Characterizing Errors in Quantum Information Processors

by

Yuval Sanders

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics (Quantum Information)

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

ERROR-free computation is an unattainable ideal, yet our world now contains many computers that appear error-free to their users. That such things are possible is explained by sophisticated theorems that demonstrate the possibility of efficiently reducing computational errors introduced by reasonably well-behaved noise.

My thesis is about the problem of determining whether noise in prototype quantum information processors is sufficiently well-behaved for fault-tolerant quantum computing to be possible. My work is divided into two themes. The first theme is the interpretation of average gate fidelity, a quantity that has become the standard performance metric for assessing progress towards fault tolerance. I have elucidated the connection between average gate fidelity and the requirements of fault-tolerant quantum computing by demonstrating the limits of fidelity as a proxy for error rate, the usual metric in fault-tolerance literature. I thereby conclude that information additional to fidelity is *required* to assess progress towards fault-tolerance. The second theme is the characterization of two-level defect systems, a particularly deleterious kind of noise that can affect superconducting-integrated-circuit-based quantum computing prototypes. I have designed statistical experimental design algorithms that can rigorously assess the influence of these defect systems, and I helped develop a proposal to mitigate their influence. I thereby demonstrate that existing experimental techniques can become much more powerful by employing advanced data collection procedures.

My work has immediate implications for current research efforts towards the first working quantum computer. Theoretical work should be directed at assessing noise sources using metrics other than average gate fidelity, and future experimental characterization techniques should become more modular in order to incorporate advanced statistical inference techniques like the ones I develop herein.

ACKNOWLEDGEMENTS

Dedicated to the Memory of Joseph Morris Sanders

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

x

LIST OF ALGORITHMS

INTRODUCTION

What casts the pall over our victory celebration? It is the curse of dimensionality, a malediction that has plagued the scientist from the earliest days.

RICHARD BELLMAN

CONFIDENCE in the possibility of large scale quantum computers is derived in great part from the existence of fault-tolerance threshold theorems. These theorems guarantee that a given quantum circuit can be efficiently simulated with arbitrary precision if noise is sufficiently well-behaved and if the rate of error falls below some fixed, non-zero threshold. Thus contemporary research efforts into quantum computing are often focussed on reducing noise to within tolerable levels.

Yet the assessment of noise in quantum computer prototypes is rapidly becoming a major challenge due to the curse of dimensionality: a complete representation of a noise operator affecting a quantum register requires the specification of a number of parameters that is an exponential function of the size of the register, i.e. the number of qubits. Without data-efficient methods of identifying the noise processes affecting quantum computer prototypes, reliable assessment of progress will eventually become impossible. In fact, state-of-the-art devices involve the manipulation of more than ten qubits; the state of such a device would ideally be represented by a $2^{10} \times 2^{10}$ density operator (a unit trace Hermitean matrix) and is therefore specified by $4^{10} - 1 = 1048575$ distinct real parameters. The curse of dimensionality is rapidly becoming a crucial challenge for continued experimental progress towards useful quantum computers.

As its title suggests, this thesis is aimed at the problem of characterizing errors in quantum information processors. The error characterization problem is a long-term challenge for both scientific research into fault-tolerant quantum computation and industrial research and development toward the first useful quantum computer. This thesis presents results on the data-efficient characterization of an important noise source, two level defect systems, found in superconducting-integrated-circuit-based quantum computer prototypes. This thesis also explains how to connect a standard performance metric, the average gate fidelity, to the requirements of fault tolerance.

This introductory chapter is divided into three sections. Section 1.1 clarifies the scope of this thesis by stating its aims and the motivation for these aims, Section 1.2 summarizes the results in this thesis, and Section 1.3 gives an outline of the remaining content of this thesis.

Motivation and Aims $\S$**1.1**

A device capable of executing quantum circuits subject to non-pathological noise can efficiently and accurately simulate ideal quantum computations. It is sufficient for noise to be non-pathological in two ways: first it cannot introduce highly correlated errors across distant gates in the circuit, and it cannot induce errors at a rate above some threshold. Thus, our progress towards fault-tolerant quantum computation can in principle be measured in terms of the rate and correlation of errors that occur in executed quantum circuits. The problem of characterizing errors is to determine, in practice, the rate and correlations of errors in quantum information processors.

Characterizing error is, in general, exponentially difficult. Because the error rate of a computational circuit depends on the choice of input and because the size of an input is measured in bits, the number of possible inputs to a circuit is the exponential (base two) of the number n of bits required to specify the input. As it is possible for a circuit to perform well on all but one of the possible input states, each input must be checked in general. To form a complete picture of the effect of error on a circuit (assuming Markovian noise), the probability of observing each of the 2^n possible outputs must be assessed for each of the 2^n possible inputs, meaning that at least 4^n probabilities need to be known in order to fully characterize the performance of a given logic circuit, be it classical or quantum.

The exponential cost of characterizing error is a challenge for the meaningful assessment of progress towards practical and fault-tolerant quantum computation. Whereas small quantum information processors can be completely characterized within a rea-

sonable amount of time, the exponential cost ensures that such characterization rapidly becomes impractical as the size of the system increases. Naturally, the problem of error characterization becomes much easier if there is prior knowledge about the behaviour of the device. For example, an n-bit string might be encoded into some physical device that is known to suffer only noise that acts independently of the coded string. If the chance of observing one altered bit is p, the chance of observing k altered bits is p^k and the chance of observing no altered bits is $(1-p)^n$. Therefore, the probability of error can be summarized by a single value p, which is independent of n. The cost of characterizing error under the string-independent noise model is no longer exponential in n. Practical solutions to the error characterization problem should therefore be directed in part at establishing the truth of simplifying assumptions such as that of string-independence.

The goal of this thesis is to analyze logical errors that occur in the technology that is widely seen as one of the most promising candidates for scalable quantum computing: the technology of superconducting integrated circuits (SICs). SICs are analogous to the silicon-based integrated circuits that underlie all modern computing technology. Perhaps the most important advantage for SICs is that the scalable manufacturing techniques used to produce silicon-based integrated circuits may also be employed to develop SICs. Impressive advances in the control of so-called superconducting qubits have led to private investment into the development of quantum computers; most notably from D-Wave Systems Inc., Google Inc. and IBM Corporation. However, experimental characterization of the frequency and kind of errors affecting SIC-based quantum computing prototypes remains an open problem.

A particularly important source of noise in SICs is that which is introduced by two-level defect systems (TLDSes). Their source and nature is mysterious, but the recent experiments of Lisenfeld et al. (2015) give a strong indication that TLDSes are structural defects in the dielectric comprising Josephson junctions, which are the primary electric circuit component in SICs. TLDSes are especially dangerous for the reliable processing of information because they can introduce logical errors with high probability. They do this by coherently absorbing and reëmitting excitations of a superconducting qubit to which they are coupled with a frequency that can be on the order of the inverse of the gate time.

Mitigating against the presence of TLDSes is possible if highly accurate descriptions about the behaviour of TLDSes are available. Although an approach to mitigating against one deleterious effect of TLDSes is discussed in Chapter 7, such strategies are not the primary aim of this thesis. The primary aim of this thesis is to give methods for characterizing TLDSes based on experimental data. The methods should be data-efficient because the effect of TLDSes can fluctuate. Indeed, it is possible for TLDS behaviour to change on the timescale of the very experiments designed to detect their presence.

Data efficiency is nonetheless possible because existing evidence supports a simple yet accurate phenomenological model of TLDSes. In this model, each TLDS is specified by two parameters, so the number of parameters needed to specify the noise model grows linearly with the system size instead of exponentially.

The secondary aim of this thesis is to determine the meaning of what has become the standard performance metric for assessing progress towards fault-tolerant quantum computing: average gate fidelity. Progress towards fault-tolerance can be assessed, in principle at least, by comparing measured performance to fault-tolerance thresholds. But fault-tolerance thresholds are threshold *error rates*, which bears no functional relationship with the average gate fidelity. That is to say, two quantum gates can have identical average fidelities but different error rates. In fact, the error rates can be vastly different. Yet the average fidelity is somewhat salvageable as a measure of gate quality; Wallman and Flammia (2014) showed that bounds can be placed on the error rate of a gate. But the lower bound is proportional to the square of the upper, meaning that the numerical values of these bounds differ by orders of magnitude when the error rate is low (i.e. the average gate fidelity is high). Though this relationship seems unnecessarily pessimistic, Chapter 3 shows that the relationship is in fact optimal (perhaps up to constant factors not far from unity).

To summarize, this thesis aims to characterize the error affecting SIC-based prototype quantum information processors. Whereas characterizing errors are a crucial aspect of research and development towards the first quantum computer, unacceptably large data sets are required for error characterization in general. Without accurate characterizations of error, progress towards fault-tolerance cannot be meaninfully assessed. This thesis presents results on the efficient characterization of TLDSes and the comparison of average gate fidelity to fault-tolerance thresholds. These results are summarized in the following section.

Results   §**1.2**

Characterizing errors in quantum information processors is a grand challenge. This thesis reports on the content of four scientific papers that constitute advances towards practical solutions to this grand challenge. Three of the four papers are published and the fourth is still in preparation. In order of appearance in this thesis, they are:

1. optimal bounds on quantum gate error rate based upon reported average fidelity, published as Sanders et al. (2016);

2. efficient detection of a TLDS using an offline data-collection algorithm, in preparation;

3. efficient estimation of TLDS parameters using an online data-collection algorithm, published as Stenberg et al. (2014); and

4. mitigating against the presence of a TLDS, published as Chasseur et al. (2015).

The first set of results, from Sanders et al. (2016), are about the connection between fault-tolerance and reported average gate fidelity, the main performance metric for assessing progress in experimental quantum computing devices. There are two results. The first result of Sanders et al. (2016) is a demonstration that the upper bound of Wallman and Flammia (2014) is asymptotically tight in the following sense. If the average gate fidelity is φ for a gate of dimension d, the upper bound to the error rate η reads

$$\eta^{\text{ub}}(\varphi, d) := \sqrt{d(d+1)(1-\varphi)}, \tag{1.1}$$

which varies approximately linearly in d for constant φ and exactly as $\sqrt{1-\varphi}$ for constant d. We show that any tighter upper bound to η must bear the same functional relationship with φ and d. Furthermore, any constant factors must not be much different from unity. The second result of Sanders et al. (2016) is a tighter bound on η in terms of φ and an additional measure of noise we call the "Pauli distance" δ^{Pauli}. We argue that such additional information is *required* in order to assess progress towards fault-tolerant quantum computing.

The second set of results, in preparation, are about the detection of TLDSes using an offline data collection algorithm. The algorithm prescribes a set of experimental data to be collected for the purpose of determining whether a TLDS is coupled to a SIC-based qubit. We argue that the algorithm is data-efficient in the sense that the number of bits of experimental data needed to detect a TLDS scales optimally (linearly) with the inverse coupling strength of the TLDS. The fact that this data collection algorithm is an offline algorithm means that it should be relatively easy to implement in a laboratory setting. We hope that experimentalists will do so in the near future.

The third set of results, published as Stenberg et al. (2014), are about the estimation of TLDS parameters using an online data collection algorithm. The distinction between online and offline algorithms is whether the algorithm can specify all the requested experiments without analyzing any data. Our online algorithm chooses the most informative experiments by analyzing data as it becomes available. As a result, our algorithm

can improve estimates of TLDS parameters from order-of-magnitude estimates to machine precision with less than a kilobit of experimental data. Our algorithm appears to work even in the presence of measurement errors and decoherence, though overcoming these experimental imperfections costs data. We again hope that experimentalists will implement this algorithm in the future, though there are more practical challenges to overcome for online algorithms.

The fourth set of results, published as Chasseur et al. (2015), describe a method for mitigating the influence of a TLDS under certain circumstances. Control of SIC-based qubits, for which the computational basis is a pair of energy eigenstates, often involves modifying the energy gap between computational basis states. But TLDSes, if present, interfere with these modifications. We provide methods to modify a qubit's energy gap so that the interaction between the qubit and the TLDS is suppressed, though our methods require an extremely accurate description of the interaction that is to be mitigated. These results motivate the main work of this thesis, which is to produce highly accurate descriptions of qubit-TLDS interactions.

To conclude, the promise of practical quantum computation can be fulfilled only when the noise affecting prototype devices is understood sufficiently well for control methods in quantum devices to be capable of delivering performance that fulfills the theoretical requirements of fault-tolerant quantum computing. This thesis describes results that constitute progress towards the grand challenge of characterizing errors in quantum information processors. This thesis shows how to connect average gate fidelity with the requirements of fault-tolerance, how to detect and describe a major source of noise in SIC-based prototypes, and how to correct for that noise if it is present. The results in this thesis should be incorporated into future research and development aimed at producing the first useful quantum computer.

Outline of the Thesis §1.3

This thesis is organized into eight chapters, including this introduction, and an appendix. Each of Chapters 3, 5, 6, and 7 describe the results of distinct research projects I undertook during the course of my PhD studentship. However, the content of Chapters 5 and 6 are closely related; this relationship is explained in Section 4.3 as two components of a solution to the 'TLDS characterization problem'.

It is useful to describe the thesis, excluding Chapter 8 (a short concluding chapter), as being composed of four parts. The first part, consisting of Chapters 1 and 2, are introductory in nature: Chapter 1 introduces the thesis and Chapter 2 introduces some necessary

context for the results presented in this thesis. The second part is Chapter 3, which describes my work on comparing average gate fidelities to fault-tolerance thresholds. The third part, consisting of Chapters 4, 5, and 6 expains my work on characterizing TLDSes using data-efficient experimental design algorithms. The fourth part, consisting of Chapter 7 and Appendix A, explains my (relatively minor) contributions to Chasseur et al. (2015), which is reprinted as Appendix A for convenience.

Part one, consisting of Chapters 1 and 2, is organized as follows. Chapter 1, the present chapter, introduces the content of the thesis in terms of its aims (Section 1.1), its results (Section 1.2), and its content (Section 1.3). Chapter 2 contains an overview of crucial background material for the results presented in this thesis. The chapter covers three independent topics. Section 2.1 explains that fault-tolerance thresholds for quantum computing are threshold error rates, which is a crucial point in Chapter 3. Section 2.2 gives a brief explanation of SIC-based quantum computing technology and the problem of TLDSes therein. Finally, Section 2.3 explains the language of statistical experimental design and the particle representation methods used to perform computations, which is necessary context for the results presented in Chapters 5 and 6.

Part two, consisting of Chapter 3, is organized as follows. Average gate fidelity is defined in Section 3.1 and is compared to the error rate of a quantum gate, defined in Section 2.1. Sections 3.2 and 3.3 present the results of the work, and Section 3.4 explains their relevance for assessing progress towards fault-tolerant quantum computing.

Part three, consisting of Chapters 4, 5, and 6, is organized as follows. Chapter 4 is subdivided into three sections: Section 4.1 derives the 'swap spectrum', Section 4.2 explains the concept of a 'coupling octave', and Section 4.3 uses this language to express the TLDS characterization problem formally. The TLDS characterization problem is thereby broken down into two simpler problems, TLDS detection and TLDS estimation, which are solved in Chapters 5 and 6.

Chapter 5 is subdivided into four sections. Section 5.1 explains the 'coupling octave sampling heuristic', which is used extensively in one of the two algorithms explained in Section 5.2 that together make up the detection algorithm. The performance of this algorithm is analyzed in Section 5.3, and the prospects for extensions and experimental implementations are discussed in Section 5.4.

Chapter 6 is organized similarly. Section 6.1 explains a heuristic, adapted from Ferrie et al. (2013), for choosing highly informative experiments based on current knowledge about likely TLDS parameters. This heuristic is the main component of Algorithm 6.1, presented in Section 6.2. The performance of the algorithm is analyzed in Section 6.3, and the prospects for experimental implementation are discussed in Section 6.4.

The fourth and final part consists of Chapter 7 and Appendix A. Appendix A is simply a reprint of Chasseur et al. (2015) together with a note explaining my contribution. This contribution is explained more thoroughly in Chapter 7. The context of the work within the remainder of the thesis is explained in Section 7.1, my contribution is described in Section 7.2, and the prospects for experimental implementation are discussed in Section 7.3.

TWO

BACKGROUND

All our knowledge begins with the senses, proceeds then to the understanding, and ends with reason. There is nothing higher than reason.

IMMANUEL KANT

E RROR characterization in quantum information processors constitutes a grand challenge that is only one component of the greater challenge of producing the world's first useful quantum computer. This thesis reports on progress towards overcoming the challenge of error characterization in SIC-based quantum computer prototypes. To explain this progress, it is necessary to explain three broad areas of active research.

The first research area is that of quantum fault-tolerance thresholds. An explanation of fault-tolerance thresholds is needed because the results described in Chapter 3 connect average gate fidelity, a common performance metric used to report progress towards fault-tolerance, with the fault-tolerance thresholds that are used as targets. Some confusion has crept into the literature about what might be called the 'units' of fault-tolerance thresholds: much as the third law of thermodynamics states that all temperature is positive *in the Kelvin scale*, so too do fault-tolerance threshold theorems show that fault-tolerance is possible if the *gate error rate* is small enough. Yet it has wrongly become common to report fault-tolerance thresholds as a threshold *average gate fidelity*. Chapter 3 is intended, in part, to explain the correct way to compare reported gate fidelities to stated fault-tolerance thresholds.

The second research area is that of SICs, superconducting integrated circuits. The results of this thesis are aimed primarily at characterizing a particular source of error in these devices: TLDSes, two-level defect systems. Their nature is not certain, but they are probably an unavoidable property of the Josephson junctions that are the essential component of SIC-based quantum information processors. Variations in the design of Josephson junctions may make their presence more or less important for quantum information processing purposes, but their influence is highly deleterious to quantum information processing and they will probably never be fully eliminated from SIC technology. An overview of the problem is therefore a necessary part of any explanation of the results presented in this thesis.

The third research area is that of algorithmic experimental design. This is a highly complex research area at the cutting edge of statistical and machine learning research, and the techniques are as varied as the problems they are designed to solve. Nonetheless, the main results of this thesis make use of some important algorithmic experimental design techniques. In particular, the main figure-of-merit for experiments is the expected posterior variance, which is calculated in practice using 'particle representations' that approximate complex probability distributions.

This chapter is separated into three sections, each of which deals with one of the above research areas. Section 2.1 explains quantum fault-tolerance thresholds, Section 2.2 explains the basic physics of SICs (Josephson junctions in particular) and the influence of TLDSes, and Section 2.3 describes the language and relevant algorithms of experimental design.

Fault-Tolerance Thresholds §2.1

Fault-tolerance thresholds are important theoretical constructs that have come to play a role as performance targets for experimental progress towards scalable quantum computing. Several facts nonetheless mitigate the prominence that these targets should have as measures of progress towards fault tolerance. Crucially, fault-tolerance thresholds are established *existentially* by threshold theorems under limited but strict assumptions about the tolerable form of noise. Realistic performance targets should therefore be based upon a recognition that the truth of noise assumptions must be established empirically.

The aim of this section is therefore to explain fault-tolerance thresholds and the kind of theorem that guarantees their existence. The content of this section is required context for the results of Chapter 3, which reports the results of Sanders et al. (2016); the content in this section is reproduced, lightly edited, from Sanders et al. (2016). Though the key

points of the derivation are well-established, the presentation of the diamond distance by induction from the total variation distance is, from what I can tell, novel. The derivation extends an operational interpretation of the error rate of a random process to that of a quantum logic gate.

Error Rates §**2.1.1**

Fault-tolerance thresholds are threshold error rates. Whereas the term 'error rate' is often used in the literature to refer to a parameter in a family of noise models, this meaning is not the one employed by threshold theorems. Much as 'distance' is not simply a parameter in a spatial coördinate system but is a measurable relationship that does not depend on the language used to express locations, so too is 'error rate' a measurable property of noisy operations whose definition does not rely on a specification of a noise model. In other words, error rates are operational even when noise models are not.

The purpose of this subsection is to define the error rate of a quantum logic gate. Though the resulting mathematical formula, expressed in Definition 2.2, is standard, the operational nature of the derivation appears to be novel. I developed it in close collaboration with Barry Sanders and Joel Wallman, with helpful input from Robin Blume-Kohout and Daniel Gottesman. The derivation is published in Sanders et al. (2016).

The definition of the error rate of a quantum logic gate builds naturally on the concept of error rate for a random process. For deterministic processes, an error has occurred if the process produces the 'wrong' output. However, no single output of a random process can be treated unambiguously as 'correct'. The rate of error is therefore defined by comparing the actual statistics of a process to its ideal statistics.

The statistics for an ideal process is described by a probability distribution p_{id} over the set of possible outputs X; the ideal probability of output $x \in X$ is $p_{\text{id}}(x)$. An error-prone process produces a different distribution p_{ac}, that represents the actual statistics over the set of possible outcomes X. The total variation distance

$$d_{\text{TV}}\left(\mu, \nu\right) \equiv \frac{1}{2} \sum_{x \in X} |\mu(x) - \nu(x)| \tag{2.1}$$

is a natural measure of the distance between two probability distributions μ and ν over a set of outcomes X.

The total variation distance $d_{\text{TV}}\left(p_{\text{ac}}, p_{\text{id}}\right)$ can be interpreted as the error rate of the process as follows. p_{ac} can be estimated by sampling the actual random process N times

11

and counting the number $n(x)$ of occurrences of each possible output x; the fraction $n(x)/N$ approaches $p_{ac}(x)$ as $N \to \infty$. By altering some fraction r of the samples so that the number of occurrences of each outcome x becomes $n'(x)$, it is possible to ensure that $n'(x)/N \approx p_{id}(x)$ rather than $p_{ac}(x)$. The fraction r is not unique, but the minimum possible value r_{min} of r must be greater than zero for large N if $p_{ac} \neq p_{id}$. By Proposition 4.7 of Levin et al. (2009), $r_{min} \to d_{TV}(p_{ac}, p_{id})$ as $N \to \infty$. Thus, $d_{TV}(p_{ac}, p_{id})$ approximates the fraction of a large sample that must be altered to ensure that the relative frequencies of each outcome match the ideal distribution p_{id}; each alteration can be interpreted as the correction of an error.

This definition can be extended to quantum logic gates in two steps. First, quantum gates act upon quantum, rather than classical, registers. Second, the error rate depends on the input to the gate and therefore must be extended somehow to an overall error rate for a gate. Because threshold theorems demand strict bounds upon error rate, both extensions involve maximization; the first maximizes over measurements and the second maximizes over inputs. The first step follows from Fuchs and van de Graaf (1999); the second is from Kitaev (1997).

Ideally, the state of a quantum register can be represented by a unit vector $|\psi\rangle$ in a d-dimensional Hilbert space $\mathscr{H}$. The register is typically treated as a collection of n qubits, in which case $\mathscr{H}$ is canonically isomorphic with the n-fold tensor product of the Hilbert space $\mathscr{Q} \cong \mathbb{C}^2$ of a single qubit: $\mathscr{H} \cong \mathscr{Q}^{\otimes n}$. In this case, $d = 2^n$. A realistic state is modelled by a density operator ρ, which is a unit trace, positive semidefinite operator on $\mathscr{H}$. A measurement of a quantum state is described by a positive operator-valued measure (POVM), which is a set of positive operators $\{E_\ell\}$ acting on $\mathscr{H}$ such that $\sum_\ell E_\ell = I$, the identity operator. The probability of observing outcome ℓ is $\text{Tr}(E_\ell \rho)$. Thus, the actual output ρ_{ac} of a gate acting on a specified input can be compared to the ideal output ρ_{id} by measuring with respect to some POVM. The error rate of this measurement is $d_{TV}(p_{ac}, p_{id})$, where $p_{ac}(\ell) = \text{Tr}(E_\ell \rho_{ac})$ and $p_{id}(\ell) = \text{Tr}(E_\ell \rho_{id})$. As demonstrated by Fuchs and van de Graaf (1999), maximizing $d_{TV}(p_{ac}, p_{id})$ over all possible choices of measurements yields

$$d_{Tr}(\rho_{ac}, \rho_{id}) := \frac{1}{2} \|\rho_{ac} - \rho_{id}\|_{Tr}, \tag{2.2}$$

where $\|A\|_{Tr} := \text{Tr}\sqrt{A^\dagger A}$ for any linear operator A. Thus, the error rate of ρ_{ac} with respect to ρ_{id} is $d_{Tr}(\rho_{ac}, \rho_{id})$.

An ideal quantum logic gate, represented by G, acts as a unitary operator on $\mathscr{H}$. Whereas the operation on a pure state can be treated as direct (i.e. $|\psi\rangle \mapsto G|\psi\rangle$), the gate can act upon a mixed state ρ. In this instance, the gate performs the action $\rho \mapsto G\rho G^\dagger$.

12

This action is represented by a quantum channel $\mathcal{G}_{\text{id}}$; explicitly,

$$\mathcal{G}_{\text{id}}(\rho) := G\rho G^\dagger. \tag{2.3}$$

This channel is compared with a non-ideal implementation $\mathcal{G}_{\text{ac}}$ that is in general not represented by unitary conjugation but is a completely positive, trace preserving linear operator on the space of density operators over $\mathcal{H}$.

As the error rate for a quantum logic gate acting on a specified input state ρ is given by $d_{\text{Tr}}\left(\mathcal{G}_{\text{ac}}(\rho), \mathcal{G}_{\text{id}}(\rho)\right)$, and as the error rate of $\mathcal{G}_{\text{ac}}$ with respect to $\mathcal{G}_{\text{id}}$ should involve maximization over inputs, the error rate of $\mathcal{G}_{\text{ac}}$ could be defined as $\max_\rho d_{\text{Tr}}\left(\mathcal{G}_{\text{ac}}(\rho), \mathcal{G}_{\text{id}}(\rho)\right)$. But such a definition is undesirable because, in general, the error rate of $\mathcal{G}_{\text{ac}} \otimes \mathbb{1}$ (where $\mathbb{1}$ represents the identity operator acting upon some ancillary space $\mathcal{H}'$) differs from that of $\mathcal{G}_{\text{ac}}$. Kitaev (1997) amends this definition by maximizing over both inputs and ancillary spaces; he defines

$$\|\mathcal{A}\|_\diamond := \sup_{\mathcal{H}'} \sup_{\rho \in \text{dens}(\mathcal{H} \otimes \mathcal{H}')} \|\mathcal{A} \otimes \mathbb{1}(\rho)\|_{\text{Tr}}, \tag{2.4}$$

where $\mathcal{A}$ is any superoperator over $\mathcal{H}$ and $\text{dens}(\mathcal{H} \otimes \mathcal{H}')$ is the set of density operators over the joint Hilbert space of the original register and some ancilla. Thus the error rate η of a gate implementation $\mathcal{G}_{\text{ac}}$ is

$$d_\diamond\left(\mathcal{G}_{\text{ac}}, \mathcal{G}_{\text{id}}\right) := \frac{1}{2}\|\mathcal{G}_{\text{ac}} - \mathcal{G}_{\text{id}}\|_\diamond. \tag{2.5}$$

For mathematical convenience, however, I use a modified but equivalent definition (Definition 2.2) for the remainder of this thesis.

Definition 2.1. If $\mathcal{G}_{\text{ac}}$ is some implementation of a gate G, define

$$\mathcal{D}_G := \mathcal{G}_{\text{ac}} \circ \mathcal{G}_{\text{id}}^{-1} \tag{2.6}$$

to be the *discrepancy channel* of G, where the channel $\mathcal{G}_{\text{id}}$ defined in Eq. (2.3) is unitary and hence invertible.

Definition 2.2. The *error rate* of an implementation of G is given by

$$\eta = d_\diamond\left(\mathcal{D}_G, \mathbb{1}\right), \tag{2.7}$$

where $\mathcal{D}_G$ is the discrepancy channel of the implementation.

The error rate as presented in Definition 2.2 is the subject of threshold theorems. In the next section, I explain these theorems and the nature of the fault-tolerance thresholds whose existence they guarantee. I take care to distinguish between thresholds that are proven to hold under general, but idealistic, assumptions about noise and those established by numerical evidence based upon simulation of the performance of a specific error-correcting code under a restricted class of noise models, which are usually chosen for ease of simulation rather than for their accurate representation of reality.

Threshold Theorems   §**2.1.2**

Fault-tolerance thresholds are, as the name suggests, threshold values of the error rate per quantum gate such that, if the actual error rate falls below this threshold, fault-tolerant quantum computing is possible. Threshold theorems guarantees that such a threshold value exists, though the theorems do not indicate *prima facie* what that value may be. This is due to two issues. First, different threshold theorems make different promises about the noise affecting quantum computers. Second, threshold values depend not only on the kind of noise present but also on the method of encoding quantum computations; different coding methods have different fault-tolerance properties. To explain these points, it is helpful to examine a threshold theorem statement. I focus on a "rough" statement of Aharonov and Ben-Or (2008) because it is relatively simple and contains a minimum of jargon.

