in one of its eigenstates, the projected state of the qutrit is assigned,

$$|\psi_{\text{proj}}\rangle = |i'\rangle \quad (\text{B.3})$$

where i' is the most likely state label according to the state assignment model. Repeated eigenstate preparation, measurement and state assignment yields statistics characterizing the readout performance conveniently visualized in the state assignment probability matrix,

$$A = \begin{pmatrix} P(g|g) & P(e|g) & P(f|g) \\ P(g|e) & P(e|e) & P(f|e) \\ P(g|f) & P(e|f) & P(f|f) \end{pmatrix} \quad (\text{B.4})$$

where the element of row i and column j corresponds to $P(s_j|s_i)$, the estimate probability of assigning a prepared state s_i to state s_j . These estimates are computed from the counts of the repeated single-shot measurements.

Under ideal conditions, A coincides with the identity matrix. However, in practice several mechanisms prevent the assignment matrix of reaching identity. For instance, T_1 decay before and during the readout, readout errors

due to finite signal to noise ratio (SNR), thermal population and state preparation errors.

We aim at achieving the highest average three-level readout and therefore choose to optimize the readout for the highest SNR between the 3 states concurrently. This corresponds to minimizing the readout overlap, ε , defined as the averaged sum of the non-diagonal elements of the state assignment probability matrix¹,

$$\varepsilon = \frac{1}{3} \sum_{i=0}^2 \sum_{j \neq i}^2 P(s_j|s_i) \quad (\text{B.5})$$

A is highly dependent on the readout pulse frequency. Therefore, we start by searching the readout frequency minimizing ε .

B.2 Readout frequency optimization

The time-response of the readout resonator depends not only on the qubit state but also the frequency of the readout pulse, f . Hence, we start by preparing the qutrit in $|g\rangle$, (respectively $|e\rangle$, $|f\rangle$) and measuring the time-integrated (with boxcar integration weights [99]) resonator response for a range of frequencies. The amplitude of each response is shown in Fig. B.1(a) as function of the readout pulse frequency.

The dominating noise sources in the readout are the amplifiers, which are independent of the measured state and subject to white noise over the considered frequency range. Hence, we model the noise for each state as frequency- and time-independent Gaussian white noise². Due to the linearity of integration, the noise for each state in the integrated IQ-plane also follows a Gaussian distribution, centered at the average time-integrated resonator response for each frequency. We provide an example for $f = 6.890\text{ GHz}$ (indicated with a dashed, black line in Fig. B.1(a) and b) in Fig. B.1(c) where the three Gaussian distributions are centered at points G , E and F respectively.

The statistically optimal approach to assign a new data point to a state in this plane is a multi-modal Gaussian mixture model (GMM) [100–102]. With this model we compute the posterior probability of a new incoming data point to be produced by each eigenstate and assign it to the state with highest posterior probability [103] (i.e. maximum likelihood state assignment). The decision boundaries are the collection of points in space with equal likelihood for different distributions.

¹Note that specific experiments could benefit from other optimizations criteria. For instance, maximizing the minimum SNR between one state and the other two, which corresponds to maximizing the minimum value of the trace of A .

² T_1 -decay introduces non-Gaussian noise which is not captured by this model. The effects of these phenomenon on the state assignment are discussed in Section B.4.

B. HIGH FIDELITY QUTRIT SINGLE-SHOT READOUT AND RESET

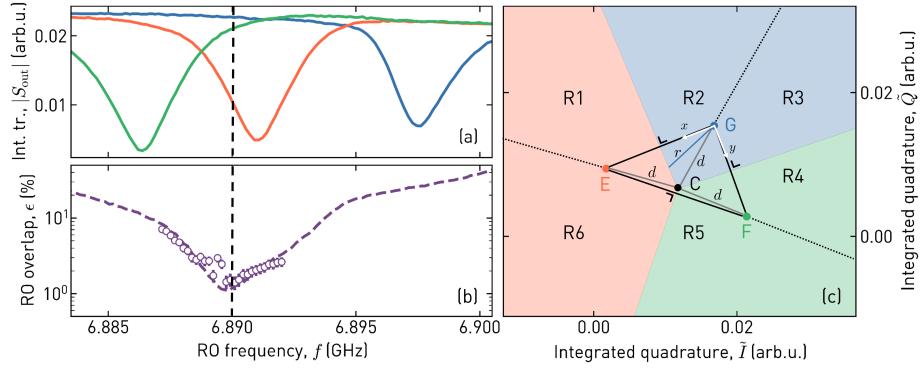


Figure B.1. Readout frequency optimization. (a) Amplitude of the integrated state-dependent resonator response (i.e. transmission) for qubit 2 prepared in $|g\rangle$ (blue), $|e\rangle$ (red), $|f\rangle$ (green). (b) Average readout overlap (in %) as predicted from the model (dashed line) and measured (scatter). (c). Integrated IQ-plane for $f = 6.890$ GHz (dashed line in (a),(b)), the frequency for which the readout overlap is minimal. Points G , E , F correspond to the average integrated resonator responds for states $|g\rangle$, $|e\rangle$ and $|f\rangle$. The shaded background color indicates which state is most likely according to a Gaussian mixture model with means centered in G , E and F , and equal and isometric covariance. R_i indicates different integration regions. C and d indicate the circumcenter and circumradius respectively.

In the case of equal and isometric covariance of the three distributions, for any measured data point, the distribution with the nearest mean (i.e. vertex of the triangle GEF) is the most likely distribution to have generated the data point. In addition, the decision boundaries coincide with the perpendicular bisectors of the triangle GEF ³ (See Fig. B.1(c)).

Hence, we compute the readout overlap ϵ_k consisting of the overlap between the 3 Gaussian distributions (one for each state) at each frequency f_k . The frequency yielding the smallest overlap between the Gaussian distributions results in highest SNR for all three states concurrently and is therefore chosen as readout frequency.

In the next section, we derive an analytical model to find the optimal readout frequency from the average time-integrated resonator responses and discuss the comparison between the model and experimental data shown in Fig. B.1(b).

B.2.1 Analytical derivation

We denote the state assignment probability matrix A_k as function of the readout frequency f_k and define the following variables,

³the perpendicular bisectors of a triangle are the collections of points for which the Euclidian distance to two vertices is equal.

- $\mathcal{N}_i^{(k)}(\mathbf{x}|\boldsymbol{\mu}_i^{(k)}, \sigma^2 \mathbf{I}) = \frac{1}{2\pi|\sigma^2 \mathbf{I}|^{1/2}} \cdot \exp\left(-(\mathbf{x} - \boldsymbol{\mu}_i^{(k)})^T (\sigma^2 \mathbf{I})^{-1} (\mathbf{x} - \boldsymbol{\mu}_i^{(k)})\right)$ the Gaussian distribution modeling measurement noise around state s_i in the two-dimensional integrated IQ-plane at frequency f_k .
- $\boldsymbol{\mu}_i^{(k)}$ the mean of the Gaussian distribution around state s_i at frequency f_k , i.e. point G , E and F in Fig. B.1(c) for states $|g\rangle$, $|e\rangle$, $|f\rangle$ at frequency $f_k = 6.890$ GHz.
- $\sigma^2 \mathbf{I}$ the equal and isometric covariance of the Gaussian distribution.
- \mathbf{x} an arbitrary point $\in \mathbb{R}^2$ (for ease of notation, we translated the complex IQ-plane to a real, two-dimensional plane).

Based on Eq. (B.5), the average readout overlap ε_k at frequency f_k is,

$$\varepsilon_k = \frac{1}{3} \sum_{i=0}^2 \sum_{j \neq i}^2 P_k(s_j|s_i) \quad (\text{B.6})$$

where $P_k(s_j|s_i)$ is the probability of assigning a measurement to state j while it was originally prepared as state i at frequency f_k , with $j \neq i$.

The minimization of the readout overlap is equivalent to a maximization of the trace of A_k over all frequencies, such that finding the optimal readout frequency results in solving,

$$f_{\text{opt}} = \arg \max_{f_k} \text{Tr}(A_k) = \arg \max_{f_k} \sum_{i=0}^2 P_k(s_i|s_i) \quad (\text{B.7})$$

At each frequency, the probability $P(s_i|s_i)$ (we omit k for clarity in the following equations) corresponds to the integral of the Gaussian density function $\mathcal{N}_i(\mathbf{x}|\boldsymbol{\mu}_i, \sigma^2 \mathbf{I})$ over the domain where the density i is higher than the density of any other state. Namely, the area of the two-dimensional IQ-plane in which an incoming point \mathbf{x} will be correctly assigned s_i if it was prepared in s_i . The integration domain are depicted in blue, red, green in Fig. B.1(c), for state $|g\rangle$, $|e\rangle$, and $|f\rangle$ respectively. The optimal readout frequency corresponds to the frequency yielding the largest sum of 3 similar integrals (one for each state)⁴.

For ease of computation, each integration domain is further divided in 2 sub-domains. By symmetry,

$$\int_{R1} \mathcal{N}_0(\mathbf{x}|\boldsymbol{\mu}_0, \sigma^2 \mathbf{I}) d\mathbf{x} = \int_{R2} \mathcal{N}_2(\mathbf{x}|\boldsymbol{\mu}_2, \sigma^2 \mathbf{I}) d\mathbf{x} = I_1 \quad (\text{B.8a})$$

⁴For the readout of a two-level system, there are only two integrals and the approximation of equal and isometric covariance allows to ignore one more dimension using an appropriate rotation. The problem reduces to minimizing the overlap between two, one-dimensional Gaussian distributions, which is inversely proportional to the distance between the two means. Hence, the typical choice of readout frequency is the frequency yielding the largest difference between the two state responses [104].

B. HIGH FIDELITY QUTRIT SINGLE-SHOT READOUT AND RESET

$$\int_{R3} \mathcal{N}_2(\mathbf{x}|\boldsymbol{\mu}_2, \sigma^2 \mathbf{I}) d\mathbf{x} = \int_{R4} \mathcal{N}_0(\mathbf{x}|\boldsymbol{\mu}_0, \sigma^2 \mathbf{I}) d\mathbf{x} = I_2 \quad (\text{B.8b})$$

$$\int_{R5} \mathcal{N}_2(\mathbf{x}|\boldsymbol{\mu}_2, \sigma^2 \mathbf{I}) d\mathbf{x} = \int_{R6} \mathcal{N}_1(\mathbf{x}|\boldsymbol{\mu}_1, \sigma^2 \mathbf{I}) d\mathbf{x} = I_3 \quad (\text{B.8c})$$

such that only 3 of the 6 sub-domain integrals need to be computed⁵,

$$\sum_{i=0}^2 P(s_i|s_i) = 2 \cdot (I_1 + I_2 + I_3) \quad (\text{B.9})$$

We provide a detailed derivation for I_1 . The other two integrals are solved in the similar way but are preceded by an appropriate rotation and translation of the coordinate system.