Theorem 2.1 (Aharonov and Ben-Or (2008))**.** *There exists a threshold $\eta_0 > 0$ such that the following holds. Let $\varepsilon > 0$. If Q is a quantum circuit operating on n input qubits for t time steps using s two- and one-qubit gates, there exists a quantum circuit Q' with depth, size, and width overheads which are polylogarithmic in n, s, t, and $1/\varepsilon$ such that, in the presence of local noise of error rate $\eta < \eta_0$, Q' computes a function which is within ε total variation distance from the function computed by Q.*

This theorem guarantees that a value $\eta_0 > 0$, called the "threshold", exists such that a quantum circuit Q can be efficiently simulated by another circuit Q' to within an arbitrary error tolerance $\varepsilon > 0$ even if Q' is subject to "local noise" at a rate $\eta < \eta_0$. Different threshold theorems are inequivalent because they assume promises about noise that are different from that of noise locality.

A logic circuit is said to experience local noise if the noise acts separately on individual logic gates. To be precise, recall that a logic circuit is defined as a directed acyclic graph with nodes labelled by elements of some set of logic gates, where arrows into a node

14

represent inputs and arrows out of a node represent outputs; this definition is standard and can be found in, for example, Savage (1997). A quantum logic circuit can also be represented by a directed acyclic graph. The noise of a logic circuit is defined to be local if it can be represented as the composition of noise processes on individual nodes of the circuit graph.

As noise is assumed to affect each gate independently, noise can be modelled by replacing the intended unitary gate G by some imperfect implementation $\mathcal{G}_{\text{ac}}$ represented as a quantum channel (i.e. a completely positive, trace-preserving linear map on density operators over the state space of the input register). Lindblad (1976) shows that such a model is reasonable for an imperfect gate if the interaction between the register space and its environment obeys the Born-Markov approximation*. An error rate η can therefore be assigned to each gate $\mathcal{G}_{\text{ac}}$ in a circuit Q', which simulates Q efficiently and accurately in the presence of local noise if $\eta < \eta_0$.

There are two important limitations to the utility of the threshold η_0. Firstly, surpassing the threshold is sufficient but not necessary for fault-tolerance: error rates larger than η_0 could be acceptable if stronger promises can be made about noise. Conversely, devices subject to noise that does not satisfy the assumptions of a threshold theorem cannot be said to be fault-tolerant based on a demonstration that error rates fall below threshold; a stronger threshold $\eta_0' < \eta_0$ could apply. The second limitation is that the choice of Q' depends in practice upon the specified quantum-error-correcting code even though threshold theorem need not demand adherence to a coding strategy. Based on the choice of code, the appropriate performance target is some lower bound η_0^{lb} to the 'true' threshold value η_0. As with the first limitation, the validity of η_0^{lb} as a performance target derives from the validity of the promises made about the noise affecting real devices.

Whereas some threshold estimates are obtained through rigorous analysis of the performance of a code in the presence of noise subject to promises of varying strength, others are obtained through numerical simulation of performance in the presence of a parametrized family of noise models. Estimates based on numerical simulation are more optimistic and are often used as performance targets for experimental fault-tolerant quantum computing research. Two prominent examples of such numerically simulated

---

*By the Born-Markov approximation, I mean three assumptions needed to justify a master equation of Lindblad form:

1. (Born approximation) the system-environment coupling is weak,

2. (Markov approximation) the correlation time of the system-environment interaction is short, and

3. the initial state of the system-environment is a product state.

thresholds are quoted by Chow et al. (2012), who report gate fidelity thresholds of between 90% and 99.5% ("depending on measurement errors"), and Barends et al. (2014), who quote a gate fidelity threshold of 99%. That these thresholds are quoted in terms of *fidelity*, rather than *error rate* (the proper metric), reflects the fact that these values are derived through simulated performance of a family of noise models that can reasonably be parametrized by a performance metric called the average gate fidelity, which is the principal subject of Chapter 3.

Analytic estimates of the threshold can be produced based on details of the proof of a threshold theorem. Aharonov and Ben-Or (2008), for example, can justify an estimate of $\eta_0^{lb} \approx 10^{-6}$ based on their choice of coding strategy. The highest rigorously established value is $\eta_0^{lb} \approx 10^{-3}$, achieved by Reichardt (2006) and Aliferis et al. (2008). Numerical estimates, by contrast, are produced by simulating the behaviour of a fault-tolerant code in the presence of an easily simulated but potentially unrealistic class of noise models. The work of Svore et al. (2005, 2006) calls into question the relationship of such estimates to thresholds of the kind established by threshold theorems. Nevertheless, these simulations are often seen as indicative of true threshold values: Fowler et al. (2012) claim that a threshold value of about 0.57% error rate is appropriate for the surface code based on simulated performance under a depolarizing noise model.

A direct comparison of the above threshold values is not justifiable because each value makes different assumptions about the behaviour of noise and the choice of code. Thus, the estimate of $\eta_0^{lb} \approx 1\%$ for the surface code does not make the surface code less desirable than the C4/C6 code of Knill (2005) even though the C4/C6 code could have a threshold as high as 3% because there are other practical reasons to prefer the surface code over the C4/C6 code. Similarly, actual gate performance should not be directly compared with these threshold values because those gates are certainly subject to noise that is not well-approximated by the simulated noise model.

Superconducting Integrated Circuits $\S$**2.2**

The insight underlying research into superconducting integrated circuits for quantum information processing is that quantum information can be encoded into the state of the supercurrent in a superconductor. Just as the state of the transistor when used as a switch can encode a bit of information, so too can a "superconducting qubit" encode a single qubit of information.

The goal of this section is to describe a problem that seems pervasive amongst all superconducting qubits: information exchange with unwanted two-level systems. Such

systems introduce a particularly dangerous form of noise for the qubit: if such a system is present and the qubit operates near-resonance with it, coherent exchange of energy between the qubit and unwanted system can ensure that a logical operation enacted on the qubit will fail with high probability. The central aim of this thesis is to devise experimentally viable algorithms for assessing the influence of these systems on the dynamics of qubits.

This section is organized as follows. Section 2.2.1 gives some necessary background on superconducting qubits, which are electric circuits constructed from superconducting materials that can be used to encode and manipulate a single qubit of quantum information. Section 2.2.2 describes the influence of unwanted two-level systems on the dynamics of a superconducting qubit and presents a simple but powerful phenomenological model for this interaction.

Superconducting Qubits §**2.2.1**

One half of the Nobel Prize in Physics 2000 was awarded to Jack Kilby "for his part in the invention of the integrated circuit". The Royal Swedish Academy of Sciences explain that the integrated circuit instigated the "electronic revolution", as they call it. The integrated circuit has played such a critical role because it enabled manufacturers to produce complex electronic circuits *en masse*. The prospect of adapting such a powerful existing technology for the purpose of creating the first quantum computer is therefore enticing.

The basic process of creating an integrated circuit is to first prepare a wafer of a material—usually silicon—and then imprint an electronic circuit directly upon it. SICs are produced similarly. The main difference is the choice of materials: if a circuit is constructed from superconducting materials and operated at near-zero temperature, it becomes possible to observe and exploit quantum mechanical effects. It is therefore possible in principle to design SIC-based quantum computers.

The defining feature of superconducting materials is that they conduct electricity with *precisely* zero resistance. The microscopic theory of Bardeen et al. (1957) (Nobel 1972) appears to fully explain the effect in the sort of materials used in modern quantum computing prototypes by demonstrating that, at very low temperatures, the collective behaviour of the charge carriers (called Cooper pairs because they are entangled pairs of electrons) are well-modelled by a single wavefunction that assigns complex scalar values (as opposed to vector values) to each spacetime coördinate. This wavefunction purports to give quantum amplitudes for a single, indivisible, fluid-like entity.

17

Such a "macroscopic" wavefunction exhibits a crucial phenomenon when interacting with thin barriers: the Josephson effect (Nobel 1973). The Josephson effect concerns the interaction of a supercurrent with a thin insulating barrier separating two superconducting materials. Whereas a thick insulating barrier functions as a capacitor, Josephson (1962) demonstrated that the current (I) and voltage (V) across a thin barrier (now called a Josephson junction) satisfy the relations

$$I = I_0 \sin \delta, \ V = \frac{\Phi_0}{2\pi} \frac{\mathrm{d}\delta}{\mathrm{d}t}, \tag{2.8}$$

where Φ_0 is the flux quantum $h/(2e)$ (h is Planck's constant and e is the charge of an electron), I_0 is the critical current of the Josephson junction (determined by details of the junction geometry and the materials involved), and δ is the phase difference of the wavefunction across the junction.

The Josephson effect can be divided into two distinct behaviours. The first is the capacative behaviour exhibited by any insulating barrier, and the second is the 'pure' Josephson effect—a nonlinear inductance. The first behaviour can be summarized by defining a capacitance C, which depends on the details of the junction geometry and materials; the second behaviour is derived by computing the time-derivative of I:

$$\frac{\mathrm{d}I}{\mathrm{d}t} = I_0 \cos \delta \frac{\mathrm{d}\delta}{\mathrm{d}t} = 2\pi \frac{I_0}{\Phi_0} V \cos \delta, \tag{2.9}$$

which, by Faraday's law of induction $V = -L\frac{\mathrm{d}I}{\mathrm{d}t}$ for some inductance L, suggest the definition of the Josephson inductance

$$L_{\mathrm{J}} = \frac{\Phi_0}{2\pi I_0 \cos \delta} = \frac{\Phi_0}{2\pi \sqrt{I_0^2 - I^2}} \ (I < I_0). \tag{2.10}$$

The relative strength of the capacative and inductive behaviour can be quantified by considering the work done by each. The work done by the Josephson junction can be calculated by integrating the power

$$P = IV = \frac{\Phi_0 I_0}{2\pi} \sin \delta \frac{\mathrm{d}\delta}{\mathrm{d}t} = -E_{\mathrm{J}} \frac{\mathrm{d} \cos \delta}{\mathrm{d}t} \tag{2.11}$$

with respect to time, where E_{J} is the Josephson energy. However, the work required to bring the charge difference across a capacitor from 0 to Q is $Q^2/(2C)$. Thus the capacitive charging energy of the Josephson junction,

$$E_{\mathrm{C}} := \frac{(2e)^2}{2C}, \tag{2.12}$$

can be compared to the Josephson energy to distinguish between two extremal operating regimes: either $E_J \gg E_C$ or $E_J \ll E_C$. In the first instance, the pure Josephson element dominates the physics; the capacitive behaviour dominates in the second.

Josephson junctions are the main component of superconducting "qubits", a term of art that should be used with some care to avoid category errors: a qubit, like a bit, is a unit of information whereas a superconducting qubit is a physical device. Each superconducting qubit is used to store and manipulate one qubit. According to Zagoskin and Blais (2007), these devices can be distinguished into three types: charge qubits, flux qubits, and phase qubits. Charge qubits, of which the currently popular transmons are one type, are superconducting qubits in which the Josephson junctions operate within the capacitive regime. Flux and phase qubits operate in the Josephson regime.[†]

Superconducting qubits are designed to behave as anharmonic oscillators. The lowest two energy levels (the 'ground' and 'excited' states) are treated as the computational basis states ($|0\rangle$ and $|1\rangle$), and information can be manipulated by applying microwave pulses carefully designed to enact various unitary operators. Thus a superconducting qubit behaves as a dipole. Crucially, superconducting qubits can couple to other dipoles in the environment. Such coupling introduces a particularly harmful kind of noise, which I describe in the next section.

Two-Level Defect Systems   §**2.2.2**

Superconducting qubits are dipoles—electric or magnetic, depending on the value of E_J/E_C. The dipole can be aligned or anti-aligned with an ambient field, or it can be in a superposition of these energy eigenstates. Thus the state of the dipole encodes a single qubit of information. Ideally, a dipole that has been initialized to some state and then left alone will remain in its initial state indefinitely. In reality, the dipole interacts with its environment.

All dipole-environment interactions are deleterious for quantum information processing using superconducting qubits, and there are many kinds of interaction to consider. In this thesis, I am focussed on one kind of interaction: the interaction between two dipoles.

[†]The distinction between flux and phase qubits is how they are *biased*. The dynamics of the Josephson junction phase variable in a phase qubit approximately follows that of a quantum in a 'tilted washboard' potential (the sum of a linear and a sinusoidal term), whose tilt is dictated by the value of a bias current. By contrast, flux qubits are voltage biased. The lowest two energy eigenstates of a flux qubit are persistent circulating currents in opposite directions whose degeneracy is lifted by a non-zero parallel voltage. For more information, see the review of Zagoskin and Blais (2007) or of Clarke and Wilhelm (2008).

Certain dipole-dipole interactions are desirable; two-qubit gates are implemented by controlling such interactions. However, it has been experimentally demonstrated that unwanted dipoles can and do influence the dynamics of superconducting qubits. A major aim of my thesis is to characterize this kind of interaction.

The source of unwanted dipoles in SICs is uncertain, but a recent investigation of Lisenfeld et al. (2015) gives compelling evidence that they are lattice defects of the dielectric that comprises the Josephson junction. Lisenfeld et al. (2015) study a sample in which the superconducting material is aluminium and the dielectric is aluminium oxide, an amorphous solid. Amorphous solids are called amorphous because the atomic nuclei maintain a highly disordered arrangement even at low temperatures in contrast to the highly ordered arrangement of nuclei in crystalline solids. The hypothesis investigated by Lisenfeld et al. (2015) is that the position of some nuclei in the dielectric are metastable and can therefore quantum-tunnel between two distinct positions. The resulting dipole can have an energy gap comparable to the qubit energy gap, and therefore the two dipoles can become coupled. The theory is illustrated in Figure 2.1.

Whatever their source, careful experiments by Neeley et al. (2008) demonstrate that these systems, which I henceforth call "two-level defect systems" (TLDSes), can and do resonate coherently[‡] with superconducting qubits. That is, a qubit can coherently exchange an excitation with a TLDS if the qubit and TLDS are on-resonance. TLDSes are a problem for SIC-based quantum information processing technology because coherent energy exchange between qubits and unwanted oscillators can wreak havoc on the time-evolution of the error probability. Such coherent energy exchange can cause the state of the qubit at time t to be highly anti-correlated with the qubit state at time t' even though the correlation between the qubit state at times t and t' is high under ideal circumstances. In other words, the probability of an error can become extremely high at certain times if a TLDS is present.

Coherent exchange between qubits and TLDSes were studied in great detail by Shalibo et al. (2010), who use a technique that Mariantoni et al. (2011) call swap spectroscopy. A sketch of the swap spectroscopy procedure is presented as Algorithm 2.1. The idea is as follows. After selecting a qubit frequency ν (so the energy gap is $\hbar\nu$) and a 'wait' time t,

1. the system is cooled to its ground state,

2. the qubit is excited,

3. the system is allowed to evolve, and

[‡] The "coherence" of a physical process is the extent to which it exhibits spatio-temporal correlations.
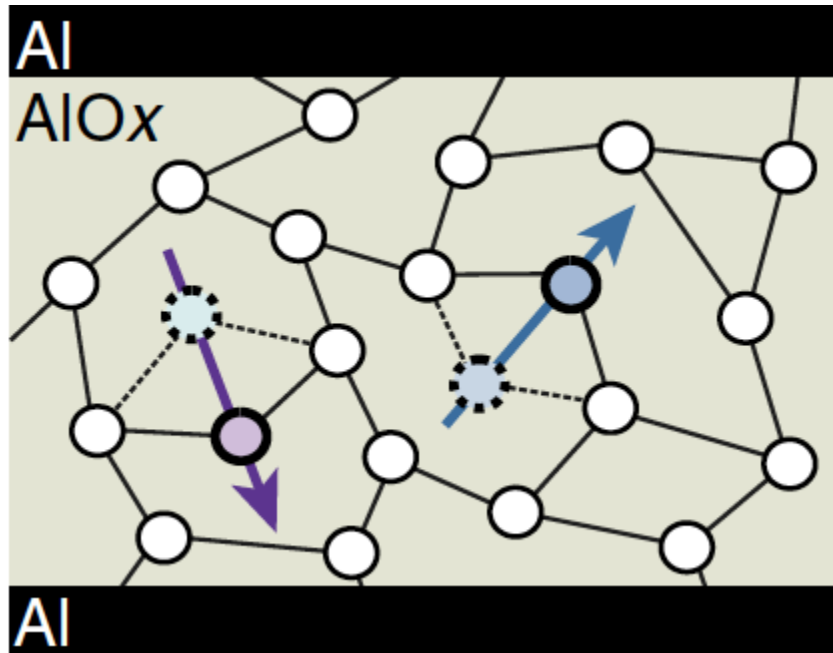
Figure 2.1: Illustration of the metastable nuclei theory of two-level defect systems. The black bands on the top and bottom represent the aluminium superconductor and the grey-green middle part represents the aluminium-oxide layer that comprises the Josephson junction. The nuclei locations are represented with white circles, with two metastable nuclei highlighted purple and blue. Even at low temperatures, the metastable nuclei could oscillate between their depicted locations and the dashed alternate locations, giving rise to an electric dipole indicated by the arrows. This image is reproduced from Figure 1 of Lisenfeld et al. (2015).

4. the qubit is measured.

This procedure exploits the high-quality control of the qubit to collect data about the interaction between the qubit and its environment, effectively treating the qubit as an environment probe. In practice, a dense grid of input values (ν, t) is chosen and Algorithm 2.1 is executed many times for each choice of (ν, t). The resulting 'clicks' are tallied, and the resulting data is plotted as in Figure 2.2. Coherent exchange of energy between the qubit and environmental oscillators is made apparent by the characteristic chevron patterns. The oscillations may represent resonance with an unwanted TLDS or with a wanted on-chip microwave resonator.

Algorithm 2.1 The swap spectroscopy protocol.

  Input: qubit frequency, ν.
  Input: wait time, t.
  Output: data bit, b.
  function SWAPSPECTROSCOPY(ν, t)
      Cool entire system to its ground state.
      Set qubit energy gap to $\hbar\nu$.
      Excite the qubit; total system has state $|\text{excited}\rangle_{\text{qubit}} |\text{ground}\rangle_{\text{environment}}$.
      Wait for time t.
      Measure the qubit in the basis $\left\{ |\text{ground}\rangle_{\text{qubit}}, |\text{excited}\rangle_{\text{qubit}} \right\}$.
      if measure $|\text{ground}\rangle_{\text{qubit}}$ **then**
         $b \leftarrow 1$ $\triangleright$ Click.
      else
         $b \leftarrow 0$ $\triangleright$ No click.
      end if
      return b
  end function

    Unlike the other algorithms in this thesis, Algorithm 2.1 cannot, strictly speaking, be run entirely on a computer. The algorithm instead specifies physical operations, rather that mathematical ones. If a computer is responsible for carrying out these operations, the computer must have interactive access to an experimental setup. Of course it is possible to *simulate* these operations on a computer; I explain how to do this in Chapter 4. The goal of such simulations is to reproduce effects like that displayed in Figure 2.2.

    Figure 2.2 demonstrates an effect that is the central motivation for this thesis. The data, provided to me by Erik Lucero, is produced via swap spectroscopy on the device
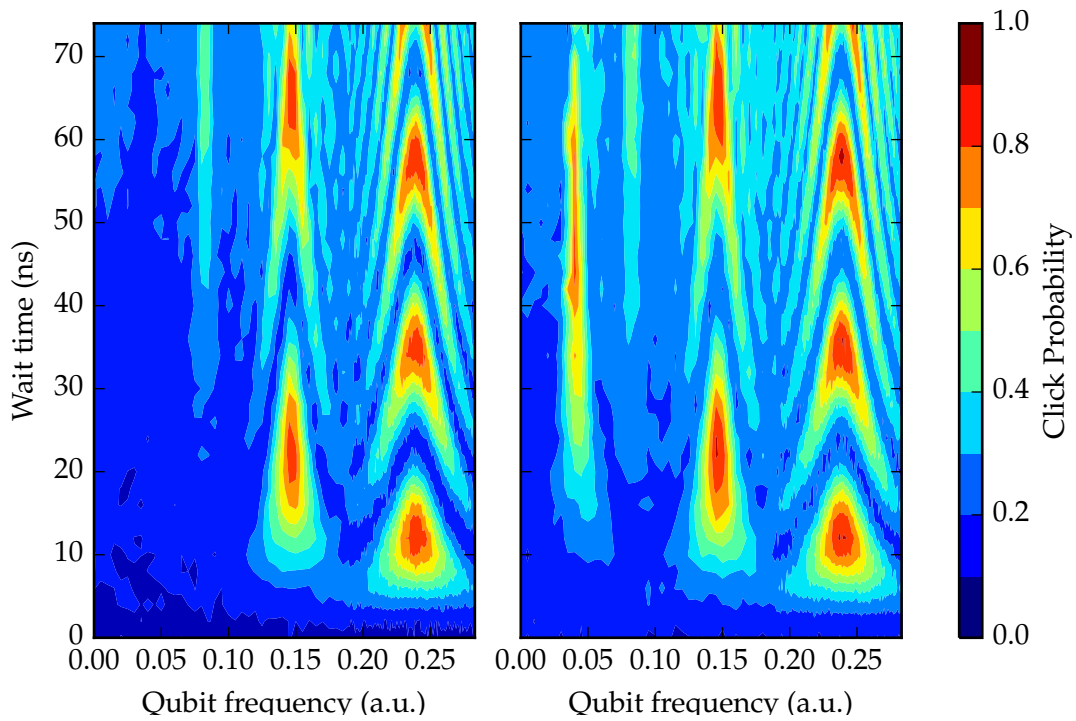
Figure 2.2: Plots of swap spectroscopy data provided to me by Erik Lucero. The data on the left was gathered from the sample investigated by Lucero et al. (2012) on 29 July 2011 (left) and on 30 July 2011 (right). The chevron patterns indicate oscillators as defined in Chapter 4, though two of these are resonances with microwave stripline resonators. The sudden appearance of the chevron patteron on the left indicates that a TLDS has appeared and is interacting with the qubit.

investigated by Lucero et al. (2012). The figure represents data collected on sequential days from the same sample. Although most of the 'swap spectrum', as I call it in Chapter 4, remains qualitatively unchanged, a new feature appears on the left. This feature represents a TLDS that has appeared seemingly from nowhere. The data shows that TLDSes can appear and disappear without warning. It is not known how frequently and how quickly this occurs.

Because TLDSes have such a deleterious effect on quantum information processing, their influence must be rigorously assessed in prototypical quantum computers. The device investigated by Lucero et al. (2012) is such a prototype, built from a kind of flux qubit called a "phase" qubit, yet no thorough statistics yet exist for TLDSes in these systems. Anecdotally, TLDSes are a frequent issue for phase qubits.

Another type of superconducting qubit called a "transmon", which is a kind of charge qubit, has become popular more recently. The promise of transmons for quantum computing has been described by Barends et al. (2013) and somewhat demonstrated by Barends et al. (2014). Transmons appear to interact with TLDSes less frequently and strongly, though there is some indication that they do occur. A central aim of my thesis is to develop techniques to detect and describe even such weakly coupled TLDSes.

Throughout this thesis, I make use of a phenomenological model for the interaction between qubits and TLDSes. In this model, qubits and TLDSes are described as aligned ferromagnetic spins. That is, each qubit or TLDS can be found in one of two energy eigenstates ('ground' or 'excited') and are in an energy eigenstate if they are aligned (both are in the ground state or both are excited). If they are not aligned, then the excitation is continually exchanged between the qubit and the TLDS. I show in Chapter 4 that this model reproduces Figure 2.2.

A crucial property of this model is that it conserves excitations. That is to say, if a qubit/TLDS in the excited state contains one excitation and zero excitations if in the ground state, then the total number of excitations remains constant under the evolution described by the ferromagnetic spin model. As the SWAPSPECTROSCOPY procedure demands that the initial number of excitations is one, the derivation of the swap spectrum in Chapter 4 proceeds by projecting the dynamics onto the single-excitation subspace of the larger Hilbert space containing all possible configurations of the qubit-TLDS system.

Experimental Design §2.3

My approach for solving the TLDS characterization problem is to provide algorithms that designs and executes useful experiments. Such evaluative judgements of exper-

iments require objective criteria by which to compare different possible experiments. These criteria should be based on *informativeness*, which is to say that good experiments appreciably increase knowledge whereas bad experiments do not. In this thesis, the goal is to learn about TLDSes; specifically, to choose a model that accurately predicts the outcome of future experiments. The utility of collected experimental data is therefore to be evaluated in terms of the performance of an *estimator* (i.e. a function from the space of possible data to the space of parameters).

The aim of this section is to explain the framework for designing and evaluating experiments. Section 2.3.1 explains how experimental data is used to upgrade knowledge about unknown model parameters, which involves the use of Bayes' theorem, and how to evaluate the result in terms of the statistics of the error in the resulting estimate. Section 2.3.2 explains the concept of particle representations and how they are used to calculate estimates of unknown model parameters based on collected experimental data. The framework of this chapter is essential for the formal problem statements found in Section 4.3 and for the results presented in Chapter 6.

Utility of Experiments §**2.3.1**

The purpose of this subsection is to define the utility of an experiment, which requires a mathematical definition of an experiment. These definitions are explained by analogy with a random process of everyday experience: the flip of a coin. The random nature of coin tosses can be modelled using a probability distribution that is specified by a bias parameter p, and then explain that a state of knowledge about an unknown value of p can be represented by a probability density functional (pdf)—specifically, a Beta distribution. The Beta distribution summarizes the state of knowledge about the bias parameter p; the knowledge can be improved by performing trials of the random process (i.e. coin flips), recording the outcomes (heads or tails), and using this data together with Bayes' theorem to update the pdf. A coin flip is thus an experiment to learn the parameter p; indeed, it is the only possible experiment for this problem. By contrast, empirical model selection problems can be solved using a variety of different experiments that are not equally useful. Here the utility of an experiment is defined to be the extent to which it is expected to reduce the variance of the posterior distribution.

A coin is a thin cylinder of metal that fits easily into the palm of a hand. The sides of the coin are called 'heads' and 'tails', so if the coin is placed on a table then the state of the coin can be represented by specifying which side, 'heads' or 'tails', is visible. Coin flipping refers to the process of throwing the coin into the air so that it revolves many

times before landing on the table. If the flipping process is chaotic enough, the outcome is indeterministic: sometimes 'heads' will show and sometimes 'tails'. The coin flip is unbiased if the probability of showing 'heads' is equal to the probability of showing 'tails'; the flip is biased if the probabilities are unequal. This kind of two-outcome trial is called a Bernoulli trial, which technically refers to a random variable with two possible outcomes, 0 or 1.

There is always uncertainty about the probability distribution that describes any real coin flip (hereafter: Bernoulli trial). Whereas the probability distribution is entirely specified by the parameter p, an infinite number of trials is necessary in principle to specify p with perfect precision. In practice, only a finite set of data is available for analysis and any such finite data set is only evidence that some values of p are more likely than others.

The problem of inferring the unknown parameter p can be solved as follows. Jeffreys (1946) argues that the initial uncertainty about p can be represented using what has come to be known as the Jeffreys prior; in this case the Jeffreys prior is Beta $\left(\frac{1}{2}, \frac{1}{2}\right)$, where the pdf of Beta$(\alpha, \beta)$ is

$$\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1 - p)^{\beta-1} \tag{2.13}$$

and Γ represents the gamma function

$$\Gamma(t) := \int_0^\infty x^{t-1} \exp(-x) \mathrm{d}x. \tag{2.14}$$

Given some set of data $[b_1, \ldots, b_N]$ ($b_k = 0$ or 1) obtained from several independent and identically distributed Bernoulli trials, the knowledge of the unknown parameter p may be updated using Bayes' theorem:

$$\Pr\left(p \middle| \mathrm{data}, \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)\right) = \frac{\Pr\left(\mathrm{data} \middle| p, \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)\right)}{\Pr\left(\mathrm{data} \middle| \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)\right)} \Pr\left(p \middle| \mathrm{Beta}\left(\frac{1}{2}, \frac{1}{2}\right)\right), \tag{2.15}$$

which turns out to be the pdf of Beta $\left(N_0 + \frac{1}{2}, N_1 + \frac{1}{2}\right)$, where

$$N_1 = \frac{1}{N} \sum_{k=1}^N b_k; \ N_0 = 1 - N_1. \tag{2.16}$$

In the above example, the 'atomic' experiment is a single flip of a coin. More generally, one could define an experiment to be N flips of a coin for any natural number N; in this

case, the random variable is binomially distributed rather than Bernoulli distributed. Clearly, a larger value of N specifies a more useful experiment. This observation can be formalized and extended by recognizing that larger values of N ensure that the posterior distribution has a smaller variance: the variance of the Beta distribution is

$$\text{VAR}(\text{Beta}(\alpha, \beta)) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)},\tag{2.17}$$

so N flips of the coin yield a posterior variance of

$$\frac{(N_0 + \alpha)(N_1 + \beta)}{(N + \alpha + \beta)^2(N + \alpha + \beta + 1)} = O\left(\frac{1}{N}\right).\tag{2.18}$$

Therefore, a larger number of coin flips is expected to produce a smaller posterior variance.

There are of course other ways to define the precision of a distribution. The reason for the focus on the variance is that it is amenable to the particle representation techniques presented in Section 2.3.2.

In general, an experiment is a random variable whose distribution contains information about an underlying model about which information is sought. The data from such experiments can be aggregated and used as input to Bayes' theorem to update knowledge about uncertain model parameters. The utility of the data can be measured by the resulting posterior variance, which can in principle be estimated by calculating the expected posterior variance of the experiment. The problem of experimental design is therefore to choose experiments that will produce data that create significant reductions in the posterior variance when that data is used as an input to Bayes' theorem.

Particle Representations §**2.3.2**

A particle representation of a pdf is specified by a set of "locations", which are possible outcomes of the random process. In the coin flip example, a location is simply a choice of p. If the output of a random process is some vector x and $\{x_1, \ldots, x_N\}$ is a set of locations, then a particle representation of the pdf would be specified by

$$\{\delta(x - x_1), \ldots, \delta(x - x_N)\},\tag{2.19}$$

where $\delta(\bullet)$ represents a Dirac delta function. The particle representation approximates
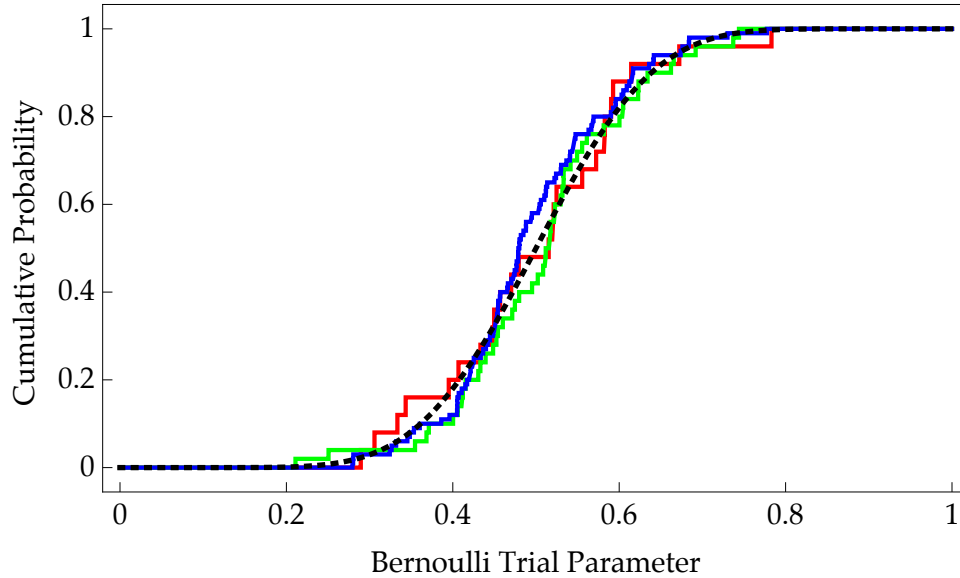
Figure 2.3: Some particle representations of the distribution Beta $\left(N_0 + \frac{1}{2}, N_1 + \frac{1}{2}\right)$, where $N_0 = N_1 = 10$. This plot depicts the cumulative distribution function; i.e. $\Pr\left(p < x \middle| p \sim \text{Beta}\left(N_0 + \frac{1}{2}, N_1 + \frac{1}{2}\right)\right)$, where x is the abscissa variable of the plot. The particle locations are chosen randomly according to the same Beta distribution. The red/green/blue curve depicts the approximation from 25/50/100 particles; the true curve is black and dotted. The pdfs corresponding to each cumulative distribution function are simply their respective derivatives; the true distribution is continuously differentiable on the interval $(0,1)$ but the derivatives of the particle-representation-based approximations are either zero or singular.

the pdf well if, for any reasonable[§] function f,

$$\int_{\text{region}} f(x)\, \text{pdf}(x)\mathrm{d}x \approx \frac{1}{N} \sum_{k=1}^{N} \int_{\text{region}} f(x)\delta\left(x - x_k\right)\mathrm{d}x, \tag{2.20}$$

where the integrals in the summand evaluate to $f(x_k)$ if x_k is in the region and evaluate to 0 otherwise.

The importance of the particle representation is that it enables computations. Whereas the usual abstract mathematical constructs of probability theory include non-algorithmic

---

[§] It is always possible to create pathological functions. Numerical analysis involving particle filters is trustworthy only if these pathological cases are highly improbable.

notions such as real numbers and function equivalence up to Lebesgue-measure-zero sets, particle representations are always finite objects and therefore enable an algorithmic approach to statistical inference. Perhaps the most important computation is also one of the simplest: the calculation of the mean value of a particle filter, as defined in Algorithm 2.2.

Algorithm 2.2 The mean location of a particle representation.

Input: Particle representation, $\{x_k\}_{k=1}^n$.
Output: Mean particle location, μ.
function MEAN($\{x_k\}$)
    $\mu \leftarrow \sum_k x_k / n$
    return μ
end function

Algorithm 2.2 defines MEAN, which is a function of particle representations that yields the average particle location μ. The average location μ serves as an estimate of the mean value of the pdf that the particle representation is meant to approximate. It is also possible to estimate the covariance matrix of a pdf by calculating the covariance matrix of its particle representation as in Algorithm 2.3.

Algorithm 2.3 The covariance matrix of a particle representation.

Input: Particle representation, $\{x_k\}_{k=1}^n$.
Output: Covariance matrix, Σ.
function COV($\{x_k\}$)
    $\mu \leftarrow$ MEAN($\{x_k\}$) $\triangleright$ Algorithm 2.2.
    $\Sigma \leftarrow \left(\sum_k x_k \cdot x_k^{\mathrm{T}} / n \right) - \mu \cdot \mu^{\mathrm{T}}$ $\triangleright$ $\bullet^{\mathrm{T}}$ represents the transpose of a vector $\bullet$.
    return Σ
end function

The next algorithm, Algorithm 2.4, is the first of two steps needed to perform updates to a particle representation based on a data set and Bayes' theorem. Bayesian updates of pdfs require a computation of the likelihood function

$$\mathcal{L}(x) := \frac{\Pr\left(\mathrm{data}|x, \mathrm{prior}\right)}{\Pr\left(\mathrm{data}|\mathrm{prior}\right)}, \tag{2.21}$$

which is proportional to $\Pr\left(\mathrm{data}|x, \mathrm{prior}\right)$. Algorithm 2.4 returns weighted likelihoods evaluated at each particle location.

Algorithm 2.4 Calculate likelihoods of particles.

> **Input:** Likelihood function, $\mathcal{L}$. $\qquad\qquad\qquad\qquad \triangleright \mathcal{L}(x) \in [0,1]$ for any x.
> **Input:** Particle representation, $\{x_k\}_{k=1}^n$.
> **Output:** Particle weights, $\{w_k\}$.
> **function** WEIGHTS$(\mathcal{L}, \{x_k\})$
>   **for** $k \leftarrow 1, \dots, n$ **do**
> $w_k \leftarrow \mathcal{L}(x_k)$
>   **end for**
> $W \leftarrow \sum_k w_k$
>   **for** $k \leftarrow 1, \dots, n$ **do**
> $w_k \leftarrow w_k / W$
>   **end for**
>   **return** $\{w_k\}$
> **end function**

The output of Algorithm 2.4 thus represents the relative likelihood of each particle in the particle representation according to some specified likelihood function. This likelihood function $\mathcal{L}(x)$ expresses the probability that a specified data set was produced by the model with parameters x, a number which derives its meaning relative to the likelihood of other parameter vectors x'. The resulting weights must ultimately be used to choose a particle representation for the posterior distribution obtained by applying Bayes' theorem to the data set. Though there are many possible resample algorithms, I use only a standard resampler of Liu and West (2001) as expressed in Algorithm 2.5.

The case of the Beta distribution is computationally tractable without the need for particle representations, but such tractability is the exception and not the norm. Usually, posterior pdfs become exceedingly complex as data accumulates. The value of particle representations is that computations such as RESAMPLE$(\{x_k\}, \text{WEIGHTS}(\mathcal{L}, \{x_k\}))$ have a space cost proportional to the size of the particle representation. This computational tractability is crucial for programs such as the one I introduce in Chapter 6.

I have not discussed one aspect of the particle representation: the possibility of sample degeneracy. If the particle representation is to accurately represent some pdf, then there should be few particles in regions where the weight of the particle turns out to be small. If too many particles are in such regions, the particle representation implies too high a value to the pdf in that region. Thus it is occasionally necessary to resample the particles.

The decision about whether to resample the particles can be made by computing the

---
Algorithm 2.5 The particle resampling algorithm of Liu and West (2001).

Input: Particle representation, $\{x_k\}_{k=1}^n$.
Input: Particle weights, $\{w_k\}_{k=1}^n$.   ▷ Algorithm 2.4
Output: New locations, $\{y_k\}$.
function RESAMPLE($\{x_k\}, \{w_k\}$)
     $a \leftarrow 0.98$ ▷ Other values are possible, but this choice is standard.
     $\mu \leftarrow$ MEAN($\{x_k\}$)   ▷ Algorithm 2.2
     $\Sigma \leftarrow (1 - a^2)$COV($\{x_k\}$)   ▷ Algorithm 2.3
     for $k \leftarrow 1, \ldots, n_{\text{ptcls}}$ **do**
         Draw ℓ with probability w_ℓ.
         $\mu_\ell \leftarrow a x_\ell + (1 - a)\mu$
         Draw y_k from normal(μ_ℓ, Σ).
     end for
     return $\{y_k\}$
end function
---

effective particle number. If the weights of the particles are $\{w_k\}_{k=1}^n$, then the effective particle number is

$$n_{\text{eff}} := \frac{1}{\sum_k w_k^2}. \tag{2.22}$$

If the effective particle number falls below some threshold, then the issue should be resolved by resampling: $\{x_k\} \leftarrow$ RESAMPLE($\{x_k\}, \{w_k\}$). Numerical results in this thesis are produced with code that prevents sample degeneracy using this technique, which is developed and published as the QInfer project of Granade et al. (2016). I have made some minor contributions to this project, mostly through testing. The numerical results of Chapters 5 and 6 are produced using QInfer.

THREE

OPTIMAL BOUNDS ON LOGICAL ERROR RATE

Error is viewed, therefore, not as an extraneous and misdirected or misdirecting accident, but as an essential part of the process under consideration.

JOHN VON NEUMANN

E XPERIMENTAL advances in quantum computing are often quoted in terms of average gate fidelities, which now reportedly[*] exceed 99.9% for single-qubit gates and 99% for two-qubit gates in SIC-based quantum computing prototypes. Here I explain how to use reported average gate fidelity to determine bounds on the error rate of a quantum computational gate and thereby compare reported fidelities to fault-tolerance thresholds.

Given that the average gate fidelity is φ, Wallman and Flammia (2014) showed that the error rate η can be bounded as follows. Magesan et al. (2012a) showed that the error rate of a gate of dimension d[†], if subject only to Pauli noise, is

$$\eta^{\text{Pauli}} := \left(1 + \frac{1}{d}\right)(1 - \varphi).\tag{3.1}$$

---

[*]These numbers are advertised by Barends et al. (2014) and anticipated by Chow et al. (2012).

[†]A gate is said to have dimension d if the dimension of the input Hilbert space is d. In particular, an n-qubit gate has dimension 2^n.

Wallman and Flammia (2014) showed that the error rate η of gates in a circuit subject to arbitrary local noise satisfies

$$\eta^{\text{Pauli}} \leq \eta \leq d\sqrt{\eta^{\text{Pauli}}}. \tag{3.2}$$

The work of Magesan et al. (2012a) shows that the lower bound is tight, but it is not known if the upper bound is tight.

The main result of this chapter, published as Sanders et al. (2016), is a demonstration that the upper bound is an asymptotically tight approximation to the tightest possible upper bound on error rate in terms of fidelity. Although this bound is unlikely to be saturated by experimental noise, we demonstrate, using explicit examples, that the bound indicates a realistic deviation between the true error rate and the reported average fidelity. This result is unfortunate because the square-root ensures that superficially impressive gate fidelities do not guarantee low error rates. A two-qubit gate with 99% fidelity is, for example, only guaranteed to have an error rate below 45%. Indeed, I present an explicit example of a two-qubit gate with fidelity 99% but an error rate slightly under 13%.

The message of this work is not that fault-tolerance thresholds remain far out of reach for quantum computer prototypes. Rather, the message is that average fidelity is not an appropriate metric for measuring progress towards fault-tolerance. Average fidelity, though experimentally accessible, fails to capture important information about the effect of noise on logical error rates: additional information about noise is *required*. We introduce the Pauli distance δ^{Pauli} as a measure of the deviation between the true error rate and η^{Pauli}, and we derive tighter bounds on the error rate in terms of φ and δ^{Pauli}.

As the first author of Sanders et al. (2016), I take primary responsibility for its content. My contributions were as follows. I recognized that average gate fidelity is meaningful as a measure of the quality of a quantum gate if it can be used to bound the error rate, which is the subject of fault-tolerance thresholds. In collaboration with Barry Sanders, I developed an operational interpretation of the error rate of a quantum logic gate and used this to argue that fault-tolerance thresholds are, as currently written, bounds on the error rate as defined in Definition 2.2. Having become aware of the bounds of Wallman and Flammia (2014), I conjectured that their upper bound is tight and I sought to prove this. Though I failed, I proved that the dimensional scaling is optimal. I discussed this result with Joel Wallman, who believed that the scaling of error rate with respect to fidelity must also be optimal. We jointly proved this optimal scaling.

I also recognized the need for additional noise measures. Being aware of the work of Magesan et al. (2012a), I reasoned that approximately Pauli noise should nearly saturate the lower bound. I therefore defined the Pauli distance and derived the bounds presented as Propositions 3.3 and 3.4. I performed the numerical simulations for Figure 3.1 and prepared the manuscript for publication.

This chapter is organized as follows. Section 3.1 presents the average gate fidelity as an experimentally convenient performance metric that nonetheless lacks a clear operational meaning. The relationship between average gate fidelity and the error rate of a quantum gate is clarified with Example 3.1, which shows that fidelity can grossly underestimate the effect of important noise sources. Section 3.2 presents Propositions 3.1 and 3.2, which demonstrate the optimal scaling of the upper bound of Wallman and Flammia (2014) with respect to φ and d, respectively, if the other is held constant. Proposition 3.1 further demonstrates that any constant prefactors must be close to unity. Section 3.3 presents the Pauli distance as a measure of the deviation between average fidelity and error rate. Finally, Section 3.4 uses the results of this chapter to give an assessment of fidelity-based reports of progress towards fault-tolerant quantum computing. The content of this chapter is adapted from Sanders et al. (2016).

Average Gate Fidelity §3.1

Average gate fidelity has become a standard performance metric for experimental efforts towards practical quantum computers. Its popularity is due in a large part to the existence of scalable estimation methods like randomized benchmarking. But average gate fidelity is quite distinct from the gate error rate, and this important distinction is not always respected in the literature. The purpose of this section is to explain average fidelity and, most importantly, to explain what it is not.

Average fidelity defines a metric, in the mathematical sense, on the space of quantum channels. That metric is topologically equivalent to the error rate, so high fidelity implies that error rates are low. But average fidelity is not a function of error rate: two gates with equal fidelities can have different error rates. Furthermore, average fidelity is not a good proxy for the error rate because the error rate of two gates with equal fidelities can differ by orders of magnitude. The sense in which high fidelity implies low error rate is a qualitative statement only. High fidelity is, at best, an *indication* that error rates are low.

The average gate fidelity is defined as the average state fidelity of the output of an imperfect quantum gate to its ideal. The fidelity of a state ρ to a pure state $|\psi\rangle$ is defined to be $\mathrm{Tr}\left(|\psi\rangle\langle\psi|\rho\right)$, so the fidelity of the output of an implemented gate $\mathcal{G}_{\mathrm{ac}}$ to the output

of the ideal gate $\mathcal{G}_{\text{id}}$ for a given input state $|\psi\rangle$ is

$$\text{Tr}\left(\mathcal{G}_{\text{id}}(|\psi\rangle\langle\psi|)\mathcal{G}_{\text{ac}}(|\psi\rangle\langle\psi|)\right) = \langle\psi|\,\mathcal{D}_G\left(|\psi\rangle\langle\psi|\right)|\psi\rangle. \tag{3.3}$$

Averaging over pure state inputs with respect to the Haar measure yields the average gate fidelity

$$\varphi := \int \mathrm{d}\mu(\psi)\,\langle\psi|\,\mathcal{D}_G\left(|\psi\rangle\langle\psi|\right)|\psi\rangle. \tag{3.4}$$

The popular randomized benchmarking protocol of Emerson et al. (2005) produces an estimate of this quantity averaged over a gate-set, though extensions such as that proposed by Magesan et al. (2012b) produce estimates of the average gate fidelity for individual gates.

The state infidelity can be interpreted as the error rate of a particular measurement as follows. First define for each pure state $|\psi\rangle$ the POVM $\{|\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi|\}$; the outcome of this measurement applied to a state ρ will be $|\psi\rangle\langle\psi|$ with probability $\langle\psi|\rho|\psi\rangle$, which is the state fidelity of ρ with respect to $|\psi\rangle$. Therefore, the state fidelity of ρ with respect to $|\psi\rangle$ is equal to the bias parameter of the Bernoulli trial specified by the POVM. The ideal bias parameter is of course 1, and the total variation distance between Bernoulli distributions with bias parameters p and q is simply $|p - q|$. Therefore, the total variation distance between the actual and ideal measurement statistics for the POVM applied to ρ is the state infidelity, $1 - \langle\psi|\rho|\psi\rangle$. However, this interpretation of the state infidelity as an error rate does not extend to the average infidelity $1 - \varphi$ because the integral of Eq. (3.4) does not admit a fixed measurement basis such as that of the POVM $\{|\psi\rangle\langle\psi|, \mathbb{1} - |\psi\rangle\langle\psi|\}$. Therefore, the average gate infidelity is not an average error rate.

To clarify the relationship between average gate fidelity and error rate, consider two noise processes on a single qubit: a depolarizing noise process

$$\mathcal{E}_r^{\text{dep}}(\rho) := (1 - r)\rho + rI/2, \tag{3.5}$$

where I is the identity operator on $\mathcal{Q}$, and a unitary noise operator

$$\mathcal{E}_\theta^{\text{U}}(\rho) := U\rho U^\dagger, \tag{3.6}$$

where U is some unitary operator on $\mathcal{Q}$ with eigenvalues $\exp(\pm i\theta)$ for $0 \leq \theta \leq \pi$. The average gate fidelity for depolarizing noise is

$$\varphi^{\text{dep}}(r) = 1 - \frac{r}{2} \tag{3.7}$$

whereas a theorem of Nielsen (2002) implies that the fidelity of unitary error is

$$\varphi^{\text{U}}(\theta) = \frac{1}{3} + \frac{2}{3}\cos^2\theta. \tag{3.8}$$

By contrast, the error rate for depolarizing noise is

$$\eta^{\text{dep}}(r) = \frac{3}{4}r, \tag{3.9}$$

which follows from the fact that depolarizing noise is Pauli, and

$$\eta^{\text{U}}(\theta) = \sin\theta \tag{3.10}$$

for unitary error by a theorem of Johnston et al. (2009). Therefore,

$$\eta^{\text{dep}} = \frac{3}{2}\left(1 - \varphi^{\text{dep}}\right) \tag{3.11}$$

for depolarizing noise but

$$\eta^{\text{U}} = \sqrt{\frac{3}{2}\left(1 - \varphi^{\text{U}}\right)} \tag{3.12}$$

for unitary error, which means that there is no single function f such that $f(\varphi) = \eta$ for every possible noise channel. As a consequence, fidelity is not a good proxy for error rate. This is demonstrated by the following example.

Example 3.1. Consider a single-qubit gate that is subject to the two noise processes of Eqs. (3.5) and (3.6): depolarizing and unitary. The depolarizing rate is $r = 10^{-3}$, with corresponding fidelity

$$\varphi^{\text{dep}} = 1 - 5.0 \times 10^{-4}, \tag{3.13}$$

whereas the unitary error has angle $\theta = 10^{-2}$, with corresponding fidelity

$$\varphi^{\text{U}} = 1 - 6.7 \times 10^{-5}. \tag{3.14}$$

The combination

$$\mathcal{D}_G := \mathcal{E}_r^{\text{dep}} \circ \mathcal{E}_\theta^{\text{U}} \equiv \mathcal{E}_\theta^{\text{U}} \circ \mathcal{E}_r^{\text{dep}} \equiv (1-r)\mathcal{E}_\theta^{\text{U}} + r\mathcal{E}_{r=1}^{\text{dep}} \tag{3.15}$$

has fidelity

$$\varphi^{\text{tot}} = (1-r)\varphi^{\text{U}} + \frac{r}{2} = 1 - 5.3 \times 10^{-4}, \tag{3.16}$$

36

so the fidelity loss seems to arise mostly from depolarizing noise. Yet the error rate due to unitary error is

$$\eta^{\mathrm{U}} = 10^{-2} \tag{3.17}$$

whereas the error rate due to depolarizing noise is

$$\eta^{\mathrm{dep}} = 7.5 \times 10^{-4}. \tag{3.18}$$

The triangle inequalities imply that the error rate of the combined noise process is

$$\eta^{\mathrm{tot}} = (1 \pm 0.08) \times 10^{-2}, \tag{3.19}$$

so the unitary error is in fact dominating over depolarizing even though the fidelity appears to imply the reverse.

Thus, information beyond fidelity is needed to assess the relative importance of various noise processes affecting a quantum computer; Section 3.3 introduces the Pauli distance as a measure of the discrepancy between fidelity and error rate. Other measures, such as the unitarity of Wallman et al. (2015), may also prove useful for quantifying this discrepancy. In the absence of such additional information, the best known bounds on the error rate in terms of fidelity are those of Wallman and Flammia (2014):

$$\left(1 + \frac{1}{d}\right)(1 - \varphi) \leq \eta \leq \sqrt{(d^2 + d)(1 - \varphi)}. \tag{3.20}$$

Whereas Magesan et al. (2012a) show that the lower bound is tight, the next section shows that the upper bound is probably tight.

Tightness of the Upper Bound §3.2

The purpose of this section is to show that the upper bound η^{ub} given by Wallman and Flammia (2014) scales optimally in each of the variables φ (Section 3.2.1) and d (Section 3.2.2) if the other is fixed. These optimal scaling results are presented as formal mathematical propositions involving asymptotic notation: Propositions 3.1 and 3.2 show that the least upper bound of error rate with respect to average fidelity (Definition 3.1) is asymptotically bounded above and below by the upper bound of Wallman and Flammia (2014). To make these statements precise, some terminology is necessary. The terminology and results are reproduced from Sanders et al. (2016).

Definition 3.1. The *least upper bound of error rate with respect to average gate fidelity* $\eta^{\text{lub}} = \eta^{\text{lub}}(\varphi, d)$ is the unique function of φ and d that satisfies the following. For any discrepancy channel $\mathcal{D}_G$ of dimension d with average gate fidelity φ and error rate η, $\eta^{\text{lub}}(\varphi, d) \geq \eta$. Furthermore, suppose that $\eta^{\text{ub}} = \eta^{\text{ub}}(\varphi, d)$ is any other function with the same property. Then $\eta^{\text{lub}}(\varphi, d) \leq \eta^{\text{ub}}(\varphi, d)$ for all φ and d.

The scaling of $\eta^{\text{lub}}(\varphi, d)$ is established as a function of each variable when the other is fixed. Notationally, I distinguish fixed from variable quantities as follows. If the dimension d is fixed but φ is variable, I write $\eta^{\text{lub}}(\varphi)|_d$; if vice versa, $\eta^{\text{lub}}(d)|_\varphi$. I use similar notation for η^{ub}. The asymptotic scaling statements involve the following two quantities that tend to infinity as the error rate tends to zero.

Definition 3.2. Define the *inverse error rate* of a quantum logic gate to be $\zeta := \eta^{-1}$, where η is the error rate of the gate. Thus, $\zeta \to \infty$ as $\eta \to 0$. Additionally,

$$\zeta^{\text{lub}} := \left(\eta^{\text{lub}} \right)^{-1}, \zeta^{\text{ub}} := \left(\eta^{\text{ub}} \right)^{-1}; \tag{3.21}$$

these are *lower* bounds to ζ.

Definition 3.3. Define the *inverse average infidelity* of a quantum logic gate to be

$$v := (1 - \varphi)^{-1}. \tag{3.22}$$

Thus, $v \to \infty$ as $\varphi \to 1$.

The objective of this section is therefore to compare $\zeta^{\text{lub}} = \zeta^{\text{lub}}(v, d)$ to

$$\zeta^{\text{ub}}(v, d) = \frac{\sqrt{v}}{d\sqrt{1 + \frac{1}{d}}}. \tag{3.23}$$

I show that

$$\zeta^{\text{lub}}(v)|_d \in \Theta(\sqrt{v}) \tag{3.24}$$

and

$$\zeta^{\text{lub}}(d)|_v \in \Theta(d^{-1}); \tag{3.25}$$

thus, ζ^{ub} has optimal scaling with respect to φ and d if the other is fixed.

Definition 3.4. Define the *generalized controlled-phase gate* by

$$\mathcal{G}_{\text{id}}(\rho) := U_\theta \rho U_\theta^\dagger; \; U_\theta := \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & e^{\mathrm{i}\theta} \end{bmatrix}, \tag{3.26}$$

where U_θ is expressed in the computational basis.

Proposition 3.1. *For fixed dimension d,*

$$\zeta^{\text{lub}}(v)|_d \in \Theta(\sqrt{v}). \tag{3.27}$$

Therefore,

$$\zeta^{\text{ub}}(v)|_d \in \Theta\left(\zeta^{\text{lub}}(v)|_d\right). \tag{3.28}$$

Furthermore,

$$\eta^{\text{lub}}(v)|_d \geq \frac{1}{2}(d-1)^{-\frac{1}{2}} \times \eta^{\text{ub}}(\varphi)|_d. \tag{3.29}$$

Proof. Consider an implementation of the generalized controlled-phase gate given by $\mathcal{G}_{\text{ac}}(\rho) = \rho$, the identity channel. By a formula of Nielsen (2002), the average gate fidelity is

$$\varphi = \frac{d + |\text{Tr}(U_{-\theta})|^2}{d + d^2} = 1 - \frac{2(d-1)}{d(d+1)}(1 - \cos\theta). \tag{3.30}$$

By Theorem 26 of Johnston et al. (2009),

$$\eta = \sqrt{\frac{1 - \cos\theta}{2}}; \tag{3.31}$$

hence,

$$\zeta = \sqrt{\frac{4(d-1)}{d(d+1)}} \times \sqrt{v}. \tag{3.32}$$

By contrast,

$$\zeta^{\text{ub}} = \sqrt{\frac{1}{d(d+1)}} \times \sqrt{v}. \tag{3.33}$$

Furthermore, ζ^{lub} is defined so that $\zeta \geq \zeta^{\text{lub}} \geq \zeta^{\text{ub}}$; thus,

$$\sqrt{\frac{4(d-1)}{d(d+1)}} \times \sqrt{v} \geq \zeta^{\text{lub}}(v)|_d \geq \sqrt{\frac{1}{d(d+1)}} \times \sqrt{v}. \tag{3.34}$$

$\square$

Example 3.2. All single-qubit unitary errors satisfy

$$\eta = \frac{1}{2}\eta^{\text{ub}} = \sqrt{\frac{3}{2}(1-\varphi)}. \tag{3.35}$$

If $\mathcal{D}_G(\rho) = U\rho U^\dagger$ for some 2×2 unitary operator U, then the eigenvalues of U can be written as $e^{\pm i\theta/2}$ for some θ. The diamond distance $d_\diamond$ and the fidelity are unitarily invariant, so the error rate of $\mathcal{D}_G$ depends only on θ. Furthermore, $\mathcal{D}_G$ is equivalent to the generalized controlled-phase gate (Definition 3.4) and hence η satisfies Eq. (3.31). Eq. (3.30) therefore implies that $\eta^{\text{ub}} = 2\eta$.