We start by a shift of coordinate system with the origin at point G and the x -axis aligned with the line segment $|GE|$ such that $\mu_0 = 0$ (see white coordinate system in Fig. B.1(c)), followed by a polar coordinate transformation. In this parametrization, the radius of integration r is a function of θ , the angle between the x -axis of the coordinate system centered in point G and r . Namely,

$$r(\theta) = \begin{cases} m/\cos\theta, & \text{if } -\pi/2 < \theta < \pi/2. \\ +\infty, & \text{otherwise.} \end{cases} \quad (\text{B.10})$$

with m being the distance between G and the midpoint of (G, E) , and assuming $\theta \in [-\pi, \pi]$. The integration bounds of θ are $\pi - \gamma$ and γ where $\gamma = \arccos d/m$.

Eq. (B.8a) becomes in the new coordinate system

$$I_1 = \int_{R2} \frac{1}{2\pi\sigma^2} e^{-\frac{\mathbf{x}^\top \mathbf{x}}{2\sigma^2}} d\mathbf{x} \quad (\text{B.11})$$

and in polar coordinates with the corresponding integration boundaries,

$$I_1 = \begin{cases} \int_{\gamma-\pi}^{\gamma} \int_0^{m/\cos\theta} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho d\rho d\theta, & \text{if } -\pi/2 < \theta < \pi/2. \\ \int_{\gamma-\pi}^{\gamma} \int_0^{\infty} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho d\rho d\theta, & \text{otherwise.} \end{cases} \quad (\text{B.12})$$

We note that the inner integral in both cases is the derivative of a Gaussian (up to a constant),

$$\int_0^a \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\rho^2}{2\sigma^2}\right) \rho d\rho d\theta = \frac{1}{2\pi} \left(1 - \exp\left(\frac{-a^2}{2\sigma^2}\right)\right) \quad (\text{B.13})$$

⁵Note that additional weighting factor to the 3 integrals in this equation allow to favor SNR of 1 or 2 of the 3 states.

and therefore simplify Eq. (B.12) to

$$I_1 = \begin{cases} \int_{\gamma-\pi}^{\gamma} \frac{1}{2\pi} \left(1 - e^{-\frac{m^2}{2\sigma^2 \cdot \cos^2 \theta}} \right) d\theta, & \text{if } -\pi/2 < \theta < \pi/2. \\ \int_{\gamma-\pi}^{\gamma} \frac{1}{2\pi} d\theta, & \text{otherwise.} \end{cases} \quad (\text{B.14})$$

Since $\int_{\gamma-\pi}^{\gamma} \frac{1}{2\pi} d\theta$ is independent of both θ , it results in a constant function integrated over π . This simplifies I_1 further,

$$I_1 = \frac{1}{2} - \begin{cases} \int_{\gamma-\pi}^{\gamma} e^{-\frac{m^2}{2\sigma^2 \cdot \cos^2 \theta}} d\theta, & \text{if } -\pi/2 < \theta < \pi/2. \\ 0, & \text{otherwise.} \end{cases} \quad (\text{B.15})$$

The 0 values for θ prevent the remaining integral to have a closed form solution, but it can easily be computed numerically.

An analogous development holds for I_2 and I_3 for each frequency, the average overlap yields,

$$\varepsilon_k = 1 - \frac{2}{3} \cdot \sum_{i=1}^3 I_{i,k} \quad (\text{B.16})$$

and the optimal readout frequency is extracted as defined in Eq. (B.7),

$$f_{\text{opt}} = \arg \max_{f_k} \sum_{i=1}^3 I_{i,k} \quad (\text{B.17})$$

In Fig. B.1(b), we show the readout overlap $\varepsilon = \varepsilon \cdot 100\%$ computed with Eq. (B.16) for the integrated resonator responses of Fig. B.1(a) with a dashed, purple line. The minimum average overlap is 1% at $f_{\text{opt}} = 6.890$ GHz (dashed black line).

B.2.2 Comparison to experimental data

We validate the model by experimentally sweeping the readout frequency, recording 20000 single-shot measurements for each eigenstate, and assigning the shots to reconstruct A_k near the frequency of interest. The total (average) experimental error is $1 - \text{Tr}(A_k)/3$. However, the latter includes not only readout overlap errors, but also errors originating from T_1 -decay and thermal population. Therefore, we use preselection on the single-shot measurements to mitigate the effect of thermal population (see Appendix A). To account for errors expected from T_1 -decay for levels $|e\rangle$ and $|f\rangle$, we subtract $1 - \exp(-t/T_1^{ge})/3$ and $1 - \exp(-t/T_1^{gf})/3$ from the total error, where t is half the readout length (200 ns). We use $T_1^{ge} = 27.7\ \mu\text{s}$ which we measured prior to the single-shot measurements and thereby also infer

$T_1^{ef} \approx 27.7/\sqrt{2}\mu\text{s}$. The estimated readout overlap errors, which we show in scattered points in Fig. B.1(b), are in good agreement with the model.

To further reduce the errors caused by readout overlap, we describe in the next section the choice of mode-matched integration weights that linearly increase the SNR. Note that, since the integration is linear, the readout frequency found in this section remains optimal with mode-matched integration weights.

B.3 Mode-matched integration weights

With the readout frequency fixed, we define the mode-matched integration weights for single-shot readout. In a two-level system, taking the complex conjugate of the difference between the average ground and excited state responses of the readout pulse as mode-matched filter coefficients is shown to provide near-optimal filter efficiency [99, 104, 105]. The mode-matched weighted integration of the time trace $S_{\text{out}}(t)$ is then:

$$x_1 = \int_0^T w_1(t) \cdot S_{\text{out}}(t) dt \quad (\text{B.18})$$

where $w_1(t)$ are the mode-matched weights described above. The index specifies that this is the complex weights corresponding to the *first* integration unit. A second integration unit is defined to extend this scheme to a three-level system [102]. More specifically, the second integration unit and corresponding weights, $w_2(t)$, are chosen such that $\{w_1(t), w_2(t)\}$ is a basis spanning the hyper-plane defined by the average response of the three states. Similarly to the two-level case, $w_2(t)$ corresponds to the complex conjugate of the difference between the average ground and *second* excited state responses.

These two sets of complex weights are then uploaded on the UHFQA to perform real-time integration of the time traces, such that each single-shot measurement results in a two-dimensional data point (one for each integration unit), x , where the integral is replaced by a sum,

$$x_i = \sum_{t=0}^T w_{i,t} \cdot S_{\text{out},t} \text{ for } i \in \{1, 2\} \quad (\text{B.19})$$

and the time interval is the sampling rate of the UHFQA (2.4 GHz).

B.4 Characterization of three-level single-shot readout

Three-level single-shot readout has been previously been demonstrated on transmon qutrits [106, 107]. This section demonstrates the ability to perform

single-shot readout on a transmon with a correct state assignment probability of 98.52%⁶. To the best of our knowledge, this exceeds other correct state assignment probabilities reported in literature for transmon qutrits. In addition, we describe how to use the readout scheme to characterize the three-level populations in any measurements and illustrate it on the T_1 measurement of the $|f\rangle$ -level.

B.4.1 Single-shot readout characterization

We use the following calibration routine to assess the performance of the qutrit readout:

1. Calibrate parameters of the R_x^π rotation pulse to prepare the qubit in state $|e\rangle$.
2. Calibrate parameters of the second excitation pulse to prepare the qubit in state $|f\rangle$.
3. Measure average, time-dependent readout responses of the resonator for each eigenstate with the readout pulse frequency f_{opt} obtained as described in Section B.2. The complex conjugate difference in IQ-plane between the ground and first excited state time-traces is taken as integration weights for the first integration unit. The second is defined by using the Gram-Schmidt algorithm [108] on the complex conjugate difference between the ground and second excited state to find an vector, ortho-normal to the first set of integration weights, and spanning the hyperspace of the 3 eigenstates.
4. Upload the two sets of integration weights, $\{w_1, w_2\}$, to the UHFQA for real-time integration.
5. Prepare each eigenstate 50000 times⁷, which after real-time integration results in a two-dimensional data point per shot.
6. Fit a GMM to the dataset of integrated shots using the expectation maximization algorithm [100, p. 434-439] with the constraint of equal covariance.
7. Assess the readout by constructing the state assignment probability matrix.

We apply this procedure to qubit 1 of the device described in Appendix A and present the results in Fig. B.2. The corresponding mode-matched integration weights, as well as the readout parameters optimization are reported in [103, Appendix B].

⁶For the comparison between three-level and two-level readout performance on the same qubit, see [103].

⁷Empirically found to be a number of shots achieving low statistical estimation errors.

B. HIGH FIDELITY QUTRIT SINGLE-SHOT READOUT AND RESET

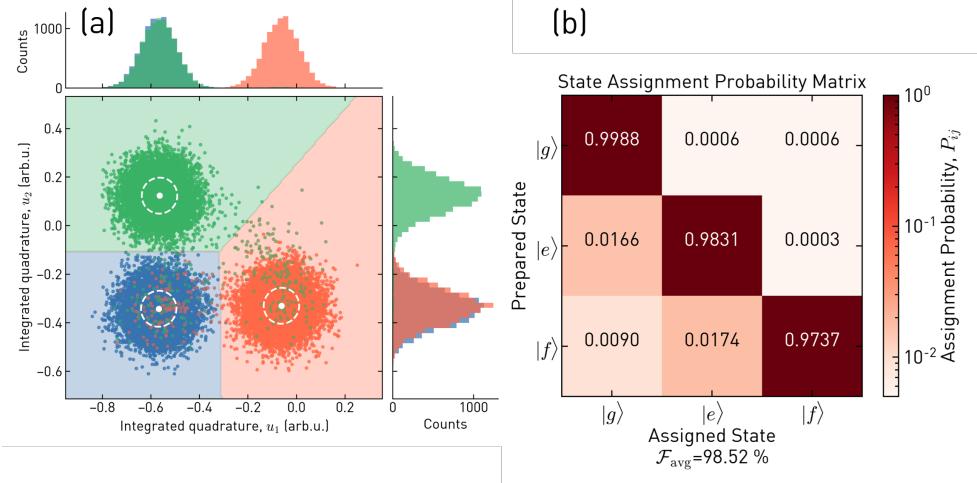


Figure B.2. Statistically optimal state assignment with mode-matched integration weights. (a) First 10000 of 50000 single-shot measurements for each eigenstate preparation of qubit 1. Preselection was performed to mitigate thermal population (see Appendix A for thermal population estimates). The boundaries correspond to the decision boundaries of the fitted GMM. The white points correspond to the position of the means of each Gaussian component and the circle to the standard deviation. In addition, marginal histogram distributions are displayed right and above the main plot. (b) State assignment probability matrix corresponding to the 3x50000 single-shot measurements presented in (a), after pre-selection. The rows correspond to the prepared states and the columns to which state the prepared eigenstates were assigned using the GMM.