Example 3.3. There exists a two-qubit gate with fidelity 99.0% but error rate 12.9%. Consider the generalized controlled-phase gate (Definition 3.4) acting on two qubits: one target qubit and one control qubit. Setting $\theta = 0.259$, we have $\varphi = 99.0\%$ by Eq. (3.30) and $\eta = 12.9\%$ by Eq. (3.31).

Scaling With Dimension §**3.2.2**

Proposition 3.2. *For fixed fidelity φ,*

$$\zeta^{\text{lub}}(d)|_\varphi \in \Theta\left(d^{-1}\right). \tag{3.36}$$

Therefore,

$$\zeta^{\text{ub}}(d)|_\varphi \in \Theta\left(\zeta^{\text{lub}}(d)|_\varphi\right). \tag{3.37}$$

Proof. Consider a special case of the generalized controlled-phase gate in which $\theta = \pi$, so the unitary U_π has an eigenvalue of -1. In this case, $\|\mathcal{G}_{\text{id}} - \mathbb{1}\|_\diamond = 2$ by Theorem 26 of Johnston et al. (2009). The implementation is

$$\mathcal{G}_{\text{ac}} := (1-\lambda)\mathcal{G}_{\text{id}} + \lambda\mathbb{1}. \tag{3.38}$$

The error rate is

$$\eta = \frac{1}{2} \|\mathcal{G}_{\text{ac}} - \mathcal{G}_{\text{id}}\|_\diamond = \lambda \times \frac{1}{2} \|\mathcal{G}_{\text{id}} - \mathbb{1}\|_\diamond = \lambda. \tag{3.39}$$

By applying the formula of Nielsen (2002) to the Kraus decomposition

$$\left\{ \sqrt{1 - \lambda} I, \sqrt{\lambda} U_\pi \right\} \tag{3.40}$$

of the discrepancy channel $\mathcal{D}_{U_\pi}$:

$$\varphi = \frac{d + (1 - \lambda)|\operatorname{Tr}(I)|^2 + \lambda |\operatorname{Tr}(U_\pi)|^2}{d + d^2} = 1 - \frac{4(d-1)}{d(d+1)} \times \lambda. \tag{3.41}$$

Combining Eq. (3.39) with Eq. (3.41) yields

$$\zeta = \frac{4(d-1)}{d(d+1)} v. \tag{3.42}$$

By definition, $\zeta^{\text{ub}} \leq \zeta^{\text{lub}} \leq \zeta$, which implies

$$\frac{1}{d} \sqrt{\frac{v}{1 + \frac{1}{d}}} \leq \zeta^{\text{lub}} \leq \frac{4(d-1)}{d(d+1)} v. \tag{3.43}$$

For fixed v, define the constants $c_1 = 2^{-\frac{1}{2}}$ and $c_2 = 4v$. Then

$$\frac{c_1}{d} \leq \zeta^{\text{lub}}(d)|_v \leq \frac{c_2}{d}, \tag{3.44}$$

hence $\zeta^{\text{lub}}(d) \in \Theta(d^{-1})$. $\qquad\qquad\square$

Bounding Error Rate With Pauli Distance §3.3

Here I derive improved bounds based on an additional promise about noise. Specifically, I provide alternative lower and upper bounds on error rate in terms of gate fidelity and a quantity called the "Pauli distance". I show that the connection between error rate and gate fidelity is strongly improved if the Pauli distance of the noise process is known. I give numerical examples for two important single-qubit noise processes: amplitude damping and unitary error.

The Pauli distance is defined to be the diamond distance between a quantum channel and its Pauli-twirl. To be precise, recall that the single-qubit Pauli matrices are

$$I := \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ X := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \ Y := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \ \text{and } Z := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}; \tag{3.45}$$

a multi-qubit Pauli matrix is defined as a tensor product of single-qubit Pauli matrices. A Pauli channel is a quantum channel that has a Kraus representation in which each Kraus operator is proportional to a Pauli matrix.

Definition 3.5. The *Pauli-twirl* of an n-qubit channel $\mathcal{E}$ (i.e. $d = 2^n$) is

$$\mathcal{E}^{\mathrm{PT}}(\bullet) := \frac{1}{4^n} \sum_{k=1}^{4^n} P_k^\dagger \, \mathcal{E}\left(P_k \bullet P_k^\dagger \right) P_k, \tag{3.46}$$

where P_k represents a choice of n-qubit Pauli matrix.

Definition 3.6. The *Pauli distance* of a gate implementation with discrepancy channel $\mathcal{D}_G$ is defined to be

$$\delta^{\mathrm{Pauli}} := d_\diamond \left(\mathcal{D}_G, \mathcal{D}_G^{\mathrm{PT}} \right), \tag{3.47}$$

where $\mathcal{D}_G^{\mathrm{PT}}$ is the Pauli-twirl of $\mathcal{D}_G$.

The Pauli-twirl of any channel is a Pauli channel, and the Pauli-twirl of a Pauli channel is the same channel. For any channel $\mathcal{E}$, $\mathcal{E}$ and $\mathcal{E}^{\mathrm{PT}}$ have the same average gate fidelity because the average gate fidelity is a linear functional that is invariant under unitary conjugation of its argument. The minimum error rate η^{min} for a set of channels with fixed fidelity φ is achieved by a Pauli channel: $\eta^{\mathrm{min}} = (1 + 2^{-n})(1 - \varphi)$, where n is the number of qubits. As Nielsen and Chuang (2000) explain, several common sources of noise, such as depolarizing error and dephasing (T_2) processes, are Pauli and therefore have $\delta^{\mathrm{Pauli}} = 0$. But many other sources of noise, such as amplitude-damping processes and unitary errors, are not Pauli and therefore have $\delta^{\mathrm{Pauli}} > 0$.

Proposition 3.3. *The error rate η of an n-qubit gate with gate fidelity φ and Pauli distance δ^{Pauli} satisfies*

$$\left| \delta^{\mathrm{Pauli}} - \eta^{\mathrm{Pauli}} \right| \leq \eta \leq \delta^{\mathrm{Pauli}} + \eta^{\mathrm{Pauli}}. \tag{3.48}$$

Proof. By the triangle inequality,

$$\frac{1}{2} \left\| \mathcal{D}_G - \mathbb{1} \right\|_\diamond = \frac{1}{2} \left\| \mathcal{D}_G - \mathcal{D}_G^{\mathrm{PT}} + \mathcal{D}_G^{\mathrm{PT}} - \mathbb{1} \right\|_\diamond \leq \frac{1}{2} \left\| \mathcal{D}_G - \mathcal{D}_G^{\mathrm{PT}} \right\|_\diamond + \frac{1}{2} \left\| \mathcal{D}_G^{\mathrm{PT}} - \mathbb{1} \right\|_\diamond. \tag{3.49}$$

The left-hand side equals η and the right-hand side equals $\delta^{\text{Pauli}} + \eta^{\text{Pauli}}$. Similarly, the reverse triangle inequality implies $\left|\delta^{\text{Pauli}} - \eta^{\text{Pauli}}\right| \leq \eta$. $\qquad\square$

Proposition 3.3 thus enables bounds to be placed on possible values of η in terms of φ and δ^{Pauli}. Indeed, a variation of this proposition, Proposition 3.4, can be applied to noise channels that have a known structure. Although Proposition 3.4 yields weaker bounds than Proposition 3.3, it might be easier in practice to estimate δ^{Pauli} for individual sources of noise rather than for the overall noise process.

Proposition 3.4. *Suppose $\mathcal{D}_G = \sum_k \mathcal{E}_k$, where each $\mathcal{E}_k$ is some quantum channel. Let δ_k^{Pauli} represent the Pauli distance of $\mathcal{E}_k$. Then the error rate η of $\mathcal{D}_G$ satisfies*

$$\eta \leq \eta^{\text{Pauli}} + \sum_k \delta_k^{\text{Pauli}}. \tag{3.50}$$

Proof. If δ^{Pauli} is the Pauli distance of $\mathcal{D}_G$, Proposition 3.3 implies that

$$\eta \leq \eta^{\text{Pauli}} + \delta^{\text{Pauli}}, \tag{3.51}$$

so only the following needs to be shown:

$$\delta^{\text{Pauli}} \leq \sum_k \delta_k^{\text{Pauli}}. \tag{3.52}$$

As the Pauli-twirl operation on quantum channels is linear, i.e.

$$\mathcal{D}_G^{\text{PT}} = \left(\sum_k \mathcal{E}_k\right)^{\text{PT}} = \sum_k \mathcal{E}_k^{\text{PT}}, \tag{3.53}$$

the triangle inequality can be applied repeatedly to obtain

$$\frac{1}{2}\left\|\mathcal{D}_G - \mathcal{D}_G^{\text{PT}}\right\|_\diamond = \frac{1}{2}\left\|\left(\sum_k \mathcal{E}_k\right) - \left(\sum_k \mathcal{E}_k^{\text{PT}}\right)\right\|_\diamond \leq \sum_k \frac{1}{2}\left\|\mathcal{E}_k - \mathcal{E}_k^{\text{PT}}\right\|_\diamond. \tag{3.54}$$

The left-hand side equals δ^{Pauli} and the right-hand side equals $\sum_k \delta_k^{\text{Pauli}}$. $\qquad\square$

Consider two examples of single-qubit noise processes in which δ^{Pauli} is non-zero: unitary error, which is a model of control error, and an amplitude damping process, which is a model of thermalization. The unitary error can be entirely specified by the

43
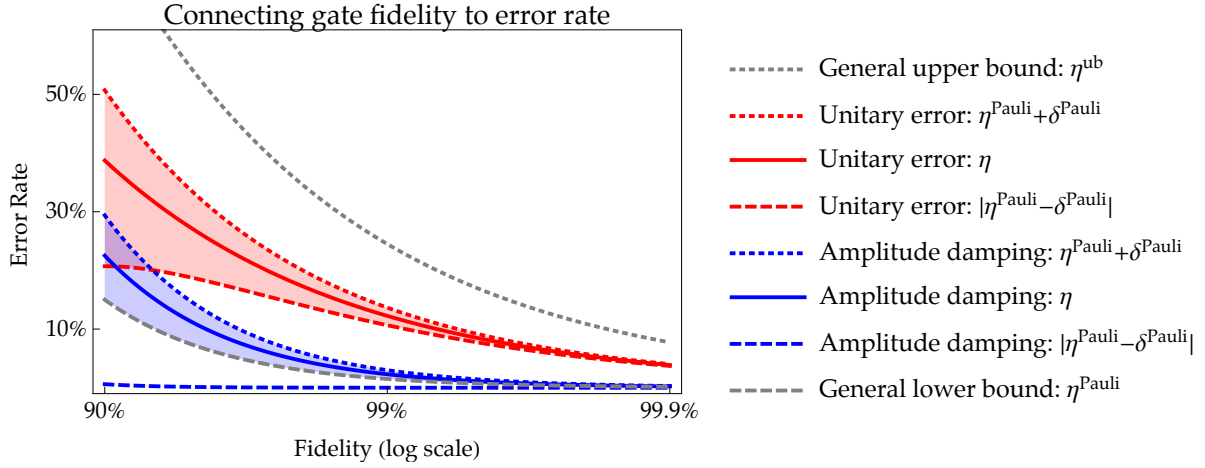
Figure 3.1: An illustration of the dichotomy between average gate fidelity and the error rate for single-qubit noise channels. The grey curves illustrate the generally applicable lower (dashed) and upper (dotted) bounds. The red curves pertain to unitary errors and the blue curves pertain to an amplitude damping ("a.d.") process. The solid red/blue curves are the numerical values of the error rate (vertical axis) given the average gate fidelity (horizontal axis) of the unitary/a.d. model. The dotted red/blue curves are the values of the Pauli-distance-based upper bound $\eta^{\mathrm{Pauli}} + \delta^{\mathrm{Pauli}}$ calculated for the unitary/a.d. model; the dashed are values for the lower bound $\left|\eta^{\mathrm{Pauli}} - \delta^{\mathrm{Pauli}}\right|$. The red/blue shading indicates region estimates for error rate based upon fidelity and Pauli distance; as $\left|\eta^{\mathrm{Pauli}} - \delta^{\mathrm{Pauli}}\right|$ falls below η^{Pauli} for the a.d. process, the blue region uses η^{Pauli} rather than $\left|\eta^{\mathrm{Pauli}} - \delta^{\mathrm{Pauli}}\right|$ as a lower bound for η. The calculations were performed using the QETLAB project of Johnston (2015). This figure is reproduced from Sanders et al. (2016).

eigenvalues $e^{\pm i\theta}$ of the unitary operator, and the amplitude damping process may be specified by a rate parameter r. Both r and θ may be expressed in terms of the observed average gate fidelity φ and thus the error rate of each can be numerically evaluated as a function of φ. The results of this numerical evaluation are displayed in Figure 3.1.

The most important observation about Figure 3.1 is that the Pauli-distance-based bounds on η yield excellent estimates of the error rate of a noise process as fidelity increases. In fact, fidelity indicates a confidence interval if δ^{Pauli} is considered as an estimate of η. Therefore, the Pauli distance can be interpreted as a measure of the 'badness' of noise in the sense that it indicates the size of the gap between fidelity and error rate.

The theoretical possibility of fault-tolerant quantum computation is guaranteed by threshold theorems, which demonstrate that a given quantum circuit can be redesigned with reasonable overhead to become tolerant of local errors that occur at a rate η below a threshold value η_0. The aim of this chapter has therefore been to convert gate fidelity φ, a commonly reported figure-of-merit for quantum logic operations, into an upper bound η^{ub} for η that can be compared, in principle, to η_0.

Of course the noise assumptions underlying fault-tolerance threshold theorems could be either weaker or stronger than reasonable assumptions about the noise of real devices, but this subtlety is often overlooked: numerical simulations such as those of Knill (2005) and Raussendorf and Harrington (2007) are often considered to be indicative of a code-specific threshold value η_0^{lb} even though both papers are clear that only one well-behaved family of noise models is being simulated. The proper interpretation of these results is, in the words of Knill (2005), as "evidence that accurate quantum computing is possible for [error rates[‡]] as high as three per cent". Thus, Knill (2005) claims not that 3% is an *estimate* of η_0^{lb}, but that it is an *upper bound*. The results of Raussendorf and Harrington (2007) should be interpreted similarly.

Whatever their actual value, fault-tolerance thresholds are often treated as performance targets for the development fault-tolerant quantum computing. If so, one must bear in mind the several caveats discussed above. Estimated thresholds on the order of $\sim 1\%$ are nevertheless commonly quoted and, I believe, reasonable performance targets despite the caveats. But the units are crucial: these estimates are of a threshold *error rate*, not average gate infidelity as some have falsely asserted in the literature. Brown et al. (2011), for example, misquote Preskill (1998) as arguing for a threshold in terms of average gate infidelity, rather than error rate. Chow et al. (2012) claim that gate fidelities of 90-99.5% ("depending on measurement errors") suffice for surface-code-based fault-tolerant quantum computation by quoting results obtained using simulations of Pauli noise, which bears a functional relationship between error rate and infidelity that is not general (see Example 3.2). The same erroneous conflation of infidelity with error rate is used to justify the claim of Barends et al. (2014), which is that current device performance surpasses the fault-tolerance threshold for surface code quantum computing. Barends et al. (2014) thereby commit a fallacy of equivocation: on the one hand they quote an estimated threshold of about 1% error rate, which is a statement of belief about the likely value of a code-specific fault-tolerance threshold, and on the other hand quote their

---

[‡]Knill (2005) calls it the "error probability per gate", or EPG.

randomized benchmarking results as an error rate, which means average gate infidelity in that context. Their argument would not be fallacious if the noise in their system was guaranteed to be Pauli, but it is not. The appendix of Barends et al. (2014) describes at least one source of non-Pauli noise: TLDSes, the main subject of this thesis, which introduce large unitary errors.

In the absence of a reliable and quantitative noise analysis in addition to the results of randomized benchmarking, the best rigorous statement that can be made about the observed error rate in the systems described by Brown et al. (2011), Chow et al. (2012), and Barends et al. (2014) is provided by the bounds of Wallman and Flammia (2014). In particular, the upper bound of Wallman and Flammia (2014) can be used to convert a threshold error rate η_0 to a threshold fidelity φ_0 as follows: a d-dimensional gate is guaranteed to have an error rate below threshold if $\varphi > \varphi_0$, where

$$\varphi_0 := 1 - \frac{\eta_0^2}{d^2 + d}. \tag{3.55}$$

A fidelity of $\varphi > 99.9995\%$, for example, is needed to ensure that a two-qubit gate ($d = 4$) has an error rate below $\eta_0 = 1\%$. In fact, Eq. (3.55) implies an important constraint on the interpretation of the average fidelity as a figure-of-merit for quantum logic gates: unless

$$\varphi > 1 - \frac{1}{d^2 + d}, \tag{3.56}$$

it is possible that the error rate of the quantum gate is 100% for some input and measurement. The triviality threshold is 83% for single-qubit gates and 95% for two-qubit gates.

Of course two-qubit gates of 95% fidelity are unlikely to be as bad in practice as Eq. (3.56) indicates. Given further information about gate performance such as the Pauli distance introduced in Section 3.3, it is possible to justify more accurate bounds on gate performance based on fidelity. But such additional information is quashed by randomized benchmarking. The results this chapter therefore imply a need to move beyond randomized benchmarking for assessing progress towards fault-tolerance. Over-reliance on randomized benchmarking has led to erroneous claims from Chow et al. (2012) and Barends et al. (2014) that current gate performance in SICs have been shown to surpass the surface code fault-tolerance threshold.

FOUR

THE SWAP SPECTRUM MODEL

The main point to realize is that all knowledge presents itself within a conceptual framework adapted to account for previous experiences and that any such frame may prove too narrow to comprehend new experiences.

NIELS BOHR

CHARACTERIZATION of TLDSes is the central challenge of my thesis. The purpose of this chapter is to explain this challenge in technical terms. Thus the main goal of this chapter is to present a formal problem statement. The problem, in technical terms, is to identify any oscillators visible to the probe system if operated under specified conditions. These oscillators are elements of a mathematical model and are specified by two parameters: a frequency ω and a coupling strength g. These oscillators are visible or invisible to a probe system depending on the operating conditions of the probe, specified by a parameter ν, and the time at which the state of the probe is measured, specified by a parameter t. The oscillators are models of TLDSes, and the probe is a model of the qubit.

I do not aim to solve the TLDS characterization problem in full generality. Rather, I give algorithms that solve the characterization problem assuming that, at maximum, one TLDS is visible. These algorithms can also be applied when multiple TLDSes are present, provided that the respective oscillator parameters obey a separation criterion. This need for well-separated oscillators motivates the definition of coupling octaves (Section 4.2),

which are order-of-magnitude specifications for oscillator parameters. Based on this definition of coupling octaves, I further subdivide the TLDS characterization problem into two simpler problems: that of *detecting* TLDSes and that of *estimating* the parameters of a TLDS given the promise that they are contained within some coupling octave.

This chapter is divided into three sections. Section 4.1 presents the definition of the swap spectrum, which is derived from a phenomenological model of TLDSes. Section 4.2 motivates and defines coupling octaves. Finally, Section 4.3 gives the formal statement of the TLDS characterization problem as well as the two simpler problems that the algorithms of Chapters 5 and 6 are designed to solve.

Deriving the Swap Spectrum §4.1

The purpose of this section is to derive and explain the swap spectrum, which is a model of the interaction between a probe and its environment. The model involves a two-level probe whose energy spacing is controllable, and an environment that can contain any number N of distinct oscillators that couple to the probe. Each oscillator is labelled by a natural number k ($k = 1, \ldots, N$) and is specified by a resonance frequency ω_k and a coupling strength g_k. Thus the model is specified by the vector

$$\sigma := [(\omega_1, g_1), \ldots, (\omega_N, g_N)]. \tag{4.1}$$

The problem of TLDS characterization is therefore to find σ.

The swap spectrum is not defined exactly as this model. Instead, I define the swap spectrum to be that which is measured by swap spectroscopy, as defined in Algorithm 2.1. There is some ambiguity about whether this swap spectrum should refer to the time-domain or frequency-domain response of the probe to its environment; I take the frequency-domain view because this matches the usual meaning of the term 'spectrum'. Thus the swap spectrum is defined to be the Laplace transform of the time-dependent amplitude of observing the probe in its excited state.

This section proceeds as follows. Section 4.1.1 explains the exchange Hamiltonian, which is a phenomenological model for the process by which a qubit exchanges energy with environmental TLDSes. Section 4.1.2 gives the derivation of the swap spectrum proper. Finally, Section 4.1.3 discusses numerical simplifications based on the promise that oscillators are 'well-separated'.

The Exchange Hamiltonian

As with most noise models in quantum information, the physical system in question is described using a tensor product of two Hilbert spaces, one for the system to be studied and one for the environment. The evolution of the system is governed by the Schrödinger equation acting upon this joint Hilbert space. Throughout this thesis, the Schrödinger equation is represented using Heisenberg matrix formalism, which is to say that the dynamics are specified by the matrix differential equation

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\ket{\psi} = H\ket{\psi}, \tag{4.2}$$

where $\ket{\psi}$ represents the physical state (using Dirac ket notation) and H is a Hermitean matrix called the Hamiltonian acting upon the joint system-environment Hilbert space.

The system is treated as a two-level atom, which is to say that it can be found in one of two distinct energy eigenstates. As superconducting qubits behave as dipoles, these eigenstates represent the cases where the dipole is either aligned or anti-aligned with the ambient field. The system Hilbert space is therefore $\mathcal{Q} \cong \mathbb{C}^2$ together with an identification of the energy eigenbasis with the eigenbasis of the Pauli Z matrix acting on $\mathcal{Q}$. The Pauli matrix acting on the system Hilbert space is called the system Hamiltonian because it specifies the energy eigenbasis of the system considered in isolation from its environment. In the swap spectrum model, the system is used as a probe for its environment; I shall hereafter refer to it as the 'probe'. The TLDS model assumes the environment is composed of N other two-level atoms, so the Hilbert space of the environment is $\mathcal{Q}^{\otimes N} \cong \mathbb{C}^{2^N}$ together with a privileged basis, $\{\ket{g}, \ket{e}\}$ for 'ground' and 'excited' states—the lower and upper energy eigenstates. The privileged bases for the Hilbert spaces for the probe and environment, called $\mathcal{H}_P$ and $\mathcal{H}_E$ respectively, suggests the following decomposition of the Hamiltonian of the total system:

$$H_{\mathrm{sys}} := H_P^{(\nu)} + H_E + H_{\mathrm{int}}, \tag{4.3}$$

where $H_P^{(\nu)}$ acts trivially on the environment Hilbert space, H_E acts trivially on the probe Hilbert space, H_{int} describes the interaction between probe and environment, and $\hbar\nu$ is the energy gap of the probe. ν has units of frequency and is called the 'probe frequency'. In swap spectroscopy, ν is a control parameter that can be chosen at will.

The phenomenological model for swap spectroscopy is as follows. By convention, the energy of the ground state of the system is set to be zero:

$$H_{\mathrm{sys}}\ket{g_P, g_E} = 0. \tag{4.4}$$

The probe and environment Hamiltonians represent the evolution of each system if it does not interact with the other:

$$H_P^{(\nu)} = \hbar\nu \, |e_P\rangle\langle e_P| \tag{4.5}$$

and, using $|e_{osc_k}\rangle$ to refer to the first excited state of oscillator k,

$$H_E = \sum_{k=1}^{N} \hbar\omega_k \, |e_{osc_k}\rangle\langle e_{osc_k}| \, . \tag{4.6}$$

Note that these projectors are implicitly acting trivially on the remainder of the total Hilbert space; that is, $|e_P\rangle\langle e_P|$ stands for $|e_P\rangle\langle e_P| \otimes I_E$. The interaction Hamiltonian describes the exchange of energy between the qubit and environmental oscillators, and is given by

$$H_{int} = \sum_{k=1}^{N} \hbar g_k \left(|g_P, e_{osc_k}\rangle\langle e_P, g_{osc_k}| + |e_P, g_{osc_k}\rangle\langle g_P, e_{osc_k}| \right) . \tag{4.7}$$

Notice that H_{sys} conserves the number of excitations in the system. If $|g\rangle$ has an excitation number of 0 and $|e\rangle$ has an excitation number of 1, define the total excitation number of any eigenstate of $H_P + H_E$, which must be a tensor product of lower and upper energy eigenstates, to be the sum of the excitation numbers of the component states. For example, the excitation number of the state $|g_P \, e_{osc}\rangle$ in a system with one oscillator equals the excitation number of $|e_P \, g_{osc}\rangle$, which is 1. In general, one can write the excitation number operator

$$E := |e_P\rangle\langle e_P| + \sum_{k=1}^{N} |e_{osc_k}\rangle\langle e_{osc_k}| \tag{4.8}$$

and observe that E commutes with H_{sys}.

Swap spectroscopy is designed to ensure that the initial excitation number of the system is equal to 1. Therefore, the state space of the system can be reduced to the 'single-excitation' subspace, which is to say the eigenspace of the operator E corresponding to eigenvalue one. By restricting to the single-excitation subspace, the system-environment Hilbert space can be reduced to one whose basis states represent the location of the excitation. The excitation starts in the probe, a situation represented by $|0\rangle$, and could

transfer to oscillator k, represented by $|k\rangle$. Thus the initial state for swap spectroscopy can be represented as

$$|\psi(0)\rangle = |0\rangle = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{4.9}$$

and the 'exchange' Hamiltonian reads

$$\frac{1}{\hbar} H_{\text{exchange}}^{(\nu)} := \begin{bmatrix} \nu & g_1 & g_2 & \cdots & g_N \\ g_1 & \omega_1 & 0 & \cdots & 0 \\ g_2 & 0 & \omega_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ g_N & 0 & 0 & \cdots & \omega_N \end{bmatrix}, \tag{4.10}$$

where ν is highlighted because it is a control parameter rather than a model parameter.

Expression for the Swap Spectrum §**4.1.2**

Put simply, the swap spectrum* of a qubit with frequency ν is defined to be the Laplace transform of the time-dependent amplitude of observing the probe in its initial state. In other words, the swap spectrum is defined to be

$$\mathcal{L}\left[\langle \psi(0)|\psi(t)\rangle\right](s), \tag{4.11}$$

where $\mathcal{L}$ is the Laplace transform operator and s is the Laplace-domain conjugate of the time variable t. This definition is suggested by the possibility of solving the Schrödinger equation by computing the Laplace transform. Let

$$|\overline{\psi}(s)\rangle := \begin{bmatrix} F_0^{(\nu)}(s) \\ F_1^{(\nu)}(s) \\ F_2^{(\nu)}(s) \\ \vdots \\ F_N^{(\nu)}(s) \end{bmatrix} \tag{4.12}$$

---

*The name references the intuition that an excitation is being 'swapped' between the probe and the environmental oscillator.

be the Laplace transform of $|\psi(t)\rangle$. Then the Laplace transform of the Schrödinger equation is

$$\left(s + \frac{i}{\hbar} H^{(\nu)}_{\text{exchange}}\right) |\overline{\psi}(s)\rangle = |\psi(0)\rangle. \tag{4.13}$$

The swap spectrum is therefore defined as

$$\mathcal{L}\left[\langle\psi(0)|\psi(t)\rangle\right](s) = \langle\psi(0)|\overline{\psi}(s)\rangle = F_0^{(\nu)}(s). \tag{4.14}$$

To compute $F_0^{(\nu)}(s)$, I perform Gaussian elimination. Notice that the row reduction

$$\begin{bmatrix} s+i\nu & ig_1 & ig_2 & \cdots & ig_N \\ ig_1 & s+i\omega_1 & 0 & \cdots & 0 \\ ig_2 & 0 & s+i\omega_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ig_N & 0 & 0 & \cdots & s+i\omega_N \end{bmatrix} \rightarrow \begin{bmatrix} s+i\nu & ig_1 & ig_2 & \cdots & ig_N \\ \frac{g_1^2}{s+i\omega_1} & -ig_1 & 0 & \cdots & 0 \\ \frac{g_2^2}{s+i\omega_2} & 0 & -ig_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{g_N^2}{s+i\omega_N} & 0 & 0 & \cdots & -ig_N \end{bmatrix}$$

$$\tag{4.15}$$

leaves the right-hand side of Eq. (4.13) unchanged. By adding all the rows and defining

$$F_P^{(\nu)}(s) := \frac{1}{s+i\nu}, \quad F_E^{(\nu)}(s) := \sum_{k=1}^{N} \frac{g_k^2}{s+i\omega_k}, \tag{4.16}$$

I calculate

$$F_0^{(\nu)}(s) = \frac{F_P^{(\nu)}(s)}{1 + F_P^{(\nu)}(s)F_E^{(\nu)}(s)}. \tag{4.17}$$

$F_0^{(\nu)}(s)$ is the Laplace transform of the time-dependent amplitude of observing the probe in its excited state; i.e. the swap spectrum. From a signal processing perspective, the terms $F_P(s)$ and $F_E(s)$ can be interpreted as transfer functions associated respectively to the probe and to the environment. In the language of feedback control, $F_P(s)$ can be interpreted as the transfer function of a combined controller-plant system, whereas $F_E(s)$ can be interpreted as the transfer function of a feedback loop. Notice that the transfer function of the system approaches $F_P(s)$ in the limit of zero coupling (i.e. $g_k \rightarrow 0$) to the environment.