Predicting the most likely state on the entire 2D-plane results in the decision boundaries indicated by the background color of the experimental data in Fig. B.2(a). These one-dimensional boundaries can be interpreted as the natural extension of the threshold (zero-dimensional) used in a two-level system for assigning a shot to the ground or excited state.

We show the corresponding state assignment probability matrix, A , in Fig. B.2(b). The average three-level correct state assignment probability is defined as

$$\mathcal{F}_{\text{avg}} = \frac{1}{3} \text{Tr}(A) \quad (\text{B.20})$$

For the presented run, it amounts to 98.52%. Notable observations for each matrix row:

- $|g\rangle$: Preselection procedure removes the majority of the thermal population (on the order of 1%, see Appendix A). We estimate 0.05% measurement-induced transition by counting the number of ground-state data points further than 5 standard deviations away from the

B.4. Characterization of three-level single-shot readout

mean⁸. remaining errors are caused by the overlap between the distributions.

- $|e\rangle$:
 - The assignment of a $|g\rangle$ state for a $|e\rangle$ preparation is dominated by T_1 -decay. Given a T_1 of $25.11\ \mu s$ (last recorded T_1) 1.11% of the prepared $|e\rangle$ shots decay during the first 280 ns of the readout⁹. This number explains an important fraction of the observed 1.66%. The overlap of distributions due to finite SNR is another source of errors.
 - The assignment of $|f\rangle$ state for a $|e\rangle$ likely arise from measurement induced transitions, amounting to approximately 0.02%. Errors originating from readout overlap errors likely explain other errors.
- $|f\rangle$: similarly, misclassification of the $|f\rangle$ state is dominated by T_1^{ef} decay. With a T_1^{ef} of $\sim 14.4\ \mu s$, 1.93% of the initial $|f\rangle$ population are expected to decay to $|e\rangle$ or $|g\rangle$. The largest fraction thereof is expected to be classified as $|e\rangle$, while a smaller fraction is expected to have decayed twice during the measurement. The remaining 0.7% are partially explained by readout overlap errors and preparation errors (which are expected to be higher than $|e\rangle$ -level preparation errors because the $|f\rangle$ -level preparation requires two pulses).

These numbers are summarized in Table B.1. Overall, the error mechanism dominating the state assignment probability matrix is T_1 - and T_1^{ef} -decays, which explain on average 1.01% of the 1.48% errors. The total readout overlap error of the Gaussian components, computed using the method described in B.2.1, amounts to 0.072%. Estimated measurement-induced transitions explain on average 0.07%. Remaining errors amount to 0.33%, which we suspect are related to fluctuations in T_1 and T_1^{ef} as measurement-induced transitions are extremely small on this qubit. This analysis also does not take into account errors originating from leakage into the $|h\rangle$ (or higher) state(s). These errors can lead to misinterpretation of misclassifications of the $|f\rangle$ -level in the observed subspace of states $\{|g\rangle, |e\rangle, |f\rangle\}$, depending on where those higher states are projected in the integrated plane. We do observe such effects when optimizing the readout amplitude (see [103, Appendix B, Fig. 11]). However, even if not completely negligible, these errors do

⁸Decision boundaries are approximately 3.3 standard deviations away from the mean. The probability that an error is caused by readout overlap is much smaller than by measurement-induced excitation at a distance of 5 standard deviation from the mean.

⁹The entire readout is 400 ns long but only decay events before the first 280 ns are expected to be misclassified. This corresponds to the time at which 50% of the information (i.e. integral of the absolute value of mode-matched weights) to distinguish $|g\rangle$ and $|e\rangle$ is acquired

B. HIGH FIDELITY QUTRIT SINGLE-SHOT READOUT AND RESET

not impact directly the subspace of interest spanned by $\{|g\rangle, |e\rangle\}$. Indeed, they mostly occur due to measurement-induced transitions on prepared $|f\rangle$ -states, which we do not use directly when running quantum algorithms in the computational subspace.

Table B.1: Error mechanisms for the measurement displayed in Fig. B.2.

Error mechanism	Error (%)
T_1 and T_1^{ef} -decay	1.01
Readout overlap	0.07
Measurement-induced transitions	0.07
Total explained errors	1.15
Total errors	1.48

This analysis suggests that improving the qubits lifetime and/or increasing the readout speed is the most effective way to further increase the correct state assignment probability. In cases where the overlap errors are higher, it might be worth exploring bi-chromatic readout pulses [89]. Indeed, this method could help increasing the SNR between the 3 eigenstates, thereby reducing overlap errors. It could also help reducing the readout amplitude, which in turn can contribute reducing measurement-induced leakage. It remains unclear whether the improvements this readout pulse may provide are significant and worth the overhead of calibrating additional parameters.

In summary, we have demonstrated the ability to perform three-level, single-shot readout with an average correct assignment probability superior to 98% on a transmon qutrit. The assignment errors are dominated by decay events during the 400 ns readout originating from the finite lifetime of the $|e\rangle$ and the $|f\rangle$ level. In the next section, we detail how to monitor the populations of the first three basis states of a transmon with this readout scheme for arbitrary measurements and illustrate the scheme on a T_1 -measurement of the $|f\rangle$ level.

B.4.2 Using the three-level readout

After calibration on reference single-shot readout measurements, the GMM enables the assignment of state labels to single-shots in any measurement. The fraction of single-shot assigned to each of the state $|g\rangle$, $|e\rangle$ and $|f\rangle$ yield (measured) estimates $\mathbf{P}^m = (P_g^m, P_e^m, P_f^m)^\top$ for the populations $\mathbf{P} = (P_g, P_e, P_f)^\top$ of Eq. (B.1) [107].

However, as discussed in Section B.4.1, the state assignment is sensitive to readout imperfections and therefore yields biased population estimates.

B.4. Characterization of three-level single-shot readout

For instance, a pure $|e\rangle$ state measurement should result in $\mathbf{P}^m = (0, 1, 0)^\top$ but might instead yield $\mathbf{P}^m = (0.017, 0.983, 0.0)^\top$, just as in the calibration measurement. To account for these imperfections, we use a latent variable model [109] to infer the (corrected) qutrit populations \mathbf{P}^c relative to the calibration measurement,

$$A^\top \cdot \mathbf{P}^c = \mathbf{P}^m \quad (\text{B.21})$$

where \mathbf{P}^c are the latent (corrected) qutrit populations (3x1), \mathbf{P}^m are the measured qutrit populations and A is the reference state assignment matrix obtained during calibration. To infer \mathbf{P}^c , we multiply the measured population by the transpose inverse of the reference state assignment probability matrix,

$$\mathbf{P}^c = (A^\top)^{-1} \cdot \mathbf{P}^m \quad (\text{B.22})$$

where $(A^\top)^{-1}$ can be seen as a linear filter acting on \mathbf{P}^m .

Note that this procedure assumes that A stays constant over time, which is not evidently fulfilled. For instances, fluctuations in T_1^{ge} and T_1^{ef} directly induce fluctuations in A , which can lead to over- or under-correction of the populations. Therefore, a reference matrix A must be recorded at the end of each measurement sequence, similarly to $|g\rangle$ and $|e\rangle$ state calibration points typically used in Rabi oscillation measurements.

In Fig. B.3, we display the time evolution of the measured qutrit populations for a T_1 -measurement of the $|f\rangle$ -level. The scattered points, which correspond to the measured populations \mathbf{P}^m , are in excellent agreement with the rate-equation evolution shown in solid lines. The rate-equation accounts for thermal population and its only free parameter is T_1^{ef} , for which we find a value of $14.96 \pm 0.09 \mu\text{s}$.

At short time scales, the qutrit is in the second excited state and progressively decays into the first excited state. P_e^m first grows exponentially as P_f^m decays with the time constant of T_1^{ef} . After approximately $20 \mu\text{s}$, the decay from the $|e\rangle$ -state to the $|g\rangle$ -state with time constant T_1^{ge} becomes more important than the remaining incoming population from the $|f\rangle$ -level. This results in a decrease of the $|e\rangle$ -level population at larger time scales. The $|g\rangle$ -state population is zero at the start of the measurement and increases progressively as the time evolving population of the $|e\rangle$ -level decays into the $|g\rangle$ -state. The population dynamics in this measurement illustrate that obtaining a good estimate for T_1^{ef} is more intricate without three-level readout because there is no way easy way to detect the cascaded decay without monitoring all states.

In the next section, we describe another application of the three-level readout scheme, namely the active reset of a qutrit to its ground state.

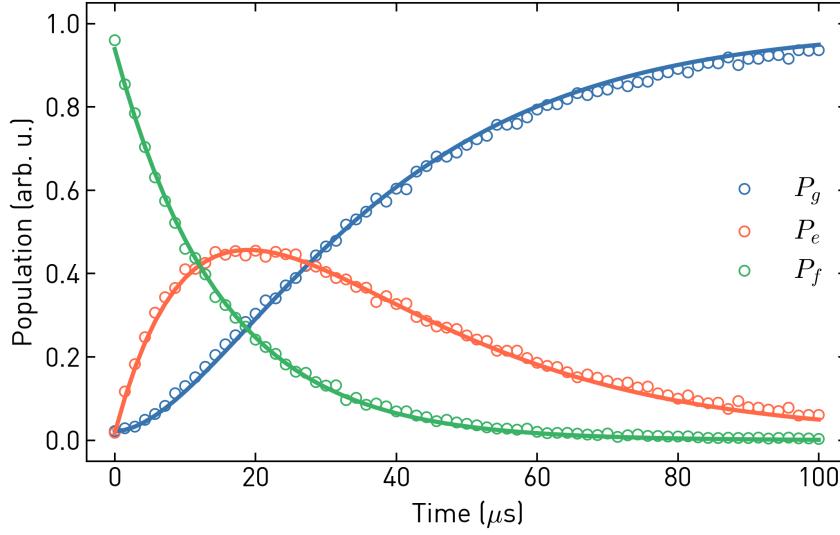


Figure B.3. Illustration measurements of the three-level readout scheme on qubit 1. T_1 measurement of the $|f\rangle$ -level. Time evolution of measured population for each eigenstate (scattered) and population extracted from a rate-equation (solid lines) that accounting for thermal population. We extract a $|f\rangle$ -level lifetime of $14.96\ \mu\text{s}$.