Though I shall say little about decoherence in this thesis that is not introduced directly by TLDSes, it is a relatively simple matter to model decoherence due to thermalization.

If, for example, the probe is not coupled to any TLDSes but is in contact with a zero-temperature bath, the probability of observing the probe in its excited state should decay towards zero exponentially in time. By including a complex probe frequency, $v \to v - i\Gamma_{\text{dec}}$, the amplitude of observing the excited state of the probe is $\exp(-ivt - \Gamma_{\text{dec}}t)$, meaning that the click probability is $\exp(-2\Gamma_{\text{dec}}t)$. Thermalization affecting TLDSes can be modelled similarly.

In the absence of decoherence, the poles of $F_0^{(v)}(s)$ are simple and purely imaginary. To see this, consider

$$\lambda^{(v)}(x) := -iF_0^{(v)}(-ix) = \frac{\frac{1}{x-v}}{1 - \frac{1}{x-v}\sum_k \frac{g_k^2}{x-\omega_k}} = \frac{\prod_{k=1}^N (x - \omega_k)}{\prod_{\ell=0}^N (x - r_\ell)} = \sum_{\ell=0}^N \frac{c_\ell}{x - r_\ell}, \quad (4.18)$$

where r_ℓ and c_ℓ represent the poles and residues of the real rational function $\lambda^{(v)}(x)$. Let

$$\text{zeroes}(x) := \prod_{k=1}^N (x - \omega_k) \quad (4.19)$$

and

$$\text{poles}(x) := \frac{\text{zeroes}(x)}{\lambda^{(v)}(x)} = (x - v)\,\text{zeroes}(x) - \left(\sum_{k=1}^N \frac{g_k^2}{x - \omega_k} \right) \text{zeroes}(x) \quad (4.20)$$

respectively represent the numerator and denominator of $\lambda^{(v)}(x)$. Then

$$\lim_{x \to +\infty} \text{sgn}(\text{poles}(x)) = +1, \quad (4.21)$$

$$\text{sgn}\left(\text{poles}\left(\omega_\ell\right)\right) = \text{sgn}\left(-g_\ell^2 \prod_{k \neq \ell} (\omega_\ell - \omega_k)\right) = (-1)^{N-\ell+1} \quad (4.22)$$

and

$$\lim_{x \to -\infty} \text{sgn}(\text{poles}(x)) = (-1)^{N+1}. \quad (4.23)$$

Thus, $\text{poles}(x)$ changes sign $N + 1$ times; by the intermediate value theorem, all $N + 1$ roots of $\text{poles}(x)$ are real. Therefore, the poles of

$$F_0^{(v)}(s) = i\lambda^{(v)}(is) = \sum_{\ell=0}^N \frac{c_\ell}{s + ir_\ell} \quad (4.24)$$

53

are purely imaginary.

The expression for the amplitude of $|\psi(0)\rangle$ as a function of time,

$$f_0^{(\nu)}(t) := \langle \psi(0)|\psi(t)\rangle = \sum_{k=0}^{N} c_k \exp(-ir_k t), \tag{4.25}$$

allows direct computation of the probability of detecting the probe in its ground state;

$$p_{\text{click}}(\nu, t) := 1 - \left| f_0^{(\nu)}(t) \right|^2. \tag{4.26}$$

I call this probability the 'click probability' because I refer to the event that the probe is detected in its ground state as a 'click'. If the probe is found in its excited state (i.e. its initial state), no click is recorded.

Oscillator Separation   §**4.1.3**

A crucial property of swap spectra is that the dynamics of the probe when far-detuned from resonance with any oscillators is indistinguishable from the dynamics of the probe if no oscillators are present. Furthermore, two oscillators that are 'well-separated' (defined below) appear isolated to the probe: the dynamics of the probe when near-resonance with one oscillator (hence far-detuned from the other) is indistinguishable from the probe dynamics if the second oscillator were simply not present. This phenomenon is crucial to the applicability of the 'coupling octave', defined in Section 4.2.

I focus first on single-oscillator swap spectra, which are specified by parameter vectors of the form $\sigma = [(\omega, g)]$. In this case, the real transfer function of Eq. (4.18) reads

$$\lambda^{(\nu)}(x) = \frac{x - \omega}{(x - \nu)(x - \omega) - g^2} = \frac{c_+}{x - r_+} + \frac{c_-}{x - r_-}, \tag{4.27}$$

where

$$r_\pm = \frac{\nu + \omega}{2} \pm \frac{1}{2}\sqrt{(\nu - \omega)^2 + 4g^2}, \; c_\pm = \frac{1}{2}\left(1 \pm \frac{\nu - \omega}{\sqrt{(\nu - \omega)^2 + 4g^2}}\right). \tag{4.28}$$

By Eqs. (4.25) and (4.26),

$$p_{\text{click}}(\nu, t) = \frac{4g^2}{(\nu - \omega)^2 + 4g^2} \sin^2\left(\frac{t}{2}\sqrt{(\nu - \omega)^2 + 4g^2}\right). \tag{4.29}$$

54

By defining the dimensionless detuning parameter

$$\delta := \frac{\nu - \omega}{2g}, \tag{4.30}$$

Eq. (4.29) reads

$$p_{\text{click}}(\nu, t) = \frac{\sin^2\left(gt\sqrt{1 + \delta^2}\right)}{1 + \delta^2} < \frac{1}{1 + \delta^2}; \tag{4.31}$$

the inequality on the right-hand side implies that $p_{\text{click}}(\nu, t) \to 0$ as $\delta \to \infty$. So the oscillator represented by (ω, g) becomes invisible to swap spectroscopy experiments if the qubit is highly detuned from resonance (i.e. $|\nu - \omega| \gg g$). This is exactly what is expected: the dynamics of a probe that is far-detuned from resonance with an oscillator does not exchange energy with the oscillator.

I define two oscillators, (ω_1, g_1) and (ω_2, g_2), to be 'well-separated' if

$$\frac{\max(g_1^2, g_2^2)}{|\omega_1 - \omega_2|} \tag{4.32}$$

is small. What this means is that the true two-oscillator swap spectrum, $\sigma_{\text{true}} = [(\omega_1, g_1), (\omega_2, g_2)]$, is well-approximated by a one-oscillator swap spectrum for all (ν, t): by $\sigma_1 = [(\omega_1, g_1)]$ if $\nu \approx \omega_1$ and by $\sigma_2 = [(\omega_2, g_2)]$ if $\nu \approx \omega_2$. To establish this fact, I first give an exact formula for the click probability for a one-oscillator swap spectrum. I then explain how to derive an exact formula for the two-oscillator case, though the resulting expression is too complicated to write explicity. I also give a method for approximating two-oscillator dynamics when oscillators are not well-separated.

Consider the two-oscillator swap spectrum as specified by Eq. (4.18): if the swap spectrum is $\sigma = [(\omega_1, g_1), (\omega_2, g_2)]$,

$$\lambda^{(\nu)}(x) = \frac{(x - \omega_1)(x - \omega_2)}{(x - \nu)(x - \omega_1)(x - \omega_2) - g_1^2(x - \omega_2) - g_2^2(x - \omega_1)}. \tag{4.33}$$

The roots of the denominator can be found via the cubic formula, and the resulting click probabilities can thereby be computed. However, the resulting expressions are too complex to be helpful intuitively. What makes things easier is to create plots similar to Figure 4.1. When this is done, as in Figure 4.2, a method of approximating two-oscillator swap spectra becomes apparent.

Suppose, without loss of generality, that $\omega_2 > \omega_1$ and $g_2 < g_1$. The characteristic 'avoided crossing' pattern shown by the poles of the single oscillator swap spectrum is
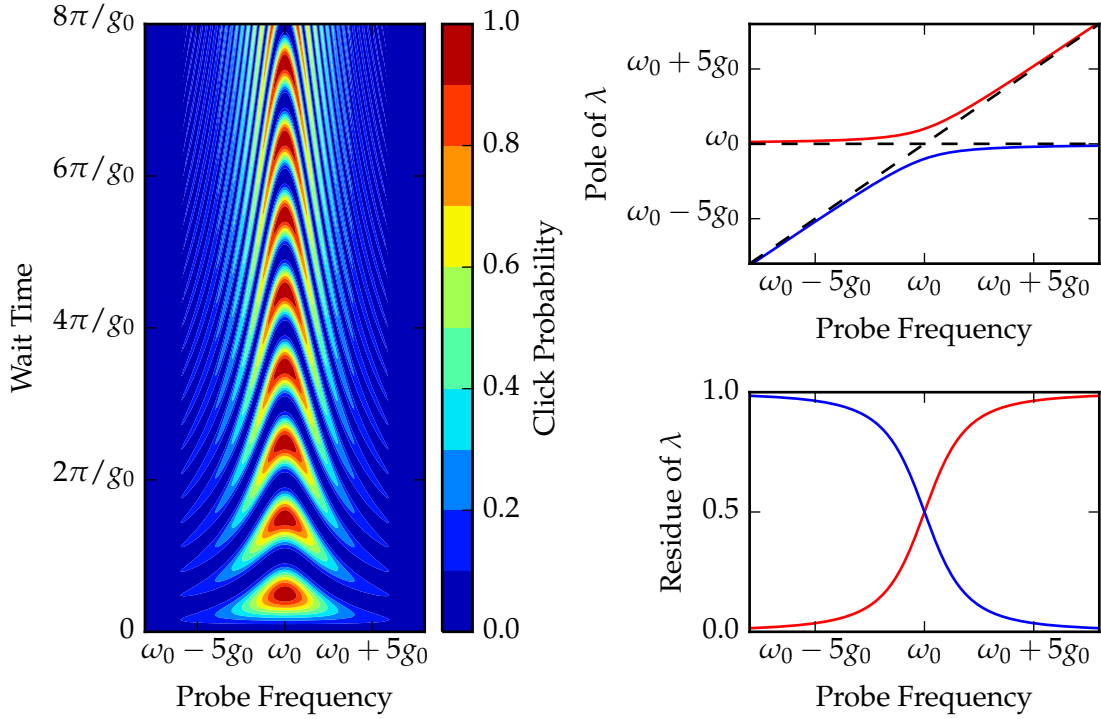
Figure 4.1: Numerics for a single-oscillator swap spectrum, specified by $\sigma = [(\omega_0, g_0)]$. The plot on the left shows the click probability as a function of the probe frequency ν and the wait time t, and shows the characteristic 'chevron' pattern. Note that the click probability becomes negligibly small if the probe is far-detuned from resonance or if $t \ll 1/g$. The poles and residues of the swap spectrum are displayed on the top right and bottom right, respectively, and are distinguished by colour: red for the positive pole and corresponding residue; blue for the negative. The dashed lines of the top right figure denote asymptotes, where the ordinate values are ω_0 and ν. Note that only one pole is visible if the probe is far-detuned from resonance.
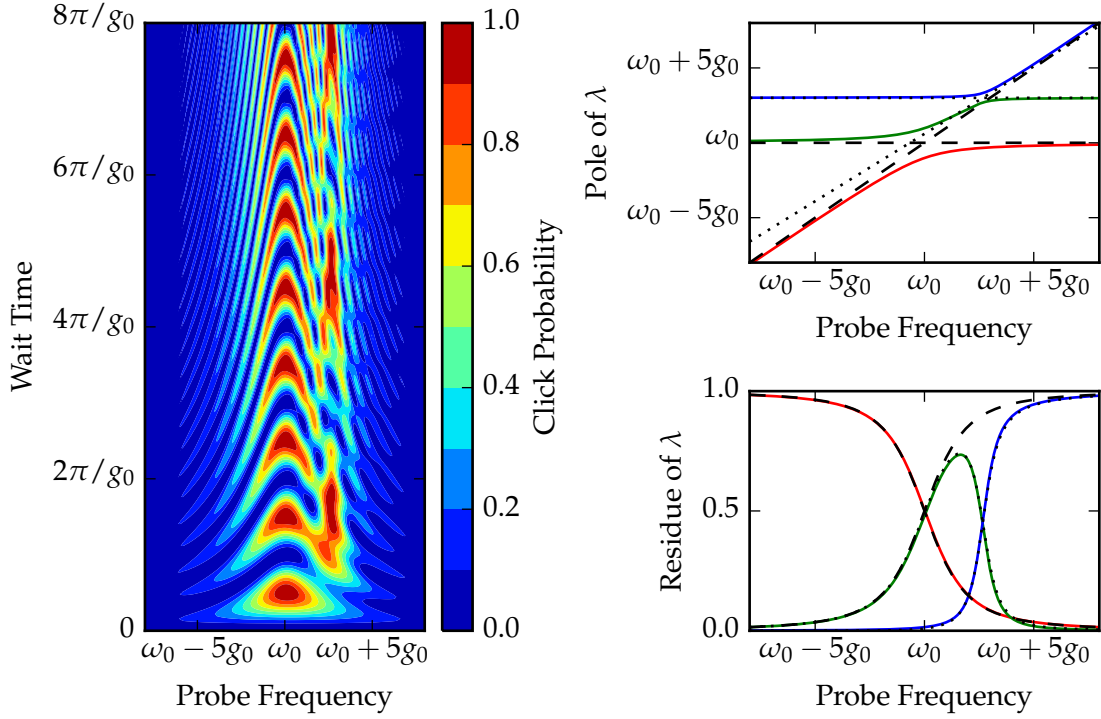
Figure 4.2: Numerics for a two-oscillator swap spectrum, specified by $\sigma = [(\omega_0, g_0), (\omega_0 + 3g_0, g_0/3)]$. The plot on the left shows the click probability as a function of the probe frequency ν and the wait time t, and shows two 'chevron' patterns, each one corresponding to an oscillator. The poles and residues of the swap spectrum are displayed on the top right and bottom right, respectively, and are distinguished by colour: red for the largest pole and corresponding residue, green for the middle, and blue for the smallest. The dashed lines of the top right figure denote asymptotes, where the ordinate values are ω_0 and ν, and the dotted lines denote the asymptotes approximated as indicated in the main text. The dashed curves of the bottom right figure denote the unperturbed residues that would be observed if only the first oscillator is present, and the dotted curves are the approximate residues of the second oscillator. The colours of the bottom right figure indicate which residue corresponds to which pole in the top right figure.

exhibited twice: by each of the horizonal asymptotes $\nu = \omega_1$ and $\nu = \omega_2$. Whereas the diagonal asymptote of the poles corresponding to the more strongly coupled oscillator (ω_1, g_1) is not obviously perturbed if $\omega_2 - \omega_1$ is sufficiently large, the second oscillator has its diagonal asymptote shifted because the diagonal asymptote is equal to the positive pole of the first oscillator. The probe resonates with the second oscillator not if $\nu = \omega_2$, but

$$\frac{\nu + \omega_1}{2} + \frac{1}{2}\sqrt{(\nu - \omega_1)^2 + 4g_1^2} = \omega_2 \Rightarrow \nu = \omega_2 - \frac{g_1^2}{\omega_2 - \omega_1} \qquad (4.34)$$

corresponds to resonance, at least approximately, with the second oscillator. Furthermore, the non-unit slope of the positive pole of the first oscillator (which turns out to be equal to c_+) 'dresses' the probe frequency and makes the resonance pattern of the second oscillator seem stronger.

Whereas the dashed lines of Figure 4.2 correspond to the pole asymptotes (top right) and residues (bottom right) shown in Figure 4.1, the dotted lines correspond to the asymptotes of the second oscillator (top right) and an approximation of the residues of the second oscillator (bottom right). The residue approximation is simply to calculate the single-oscillator residues of the second oscillator, as defined in Eq. (4.28), and multiply them to the residue corresponding to the positive pole of the first oscillator, again as defined in Eq. (4.28).

Based on this analysis, I define 'well-separated' oscillators to be those with the property that

$$\frac{\max(g_1^2, g_2^2)}{|\omega_1 - \omega_2|} \ll \min(g_1, g_2). \qquad (4.35)$$

The reason I compare the correction term of Eq. (4.34) to the smaller coupling strength is that the width of the 'chevron' pattern associated to oscillators in click probability plots is comparable to the coupling strength of the oscillator. Therefore, a shift induced by another oscillator is noticeable if it is greater than or comparable to the coupling strength. The smaller the correction term, the better the relevant approximations. If oscillators are well-separated by this definition, the click probability can be approximated by the product of single-oscillator click probabilities and therefore the swap spectrum can be approximated by the convolution of single-oscillator swap spectra.

Coupling Octaves §4.2

The purpose of this section is to introduce the concept of a coupling octave. Coupling octaves are regions of $\mathbb{R} \times \mathbb{P}$, where $\mathbb{R}$ is the set of real numbers and $\mathbb{P}$ is the set of positive

real numbers, specified by a frequency ω_{oct} and coupling strength g_{oct}. Specifically, $\text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$ is the set of pairs (ω, g) such that

$$\omega_{\text{oct}} - g_{\text{oct}} \leq \omega \leq \omega_{\text{oct}} + g_{\text{oct}} \text{ and } \frac{g_{\text{oct}}}{2} \leq g \leq g_{\text{oct}}. \tag{4.36}$$

The reason for such a definition is to capture the idea that oscillators whose parameters are 'far away' from one another have independent dynamics. From the point of view of the probe system, some oscillators are invisible.

The need for a definition like that of a coupling octave arises from the observation that $p_{\text{click}}(\nu, t) \approx 0$ for many values of (ω, g). Under these circumstances, the output of SWAPSPECTROSCOPY(ν, t) is almost indistinguishable from that which would have been produced had no oscillator been present at all. This is important because this data is almost entirely uninformative if an oscillator is known to be present but has parameters that ensure the click probability is near zero.

This issue of near-zero click probabilities can be anticipated if order-of-magnitude estimates are available for oscillator parameters. For example, Eq. (4.31) implies that $p_{\text{click}} \approx 0$ if $\log \left| \frac{\nu - \omega}{2g} \right| \gg 1$. Furthermore, the small angle approximation applied to the same equation implies that $p_{\text{click}} \approx 0$ if $gt \ll 1$ (equivalently, $\log(t) \ll -\log(g)$) even if ν is comparable to ω. Thus, estimates of $\log(g)$ and $\log(\omega)$ can prove invaluable. Such estimates can also be used to ensure that oscillators are well-separated.

I focus on estimates of the *logarithm* of these parameters because errors on the order of a few percent are acceptable for ensuring that the click probability is reasonably large even though they correspond to rather large errors in the estimates of the parameters themselves. The central insight is that the techniques needed for establishing order-of-magnitude estimates for oscillator parameters are quite different from establishing highly accurate estimates of the same. Whereas highly accurate estimates can be obtained using the techniques described in Section 2.3, these techniques are guaranteed to function properly only if order-of-magnitude estimates of the relevant parameters have been established. Establishing order-of-magnitude estimates is more like a search problem than a parameter estimation problem.

An important feature of $\text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$ is that its width, the size of its projection onto the ω subspace, is equal to $2g_{\text{oct}}$; that is, the width is independent of ω_{oct} but proportional to g_{oct}. This feature is important because it reflects the intuitively apparent fact that weakly coupled oscillators are harder to 'find'. To make this clearer, suppose that one is asked to choose a coupling octave that contains an oscillator with a known coupling strength g but an unknown frequency ω that is promised to fall within the

interval $(\omega_{\min}, \omega_{\max})$, which is defined so that $g < \frac{1}{2}(\omega_{\max} - \omega_{\min})^\dagger$. To accomplish this task, access to an oracle $\mathcal{O}$ is provided: $\mathcal{O}(\omega_{\text{oct}}, g_{\text{oct}}) = 1$ if $(\omega, g) \in \text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$; $\mathcal{O}(\omega_{\text{oct}}, g_{\text{oct}}) = 0$ otherwise.

The task of finding the oscillator can be accomplished as follows. First, choose g_{oct} so that $g \in (g_{\text{oct}}/2, g_{\text{oct}})$. This can be accomplished by setting

$$g_{\text{root}} = \frac{1}{2}(\omega_{\max} - \omega_{\min}) \tag{4.37}$$

so that $g_{\text{oct}} = g_{\text{root}}/2^n$, where

$$n := \left\lfloor \log_2 \left(\frac{g_{\text{root}}}{g} \right) \right\rfloor = \lfloor \log_2(g_{\text{root}}) - \log_2(g) \rfloor. \tag{4.38}$$

The integer n is called the *level* of g with respect to the interval $(\omega_{\min}, \omega_{\max})$. The oracle is queried with 2^n octave frequencies, each of which correspond to a 'shift' specified by an n-bit string. The shift value is defined in Algorithm 4.1. Setting $s_b = \text{SHIFT}(b)$ for each n-bit string b, define

$$g_b := g_{\text{root}}/2^n; \; \omega_b = \omega_{\min} + 2s_b g_b. \tag{4.39}$$

Then perform $\mathcal{O}(\omega_b, g_b)$ for each choice of n-bit string b. The oracle will return 1 for at least one choice of $b^\ddagger$.

Notice the query complexity of the above algorithm. In the worst case, between $\frac{g_0}{2g} - 1$ and $\frac{g_0}{g} - 1$ queries are needed to find an octave containing the oscillator, and half of that is required in the average case; the query complexity is $\Theta(1/g)$. Compare this to the query complexity of unstructured search: for an unordered list of length k containing exactly one marked item, $\Theta(k)$ queries are needed to find that marked item in both the average and worst cases. As the number of octaves that can fit in a given frequency range is inversely proportional to g, the oracle-based algorithm for finding an octave containing (ω, g) for a known g is optimal.

Now suppose that g is not known but is instead promised to be larger than some minimum value $g_{\min}$; g is also promised to be smaller than g_{root} as before. Set

$$\omega_{\text{root}} := \frac{1}{2}(\omega_{\max} - \omega_{\min}), \tag{4.40}$$

$\dagger$This restriction circumvents a trivial case. If $g \geq \frac{1}{2}(\omega_{\max} - \omega_{\min})$, one could choose $g_{\text{oct}} = g$ and ω_{oct} uniformly at random from the interval $(\omega_{\min}, \omega_{\max})$.

$\ddagger$The oracle could return 1 twice if $\omega = \omega_b$ for any n-bit string b—except for $b = [0, \ldots, 0]$ or $[1, \ldots, 1]$, for which $\omega_b = \omega_{\min}$ or $\omega_{\max}$ respectively.

Algorithm 4.1 The shift value of an octave key.

Input: Bit string, $b = [b_0, \dots, b_{n-1}]$.   ▷ b could be empty; i.e. $n = 0$.
Output: Shift value, s.
function SHIFT(b)
   if $n = 0$ **then**
      $s \leftarrow 1/2$
   else
      $s \leftarrow 1/2 + \sum_{k=0}^{n-1} 2^k b_k$
   end if
   return s
end function

---

which corresponds to $\omega_\varnothing$. Define the 'root' octave to be Octave $(\omega_{\text{root}}, g_{\text{root}})$ and the 'b-child' octave ($b \neq \varnothing$) to be Octave (ω_b, g_b). The oscillator can be found by querying the root octave and each child; i.e., execute $\mathcal{O}(\omega_b, g_b)$ for each bit string b of length less than or equal to

$$n_{\max} := \lfloor \log_2(g_{\text{root}}) - \log_2(g_{\min}) \rfloor. \tag{4.41}$$

There are a total of $2^{n_{\max}+1} - 1$ bit strings to query, meaning that the query complexity of finding an oscillator in a fixed interval with coupling strength greater than some minimum value $g_{\min}$ is $\Theta(1/g_{\min})$.

The coupling octaves defined in this section play a central role in the solution to the detection problem (Problem 2) and the statement of the estimation problem (Problem 3). This is because a specification of a coupling octave constitutes a dyadic-order-of-magnitude estimate of an oscillator's parameters. The coupling octaves capture concerns about the visibility of an oscillator with respect a probe with parameters (ν, t). Furthermore, the octave query model gives a strong indication of the optimal data cost for the problem of finding an oscillator.

Formal Problem Statements   §**4.3**

The discussion of the preceding sections in this chapter enables a formal statement of the central problem of this thesis. That problem is what I call the TLDS characterization problem, which can be roughly stated as the problem of finding and identifying any TLD-Ses that are coupled to a superconducting qubit. This formal problem statement assumes that the underlying physics is perfectly described by the swap spectrum model. Thus

the TLDS characterization problem is to identify the swap spectrum model parameters correctly.

This section proceeds as follows. Section 4.3.1 states the general TLDS characterization problem; a problem that serves as a context for the two simpler problems of TLDS detection (Section 4.3.2) and estimation (Section 4.3.3). Though the TLDS detection and estimation problems combined do not reproduce the TLDS characterization problem, their solutions can be combined to form a powerful technique for finding and describing TLDSes.

The General Characterization Problem   §**4.3.1**

The general TLDS characterization problem is, simply put, to correctly identify the swap spectrum parameters σ. Yet the notion of 'correctness' is ambiguous: the swap spectrum model is susceptible to overfitting with weakly coupled oscillators and far-detuned oscillators. To account for such difficulties, the formal statement of the general TLDS characterization problem includes a restriction on the input parameters (ν, t) for swap spectroscopy.

Problem 1 (TLDS Characterization). Suppose that the results of any swap spectroscopy experiment are perfectly modelled by a swap spectrum with parameter vector σ_{true}. If $\varepsilon > 0$ is a target error tolerance and the swap spectroscopy control parameters (ν, t) are confined to the intervals $(\nu_{\min}, \nu_{\max})$ and $(0, t_{\max})$ for ν and t respectively, choose a parameter vector σ_{est} such that

$$\sup_{\substack{\nu \in (\nu_{\min}, \nu_{\max}) \\ t \in (0, t_{\max})}} \left| p_{\text{click}}^{\text{true}}(\nu, t) - p_{\text{click}}^{\text{est}}(\nu, t) \right| < \varepsilon, \tag{4.42}$$

where $p_{\text{click}}^{\text{true}}$ and $p_{\text{click}}^{\text{est}}$ represent the click probabilities of the swap spectra specified respectively by the parameter vector σ_{true} and σ_{est}.

So posed, the TLDS characterization problem is not to find the 'true' model parameters σ_{true} but rather model parameters σ_{est} that yield an approximation to the true dynamics. The TLDS characterization problem can of course be solved if a great deal of data is collected via swap spectroscopy. However, a high cost of data collection prevents investigation of rapid changes to the qubit-environment interaction. The goal of this thesis is not simply to solve the TLDS characterization problem but to do so in a data-efficient manner. That data-efficiency enables investigation of transient TLDSes.

The difficulty of the TLDS characterization problem depends on the number of oscillators present, how strongly or weakly the TLDSes are coupled to the qubit, and the extent to which the TLDSes are spectrally crowded. Solving the TLDS characterization problem if TLDSes are not well-separated is especially challenging, and there is no solution in this thesis beyond applying a naïve data collection algorithm and fitting a multi-oscillator model accordingly. Instead, this thesis presents data-efficient algorithms for solving a reduced version of the general TLDS characterization problem. The reduction is to assume that there is no more than one visible TLDS, though the separation argument of Section 4.1.3 implies that solutions to this problem can be used on a swap spectrum with many well-separated oscillators by simply subdividing the range of possible probe frequencies into smaller disjoint intervals.