B.5 Three-level, measurement-based active reset

Low-latency analysis of quantum states is key for many quantum algorithms, including quantum error correction [110], quantum teleportation [111], quantum deep learning [94] and active qubit reset [107, 112]. These algorithms employ the (assigned) projected state to condition further operations in the quantum circuit. To ensure success of the algorithm, the analysis and feedback operation must occur much faster than the coherence time of the system.

In this section, we adapt the three-level readout introduced in Section B.1 to perform low-latency state assignment as required for feedback experiments. Next, we use the three-level state assignment in a feedback loop with a transmon to actively reset a qutrit to its ground state.

B.5.1 Setup and procedure for low-latency analysis of quantum states

We show simplified schematics of our experimental setup in Fig. B.4(a) (see Appendix A for detailed schematics). We use Zurich Instruments' UHFQA for readout combined with Zurich Instruments' HDAWG for drive and feedback pulses generation.

B.5. Three-level, measurement-based active reset

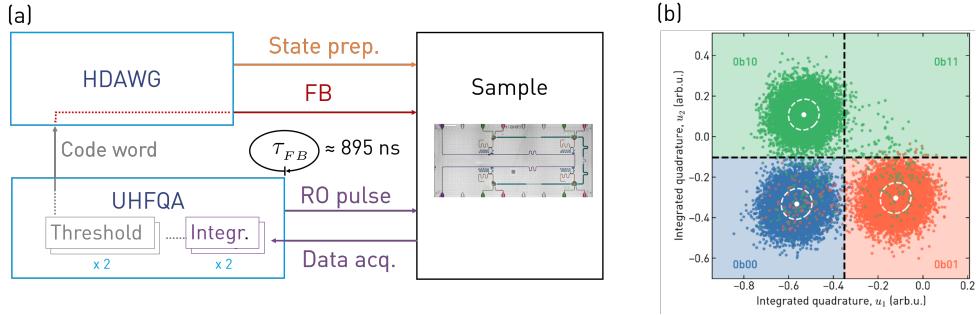


Figure B.4. (a) Simplified schematics of the experimental setup: the HDAWG is used for pulse generation and the UHFQA for readout pulse generation, data acquisition, integration, state assignment, and code word generation. (b) Calibration measurement of the three-level, single-shot readout on qubit 1 to determine the threshold required by the UHFQA for real-time state assignment. The two thresholds are indicated with black, dashed lines. The background color corresponds to the state associated with the code word for each region. The corresponding code words for each regions are displayed in the corners. Blue, red, green correspond to states $|g\rangle$, $|e\rangle$ and $|f\rangle$ respectively. The white dot and circle correspond to the mean and standard deviations of each distribution.

A feedback experiment typically consists of four steps: state preparation, readout, state assignment and feedback operation(s). First, we send microwave pulses generated by the HDAWG to prepare the quantum state. Next, we send a readout pulse and record the transmission line response with the UHFQA. As described in Section B.3, we integrate¹⁰ the response with mode-matched integration weights yielding a two-dimensional data point for each single-shot measurement. We assign the point to a state in about 400 ns with predefined thresholds on each integration unit axis, as pictured in Fig. B.4(b). We use a decision tree classifier [113] of depth 2 fitted with the CART algorithm [113] to find thresholds on each axis that minimizes the Gini impurity [100, 113]. In the limit of average SNR larger than 7, the expected difference in average correct state assignment probability between decision tree classifier and a GMM is on the order of 1% (see [103] for more details about when this comparison is valid). For the instance presented in Fig. B.4(b), the decision tree classifier in fact outperforms the GMM by 1% because it exploits the non-Gaussian distribution of decaying events. The UHFQA generates binary a two-bit code word (one for each integration channel) defining the location of the data point: the bit is zero if the data point is smaller than the threshold on that axis and 1 if it is larger. The code word is sent over a digital input/output (DIO) port to the HDAWG that sends different pulses to the device depending on the code

¹⁰The UHFQA includes a field programmable gate array (FPGA) that integrates the response in real-time.

word value.

B.5.2 Demonstration of three-level active reset

We use this experimental framework to actively reset a qutrit to its ground state. We characterize the reset procedure for a qutrit prepared in each of the three eigen states with the pulse scheme presented in Fig. B.4(a). We prepare the qutrit in $|g\rangle$, $|e\rangle$, or $|f\rangle$ and then perform N cycles of readout, state assignment and feedback pulses. For each cycle, no pulse is sent to the qutrit if the assigned state is $|g\rangle$. By contrast, if the assigned state is $|e\rangle$ ($|f\rangle$), single qutrit control pulse(s) R_{ge}^π (R_{ef}^π and subsequently R_{ge}^π) are employed to bring the qutrit back to $|g\rangle$. We perform a last readout after the N cycles to characterize the effect of the last feedback pulse.

Each cycle takes approximately 895 ns; 350 ns for the readout, 440 ns for state assignment, code word generation and logical branching in the HDAWG, 100 ns to apply the feedback operations and 5 ns buffer before the next readout. We fix the time-window reserved for the feedback operations to 100 ns (which corresponds to the longest possible feedback sequence: two, 50 ns long single qutrit gates) independently of the assigned state because the state is known only at run time but all triggers are compiled before the start of the measurement.

We present the evolution of the readout-corrected populations $P_{g,e,f}^c$ over time after state preparation in Fig. B.5(b)-(d). For reference, we perform the exact same measurement without reset pulses and display the corresponding populations $P_{g,e,f}^{NR}$ with translucent colors. We define the excited state population, $P_{\text{exc}} = 1 - P_g$ and indicate with a red dashed-line the average, steady-state excited population when no reset is applied,

$$\langle P_{\text{exc, ss}}^{NR} \rangle = \frac{1}{N+1} \sum_{n=0}^N 1 - P_{\text{exc},n}^{NR} \quad (\text{B.23})$$

where $P_{\text{exc},n}^{NR}$ is the excited state population at the n -th readout for a qutrit prepared in $|g\rangle$, and $N+1$ is the total number of readouts.

For a qutrit prepared in $|g\rangle$, about 99% of the population is measured in $|g\rangle$ at $t = 0$ (i.e. before applying the first feedback pulse), see Fig. B.5(b). The residual excited population originating from thermal excitation amounts to approximately 1% and is predominantly in the $|e\rangle$ -state. After a single feedback cycle, the residual excited population is on the order of 1%, i.e. an order of magnitude smaller, and stays approximately constant thereafter.

Similarly, a qutrit prepared in $|e\rangle$ starts with approximately 99% of the population in $|e\rangle$, but requires 2 feedback cycles (1.8 μ s) to reduce the residual excited population below $\langle P_{\text{exc}}^{NR} \rangle$ and 3 cycles (2.7 μ s) to reach a steady-state

B.5. Three-level, measurement-based active reset

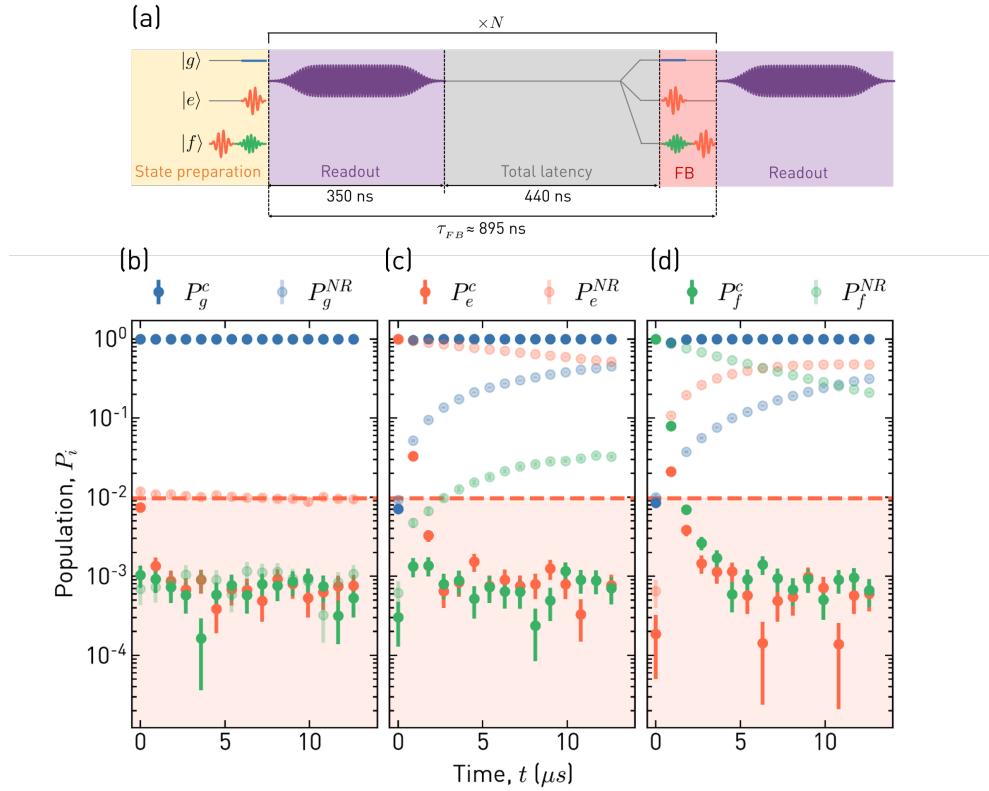


Figure B.5. (a) Pulse scheme used to test the active reset protocol. We prepare a qutrit in one of its eigenstate and subsequently apply N feedback cycles consisting of a readout pulse, a waiting time to assign a state to the acquired data and generate a code word for the HDAWG, and a feedback sequence to reset the qutrit to the ground state dependent on the code word. Finally, we apply a last readout pulse to evaluate the impact of the last feedback cycle. Single-qubit identity gates I (i.e. wait time) are shown in blue, single-qubit x-rotation in the ge -subspace R_{ge}^π are indicated in red and in red and single-qubit x-rotation in the ef -subspace R_{ef}^π are indicated in green. (b-d) Evolution of the readout-corrected populations $P_{g,e,f}^c$ over time after a state preparation in $|g\rangle$ (b), $|e\rangle$ (c) and $|f\rangle$ (d) of qubit 1. The n -th time step corresponds to the n -th readout of the reset scheme. The reference measurement recording populations $P_{g,e,f}^{NR}$ without reset pulse is shown with translucent colors in each panel. We indicate the average, steady-state excited population $\langle P_{\text{exc}}^{NR} \rangle$ with a dashed red line, and shade the region below this threshold in light red.

of approximately 1.5%, see Fig. B.5(c). This population transfer occurs about 100 times faster than it would through the natural decay of the qutrit (see translucent data points for comparison). The increase in P_f^{NR} is due to the accumulation of measurement-induced transitions. The ability to readout the $|f\rangle$ -state and correct for these excitation results in a constant $|f\rangle$ -state population of approximately 1% when the feedback is activated.

B. HIGH FIDELITY QUTRIT SINGLE-SHOT READOUT AND RESET

Finally, we also demonstrate the ability to reset a qutrit prepared in $|f\rangle$ below $\langle P_{\text{exc}}^{\text{NR}} \rangle$ with 3 feedback cycles ($2.7 \mu\text{s}$), see Fig. B.5(d).

In Fig. B.6, we present the residual excited state population for a qutrit prepared in $|e\rangle$ and $|f\rangle$, which we fit to a exponential decaying model to extract the reset rate, Γ , and the residual, steady-state excited population $P_{\text{exc}}^{\text{ss}}$,

$$P_{\text{exc}} = P_{\text{exc}}^{\text{ss}} + (1 - P_{\text{exc}}^{\text{ss}}) \cdot e^{-2\pi\Gamma t} \quad (\text{B.24})$$

We extract a reset rate of 0.60 MHz (0.41 MHz) and a steady-state, residual excited population of 1.5% (1.7%) for a qutrit prepared in $|e\rangle$ ($|f\rangle$).

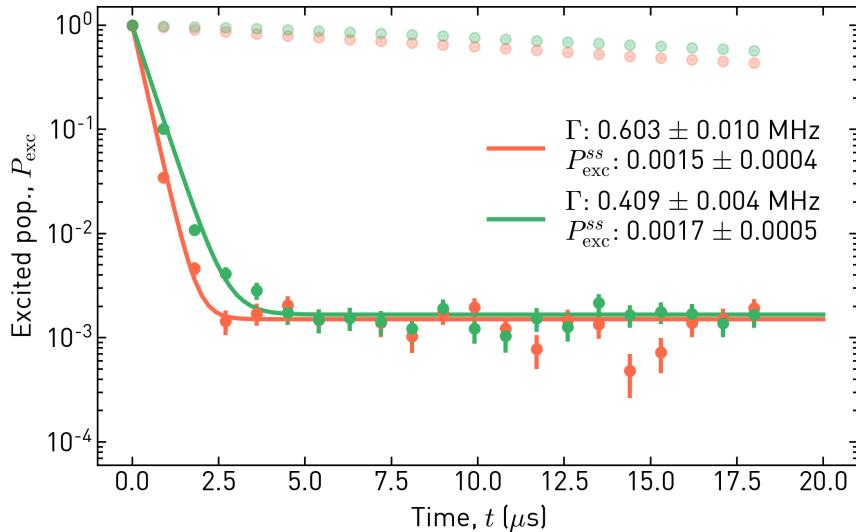


Figure B.6. Excited state population for a qutrit prepared in $|e\rangle$ (red) and $|f\rangle$ (green). The solid lines correspond to a fit to Eq. (B.24) to extract the reset rate and the residual, steady-state excited population. For reference, the excited state population when no reset is applied is shown in translucent colors.

Comparison to literature

In Fig. B.7, we extend the comparison of different reset protocols presented in Ref. [107] in terms of reset rate and residual, steady-state population to include this work (reset on $|e\rangle$ only, because not all other reset protocols include reset of $|f\rangle$).

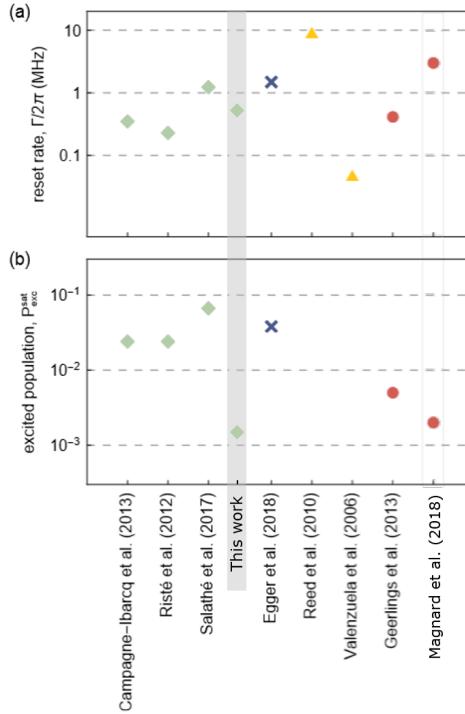


Figure B.7. Comparison of reset rates Γ (a) and residual, steady-state excited populations P_{exc}^{ss} (b) of various implementations of qubit reset protocols. The protocol represented with green squares are, similarly to this work, based on qubit measurement and feedback control. We compare additional protocols based on: sequential π -pulses to a dissipative state (blue cross), qubit frequency tuning via flux pulses (yellow triangles) and all-microwave drive induced dissipation (red circles). Figure and legend reproduced and adapted from [107, Supplementary Material].

Limitations

There are two main limitations preventing higher reset rates. The first is the length of the feedback cycle. Indeed, the rates are inversely proportional to the time interval between two readouts: a reduction of 25% of the feedback cycle yields an increase of 25% of the reset rate, assuming the proportion of states assigned correctly at each time step stays constant (in fact, the proportion grows, see next paragraphs).

The second limitation on a qutrit prepared in $|e\rangle$ is the decay between the second half of the readout pulse and the arrival of the feedback pulse. If the qutrit state was correctly labeled as $|e\rangle$ but has decayed before the feedback pulse is applied, the feedback pulse will set the qutrit state back to $|e\rangle$. We expect 2.5-3% of prepared states to decay during the above-mentioned time interval (615 ns) for T_1 of 20-24 μ s which we typically observe for qubit 1. This percentage is consistent with the 3.3% $|e\rangle$ -state population observed

after 1 feedback cycle for a qutrit prepared in $|e\rangle$.

The same phenomenon limits the reset rate on a qutrit prepared in $|f\rangle$, however, the shorter $|f\rangle$ -level lifetime results in more decay events within the time interval and hence a slower reset rate.

Hence, further increase the reset rate requires a reduction of decay events which can be obtained by an increase in qutrit lifetime, or a reduction of the total latency before the feedback pulses are applied, i.e. a faster readout or an optimized UHFQA firmware capable of assigning states faster.

The main factor limiting the residual, steady-state population is the readout infidelity of the ground state, i.e. the probability of assigning $|e\rangle$ or $|f\rangle$ to a qutrit actually in $|g\rangle$. For qubit 1, the readout infidelity is 1.6 % which is in excellent agreement with the observed $P_{\text{exc}}^{\text{ss}}$ of 1.5 % and 1.7 %.

We can decrease this readout infidelity by shifting the decision thresholds further away from the mean of the ground state distribution to increase the size of the region where a state is labeled as $|g\rangle$ and decrease errors caused by ground state data point located near the decision boundaries. The ultimate population limit then becomes the re-thermalization occurring between the readout and the feedback pulse, which amounts to 0.2-0.4 % for qubit 1 on this device.

Nevertheless, using more conservative thresholds for the ground state comes at a expense of increasing the probability of assigning $|g\rangle$ to a qutrit actually in $|e\rangle$ or $|f\rangle$, which in turn decreases the reset rate.

Conclusion

In this section, we have implemented a three-level, measurement-based active reset protocol yielding a reset rate of 0.60 MHz (0.41 MHz) and a residual, steady-state excited population of 1.5 % (1.7 %) for a qutrit prepared in $|e\rangle$ ($|f\rangle$).

In addition, this reset protocol can be used at the end of any measurement sequence to reset the qutrit to its ground state, and thereby increase the measurement repetition rate by a factor of 10 compared to waiting for the natural decay of the qutrit.

The main limitation of the reset protocol is the delay between readout and the application of the feedback pulse.

Nevertheless, this reset protocol is competitive with respect to other state-of-the-art protocols and further improvement of readout circuitry, as well as electronics firmware optimization potentially allow yet better performance.

Acknowledgments

I wish to express my gratitude to Prof. Dr. Andreas Wallraff and Dr. Christopher Eichler for the opportunity of performing my master thesis at the Quantum Devices Lab, as well as for the continuous feedback on my work. These months have been a wonderful journey during which I acquired a significant amount of knowledge, and met many passionate, friendly and helpful people.

I particularly thank Dr. Christian Andersen for his precious supervision and encouragements throughout my thesis. He introduced me to experiments with quantum computers, guided many of my steps, provided detailed feedback on my writings, and always found time to answer my countless theoretical and conceptual questions. The achievements of this thesis would not have been possible without his guidance.

I extend my gratitude to Michele Collodo for his insightful remarks on many plots presented in this thesis, for his detailed feedback on the entire manuscript of this thesis, and for the myriad of quantum physics discussions.

I appreciate the valuable contribution of Dr. Christoph Hellings for the QAOA project since he joined in November 2019, which include but are not limited to improvements of the QAOA measurement code, insights in understanding QAOA landscape symmetries and tuning up the 7-qubit device in preparation of larger scale QAOA experiments.

I thank Ants Remm and Stefania Lazar for the considerable amount of their time they spent to introduce me to pycQED, the measurement framework used for these experiments. I also acknowledge the rest of the quantum computing team for their valuable input during Monday team meetings.

From the SuperQNet team, I particularly thank Paul Magnard and Kevin Reuer for relevant discussions on three-level, single-shot readout and active reset.

ACKNOWLEDGMENTS

I extend my thanks to Carlotta Ruppert, Bruno Lacroix, Gauthier Muguerza and Vadim Lacroix for their proofreading and feedback of the manuscript of this thesis.

Finally, I extremely grateful to Carlotta Ruppert and my family for their continuous support throughout this thesis.