The TLDS characterization problem for a single TLDS is reduced further. As explained in Section 4.2, the problem of 'finding' a TLDS is a separate problem from that of producing highly accurate estimates of its parameters. These are the TLDS detection and estimation problems and are presented as Problems 2 and 3, respectively. These two reduced problems are the real subject of the thesis and are solved in Chapters 5 and 6.

The Detection Problem §**4.3.2**

The aim of this thesis is not to solve the TLDS characterization problem in its full generality. Rather, the aim is to solve a restricted version: the swap spectrum is promised to contain no more than one oscillator. Difficulties arise even under these limited circumstances; the most important of which is to distinguish between swap spectra that contain no oscillators and swap spectra with a weakly coupled oscillator. This is the TLDS detection problem.

Put more formally, the TLDS detection problem is a hypothesis testing problem. Instead of being concerned with the parameters of an oscillator, if present, the problem is to correctly decide whether an oscillator is present at all. In the language of hypothesis testing, a distinction is drawn between the hypothesis of zero oscillators (the 'null' hypothesis) and the hypothesis of one oscillator (the 'alternate' hypothesis). Solving the detection problem requires data from swap spectroscopy carried out over a wide range of probe parameters (v, t) due to the visibility considerations described in Section **??**.

Problem 2 (TLDS Detection). Suppose that the swap spectroscopy control parameters (v, t) are confined to the intervals $(v_{\min}, v_{\max})$ and $(0, t_{\max})$ for v and t, respectively. The true model is promised to be either $\sigma_{\text{true}} = \varnothing$ (i.e. no oscillators are present), which is

the *null hypothesis*, or $\sigma_{\text{true}} = [(\omega_{\text{true}}, g_{\text{true}})]$, where

$$\nu_{\text{min}} < \omega_{\text{true}} < \nu_{\text{max}} \tag{4.43}$$

and

$$\frac{1}{2}(\nu_{\text{max}} - \nu_{\text{min}}) < g_{\text{true}} < \frac{\pi}{2t_{\text{max}}}. \tag{4.44}$$

The problem is to correctly accept or reject the null hypothesis.

Problem 2 places restrictions on the allowable parameters of any oscillators to account for visibility. Oscillators with parameters far away from this specified region are invisible to swap spectroscopy experiments carried out with the probe parameters (ν, t). So posed, this problem can be solved by using the oracular algorithm of Section 4.2, though of course there is no reason to believe that such an oracle can be constructed using swap spectroscopy with any kind of data-efficency. The oracular algorithm closely informs the data-efficient solution to this problem that is presented in Chapter 5.

The Estimation Problem §**4.3.3**

Once the existence of a TLDS with parameters (ω, g) has been established, the remainder of the TLDS characterization problem is to produce an accurate estimate of those parameters. This is the TLDS (parameter) estimation problem. It can be solved using an online experimental design algorithm, such as that described by Granade et al. (2012). However, these techniques are effective only when the oscillator parameters are reasonably well established.

Online experimental design algorithms function by choosing the most informative possible experiments. But if the oscillator parameters are highly uncertain, all swap spectroscopy experiments are rather uninformative. If SWAPSPECTROSCOPY(ν, t) returns no click, it could be that $\log(t) \ll -\log(g)$, that $|\nu - \omega| \gg g$, or simply that (ν, t) falls in one of the many valleys of p_{click}.

The TLDS estimation problem therefore presumes that the oscillator parameters are promised to fall within a particular coupling octave. This restriction ensures that the online experimental design algorithm described in Chapter 6 and published in Stenberg et al. (2014) functions correctly. Without this restriction, the issue of 'outliers' arises. This is discussed in detail in Chapter 6.

Problem 3 (TLDS Parameter Estimation)**.** Suppose that the results of any swap spectroscopy experiment are perfectly modelled by a swap spectrum with parameter vector

$\sigma_{\text{true}} = [(\omega_{\text{true}}, g_{\text{true}})]$. If $(\omega_{\text{true}}, g_{\text{true}}) \in \text{Octave}(\omega_0, g_0)$ for some reference oscillator parameters (ω_0, g_0) and $\varepsilon > 0$ is a target error tolerance, choose a parameter vector $\sigma_{\text{est}} = [(\omega_{\text{est}}, g_{\text{est}})]$ such that

$$(\omega_{\text{est}} - \omega_{\text{true}})^2 + (g_{\text{est}} - g_{\text{true}})^2 < \varepsilon. \tag{4.45}$$

There are two crucial distinctions to consider between Problem 3 and Problems 2 and 1. The first distinction is that Problem 3 contains no restriction on the choice of experiment parameters (v, t). This is because, as Ferrie et al. (2013) argue, wait times t should be roughly the inverse of the standard deviation of the coupling strength estimate. Restricting the maximum value of t also restricts the data-efficiency of producing an accurate estimate of the oscillator parameters. This point is explained further in Chapter 6.

The second distinction between the estimation problem and Problem 1 is the loss function (there is no explicit loss function for the detection problem). Whereas the motivation for using the supremum metric as a loss function in Problem 1 is based on the need to predict the outcome of experiments that have not been performed, the use of the (squared) Euclidean metric as a loss function for Problem 3 is based on computational needs. This change is not problematic because the supremum metric on the click probabilities can be made arbitrarily small by ensuring that the Euclidean distance between the estimated and true oscillator parameters is small enough.

The solutions to the detection and estimation problems, presented in Chapters 5 and 6, can be combined to produce extremly accurate estimates of unknown oscillator parameters with remarkably little data. The purpose of this section has been to explain what these algorithms are *designed* to do as opposed to what they are *capable* of doing.

DETECTING DEFECTS

You can observe a lot by watching.

C HARACTERIZING TLDSes requires a combination of techniques, each of which is designed to achieve a more limited objective. This thesis is focussed on the more limited challenge of characterizing just one TLDS, though the techniques are applicable for characterizing many well-separated TLDSes. In Section 4.3, the problem of characterizing a TLDS is further reduced to two easier problems: detection and estimation. This chapter describes a solution to the detection problem, and Chapter 6 describes a solution to the estimation problem. This chapter is entirely my own work, although Frank Wilhelm has provided helpful guidance and input at all stages.

In this chapter, TLDSes are detected in a manner analogous to the simplistic oracular approach described in Section 4.2. In that approach, access to an oracle $\mathcal{O}$ is presumed. This oracle is capable of determining with certainty whether the true oscillator parameters (ω, g) are contained in $\text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$: $\mathcal{O}(\omega_{\text{oct}}, g_{\text{oct}}) = 1$ if $(\omega, g) \in \text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$; $\mathcal{O}(\omega_{\text{oct}}, g_{\text{oct}}) = 0$ otherwise. It is possible to certify with $O(1/g)$ oracle queries that no oscillator with coupling strength greater than g is present within a given frequency range. This is optimal.

The main insight in this chapter is that a particular data collection heuristic, called the *octave sampling heuristic*, produces data that enables the design of a decision procedure

that behaves like the oracle $\mathcal{O}$. The decision procedure is not as reliable as $\mathcal{O}$, but the similarities are great enough that the resulting detection algorithm in this chapter requires $O(1/g)$ samples to produce a result similar to that which is produced by the oracle $\mathcal{O}$. The octave heuristic is motivated in part by an observation of Cole et al. (2005), wherein it was noted (in Figure 3(a)) that frequency information can be reconstructed from a time-dependent pdf even when only one bit is collected from each element of a chosen grid of times. We showed that this insight holds for our model in Figure 1(b) of Stenberg et al. (2014).

The purpose of this chapter is to present an algorithm for detecting oscillators in the swap spectrum model. The algorithm is the composition of two algorithms, Algorithm 5.2 and Algorithm 5.3. Algorithm 5.2 gathers data according to the octave sampling heuristic applied to several coupling octaves, and Algorithm 5.3 compares the resulting data from each octave to a threshold condition to decide whether an oscillator is present.

This chapter is organized as follows. Section 5.1 describes the octave sampling heuristic and its similarity with the oracle $\mathcal{O}$, described in Section 4.2. Section 5.2 presents the main algorithm as the composition of Algorithm 5.2 with Algorithm 5.3. Section 5.3 analyzes the performance of the detection program using both theoretical and numerical arguments. Finally, Section 5.4 discusses the potential for implementing this work in experiment to detect the influence of TLDSes and the statistics of their appearance and disappearance.

The Octave Sampling Heuristic §5.1

The purpose of this section is to explain the octave sampling heuristic, presented as Algorithm 5.1 and to derive an expression for the 'octave heuristic click probability', which is needed to analyze the performance of the detection program. The heuristic chooses experiment parameters (ν, t) in order to determine whether an oscillator is present based on the assumption that the oscillator, if present, has parameters in a particular octave. The result of SWAPSPECTROSCOPY(ν, t) is a click with a probability that is determined ultimately by the input to the octave heuristic and is called the 'octave heuristic click probability'.

The difference between the octave heuristic click probability in the presence or absence of an oscillator becomes negligible if the oscillator parameters are far away from the selected octave, so the value of Algorithm 5.1 is in its ability to choose experiments for which the octave heuristic click probability is markedly different depending on whether an oscillator is present with parameters in or near the selected octave.

The heuristic presumes that the oscillator parameters are chosen uniformly at random from within the coupling octave. In effect, the heuristic makes such a random selection and then acts accordingly. Suppose for the sake of argument that the goal is not to detect an oscillator whose parameters might belong to some parameter range. Instead, suppose that the goal is to determine whether there is (a) no oscillator, or (b) a single oscillator whose parameters are exactly (ω, g). In this case, the click probability for experiment parameters (ν, t) reads

$$p_{\text{click}}(\nu, t) = \frac{4g^2}{(\nu - \omega)^2 + 4g^2} \sin^2\left(\frac{1}{2}t\sqrt{(\nu - \omega)^2 + 4g^2}\right) \tag{5.1}$$

if hypothesis (b) is true. Otherwise, $p_{\text{click}}(\nu, t) \equiv 0$. The central observation of the octave heuristic is that, if hypothesis (b) is true, $p_{\text{click}}(\omega, \frac{\pi}{2g}) = 1$. Therefore, setting the experiment parameters so that $\nu = \omega$ and $t = \frac{\pi}{2g}$ ensures that a single bit of data suffices to distinguish between hypotheses (a) and (b).

Now suppose that hypothesis (b) is replaced with the promise that the oscillator parameters (ω, g) are drawn uniformly at random from $\text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$. The choice of measurement ought to remain the same: $\nu = \omega$ and $t = \frac{\pi}{2g}$. To account for the lack of knowledge of (ω, g), one could simply choose the pair uniformly at random from the same octave. Thus ν would be drawn uniformly at random from the interval $(\omega_{\text{oct}} - g_{\text{oct}}, \omega_{\text{oct}} + g_{\text{oct}})$ and $1/t$ drawn uniformly at random from $(g_{\text{oct}}/\pi, 2g_{\text{oct}}/\pi)$; t is distributed according to the inverse uniform distribution, whose pdf is proportional to $1/t^2$.

Algorithm 5.1 The octave sampling heuristic.

Input: octave frequency, ω_{oct}
Input: octave coupling strength, g_{oct}
Output: experiment parameters, (ν, t)
function OCTAVEHEURISTIC($\omega_{\text{oct}}, g_{\text{oct}}$)
    Draw ν from unif($\omega_{\text{oct}} - g_{\text{oct}}, \omega_{\text{oct}} + g_{\text{oct}}$).   ▷ The uniform distribution.
    Draw t from unif^{-1}($g_{\text{oct}}/\pi, 2g_{\text{oct}}/\pi$).   ▷ $\frac{\pi}{2g_{\text{oct}}} \leq t \leq \frac{\pi}{g_{\text{oct}}}$.
    return (ν, t)
end function

OCTAVEHEURISTIC ($\omega_{\text{oct}}, g_{\text{oct}}$) returns experimental parameters (ν, t) that can serve as input for SWAPSPECTROSCOPY: SWAPSPECTROSCOPY(ν, t) returns 1 (i.e. a click) with probability $p_{\text{click}}(\nu, t)$. In other words, SWAPSPECTROSCOPY ∘ OCTAVEHEURISTIC (the

composition of the two functions) returns a click with a probability depending on $(\omega_{\text{oct}}, g_{\text{oct}})$ as well as the true oscillator parameters (ω, g) if an oscillator is present. This is the octave heuristic click probability,

$$p_{\text{ohc}}\left(\omega, g; \omega_{\text{oct}}, g_{\text{oct}}\right) := \frac{\pi}{2g_{\text{oct}}^2} \int_{\omega_{\text{oct}} - g_{\text{oct}}}^{\omega_{\text{oct}} + g_{\text{oct}}} \int_{\frac{\pi}{2g_{\text{oct}}}}^{\frac{\pi}{g_{\text{oct}}}} \frac{p_{\text{click}}(\nu, t)}{t^2} \mathrm{d}t \mathrm{d}\nu. \tag{5.2}$$

To make the expression for p_{ohc} more manageable, the integral can be simplified to

$$\frac{\pi g^2}{2g_{\text{oct}}^2} \int_{\omega_{\text{oct}} - g_{\text{oct}}}^{\omega_{\text{oct}} + g_{\text{oct}}} \int_{\frac{\pi}{2g_{\text{oct}}}}^{\frac{\pi}{g_{\text{oct}}}} \operatorname{sinc}^2\left(\frac{t}{2}\sqrt{(\nu - \omega)^2 + 4g^2}\right) \mathrm{d}t \mathrm{d}\nu, \tag{5.3}$$

which can be further simplified by making the following substitutions:

$$\tau := gt; \ \gamma := \frac{g}{g_{\text{oct}}}; \ \delta := \frac{\nu - \omega}{2g}; \ \delta_{\text{oct}} := \frac{\omega_{\text{oct}} - \omega}{2g_{\text{oct}}}. \tag{5.4}$$

Thus the octave heuristic click probability is

$$p_{\text{ohc}}\left(\omega, g; \omega_{\text{oct}}, g_{\text{oct}}\right) = \pi \gamma^2 \int_{\frac{\delta_{\text{oct}} - \frac{1}{2}}{\gamma}}^{\frac{\delta_{\text{oct}} + \frac{1}{2}}{\gamma}} \int_{\frac{1}{2}\pi\gamma}^{\pi\gamma} \operatorname{sinc}^2\left(\tau\sqrt{1 + \delta^2}\right) \mathrm{d}\tau \mathrm{d}\delta \tag{5.5}$$

if the true oscillator parameters are (ω, g).

Although there does not seem to be an analytic formula for $p_{\text{ohc}}(\omega, g; \omega_{\text{oct}}, g_{\text{oct}})$, the octave heuristic click probability can be bounded by observing that

$$\int_{\frac{\delta_{\text{oct}} - \frac{1}{2}}{\gamma}}^{\frac{\delta_{\text{oct}} + \frac{1}{2}}{\gamma}} \int_{\frac{1}{2}\pi\gamma}^{\pi\gamma} \operatorname{sinc}^2\left(\tau\sqrt{1 + \delta^2}\right) \mathrm{d}\tau \mathrm{d}\delta < \int_{\frac{\delta_{\text{oct}} - \frac{1}{2}}{\gamma}}^{\frac{\delta_{\text{oct}} + \frac{1}{2}}{\gamma}} \int_{\frac{1}{2}\pi\gamma}^{\pi\gamma} \frac{\mathrm{d}\tau \mathrm{d}\delta}{\tau^2 \left(1 + \delta^2\right)}. \tag{5.6}$$

The right hand side of this inequality is easily calculated:

$$\int_{\frac{1}{2}\pi\gamma}^{\pi\gamma} \frac{\mathrm{d}\tau}{\tau^2} = \frac{1}{\pi\gamma}; \ \int_{\frac{\delta_{\text{oct}} - \frac{1}{2}}{\gamma}}^{\frac{\delta_{\text{oct}} + \frac{1}{2}}{\gamma}} \frac{\mathrm{d}\delta}{1 + \delta^2} = \arctan\left(\frac{\delta_{\text{oct}} + \frac{1}{2}}{\gamma}\right) - \arctan\left(\frac{\delta_{\text{oct}} - \frac{1}{2}}{\gamma}\right) \tag{5.7}$$

Therefore,

$$p_{\text{ohc}}\left(\omega, g; \omega_{\text{oct}}, g_{\text{oct}}\right) < \gamma \left(\arctan\left(\frac{\delta_{\text{oct}} + \frac{1}{2}}{\gamma}\right) - \arctan\left(\frac{\delta_{\text{oct}} - \frac{1}{2}}{\gamma}\right)\right). \tag{5.8}$$

Based on the above bound, the octave click probability proves negligible in two important limits: $|\delta_{\text{oct}}| \to \infty$ and $\gamma \to 0$. These limits are important because they represent the possibility that oscillators are present but cannot be detected by sampling from a fixed octave. The limit $|\delta_{\text{oct}}| \to \infty$ represents the case that the oscillator's frequency ω is far from ω_{oct}, and $\gamma \to 0$ represents the case that the oscillator's coupling strength g is negligible compared to g_{oct}. If $|\delta_{\text{oct}}| > \frac{1}{2}$, we can use the identity

$$\arctan(\alpha) - \arctan(\beta) = \arctan\left(\frac{\alpha - \beta}{1 + \alpha\beta}\right) \tag{5.9}$$

to compute

$$p_{\text{ohc}}\left(\omega, g; \omega_{\text{oct}}, g_{\text{oct}}\right) < \gamma \arctan\left(\frac{\gamma}{\gamma^2 + \delta_{\text{oct}}^2 - \frac{1}{4}}\right). \tag{5.10}$$

Clearly, $p_{\text{ohc}}\left(\omega, g; \omega_{\text{oct}}, g_{\text{oct}}\right) \to 0$ as $|\delta_{\text{oct}}| \to \infty$. The limit $\gamma \to 0$ can easily be computed by using the identity

$$\arctan\left(\frac{\delta_{\text{oct}} + \frac{1}{2}}{\gamma}\right) - \arctan\left(\frac{\delta_{\text{oct}} - \frac{1}{2}}{\gamma}\right) < \arctan(+\infty) - \arctan(-\infty) = \pi, \tag{5.11}$$

which holds because arctan is a monotonically increasing function. Thus

$$p_{\text{ohc}}\left(\omega, g; \omega_{\text{oct}}, g_{\text{oct}}\right) < \pi\gamma; \tag{5.12}$$

therefore, $p_{\text{ohc}}\left(\omega, g; \omega_{\text{oct}}, g_{\text{oct}}\right) \to 0$ as $\gamma \to 0$. These limits for the octave heuristic click probability demonstrate that some oscillators are simply invisible if only experiments from that one octave are considered. In order to find such oscillators, samples must be gathered from other octaves.

The importance of these expressions for p_{ohc} is that they define the 'visibility' of an oscillator from the point of view of samples gathered from the octave heuristic applied to a given octave. Many clicks will be recorded if the oscillator is 'close' to the input octave and few clicks will be recorded otherwise, so the number of observed clicks can be compared to some threshold value; a below-threshold number of clicks constitutes evidence that any oscillator parameters are far from the specified octave. Thus, the composition of SWAPSPECTROSCOPY with OCTAVEHEURISTIC produces data that can be used to make an accurate decision about whether an oscillator is present with parameters comparable to the given octave.

The Detection Algorithm §**5.2**

The detection algorithm is a composition of two other algorithms: an algorithm that collects data and an algorithm that makes a decision, based on that data, about whether an oscillator is present. The data collection algorithm, Algorithm 5.2, gathers a prescribed amount of data via the octave sampling heuristic applied to a root octave and many of its children, to use the terminology of Section 4.2. The decision algorithm simply evaluates this data to determine whether any of the sampled octaves have returned a number of clicks greater than a specified threshold value; if not, the null hypothesis is accepted.

Algorithm 5.2 prescribes a method for collecting data from each of many coupling octaves and storing that data in a structured fashion. The resulting data structure, represented as DATA, is an associative array whose keys are binary strings and whose values are data sets of length n_{mmts}, which is an input to the algorithm. The keys have a maximum length $\lambda_{\max}$ (also an input), which ensures that no experiments with $t > t_{\max}$ are performed. Indeed, an octave key b with length λ specifies an octave that could choose t to be as large as

$$2^{\lambda+1} \times \frac{\nu_{\max} - \nu_{\min}}{\pi}. \tag{5.13}$$

To ensure $t < t_{\max}$, set

$$\lambda_{\max} < \log_2 \left(\frac{\pi t_{\max}}{\nu_{\max} - \nu_{\min}} \right) - 1. \tag{5.14}$$

The output of DATACOLLECTOR, the data structure DATA, is most easily understood through Figure 5.1. In this figure, the click probabilities of several octaves are displayed in two different views. The image on the left represents the control-space view, which is to say that it displays $\text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$ as a rectangle specified by the parameter ranges

$$\omega_{\text{oct}} - g_{\text{oct}} < \nu < \omega_{\text{oct}} + g_{\text{oct}}, \quad \frac{\pi}{2g_{\text{oct}}} < t < \frac{\pi}{g_{\text{oct}}}. \tag{5.15}$$

The longer the octave key, the taller and thinner the rectangle and the higher the rectangle on the t-axis. The background image is the true swap spectrum; notice that the darkest rectangles correspond to the bluest backgrounds. The image on the right displays the parameter-space view, which is to say that $\text{Octave}(\omega_{\text{oct}}, g_{\text{oct}})$ corresponds to the rectangle specified by the parameter ranges

$$\omega_{\text{oct}} - g_{\text{oct}} < \omega < \omega_{\text{oct}} + g_{\text{oct}}, \quad \frac{g_{\text{oct}}}{2} < g < g_{\text{oct}}. \tag{5.16}$$

Algorithm 5.2 The data collection procedure.

Input: minimum value of ν, ν_{min}
Input: maximum value of ν, ν_{max}
Input: depth, λ_{max}
Input: measurement shots per octave, n_{mmts}
Output: structured data, DATA
function DATACOLLECTOR($\nu_{min}, \nu_{max}, \lambda_{max}, n_{mmts}$)
$\quad g_0 \leftarrow \frac{1}{2}\left(\nu_{max} - \nu_{min}\right)$
$\quad$**for** $\lambda \leftarrow 0, \ldots, \lambda_{max}$ **do**
$\quad\quad g_{oct} \leftarrow g_0/2^{\lambda}$
$\quad\quad$**for** $b \in$ BITSTRINGS(λ) **do** $\qquad\qquad$ ▷ b ranges over all bit strings of length λ
$\quad\quad\quad s \leftarrow$ SHIFT(b) $\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Algorithm 4.1.
$\quad\quad\quad \omega_{oct} \leftarrow \nu_{min} + 2sg_{oct}$
$\quad\quad\quad$**for** $i \leftarrow 1 \ldots n_{mmts}$ **do**
$\quad\quad\quad\quad (\nu_i, t_i) \leftarrow$ OCTAVEHEURISTIC(ω_{oct}, g_{oct}) $\qquad$ ▷ Algorithm 5.1.
$\quad\quad\quad\quad d_i \leftarrow$ SWAPSPECTROSCOPY (ν_i, t_i) $\qquad\qquad$ ▷ Algorithm 2.1.
$\quad\quad\quad$**end for**
$\quad\quad\quad$DATA$(b) \leftarrow \left[(\nu_1, t_1, d_1), \ldots, (\nu_{n_{mmts}}, t_{n_{mmts}}, d_{n_{mmts}})\right]$
$\quad\quad$**end for**
$\quad$**end for**
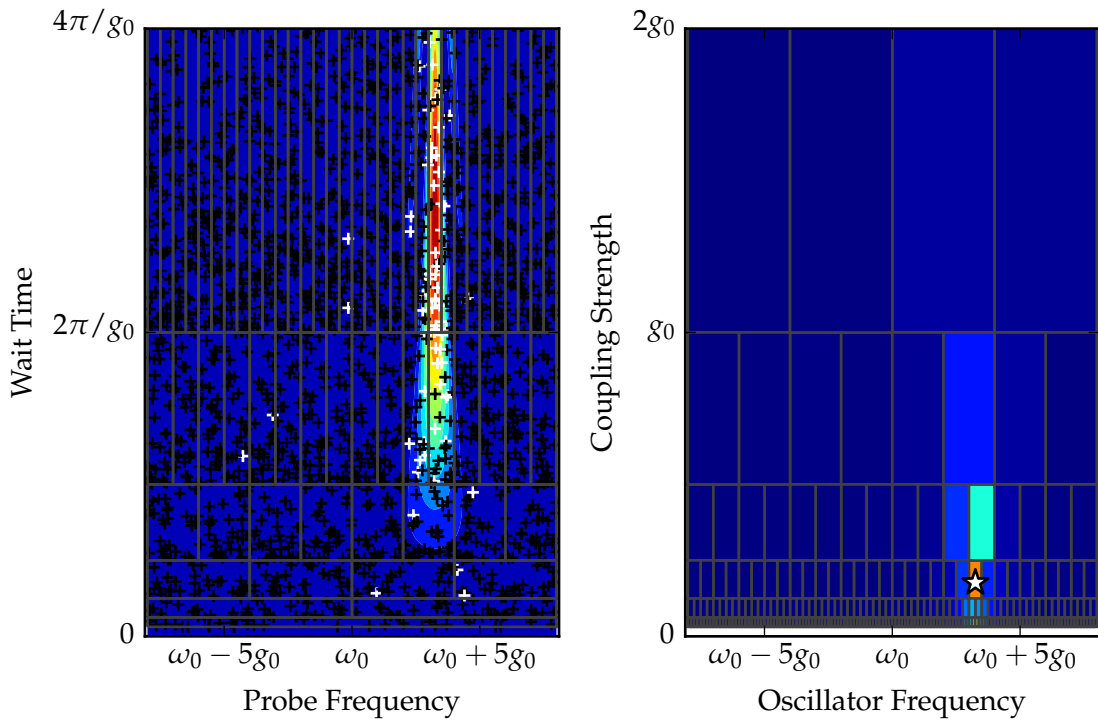$\quad$**return** DATA
end function

Figure 5.1: An illustration of the data collection algorithm. The figure on the left demonstrates a simulated data set (32 bits per octave) gathered from a single-oscillator swap spectrum; $\omega_{\text{true}} = \omega_0 + \frac{13}{4}g_0$ and $g_{\text{true}} = \frac{1}{\sqrt{32}}g_0$. The '+' markers indicate the values of (ν, t) and the colour indicates the outcome: white for a click, black otherwise. The gray lines outline the support of various octaves used in the data collection algorithm. The plot is superposed on the true click probability. On the right, the octaves are displayed as a function of the oscillator parameter ranges about which they report. The colour of each box is the average number of clicks after 1000 trials, and the white star represents $(\omega_{\text{true}}, g_{\text{true}})$.

The insight underlying the decision procedure, Algorithm 5.3, is made apparent by Figure 5.1 as well as the analysis of p_{ohc} in Section 5.1. Clearly, the octave heuristic click probability is small if the oscillator is 'far away' from the specified octave; that is, the oscillator is far-detuned from the octave or weakly coupled compared to the octave coupling strength. Under ideal conditions (no decoherence and no measurement error), even a single click indicates that an oscillator is present. But of course this is not true if any experimental imperfections hold, and it does not determine which octave is likely to contain the oscillator. A more robust approach is to compare the number of clicks in a given octave to some threshold click probability. Algorithm 5.3 therefore counts the number of clicks observed for each octave key in the output of Algorithm 5.2 and compares it to some input threshold click probability κ. Algorithm 5.3 returns a decision to accept the null hypothesis if and only if the number of clicks in every octave failed to surpass threshold.

Algorithm 5.3 The decision procedure.

Input: Data structure, DATA.
Input: Threshold click probability, κ.
Output: Decision to ACCEPT or REJECT the null hypothesis.
function DECISION(DATA, κ)
    for $b \in$ KEYS(DATA) **do**   ▷ b represents a bit string.
        $n_{mmts} \leftarrow$ LENGTH(DATA(b))
        $[(\nu_1, t_1, d_1), \ldots, (\nu_{n_{mmts}}, t_{n_{mmts}}, d_{n_{mmts}})] \leftarrow$ DATA(b)
        $n_{clicks} \leftarrow \sum_k d_k$
        if $n_{clicks} \geq \kappa n_{mmts}$ **then**
            return REJECT
        end if
    end for
    return ACCEPT
end function

The TLDS detection algorithm is DECISION ∘ DATACOLLECTOR, the composition of Algorithm 5.3 with Algorithm 5.2. The input to the combined algorithm is a range of frequencies (ν_{min}, ν_{max}) and a maximum depth λ_{max} of children of the root octave specified by the range (ν_{min}, ν_{max}). The output is a decision about whether to accept or reject the null hypothesis, which is that there are no oscillators with parameters (ω, g) such that $\omega \in (\nu_{min}, \nu_{max})$ and $g \geq (\nu_{max} - \nu_{min})/2^{\lambda_{max}+1}$ (where it is promised that $g < (\nu_{max} - \nu_{min})/2$ even if there is an oscillator present). The performance of this algorithm is analyzed in the next section.