Bibliography

- [1] R. P. Feynman, "Simulating physics with computers", *International Journal of Theoretical Physics*, vol. 21, no. 6-7, pp. 467–488, Jun. 1982, ISSN: 00207748. doi: 10.1007/BF02650179 (cit. on p. 1).
- [2] E. Fredkin and T. Toffoli, "Conservative logic", *International Journal of Theoretical Physics*, vol. 21, no. 3-4, pp. 219–253, Apr. 1982, ISSN: 00207748. doi: 10.1007/BF01857727 (cit. on p. 1).
- [3] D. Deutsch and R. Jozsa, "Rapid Solution of Problems by Quantum Computation", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 439, no. 1907, pp. 553–558, Dec. 1992, ISSN: 1364-5021. doi: 10.1098/rspa.1992.0167 (cit. on p. 1).
- [4] P. W. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer", *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1484–1509, Aug. 1995. doi: 10.1137/S0097539795293172 (cit. on p. 1).
- [5] L. K. Grover, "A fast quantum mechanical algorithm for database search", in *Proceedings of the Annual ACM Symposium on Theory of Computing*, vol. Part F129452, Association for Computing Machinery, Jul. 1996, pp. 212–219, ISBN: 0897917855. doi: 10.1145/237814.237866 (cit. on p. 1).
- [6] D. R. Simon, "On the power of quantum computation", *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1474–1483, Oct. 1997, ISSN: 00975397. doi: 10.1137/S0097539796298637 (cit. on p. 1).
- [7] E. Bernstein and U. Vazirani, "Quantum complexity theory", *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1411–1473, Jul. 1997, ISSN: 00975397. doi: 10.1137/S0097539796300921 (cit. on p. 1).

BIBLIOGRAPHY

- [8] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, "Strengths and weaknesses of quantum computing", *SIAM Journal on Computing*, vol. 26, no. 5, pp. 1510–1523, Jan. 1997, ISSN: 00975397. doi: 10.1137/S0097539796300933 (cit. on p. 1).
- [9] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th ed. Cambridge University Press, 2000, ISBN: 9781107002173 (cit. on pp. 3, 8, 19).
- [10] C. Monroe, D. M. Meekhof, B. E. King, W. M. Itano, and D. J. Wineland, "Demonstration of a fundamental quantum logic gate", *Physical Review Letters*, vol. 75, no. 25, pp. 4714–4717, Dec. 1995, ISSN: 00319007. doi: 10.1103/PhysRevLett.75.4714 (cit. on p. 4).
- [11] D. Loss and D. P. DiVincenzo, "Quantum computation with quantum dots", *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 57, no. 1, pp. 120–126, Jan. 1998, ISSN: 10941622. doi: 10.1103/PhysRevA.57.120 (cit. on p. 4).
- [12] E. Knill, R. Laflamme, and G. J. Milburn, "A scheme for efficient quantum computation with linear optics", *Nature*, vol. 409, no. 6816, pp. 46–52, Jan. 2001, ISSN: 00280836. doi: 10.1038/35051009 (cit. on p. 4).
- [13] A. Blais, R. S. Huang, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, "Cavity quantum electrodynamics for superconducting electrical circuits: An architecture for quantum computation", *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 69, no. 6, p. 062320, Jun. 2004, ISSN: 10502947. doi: 10.1103/PhysRevA.69.062320 (cit. on pp. 4, 58).
- [14] 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*, vol. 11, no. 1, May 2019 (cit. on p. 4).
- [15] D. Vion, *Quantum Entanglement and Information Processing*, D. Estbve, J. Raimond, and J. Dalibard, Eds. 2003, pp. 443–485, ISBN: 0444517286 (cit. on pp. 4, 15).
- [16] V. Bouchiat, D. Vion, P. Joyez, D. Esteve, and M. H. Devoret, "Quantum coherence with a single cooper pair", *Physica Scripta T*, vol. 76, no. T76, pp. 165–170, 1998, ISSN: 02811847. doi: 10.1238/physica.topical.076a00165 (cit. on p. 4).
- [17] Y. Nakamura, Y. A. Pashkin, and J. S. Tsai, "Coherent control of macroscopic quantum states in a single-Cooper-pair box", *Nature*, vol. 398, no. 6730, pp. 786–788, Apr. 1999, ISSN: 00280836. doi: 10.1038/19718 (cit. on p. 4).

Bibliography

- [18] 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", doi: 10.1103/PhysRevA.76.042319 (cit. on pp. 5, 14).
- [19] W. D. Oliver and P. B. Welander, "Materials in superconducting quantum bits", *MRS Bulletin*, vol. 38, no. 10, pp. 816–825, Oct. 2013, issn: 08837694. doi: 10.1557/mrs.2013.229 (cit. on p. 5).
- [20] W. K. Wootters and W. H. Zurek, "A single quantum cannot be cloned", *Nature*, vol. 299, no. 5886, pp. 802–803, 1982, issn: 00280836. doi: 10.1038/299802a0 (cit. on p. 5).
- [21] D. Gottesman, "An introduction to quantum error correction and fault-tolerant quantum computation", in, Apr. 2010, pp. 13–58. doi: 10.1090/psapm/068/2762145 (cit. on pp. 5, 57).
- [22] D. P. DiVincenzo and P. W. Shor, "Fault-Tolerant Error Correction with Efficient Quantum Codes", *Physical Review Letters*, vol. 77, no. 15, pp. 3260–3263, May 1996. doi: 10.1103/PhysRevLett.77.3260 (cit. on pp. 5, 57).
- [23] A. G. Fowler, M. Mariantoni, J. M. Martinis, and A. N. Cleland, "Surface codes: Towards practical large-scale quantum computation", *PHYSICAL REVIEW A*, vol. 86, p. 32324, 2012. doi: 10.1103/PhysRevA.86.032324 (cit. on pp. 5, 57).
- [24] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. 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. Michelsen, 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*, vol. 574, no. 7779, pp. 505–510, Oct. 2019, issn: 14764687. doi: 10.1038/s41586-019-1666-5 (cit. on pp. 5, 6).
- [25] J. Preskill, "Quantum Computing in the NISQ era and beyond", *Quantum*, vol. 2, p. 79, Aug. 2018, issn: 2521-327X. doi: 10.22331/q-2018-08-06-79 (cit. on p. 6).

BIBLIOGRAPHY

- [26] N. Moll, P. Barkoutsos, L. S. Bishop, J. M. Chow, A. Cross, D. J. Egger, S. Filipp, A. Fuhrer, J. M. Gambetta, M. Ganzhorn, A. Kandala, A. Mezzacapo, P. Müller, W. Riess, G. Salis, J. Smolin, I. Tavernelli, and K. Temme, "Quantum optimization using variational algorithms on near-term quantum devices", Oct. 2017. doi: 10.1088/2058-9565/aab822 (cit. on pp. 6, 7, 33).
- [27] A. Peruzzo, J. McClean, P. Shadbolt, M. H. Yung, X. Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, "A variational eigenvalue solver on a photonic quantum processor", *Nature Communications*, vol. 5, Jul. 2014, issn: 20411723. doi: 10.1038/ncomms5213 (cit. on p. 7).
- [28] E. Farhi, J. Goldstone, and S. Gutmann, "A Quantum Approximate Optimization Algorithm", Nov. 2014 (cit. on pp. 7, 27, 29, 32).
- [29] S. Lloyd, "Universal quantum simulators", *Science*, vol. 273, no. 5278, pp. 1073–1078, Aug. 1996, issn: 00368075. doi: 10.1126/science.273.5278.1073 (cit. on pp. 8, 33).
- [30] A. Lucas, "Ising formulations of many NP problems", *Frontiers in Physics*, vol. 2, pp. 1–14, 2014, issn: 2296424X. doi: 10.3389/fphy.2014.00005 (cit. on pp. 8, 29, 35).
- [31] M. Ganzhorn, D. J. Egger, P. Barkoutsos, P. Ollitrault, G. Salis, N. Moll, M. Roth, A. Fuhrer, P. Mueller, S. Woerner, I. Tavernelli, and S. Filipp, "Gate-Efficient Simulation of Molecular Eigenstates on a Quantum Computer", *Physical Review Applied*, vol. 11, no. 4, p. 044092, Apr. 2019, issn: 23317019. doi: 10.1103/PhysRevApplied.11.044092 (cit. on pp. 9, 10).
- [32] D. M. Abrams, N. Didier, B. R. Johnson, M. P. da Silva, and C. A. Ryan, "Implementation of the XY interaction family with calibration of a single pulse", Dec. 2019 (cit. on pp. 9, 10, 49).
- [33] B. Foxen, C. Neill, A. Dunsworth, P. Roushan, B. Chiaro, A. Megrant, J. Kelly, Z. Chen, K. Satzinger, R. Barends, F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, S. Boixo, D. Buell, B. Burkett, Y. Chen, R. Collins, E. Farhi, A. Fowler, C. Gidney, M. Giustina, R. Graff, M. Harrigan, T. Huang, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, P. Klimov, A. Korotkov, F. Kostritsa, D. Landhuis, E. Lucero, J. McClean, M. McEwen, X. Mi, M. Mohseni, J. Y. Mutus, O. Naaman, M. Neeley, M. Niu, A. Petukhov, C. Quintana, N. Rubin, D. Sank, V. Smelyanskiy, A. Vainsencher, T. C. White, Z. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Demonstrating a Continuous Set of Two-qubit Gates for Near-term Quantum Algorithms", Jan. 2020 (cit. on pp. 9, 10, 50).

Bibliography

- [34] D. C. McKay, S. Philipp, A. Mezzacapo, E. Magesan, J. M. Chow, and J. M. Gambetta, "Universal Gate for Fixed-Frequency Qubits via a Tunable Bus", *Physical Review Applied*, vol. 6, no. 6, p. 064007, Dec. 2016, issn: 23317019. doi: 10.1103/PhysRevApplied.6.064007 (cit. on p. 9).
- [35] 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", *Physical Review Letters*, vol. 113, no. 22, Feb. 2014. doi: 10.1103/PhysRevLett.113.220502 (cit. on p. 9).
- [36] 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*, vol. 4, no. 2, eaao3603, Feb. 2018, issn: 23752548. doi: 10.1126/sciadv.aao3603 (cit. on p. 9).
- [37] J. S. Otterbach, R. Manenti, N. Alidoust, A. Bestwick, M. Block, B. Bloom, S. Caldwell, N. Didier, E. S. Fried, S. Hong, P. Karalekas, C. B. Osborn, A. Papageorge, E. C. Peterson, G. Prawiroatmodjo, N. Rubin, C. A. Ryan, D. Scarabelli, M. Scheer, E. A. Sete, P. Sivarajah, R. S. Smith, A. Staley, N. Tezak, W. J. Zeng, A. Hudson, B. R. Johnson, M. Reagor, M. P. da Silva, and C. Rigetti, "Unsupervised Machine Learning on a Hybrid Quantum Computer", Dec. 2017 (cit. on pp. 9, 10, 39).
- [38] G. Pagano, A. Bapat, P. Becker, K. S. Collins, A. De, P. W. Hess, H. B. Kaplan, A. Kyprianidis, W. L. Tan, C. Baldwin, L. T. Brady, A. Deshpande, F. Liu, S. Jordan, A. V. Gorshkov, and C. Monroe, "Quantum Approximate Optimization with a Trapped-Ion Quantum Simulator", Jun. 2019 (cit. on pp. 9, 10, 37, 40).
- [39] A. Bengtsson, P. Vikstål, C. Warren, M. Svensson, X. Gu, A. F. Kockum, P. Krantz, C. Križan, D. Shiri, I.-M. Svensson, G. Tancredi, G. Johansson, P. Delsing, G. Ferrini, and J. Bylander, "Quantum approximate

- optimization of the exact-cover problem on a superconducting quantum processor”, Dec. 2019 (cit. on pp. 9, 10, 37, 40).
- [40] 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*, vol. 460, no. 7252, pp. 240–244, Jul. 2009. doi: 10.1038/nature08121 (cit. on pp. 10, 11, 15).
 - [41] 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”, *Physical Review Letters*, vol. 91, no. 16, 2003, ISSN: 10797114. doi: 10.1103/PhysRevLett.91.167005 (cit. on p. 11).
 - [42] C. K. Andersen, A. Remm, S. Balasius, S. Krinner, J. Heinsoo, J.-C. Besse, M. Gabureac, A. Wallraff, and C. Eichler, “Entanglement Stabilization using Parity Detection and Real-Time Feedback in Superconducting Circuits”, Tech. Rep., 2019 (cit. on pp. 16, 54, 55).
 - [43] C. K. Andersen, A. Remm, S. Balasius, A. Wallraff, and C. Eichler, “Improved Quantum Device”, *QuSurf Monthly Report*, vol. 33, 2019 (cit. on pp. 16, 53, 55).
 - [44] C. K. Andersen, S. Lazar, and A. Remm, “Sample Report: ATC40 M113 B4QP3”, 2018 (cit. on p. 18).
 - [45] E. Magesan, J. M. Gambetta, and J. Emerson, “Scalable and robust randomized benchmarking of quantum processes”, *Physical Review Letters*, vol. 106, no. 18, May 2011, ISSN: 00319007. doi: 10.1103/PhysRevLett.106.180504 (cit. on p. 18).
 - [46] E. Magesan, J. M. Gambetta, and J. Emerson, “Characterizing quantum gates via randomized benchmarking”, *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 85, no. 4, Apr. 2012, ISSN: 10502947. doi: 10.1103/PhysRevA.85.042311 (cit. on p. 18).
 - [47] I. L. Chuang and M. A. Nielsen, “Prescription for experimental determination of the dynamics of a quantum black box”, *Journal of Modern Optics*, vol. 44-11, no. 12, pp. 2455–2467, 1997, ISSN: 13623044. doi: 10.1080/09500349708231894 (cit. on p. 19).
 - [48] J. F. Poyatos, J. I. Cirac, and P. Zoller, “Complete characterization of a quantum process: The two-bit quantum gate”, *Physical Review Letters*, vol. 78, no. 2, pp. 390–393, Jan. 1997, ISSN: 10797114. doi: 10.1103/PhysRevLett.78.390 (cit. on p. 19).
 - [49] M. Reed, “Entanglement and Quantum Error Correction with Superconducting Qubits”, Nov. 2013 (cit. on p. 19).
 - [50] J. Heinsoo, “Digital quantum computation with superconducting qubits”, PhD thesis (cit. on p. 19).

Bibliography

- [51] B. Schumacher, "Sending entanglement through noisy quantum channels", *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 54, no. 4, pp. 2614–2628, 1996, ISSN: 10941622. doi: 10.1103/PhysRevA.54.2614 (cit. on p. 19).
- [52] C. J. Wood and J. M. Gambetta, "Quantification and characterization of leakage errors", *Physical Review A*, vol. 97, no. 3, Mar. 2018, ISSN: 24699934. doi: 10.1103/PhysRevA.97.032306 (cit. on p. 19).
- [53] M. Baur, "Realizing Quantum Gates and Algorithms with Three Superconducting Qubits", Tech. Rep. (cit. on p. 19).
- [54] R. C. Bialczak, M. Ansmann, M. Hofheinz, E. Lucero, M. Neeley, A. D. OConnell, D. Sank, H. Wang, J. Wenner, M. Steffen, A. N. Cleland, and J. M. Martinis, "Quantum process tomography of a universal entangling gate implemented with Josephson phase qubits", *Nature Physics*, vol. 6, no. 6, pp. 409–413, 2010, ISSN: 17452481. doi: 10.1038/nphys1639 (cit. on p. 19).
- [55] A. N. Korotkov, "Error matrices in quantum process tomography", Sep. 2013 (cit. on p. 21).
- [56] R. Jozsa, "Fidelity for mixed quantum states", *Journal of Modern Optics*, vol. 41, no. 12, pp. 2315–2323, 1994, ISSN: 13623044. doi: 10.1080/09500349414552171 (cit. on p. 22).
- [57] J.-M. Reiner, S. Zanker, I. Schwenk, J. Leppäkangas, F. Wilhelm-Mauch, G. Schön, and M. Marthaler, "Effects of gate errors in digital quantum simulations of fermionic systems", *Quantum Science and Technology*, vol. 3, no. 4, Apr. 2018. doi: 10.1088/2058-9565/aad5ba (cit. on p. 22).
- [58] 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", Jul. 2019 (cit. on p. 22).
- [59] J. Butscher, S. Krinner, and A. Wallraff, "Shaping of Fast Flux Pulses for Two-Qubit Gates Inverse Filtering", PhD thesis, ETH Zurich, 2018 (cit. on p. 22).
- [60] A. Sbihi and R. W. Eglese, "Combinatorial optimization and Green Logistics. 4OR: A", *Quarterly Journal of Operations Research*, vol. 5, no. 2, 2007. doi: 10.1007/s10288 (cit. on p. 27).

BIBLIOGRAPHY

- [61] P. J. Coles, S. Eidenbenz, S. Pakin, A. Adedoyin, J. Ambrosiano, P. Anisimov, W. Casper, G. Chennupati, C. Coffrin, H. Djidjev, D. Gunter, S. Karra, N. Lemons, S. Lin, A. Lokhov, A. Malyzhenkov, D. Masecarenas, S. Mniszewski, B. Nadiga, D. O’Malley, D. Oyen, L. Prasad, R. Roberts, P. Romero, N. Santhi, N. Sinitsyn, P. Swart, M. Vuffray, J. Wendelberger, B. Yoon, R. Zamora, and W. Zhu, “Quantum Algorithm Implementations for Beginners”, Apr. 2018 (cit. on p. 27).
- [62] V. T. Paschos, Ed., *Applications of Combinatorial Optimization*. Hoboken, NJ, USA: John Wiley & Sons, Inc., Jul. 2014, ISBN: 9781119005384. doi: 10.1002/9781119005384 (cit. on p. 27).
- [63] J. Cook, S. Eidenbenz, and A. Bärtschi, “The Quantum Alternating Operator Ansatz on Max-k Vertex Cover”, Oct. 2019 (cit. on p. 31).
- [64] Z. Wang, N. C. Rubin, J. M. Dominy, and E. G. Rieffel, “\$XY\$-mixers: analytical and numerical results for QAOA”, Apr. 2019. doi: 10.1103/PhysRevA.101.012320 (cit. on p. 31).
- [65] *Trotter product formula - Encyclopedia of Mathematics* (cit. on pp. 32, 33).
- [66] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, “Quantum Computation by Adiabatic Evolution”, Jan. 2000 (cit. on p. 32).
- [67] S. Weinberg, *The Quantum Theory of Fields*, 1 Supplement. Cambridge University Press, 1995, vol. 1, pp. 1–632. doi: 10.1007/BF02745513 (cit. on p. 32).
- [68] D. Poulin, A. Qarry, R. Somma, and F. Verstraete, “Quantum simulation of time-dependent hamiltonians and the convenient illusion of hilbert space”, *Physical Review Letters*, vol. 106, no. 17, Apr. 2011, ISSN: 00319007. doi: 10.1103/PhysRevLett.106.170501 (cit. on p. 32).
- [69] M. Heyl, P. Hauke, and P. Zoller, “Quantum localization bounds Trotter errors in digital quantum simulation”, Jun. 2018 (cit. on p. 33).
- [70] M. R. Garey and D. S. Johnson, *Computers and Intractability; A Guide to the Theory of NP-Completeness*. USA: W. H. Freeman & Co., 1990, ISBN: 0716710455 (cit. on p. 34).
- [71] B. Barak, A. Moitra, R. O’Donnell, P. Raghavendra, O. Regev, D. Steurer, L. Trevisan, A. Vijayaraghavan, D. Witmer, and J. Wright, “Beating the random assignment on constraint satisfaction problems of bounded degree”, May 2015 (cit. on p. 34).
- [72] M. B. Hastings, “CLASSICAL AND QUANTUM BOUNDED DEPTH APPROXIMATION ALGORITHMS”, Tech. Rep. 14, 2019, pp. 1116–1140 (cit. on p. 34).
- [73] S. Bravyi, A. Kliesch, R. Koenig, and E. Tang, “Obstacles to State Preparation and Variational Optimization from Symmetry Protection”, Oct. 2019 (cit. on p. 34).