74

The aim of this section is assess the performance of the detection algorithm. This section begins with a discussion of the trade-off between the data cost and the failure probability of the algorithm. Input values are suggested based on an analysis of this trade-off, summarized by Figure 5.2. Finally, the statistics of several exemplary tests of the algorithm are presented in Table 5.1.

As the purpose of the algorithm is to accept or reject the null hypothesis that no oscillators are present, the performance assessment should be based partly on the probability that the algorithm fails to make the correct decision. Of course the performance of the algorithm depends on several inputs; namely the number of measurements per octave, n_{mmts}, the maximum depth of octaves to be sampled, $\lambda_{\max}$, and the threshold click probability, κ. Whereas n_{mmts} and $\lambda_{\max}$ specify the data cost of the detection algorithm, the choice of κ dictates the probability with which the algorithm makes the wrong decision.

The choice of κ should find a balance between the two competing goals of minimizing Type I and Type II error probabilities. In the language of statistical hypothesis testing, Type I and Type II error probabilities refer to the probability that the null hypothesis, if true, is rejected (Type I) and that the null hypothesis, if false, is accepted (Type II). These two goals compete because increasing the value of κ decreases the Type I error probability but increases the Type II error probability.

To balance between these priorities, I choose $\kappa = 3/16$. The justification for this choice is summarized in Figure 5.2. The first observation is extracted from the plot on the left of Figure 5.2, which demonstrates that $p_{\mathrm{ohc}}(\omega, g; \omega_0, g_0) \gtrsim 0.4$ if $(\omega, g) \in \mathrm{Octave}(\omega_0, g_0)$. Supposing that the oscillator (if present) is found $\mathrm{Octave}(\omega_0, g_0)$, the Type II error probability can be approximately bounded above by presuming that $p_{\mathrm{ohc}}(\omega, g; \omega_0, g_0) = 0.4$. Using this assumption, the probability of observing an above-threshold number of clicks can be calculated by computing the cumulative distribution function of the Binomial distribution with success probability 0.4 and number of trials equal to n_{mmts} at the value κn_{mmts}; that is, the Type II error probability is bounded above by

$$\sum_{\ell=0}^{\lfloor \kappa n_{\mathrm{mmts}} \rfloor} \binom{n_{\mathrm{mmts}}}{\ell} p_{\mathrm{ohc}}^{\ell} (1 - p_{\mathrm{ohc}})^{n_{\mathrm{mmts}} - \ell} \tag{5.17}$$

for $p_{\mathrm{ohc}} = 0.4$. The top right plot of Figure 5.2 plots this value as a function of κ for $n_{\mathrm{mmts}} = 16$, 32, and 64. The value $\kappa = 3/16$ ensures that the Type II error probability is less than 5% for $n_{\mathrm{mmts}} = 16$, less than 1% for $n_{\mathrm{mmts}} = 32$, and less than .01% for $n_{\mathrm{mmts}} = 64$.
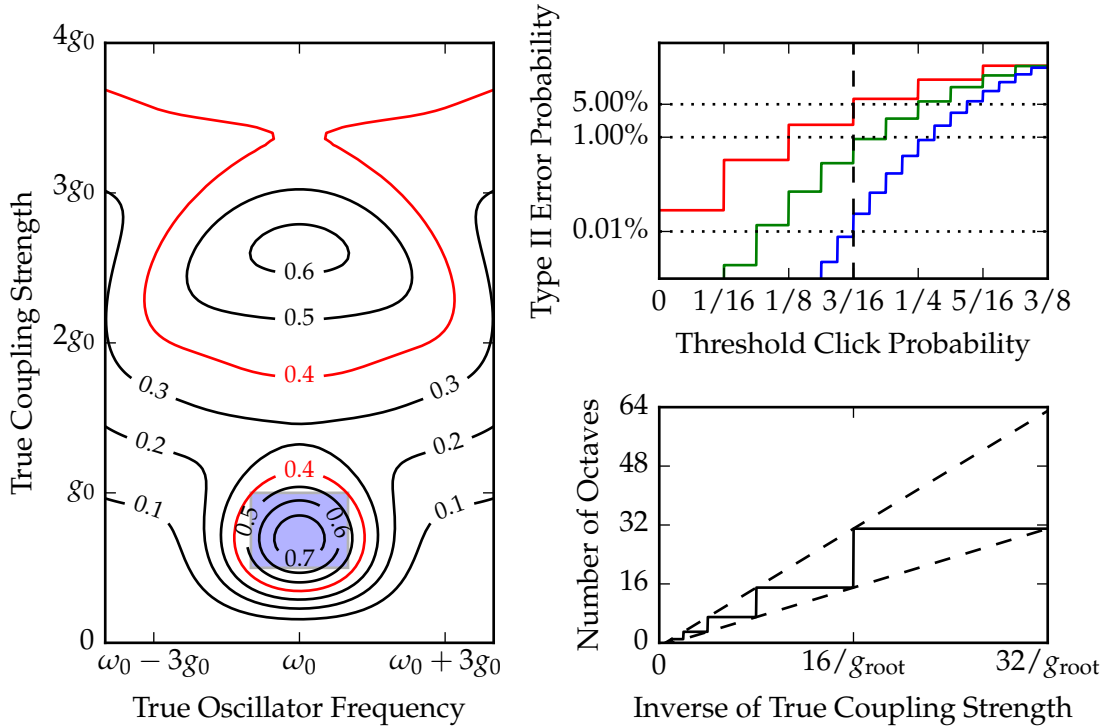
Figure 5.2: A summary of the data costs in the detection algorithm. The plot on the left depicts $p_{\mathrm{ohc}}(\omega, g; \omega_0, g_0)$; ω is the abscissa and g the ordinate. The contours depict level sets of p_{ohc} of the indicated values. The transparent blue rectangle depicts the range of parameters corresponding to Octave(ω_0, g_0). The red contour highlights the level set $p_{\mathrm{ohc}}(\omega, g; \omega_0, g_0) = 0.4$, which is selected for further analysis in the plot on the top right. The plot on the top right analyzes different choices of the threshold click probability κ input to Algorithm 5.3 based on the presumption that $p_{\mathrm{ohc}}(\omega, g; \omega_0, g_0) = 0.4$, selected because the plot on the left indicates that 0.4 is an approximate lower bound to $p_{\mathrm{ohc}}(\omega, g; \omega_0, g_0)$ if $(\omega, g) \in$ Octave(ω_0, g_0). The abscissa value represents a choice of κ, and the ordinate value indicates the probability that the number of clicks will fail to surpass κn_{mmts} if $n_{\mathrm{mmts}} = 16$ (red), $n_{\mathrm{mmts}} = 32$ (green), and $n_{\mathrm{mmts}} = 64$ (blue). The plot on the bottom right indicates the number of octaves that must be sampled by Algorithm 5.2 to ensure that an oscillator with the indicated inverse coupling strength is in an octave that has been sampled. The horizontal axis is the ratio g_{root}/g, where $g_{\mathrm{root}} := \frac{1}{2}(\omega_{\max} - \omega_{\min})$ is the coupling strength covered by the 'root' octave of the data collection algorithm. The vertical axis is the number of octaves specified by Eq. (5.18). The solid lines indicate the number of octaves n_{oct} that contain oscillators of coupling strength at least g, and the dashed lines represent the linear bounds $1/g - 1$ and $2/g - 1$.

The data cost of the detection algorithm can be bounded as follows. The number of bits of data produced by DATACOLLECTOR is $n_{\text{bits}} = n_{\text{mmts}} \times n_{\text{oct}}$, where $n_{\text{oct}} = 2^{\lambda_{\max}+1} - 1$ is the number of octaves from which data was collected. Using the assumption that $p_{\text{ohc}} = 0.4$ if the oscillator is within the specified coupling octave and $p_{\text{ohc}} = 0$ otherwise, a decision to reject the null hypothesis based on n_{bits} bits of data is a decision to accept the proposition that an oscillator is present with $g \geq g_{\min} := g_0/2^{\lambda_{\max}+1}$ and $\omega \in (\nu_{\min}, \nu_{\max})$. Therefore, the decision to accept the null hypothesis is a decision to accept the proposition that there is no oscillator with these parameters. This decision is made at a cost of n_{bits} bits of data. Notice that the number of octaves depends only on $\lambda_{\max}$ and that this number determines the minimal coupling strength of detectible oscillators: the number of octaves needed to detect an oscillator within a frequency range is inversely proportional to the coupling strength of g. As in Section 4.2,

$$n_{\text{oct}} = 2^{\lceil \log_2(g_0) - \log_2(g) \rceil} - 1. \tag{5.18}$$

Thus, $n_{\text{bits}} \in \Theta(n_{\text{mmts}}/g)$ are needed to detect an oscillator with coupling strength g in some specified frequency range.

Theoretical analysis not withstanding, the performance of the detection algorithm is easily seen through simulation. Table 5.1 depicts the results of such simulations for nine representative choices of oscillator parameters, distributed evenly around a specific octave. Performance like this can be expected as long as the maximum wait time remains well below the characteristic decoherence time and as long as the probability of measurement error is well below κ.

Ideally, the Type I error probability is zero: if there is no oscillator and no decoherence, the click probability is always zero. More realistically, a non-zero decoherence rate should be assumed; in this case, the octave heuristic click probability is

$$p_{\text{ohc}}(\Gamma; \omega_{\text{oct}}, g_{\text{oct}}) := \frac{\pi}{2g_{\text{oct}}^2} \int_{\omega_{\text{oct}}-g_{\text{oct}}}^{\omega_{\text{oct}}+g_{\text{oct}}} \int_{\frac{\pi}{2g_{\text{oct}}}}^{\frac{\pi}{g_{\text{oct}}}} \frac{\exp(-\Gamma t)}{t^2} \, dt \, d\nu, \tag{5.19}$$

where Γ is a decoherence rate affecting the qubit. By setting $\Delta := \pi\Gamma/g_{\text{oct}}$, this reduces to

$$p_{\text{ohc}}(\Gamma; \omega_{\text{oct}}, g_{\text{oct}}) = 1 - (1 - \exp(-\Delta/2))^2 - \Delta(\text{Ei}(-\Delta) - \text{Ei}(-\Delta/2)) \tag{5.20}$$

It turns out that

$$p_{\text{ohc}}(\Gamma; \omega_{\text{oct}}, g_{\text{oct}}) \approx \exp(-\mu\Delta), \tag{5.21}$$

for $\mu = 0.67286$, a value chosen by linear regression. This expression is of course negligibly small if $\Delta \ll 1$. Therefore, even the observation of a single click if $\Delta \ll 1$ is

77

$(\omega_{\text{true}} - \omega_0)/g_0$	0	0	0	1/8	1/8	1/8	1/4	1/4	1/4
g_{true}/g_0	1/16	3/32	1/8	1/16	3/32	1/8	1/16	3/32	1/8
root	0.0	0.08	1.57	0.01	0.09	1.4	0.0	0.11	1.4
0	0.33	8.55	43.71	0.05	3.86	24.27	0.05	1.1	11.09
00	0.0	0.01	0.15	0.0	0.0	0.01	0.0	0.0	0.01
000	0.0	0.01	0.1	0.0	0.0	0.05	0.0	0.0	0.0
001	0.02	0.46	4.3	0.0	0.04	0.52	0.0	0.0	0.12
01	43.55	96.35	99.92	10.98	64.08	92.46	0.51	10.48	34.83
010	0.2	3.32	20.47	0.01	1.41	11.47	0.01	0.48	4.12
011	99.94	100.0	99.69	34.01	59.5	58.57	0.18	3.68	20.91
1	0.23	7.93	43.15	0.47	15.17	59.31	0.99	21.08	71.21
10	42.71	96.03	99.95	72.13	99.89	100.0	79.06	100.0	100.0
100	99.93	99.99	99.68	100.0	100.0	100.0	99.96	100.0	99.57
101	0.12	3.46	20.65	34.68	59.41	57.36	99.95	100.0	99.76
11	0.0	0.03	0.2	0.01	0.34	2.55	0.64	10.7	34.08
110	0.0	0.45	4.13	0.02	1.22	12.13	0.15	3.66	20.15
111	0.0	0.01	0.06	0.0	0.05	0.56	0.0	0.4	4.76

Table 5.1: The null hypothesis rejection probabilities of the detection algorithm for each octave of collected data and for each of several true oscillator parameters. The root octave is $\text{Octave}(\omega_0, g_0)$, and the children octaves are labelled accordingly. The nine true oscillator parameters are evenly distributed around the '100' octave, which corresponds to $\text{Octave}(\omega_0 + g_0/8, g_0/8)$. The entries of the table represent the percentage of 10000 trials that return a decision to reject the null hypothesis. The null hypothesis is rejected if any one octave is rejected. The null hypothesis was rejected during each of the 10000 runs, and no run has ever accepted the null hypothesis—a strong indication that the Type II error probability is well below 10^{-5}.

ample evidence to distinguish between the null hypothesis that no oscillators are present and the alternate hypothesis that at least one oscillator is present. If n_{mmts} samples are to be taken from this octave and the resulting number of clicks is to be compared with a threshold click probability κ, the approximate chance of (falsely) rejecting the null hypothesis based on the data collected from one octave is

$$1 - \sum_{\ell=0}^{\lfloor \kappa n_{\mathrm{mmts}} \rfloor} \binom{n_{\mathrm{mmts}}}{\ell} \exp(-\ell \mu \Delta) \, (1 - \exp(-\mu \Delta))^{n_{\mathrm{mmts}} - \ell}. \qquad (5.22)$$

The overall probability of falsely rejecting the null hypothesis given all the data produced from Algorithm 5.2 is of course much smaller than this, so the Type I error probability remains negligible provided Δ is large for each octave.

Outlook   §5.4

The detection algorithm of this chapter is composed of a carefully designed data collection algorithm and an extremely simple decision algorithm. I have separated the detection algorithm into two pieces because I believe the data collection algorithm is more powerful than this thesis indicates. There are several reasons why I believe this.

Perhaps the most important trait of the data collection algorithm from a practical point of view is that it does not require interactive access with an experiment, unlike the main algorithm of Chapter 6. Latency (Section 6.4) is therefore not an issue. In principle, the 'offline' nature of the algorithm should impose a cost to data efficiency, but the argument of Sections 4.2 and 5.3 gives a strong indication that the data cost of the algorithm scales optimally as a function of the inverse coupling strength of detectable oscillators. So the data collection algorithm is more practical than the estimation algorithm of Chapter 6 and yet this practicality comes at no appreciable performance cost. This is not to say that one algorithm is better than the other—they accomplish different tasks—but that the prospects for immediate experimental execution of the data collection algorithm are better than for Algorithm 6.1.

Practicalities notwithstanding, the data collection algorithm is powerful from a theoretical point of view because the resulting data can be used as an input not only to Algorithm 5.3 but to other inference algorithms. In principle, the data could be used to evaluate a posterior distribution using particle representation methods. However, I suggest a much simpler approach for producing an estimate of unknown oscillator parameters in Algorithm 5.4.

Algorithm 5.4 Procedure to establish state of knowledge.

Input: data structure, DATA
Input: minimum value of v, $v_{\min}$
Input: maximum value of v, $v_{\max}$
Input: number of particles, n_{ptcls}
Output: particle locations, $\{x_k\}$
for $b \in \mathrm{KEYS}(\mathrm{DATA})$ **do** $\triangleright$ b represents a bit string.
    $[(v_1, t_1, d_1), \ldots, (v_{n_{\mathrm{trials}}}, t_{n_{\mathrm{trials}}}, d_{n_{\mathrm{trials}}})] \leftarrow \mathrm{DATA}(b)$
    $n_{\mathrm{clicks}} \leftarrow \sum_k d_k$
    $w_b \leftarrow n_{\mathrm{clicks}} / n_{\mathrm{trials}}$
end for
$w_{\mathrm{tot}} \leftarrow \sum_b w_b$
for $k \leftarrow 1 \ldots n_{\mathrm{ptcls}}$ **do**
    draw b with probability w_b / w_{tot}
    $\lambda \leftarrow \mathrm{LENGTH}(b)$
    $s \leftarrow \mathrm{SHIFT}(b)$
    $g_{\mathrm{oct}} \leftarrow 2^{-\lambda-1}(v_{\max} - v_{\min})$
    $\omega_{\mathrm{oct}} \leftarrow v_{\min} + 2sg_{\mathrm{oct}}$
    draw x_k uniformly at random from $\mathrm{Octave}(\omega_{\mathrm{oct}}, g_{\mathrm{oct}})$
end for
return $\{x_k\}$

Algorithm 5.4 produces a particle representation that is a crude estimate of likely oscillator parameters. The algorithm uses the fact that the octave heuristic click probability for Octave($\omega_{\text{oct}}, g_{\text{oct}}$) is high if and only if the oscillator parameters (ω, g) satisfy $|\omega - \omega_{\text{oct}}| \lesssim 2g_{\text{oct}}$ and $g/g_{\text{oct}} \lesssim 2$. This is apparent in Figure 5.1. The idea of Algorithm 5.4 is therefore to treat the right hand side of Figure 5.1 as a distribution of likely oscillator parameters. Of course one could attempt to directly evaluate the likelihood function of the data collected by Algorithm 5.2, but this likelihood function will almost certainly have a tall and narrow peak when compared to the wide range of parameters initially being considered. Particle representations run a substantial risk of failing on such distributions because the effective sample size of the representation is expected to be small. Algorithm 5.4 sacrifices accuracy for reliability.

The output of Algorithm 5.4 can be used as the input to the online estimation algorithm I introduced with my colleagues in Stenberg et al. (2014). Figure 5.3 depicts a crude test. For randomly chosen oscillators from octaves at level four, the data from a depth-7, 32-bit-per-octave call of DATACOLLECTOR is used as input for Algorithm 5.4 and the resulting particle representation is used as input to Algorithm 6.1, the algorithm of Stenberg et al. (2014). The result is analogous to Figure 6.3.

There is also a clear possibility of using data from DATACOLLECTOR to detect many oscillators. Provided that oscillators are well-separated (defined in Section 4.1.3) and that no oscillators occupy octaves of a high depth, the self-similarity of the data collection algorithm can be exploited to detect multiple oscillators. This is most easily explained by example. First, collect data for $\lambda_{\text{max}} = 10$:

$$\text{DATA} \leftarrow \text{DATACOLLECTOR} \left(\nu_{\text{min}}, \nu_{\text{max}}, 10, n_{\text{mmts}} \right). \tag{5.23}$$

Then define $\text{DATA}_{\text{upper}}$ as follows:

$$\text{DATA}_{\text{upper}}(\boldsymbol{b}) \leftarrow \begin{cases} \text{DATA}(\boldsymbol{b}) & \text{LENGTH}(\boldsymbol{b}) < 5, \\ \varnothing & \text{otherwise.} \end{cases} \tag{5.24}$$

Supposing that $\text{DECISION}\left(\text{DATA}_{\text{upper}}, \kappa\right)$ returns ACCEPT, proceed as follows. For each length-five bit string $\boldsymbol{b}_5$,

$$\text{DATA}_{\boldsymbol{b}_5}\left(\boldsymbol{b}_{<5}\right) \leftarrow \text{DATA}\left(\boldsymbol{b}_5 \oplus \boldsymbol{b}_{<5}\right), \tag{5.25}$$

where $\boldsymbol{b}_{<5}$ is any bit string of length less than 5 and $\boldsymbol{b}_\alpha \oplus \boldsymbol{b}_\beta$ represents string concatenation rather than bit-wise sum. Then the decision algorithm applied to each one of $\text{DATA}_{\boldsymbol{b}_5}$ can return a separate decision about whether an oscillator is present in each of
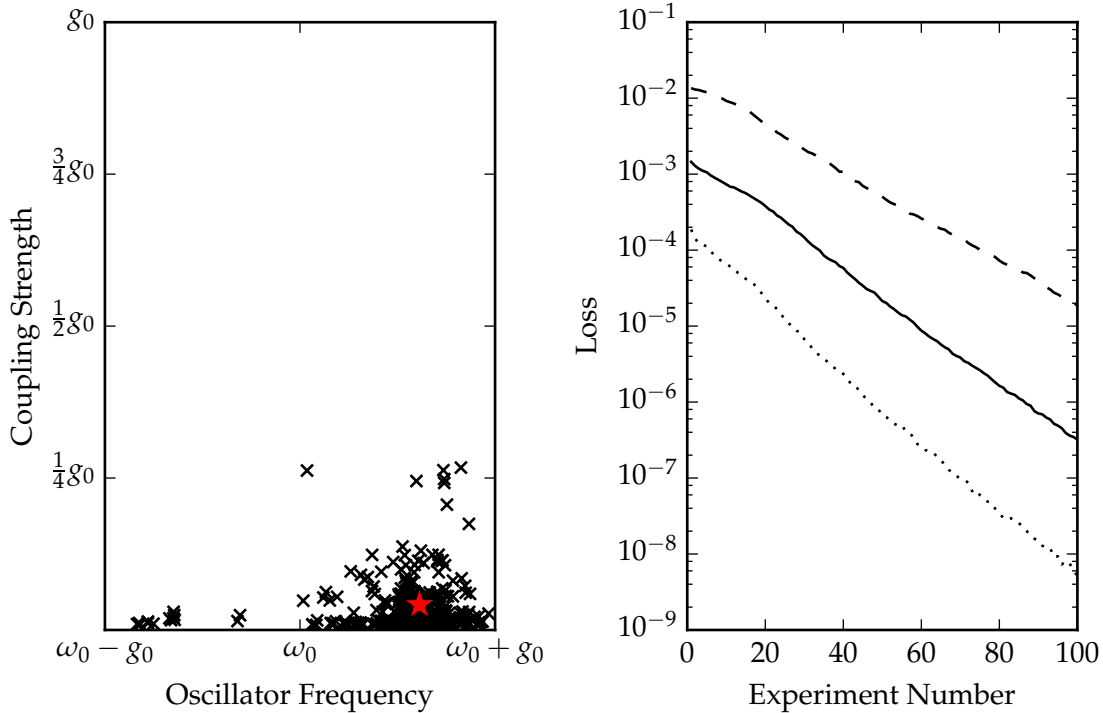
Figure 5.3: A performance test of Algorithm 5.4, which can be used to combine the detection algorithm with the estimation algorithm presented in Chapter 6. The figure on the left gives an example output ($n_{ptcls} = 500$) of Algorithm 5.4 for which the input is the output of Algorithm 5.2 with $n_{mmts} = 32$, $\lambda_{max} = 7$, $\nu_{min} = \omega_0 - g_0$, and $\nu_{max} = \omega_0 + g_0$ (so the root octave is Octave(ω_0, g_0)). Each black cross represents a particle location and the red star indicates the true oscillator parameters, which was randomly chosen from depth-4 octaves. The figure on the right is a decile plot of the kind shown on the bottom left of Figure 6.2. Here Algorithm 6.1 is composed with Algorithm 5.4 and Algorithm 5.2, respectively, where n_{mmts}, λ_{max}, ν_{min}, and ν_{max} are unchanged from the values used to produce the plot on the left but n_{ptcls} is set to 10000. The horizontal axis represents the number of experiments performed and the veritical axis represents the loss defined in Problem 3. The figure summarizes the data from 5000 trials for which the true oscillator parameters are drawn randomly from depth-4 octaves. The solid line depicts the median of the estimates (i.e. the fifth decile), the dashed curve is the ninth decile, and the dotted is the first decile. That is, 90% of the 5000 trials fall below the dashed curve, 50% below the solid curve, and 10% below the dotted curve.

the thirty-two distinct parameter ranges specified by a length-five bit string. Provided that oscillators are well-separated enough, this method should work to detect multiple weakly coupled oscillators.

The main point is that DATACOLLECTOR 'plugs in' to a wide variety of useful algorithms for assessing the swap spectrum. Although the main argument of this chapter is that the composition of DATACOLLECTOR with DECISION solves the detection problem of Section 4.3.2, this is far from the only use case for DATACOLLECTOR. As DATACOLLECTOR is an offline data collection algorithm, the prospects for implementing it in experiment are quite good. Future work should therefore aim at implementing it and designing algorithms to answer questions about the influence of TLDSes in SICs.

ESTIMATING DEFECT PARAMETERS

It appears to be a quite general principle that, whenever there is a randomized way of doing something, then there is a nonrandomized way that delivers better performance but requires more thought.

EDWIN JAYNES

THE problem of characterizing TLDSes, if posed as a parameter estimation problem, is amenable to modern online statistical inference techniques of the kind described by Granade et al. (2012). To apply these techniques, two things are needed: reasonably informative prior knowledge about the parameters to be estimated, and a method for choosing experiments based on current knowledge. I have co-authored a paper, published as Stenberg et al. (2014), that explains how to apply the techniques of Granade et al. (2012) to the problem of estimating the parameters of a single TLDS. In this chapter, I describe the result.

My contribution was to recognize the applicability of Granade et al. (2012) to the problem of characterizing TLDSes in superconducting qubits. I explained that the model constitutes a generalization of a toy problem considered by Ferrie et al. (2013) and is therefore amenable to impressive data-efficiency if experiments are selected according to a generalization of the heuristic proposed by Ferrie et al. (2013). In the ideal case, our algorithm appears to provide estimates of the oscillator parameters that improve exponentially in the number of bits of data that have been collected; machine precision is

typically achieved with 100-200 bits of experimental data. By contrast, standard swap spectroscopy collects 300-900 bits of data to produce one pixel in an image such as Figure 2.2; an image which contains 5054 pixels.

The main result of this chapter is Algorithm 6.1. The objective accomplished by the algorithm is to quickly improve a crude initial estimate of the TLDS parameters through online sampling. By "online" sampling, I mean that the algorithm uses SWAP-SPECTROSCOPY (Algorithm 2.1) in a for-loop that includes an update to the particle representation expressing the current knowledge about TLDS parameters. Thus the algorithm requires the ability to choose which experiment to run as it processes data from previous experiments.

This chapter also contains some unpublished work that explains an issue that my co-authors and I discussed in Stenberg et al. (2014). We referred to this issue as the problem of "outliers", which are instances of the estimation problem where our algorithm fails to perform satisfactorily. In this chapter, I show that the issue seems to occur when oscillators are weakly coupled.

This chapter is organized as follows. In Section 6.1, I describe the heuristic that forms the core of Algorithm 6.1, which is the main result and is presented in Section 6.2. I analyze the performance of Algorithm 6.1 and resolve the problem of outliers in Section 6.3. I conclude the chapter in Section 6.4 by discussing the possible implementation of Algorithm 6.1 in experiment.

The Estimation Heuristic §6.1

The purpose of this section is to explain the reasoning behind the experimental design heuristic that is the core of Algorithm 6.1. I explain that our heuristic is a combination of four heuristics that are distinguished by each of two distinctions between objectives. One distinction is between the objective of improving estimates based on the assumption that the prior distribution is approximately normal and the objective of ensuring that the prior distribution is approximately normal, and the other distinction is between choosing ν and choosing t.

Our heuristic is *greedy*, which is to say that each measurement choice is intended to minimize the posterior distribution obtained by incorporating the result of only that measurement. By contrast, a non-greedy strategy might sacrifice short-term gains in order to achieve longer-term gains. We choose ν and t separately and greedily: we choose $\nu \approx \omega_{\text{est}}$ because the oscillator is most visibile if $\nu \approx \omega_{\text{true}}$, and we choose $t \approx 1/\sigma_g$
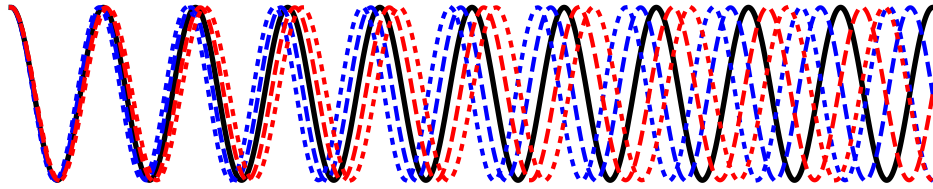
Figure 6.1: Illustration of the intuition underlying the frequency estimation heuristic of Ferrie et al. (2013). A reference sinusoid (solid black curve) is compared to four other sinusoids of approximately, but not exactly, equal frequencies. The blue dashed/dotted curves represent sinusoids with 2.5%/5% higher frequency, and the red dashed/dotted curve represent sinusoids with 2.5%/5% lower frequency.