- [74] Z. Jiang, E. G. Rieffel, and Z. Wang, "Near-optimal quantum circuit for Grover's unstructured search using a transverse field", *Physical Review A*, vol. 95, no. 6, Jun. 2017, ISSN: 24699934. doi: 10.1103/PhysRevA.95.062317 (cit. on p. 34).
- [75] M. Y. Niu, S. Lu, and I. L. Chuang, "Optimizing QAOA: Success Probability and Runtime Dependence on Circuit Depth", May 2019 (cit. on p. 34).
- [76] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost-Function-Dependent Barren Plateaus in Shallow Quantum Neural Networks", Jan. 2020 (cit. on p. 34).
- [77] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, "Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices", Tech. Rep. (cit. on pp. 34, 40).
- [78] R. M. Karp, "Reducibility among Combinatorial Problems", in *Complexity of Computer Computations*, Boston, MA: Springer US, 1972, pp. 85–103. doi: 10.1007/978-1-4684-2001-2__9 (cit. on p. 34).
- [79] E. W. Weisstein, "Incidence Matrix", (cit. on p. 35).
- [80] P. Vikstål, M. Grönkvist, M. Svensson, M. Andersson, G. Johansson, and G. Ferrini, "Applying the Quantum Approximate Optimization Algorithm to the Tail Assignment Problem", Dec. 2019 (cit. on p. 35).
- [81] J. R. Johansson, P. D. Nation, and F. Nori, "QuTiP 2: A Python framework for the dynamics of open quantum systems", *Computer Physics Communications*, vol. 184, no. 4, pp. 1234–1240, Apr. 2013, ISSN: 00104655. doi: 10.1016/j.cpc.2012.11.019 (cit. on p. 38).
- [82] J. Mockus, *Global Optimization and the Bayesian Approach*. Springer Netherlands, 1989, vol. 37, ISBN: 978-94-009-0909-0. doi: 10.1007/978-94-009-0909-0__1 (cit. on p. 39).
- [83] J. A. Nelder and R. Mead, "A Simplex Method for Function Minimization", *The Computer Journal*, vol. 7, no. 4, pp. 308–313, Jan. 1965, ISSN: 0010-4620. doi: 10.1093/comjn1/7.4.308 (cit. on p. 40).
- [84] N. Hansen, "The CMA Evolution Strategy: A Tutorial", Apr. 2016 (cit. on p. 40).
- [85] G. Verdon, M. Broughton, J. R. McClean, K. J. Sung, R. Babbush, Z. Jiang, H. Neven, and M. Mohseni, "Learning to learn with quantum neural networks via classical neural networks", Jul. 2019 (cit. on p. 40).
- [86] S. Bhatnagar, H. Prasad, and L. Prashanth, *Stochastic Recursive Algorithms for Optimization*. 2013, vol. 434, p. 280, ISBN: 978-1-4471-4284-3. doi: 10.1007/978-1-4471-4285-0 (cit. on p. 40).

BIBLIOGRAPHY

- [87] M. Alam, A. Ash-Saki, and S. Ghosh, "Analysis of Quantum Approximate Optimization Algorithm under Realistic Noise in Superconducting Qubits", Jul. 2019 (cit. on p. 41).
- [88] C. Xue, Z.-Y. Chen, Y.-C. Wu, and G.-P. Guo, "Effects of Quantum Noise on Quantum Approximate Optimization Algorithm", Sep. 2019 (cit. on p. 42).
- [89] M. C. Collodo, J. Herrmann, N. Lacroix, C. K. Andersen, A. Remm, S. Laz, J.-c. Besse, T. Walter, A. Wallraff, and C. Eichler, "Implementation of a Conditional-Phase Gate by using in-situ tunable ZZ-Interactions", *(in preparation)*, pp. 1–2, (cit. on pp. 49, 68).
- [90] 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", (cit. on p. 50).
- [91] M. Kjaergaard, M. E. Schwartz, A. Greene, G. O. Samach, A. Bengtsson, M. O'Keeffe, 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, K. Obenland, S. Lloyd, T. P. Orlando, I. Marvian, S. Gustavsson, and W. D. Oliver, "A Quantum Instruction Set Implemented on a Superconducting Quantum Processor", Jan. 2020 (cit. on p. 50).
- [92] 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", Dec. 2019 (cit. on p. 50).
- [93] Z. Jiang, K. J. Sung, K. Kechedzhi, V. N. Smelyanskiy, and S. Boixo, "Quantum Algorithms to Simulate Many-Body Physics of Correlated Fermions", *Physical Review Applied*, vol. 9, no. 4, p. 044036, Apr. 2018, ISSN: 23317019. doi: 10.1103/PhysRevApplied.9.044036 (cit. on p. 51).
- [94] Y. Cao, G. G. Guerreschi, and A. Aspuru-Guzik, "Quantum Neuron: an elementary building block for machine learning on quantum computers", Tech. Rep., 2017 (cit. on pp. 51, 70).
- [95] J. Ghosh and A. G. Fowler, "Leakage-resilient approach to fault-tolerant quantum computing with superconducting elements", *RAPID COMMUNICATIONS PHYSICAL REVIEW A*, vol. 91, p. 20302, 2015. doi: 10.1103/PhysRevA.91.020302 (cit. on p. 57).

Bibliography

- [96] D. Sank, Z. Chen, M. Khezri, J Kelly, R Barends, B Campbell, Y Chen, B Chiaro, A Dunsworth, A Fowler, E Jeffrey, E Lucero, A Megrant, J Mutus, M Neeley, C Neill, P. J. J. O'malley, C Quintana, P Roushan, A Vainsencher, T White, J Wenner, A. N. Korotkov, and J. M. Martinis, "Measurement-Induced State Transitions in a Superconducting Qubit: Beyond the Rotating Wave Approximation", *Tech. Rep.*, 2016 (cit. on p. 57).
- [97] S. S. Elder, C. S. Wang, P. Reinhold, C. T. Hann, K. S. Chou, B. J. Lester, S. Rosenblum, L. Frunzio, L. Jiang, and R. J. Schoelkopf, "High-fidelity measurement of qubits encoded in multilevel superconducting circuits", *Tech. Rep.* (cit. on p. 57).
- [98] A. Wallraff, D. I. Schuster, A. Blais, L. Frunzio, J. Majer, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, "Approaching unit visibility for control of a superconducting qubit with dispersive readout", *Physical Review Letters*, vol. 95, no. 6, p. 060501, Aug. 2005, ISSN: 00319007. doi: 10.1103/PhysRevLett.95.060501 (cit. on p. 58).
- [99] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, "Protocols for optimal readout of qubits using a continuous quantum nondemolition measurement", *Physical Review A - Atomic, Molecular, and Optical Physics*, vol. 76, no. 1, pp. 1–11, 2007, ISSN: 10502947. doi: 10.1103/PhysRevA.76.012325 (cit. on pp. 59, 64).
- [100] C. M. Bishop, *Pattern Recognition and Machine Learning*. SPRINGER, 2006, ISBN: 9780387310732 (cit. on pp. 59, 65, 71).
- [101] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. SPRINGER, 2017, ISBN: 978-0-387-84857-0. doi: 10.1007/b94608 (cit. on p. 59).
- [102] K. Reuer, S. P. Magnard, K. Akin, and P. A. Wallraff, "Real time , single shot , dispersive readout of superconducting qubits using a field programmable gate array", pp. 1–31, 2018 (cit. on pp. 59, 64).
- [103] N. Lacroix, C. K. Andersen, and A. Wallraff, "High Fidelity Qutrit Single Shot Readout", *Internal report*, 2019 (cit. on pp. 59, 65, 67, 71).
- [104] 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", *Physical Review Applied*, vol. 10, p. 34040, 2018. doi: 10.1103/PhysRevApplied.10.034040 (cit. on pp. 61, 64).
- [105] C. C. Bultink, B. Tarasinski, N. Haandbæk, 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", *Applied Physics Letters*, vol. 112, no. 9, pp. 1–10, 2018, ISSN: 00036951. doi: 10.1063/1.5015954 (cit. on p. 64).

BIBLIOGRAPHY

- [106] P Kurpiers, P Magnard, B Royer, M Pechal, J Heinsoo, Y Salathé, A Akin, S Storz, J.-c. Besse, S Gasparinetti, A Blais, and A Wallraff, "Deterministic quantum state transfer and remote entanglement using microwave photons", 2018. doi: 10.1038/s41586-018-0195-y (cit. on p. 64).
- [107] P. Magnard, P. Kurpiers, B. Royer, T. Walter, J. C. Besse, S. Gasparinetti, M. Pechal, J. Heinsoo, S. Storz, A. Blais, and A. Wallraff, "Fast and Unconditional All-Microwave Reset of a Superconducting Qubit", *Physical Review Letters*, vol. 121, no. 6, p. 060 502, Aug. 2018, ISSN: 10797114. doi: 10.1103/PhysRevLett.121.060502 (cit. on pp. 64, 68, 70, 74, 75).
- [108] Björck, "Numerics of Gram-Schmidt orthogonalization", *Linear Algebra and Its Applications*, vol. 197-198, no. C, pp. 297–316, Jan. 1994, ISSN: 00243795. doi: 10.1016/0024-3795(94)90493-6 (cit. on p. 65).
- [109] L. Cai, "Latent variable modeling", *Shanghai Archives of Psychiatry*, vol. 24, no. 2, pp. 118–120, 2012, ISSN: 10020829. doi: 10.3969/j.issn.1002-0829.2012.02.010 (cit. on p. 69).
- [110] M. D. Reed, L. Dicarlo, S. E. Nigg, L. Sun, L. Frunzio, S. M. Girvin, and R. J. Schoelkopf, "Realization of three-qubit quantum error correction with superconducting circuits", *Nature*, vol. 482, no. 7385, pp. 382–385, Feb. 2012, ISSN: 00280836. doi: 10.1038/nature10786 (cit. on p. 70).
- [111] L Steffen, Y Salathe, M Oppliger, P Kurpiers, M Baur, C Lang, C Eichler, G Puebla-Hellmann, A Fedorov, and A Wallraff, "Deterministic quantum teleportation with feed-forward in a solid state system", *Nature*, vol. 500, 2013. doi: 10.1038/nature12422 (cit. on p. 70).
- [112] Y. Salathé, P. Kurpiers, T. Karg, C. Lang, C. K. Andersen, A. Akin, S. Krinner, C. Eichler, and A. Wallraff, "Low-Latency Digital Signal Processing for Feedback and Feedforward in Quantum Computing and Communication", *Physical Review Applied*, vol. 9, no. 3, p. 034 011, Mar. 2018, ISSN: 23317019. doi: 10.1103/PhysRevApplied.9.034011 (cit. on p. 70).
- [113] L. Breiman, *Classification and regression trees*. Wadsworth International Group, 1984, p. 358, ISBN: 0534980546 (cit. on p. 71).

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Variational Quantum Algorithms with Controlled Arbitrary Phase Gates

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

Lacroix

First name(s):

Nathan

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the '[Citation etiquette](#)' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

Zurich, April 17th 2020

Signature(s)



For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.