(where σ_g is the standard deviation of the prior distribution of g) based on an argument of Ferrie et al. (2013).

Ferrie et al. (2013) consider the problem of inferring an unknown frequency using the kind of data that is gathered in swap spectroscopy. The problem considered by Ferrie et al. (2013) is a special case of Problem 3, with the additional restriction that ω_{true} is known and $\nu = \omega_{\text{true}}$ is fixed. The heuristic recommended by Ferrie et al. (2013) is to choose $t = 1/\sigma_g$, which approximately* minimizes the expected variance of the posterior distribution if the prior distribution is normal.

The intuition underlying the heuristic of Ferrie et al. (2013) is illustrated in Figure 6.1, in which I have plotted several sinusoids of approximately equal frequencies. Small differences in sinusoid frequencies ensure that measurements at small abscissa values do not effectively distinguish between the sinusoids. However, measurements at large abscissa values are not effective at reducing the posterior variance: such measurements cause the posterior distribution to become multimodal, which is a problem because the resulting posterior variance is not significantly smaller than the prior variance.

Our heuristics for choosing t are thus a variant of the Ferrie et al. (2013) strategy. There are two heuristics, rather than one, because we need to ensure the normality of the prior distribution over g. Our numerical tests, as well as those of Ferrie et al. (2013), indicate that as few as fifteen measurements, chosen uniformly at random, suffice to ensure normality of the distribution. Thus we employ a different heuristic for the first fifteen measurements that chooses t more randomly than does our main heuristic.

---

*Ferrie et al. (2013) show that $t = 1/\sigma_g$ minimizes a loss *envelope*, not the loss. The loss function itself oscillates at a frequency roughly equal to t. The value of t that minimizes the loss function (rather than the envelope) is *approximately*, but not exactly, $1/\sigma_g$.

When attempting to ensure normality, we choose t uniformly at random on the interval $\left(0, \frac{\pi}{2g_{est}}\right)$. Otherwise, we choose t to be

$$\left|\frac{\pi + z}{2g_{est}}\right|. \tag{6.1}$$

where z is a normally distributed random variable. The absolute value is taken in the unlikely event that the resulting choice of t is negative.

Our heuristic for choosing ν is simpler. When attempting to ensure normality, we choose ν uniformly at random from the interval $(\omega_{est} - g_{est}/2, \omega_{est} + g_{est}/2)$. Otherwise, we choose ν uniformly at random from the interval $(\omega_{est} - 3\sigma_g, \omega_{est} + 3\sigma_g)$. The intuition here is clear: early choices of ν are made uniformly from an interval suggested by the current estimate, and late choices of ν are within a three-sigma distance from the estimated oscillator frequency.

This heuristic is the main component of Algorithm 6.1, which is described more fully in the following section. The performance of the algorithm is comparable to the performance of the adaptive estimation results of Ferrie et al. (2013), of which this work should be seen as a generalization.

The Estimation Algorithm §6.2

The purpose of this section is to explain Algorithm 6.1, which delivers highly accurate estimates of swap spectrum oscillator parameters using a near-minimum of experimental data. The algorithm works by maintaining a state of knowledge that is continually updated as new experimental information becomes available. The current state of knowledge is always used to determine the swap spectroscopy parameters (ν, t), so the algorithm requires 'interactive' access to the experiment. Although such interactive access gives the algorithm its power, modifying existing experimental setups to grant such interative access remains an outstanding challenge for SIC-based quantum computer prototypes.[†]

The main computational cost of Algorithm 6.1 comes from maintaining and updating the state of knowledge about probable oscillator parameters, which is done with a particle representation. The computational cost increases linearly with the number of particles, but more particles mean better approximations to the posterior distribution.

[†]This information may soon be out of date given the rapidity of technological advances.

Algorithm 6.1 The online inference procedure.

Input: Particle representation, $\{x_k\}_{k=1}^n$. $\triangleright\ x_k = (\omega_k, g_k)$
Input: Number of measurement shots, M.
Output: Estimate, $(\omega_{\text{est}}, g_{\text{est}})$.
$a \leftarrow 1.57, b \leftarrow .518, c \leftarrow 3.00, M_0 \leftarrow 15$
for $\ell \leftarrow 1, \ldots, M$ **do**
    Draw r_1 uniformly at random from the interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$.
    Draw r_2 uniformly at random from the interval $(0, 1)$.
    Draw z from the standard normal distribution.
    $(\mu_\omega, \mu_g) \leftarrow \text{MEAN}(\{x_k\})$ $\triangleright$ Algorithm 2.2.
    $\Sigma \leftarrow \text{COV}(\{x_k\})$ $\triangleright$ Algorithm 2.3.
    $(\sigma_\omega^2, \sigma_g^2) \leftarrow \text{DIAG}(\Sigma)$ $\triangleright$ The diagonal elements of Σ.
    if $\ell \leq M_0$ **then** $\triangleright$ Attempt to ensure normality.
        $\nu \leftarrow \mu_\omega + r_1 \mu_g$
        $t \leftarrow a r_2 / \sigma_g$
    else $\triangleright$ Attempt to improve the estimates rapidly.
        $\nu \leftarrow \mu_\omega + c r_1 \sigma_g$
        $t \leftarrow |a + bz| / \sigma_g$
    end if
    $b \leftarrow \text{SWAPSPECTROSCOPY}(\nu, t)$ $\triangleright$ Algorithm 2.1
    if $b = 1$ **then**
        $\mathcal{L} \leftarrow p_{\text{click}}(\nu, t)$ $\triangleright$ The click probability, hence $\mathcal{L}$, is a function of (ω, g).
    else
        $\mathcal{L} \leftarrow 1 - p_{\text{click}}(\nu, t)$
    end if
    $\{w_k\} \leftarrow \text{WEIGHTS}(\mathcal{L}, \{x_k\})$ $\triangleright$ Algorithm 2.4.
    $\{x_k\} \leftarrow \text{RESAMPLE}(\{x_k\}, \{w_k\})$ $\triangleright$ Algorithm 2.5.
end for
$(\omega_{\text{est}}, g_{\text{est}}) \leftarrow \text{MEAN}(\{x_k\})$ $\triangleright$ Algorithm 2.2.
return $(\omega_{\text{est}}, g_{\text{est}})$

My numerical investigations have indicated that 5000 particles suffice for this algorithm, though most numerical results in this chapter are performed with 10000 particles and the simulations in Stenberg et al. (2014) were performed with 50000 particles and a slightly different resampler.[‡]

The maintenance of the particle representation is needed to execute the heuristic described in Section 6.1. In Algorithm 6.1, we introduce several tunable constants (a, b, c, and M_0) that can be optimized to possibly yield better performance; our values are choices that seem to work but are not produced via a comprehensive optimization strategy. In particular, the constants a and c are chosen to reflect our intuition about the problem: $a \approx \pi/2$ by analogy with the discussion of the octave sampling heuristic described in Section 5.1, and $c = 3.00$ reflects the intuition that the choice of ν should not be too far from ω_{est} but significant randomness should nonetheless be included to ensure normality of the posterior pdf. The cross-over point M_0 is also chosen based on the report of Ferrie et al. (2013) that fifteen random samples suffice to ensure approximate normality of the pdf. The constant $b = 0.518$ is, by contrast, somewhat optimized.

The power of Algorithm 6.1 can be demonstrated through simulation, presented in the following section. That performance analysis goes beyond that which was presented in Stenberg et al. (2014) because a careful statement of the TLDS estimation problem eliminates an issue we called the 'outlier' problem. I show that this problem occurs predictably if the true coupling strength is much smaller than the estimated coupling strength—even if such weakly coupled oscillators are consistent with the prior distribution.

Performance Analysis §6.3

This section presents numerical simulation results for Algorithm 6.1. These results are summarized in four figures:

1. Figure 6.2 shows a typical run of Algorithm 6.1 under ideal conditions (i.e. no decoherence or measurement error),

2. Figure 6.3 shows performance statistics for Algorithm 6.1, again under ideal conditions,

3. Figure 6.4 shows performance statistics for Algorithm 6.1 under two simple models for experimental imperfections (decoherence and measurement errors, respectively), and

---

[‡]Resampling for the coupling strength axis is performed *log-normally* instead of normally in our paper.

4. Figure 6.5 resolves an issue my co-authors and I presented in Stenberg et al. (2014) that we called the problem of 'outliers'.

Figure 6.2 shows that the quality of the estimate produced by Algorithm 6.1 improves exponentially quickly in the number of measurements taken, albeit at a cost of exponentially increasing values of the wait time t. This is unsurprising. As the most informative choice of t is inversely proportional to the uncertainty σ_g of the coupling strength g, an exponentially shrinking σ_g ensures that the most informative choice fo t grows exponentially. Algorithm 6.1 can still be considered data-efficient if one counts only the number of bits of experimental data and not the time-cost of each bit of data. In other words, Figure 6.2 indicates data-efficiency if each choice of (v, t) is considered to have equal cost. This is clearly unrealistic for large values of t. However, the twin practical considerations of decoherence (Figure 6.4) and latency (Section 6.4) imply that the exponentially increasing values of t do not *ipso facto* affect the practical applicability of Algorithm 6.1.

Figure 6.3 shows that the performance exhibited by Figure 6.2 really is typical. To generate the data, Algorithm 6.1 is executed 5000 times on random instances of Problem 3, again under ideal conditions. The exponential improvement of estimates as a function of the number of bits of collected experimental data is apparent, at least for one hundred bits of experimental data. The algorithm shows every indication of working even beyond this artificial limit to the number of bits of experimental data; the cutoff simply ensures that floating point errors do not affect the data. The loss of the estimates after 100 experiments, displayed as a histogram, appears at first to be log-normally distributed but actually is not—the data resoundingly fails several (log-)normality tests, including Anderson-Darling, Kolmogorov-Smirnov, and Pearson χ^2.

Visual inspection of the histogram reveals that the distribution is taller, thinner, and has fatter tails than a normal distribution with the same mean and variance as the (logarithm of the) loss data. This observation suggests consideration of the so-called stable distributions, famously used by Mandelbrot (1963) to model price movements in commodity markets. Such a distribution is specified by four parameters; namely, an index of stability $0 < \alpha \leq 2$, a skewness parameter $-1 \leq \beta \leq 1$ (distinct from the third moment of the distribution, which does not exist), a scale parameter $c > 0$, and a location parameter $\mu \in \mathbb{R}$. If $\alpha = 2$, the distribution is simply the normal distribution with mean μ and variance $2c^2$ regardless of β. If $\alpha = 1$ and $\beta = 0$, the distribution is the Cauchy distribution. I used Mathematica to find the stable distribution of best fit to the logarithmic (base ten) loss data of Figure 6.3 and found the parameters

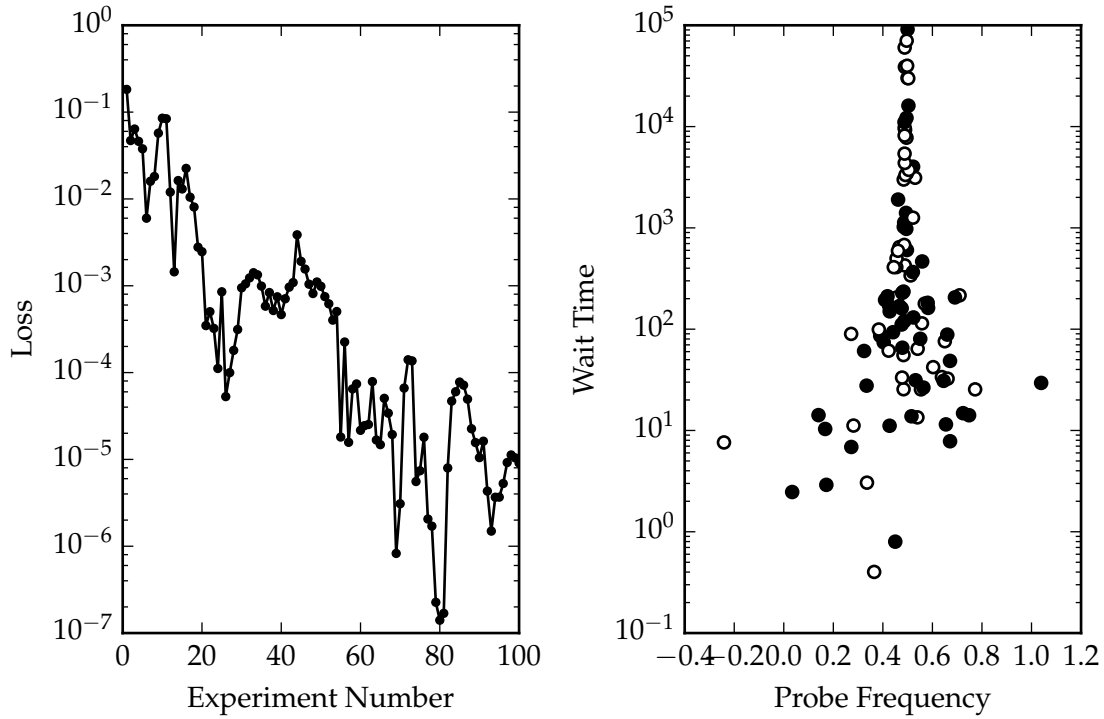$$\alpha = 1.83644, \ \beta = -0.209163, \ \mu = -4.75072, \ c = 0.866038 \tag{6.2}$$

Figure 6.2: Typical performance of Algorithm 6.1. The figure on the left gives the loss of the estimate as a function of the number of experiments; the square root of the ordinate value represents the Euclidean distance between the estimate $(\omega_{\text{est}}, g_{\text{est}})$ and the true value $(\omega_{\text{true}}, g_{\text{true}}) = (\omega_0 + g_0/2, 3g_0/4)$; here $\omega_0 := 0$ and $g_0 := 1$, and the particle representation contains 10^5 particles. The figure on the right shows the recorded data. The abscissa and ordinate values of the points represent the probe frequency ν and wait time t, respectively, and the colour of the point represents the outcome: white for clicks, black otherwise.

pass the Anderson-Darling, Kolmogorov-Smirnov, and Pearson χ^2 tests.

Figure 6.4 shows performance statistics for Algorithm 6.1 in the presence of two simple models for sources of experimental error. The figure on the left depicts the effect on performance of decoherence on the probe system (described briefly in Section 4.1.2) and the figure on the right depicts the effect of symmetric bit-flip measurement errors on the output of swap spectroscopy. Measurement errors do not seem to affect the overall exponential improvement of the estimates and only lowers the base of the exponential. Decoherence imposes a cutoff to the quality of estimates that can be obtained, because large values of t become uninformative: decoherence reduces the contrast between possible outcomes of experiments.

These results constitute strong evidence that our algorithm performs well in terms of the data needed to solve Problem 3 and that this performance is likely replicable in experiment. The algorithm delivers a highly accurate estimate of the oscillator parameters with high confidence with what seems to be optimal data-cost scaling: the estimates improve exponentially quickly if no decoherence is present, which reproduces the scaling observed by Ferrie et al. (2013) for their simpler problem. Finite decoherence clearly suppresses the data-cost scaling below exponential, but this is expected behaviour because decoherence eventually limits the contrast between local maxima and minima of the click probability.

But there is one possible area of concern: outliers. The issue is as follows. We showed in Stenberg et al. (2014) that Algorithm 6.1 has a greater-than-expected tendency to get 'stuck' on bad estimates sometimes. We gave a simple method for resolving this issue by 'reheating' the particle filter, which is to say that the uncertainty of the estimate (i.e. the covariance of the distribution) is reset to its initial value, although the mean is not. Over many repetitions, this reheating process appears to ensure that the bad estimates become 'unstuck' eventually. But these outliers occurred when we simulated our algorithm under a different prior than the one I have specified in Problem 3: in Stenberg et al. (2014), we used a normal distribution as a prior instead of the uniform distribution used here.

By changing this prior, the outlier problem is no longer apparent: the histogram of Figure 6.3 is well-shaped. But the outlier problem is made apparent in Figure 6.5. Here I perform a simulation analogous to that used to produce Figure 6.3, but with a prior that includes weak coupling (i.e. $g \in (0, g_0)$ rather than $(g_0/2, g_0)$) and with specified true oscillator parameters. The histogram shows a multimodality that indicates a persistently high probability that an estimate refuses to converge quickly to its true value. Even after running the algorithm for some time, there appears to be a constant lower limit to the eventual loss of the estimate. I think these numerical results demonstrate that our
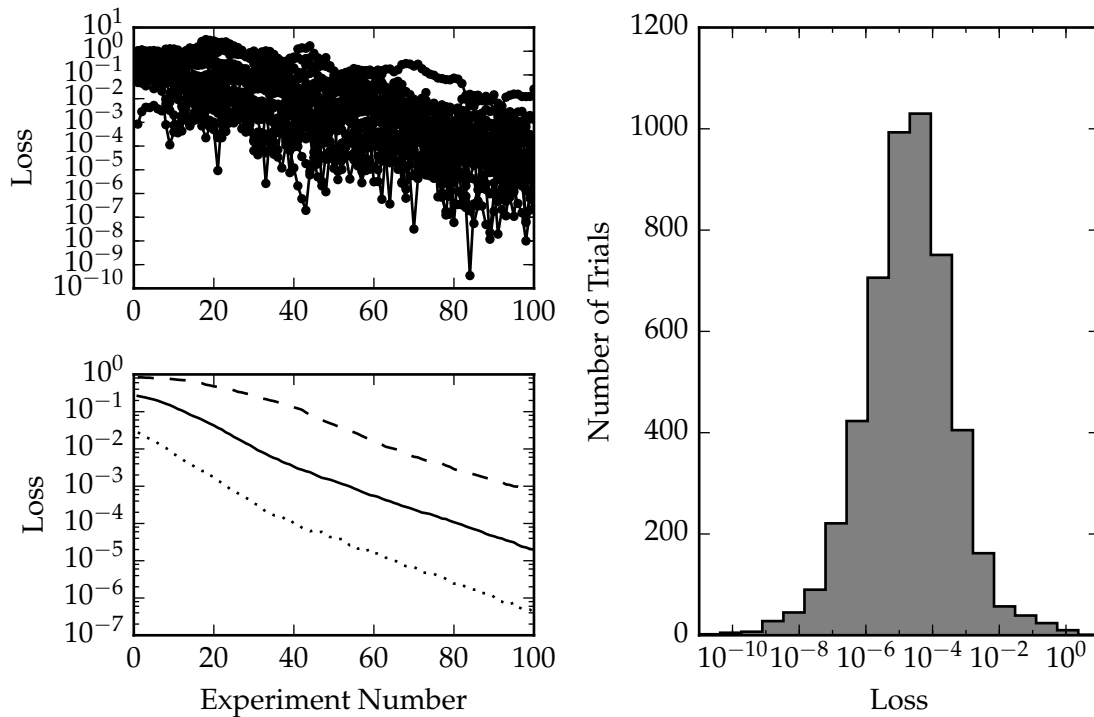
Figure 6.3: The results of numerical testing of Algorithm 6.1. Here I have set $\omega_0 := 0$ and $g_0 := 1$, and I use particle representations with 10^4 particles. The algorithm is run 5000 times; a unique ω_{true} and g_{true} randomly chosen according to the prior distribution (i.e. the uniform distribution on the rectangle $(\omega_0 - g_0, \omega_0 + g_0) \times (g_0/2, g_0)$) for each run. The figure on the top left shows the loss of 25 of these trials as a function of the number of experiments performed, and the figure on the bottom left summarizes all 5000 trials in terms of deciles. Each point of the solid black curve represents the median value (i.e. the fifth decile) across the 5000 trials, the dashed curve is the ninth decile, and the dotted is the first decile. That is, 90% of the 5000 trials fall below the dashed curve, 50% below the solid curve, and 10% below the dotted curve. The loss after 100 experiments is displayed in the histogram on the right.
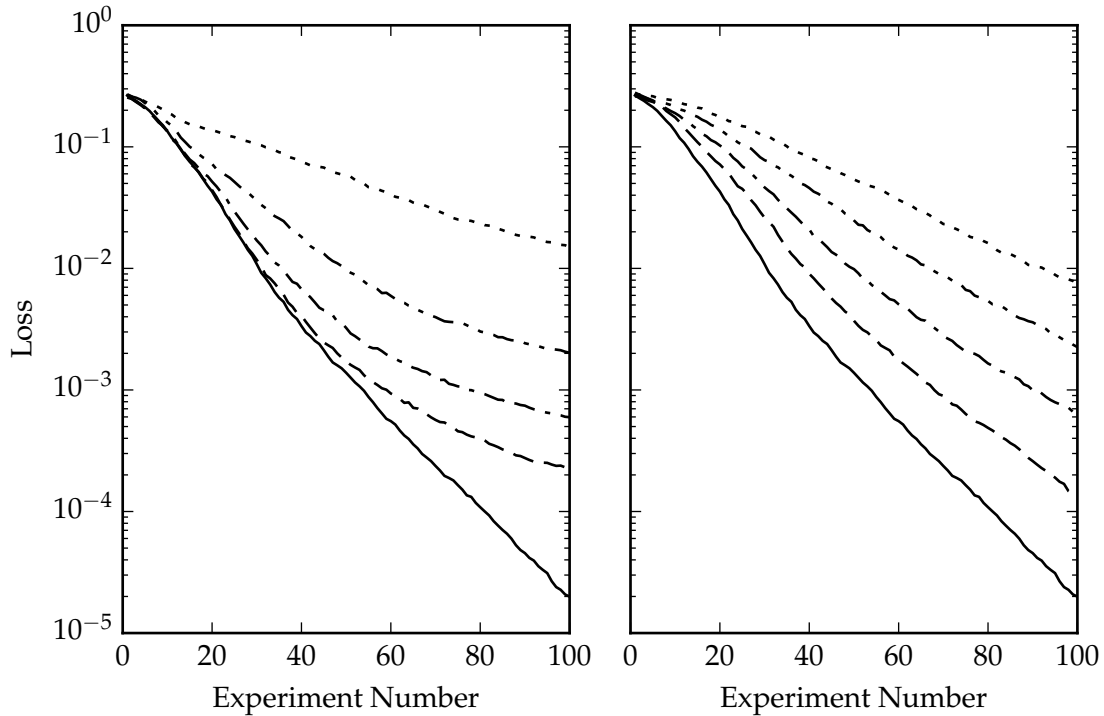
Figure 6.4: The results of numerical testing of Algorithm 6.1 in the presence of non-zero decoherence (left) and symmetric bit-flip errors (right). The solid line in both plots is the median curve from Figure 6.3, and the dashed curves give the median curve of 5000 trials in the presence of decoherence (left) and bit-flip errors (right). On the left, I have simulated decoherence rates of $g_0/1600$ (dashed), $g_0/400$ (dash-dotted), $g_0/100$ (dash-double-dotted), and $g_0/25$ (dotted). On the right, I have simulated symmetric bit-flip error probabilities of 0.05 (dashed), 0.1 (dash-dotted), 0.15 (dash-double-dotted), and 0.2 (dotted). Each particle representation contains 10^4 particles.

algorithm performs well so long as the ratio of the true coupling strength to its initial estimate is not too small. The detection algorithm of Chapter 5 is able to handle this failure mode of the algorithm from Stenberg et al. (2014).

I have thus given an analysis of the performance of Algorithm 6.1 that demonstrates its promise for producing highly accurate estimates of the frequency and coupling strength of a TLDS in experiment. Though we discussed the issue of outliers in the published paper, here I have shown that these outliers are present due to an imperfect statement of the problem of TLDS parameter estimation; the issue is corrected if the prior distribution of oscillator parameters has support inside a given coupling octave. In the next subsection, I discuss the challenges for implementing this algorithm in experiment.

Outlook §**6.4**

Algorithm 6.1 chooses parameters for swap spectroscopy, a standard experimental technique. However, the algorithm requires the data from each requested execution of swap spectroscopy before it specifies the next experimental parameters. Algorithm 6.1 is assuming that two crucial costs are equal to zero: the latency, which is the time taken to execute SWAPSPECTROSCOPY, and the switching time, which is the time taken to change swap spectroscopy input parameters.

Private conversations with researchers in the research group of John Martinis some years ago indicated that the latency is between one hundred microseconds, which is the time needed to execute phase-qubit-based dispersive readout, and thirty milliseconds, which is the time needed to transfer information from the FPGA controller and the laboratory computer through a LAN cable. These numbers are not official and are subject to change as technology improves; the numbers should only be seen as indications of reasonable values. But the fact remains that the time taken to select experiment parameters (less than a microsecond with well-designed code and not too many particles) could be much smaller than the time taken to obtain the experimental data.

That changing swap spectroscopy input parameters carries a time-cost is apparent if one recognizes that the parameters (v, t) specify a control pulse that needs to be computed and communicated to an arbitrary waveform generator. This cost can be ameliorated by performing the same experiment many times. Private conversations with the Martinis group indicate that about thirty identical experiments can be performed in the time it takes to switch from input parameters (v, t) to (v', t'), though this figure is again some years old and should not be taken as official.
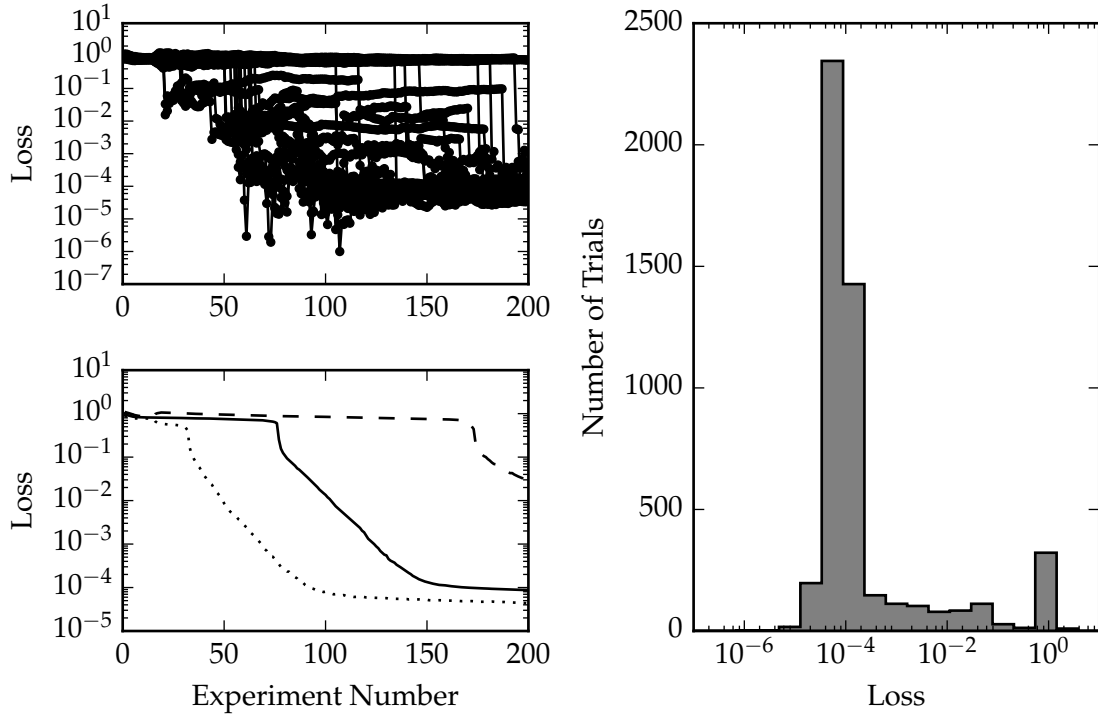
Figure 6.5: The results of numerical testing of Algorithm 6.1 when the oscillator is weakly coupled. Here the prior distribution is the uniform distribution over the rectangle $(\omega_0 - g_0, \omega_0 + g_0) \times (0, g_0)$ and the true parameters are $\omega_{\text{true}} := \omega_0 + 9g_0/10$ and $g_{\text{true}} := g_0/100$; $\omega_0 := 0$ and $g_0 := 1$. Again, I use particle representations containing 10^4 particles. The algorithm is run 5000 times and the data is plotted analogously to Figure 6.3. The figure on the top left shows the loss of 25 trials as a function of the number of experiments performed, and the figure on the bottom left summarizes all 5000 trials in terms of deciles. Each point of the solid black curve represents the median value (i.e. the fifth decile) across the 5000 trials, the dashed curve is the ninth decile, and the dotted is the first decile. That is, 90% of the 5000 trials fall below the dashed curve, 50% below the solid curve, and 10% below the dotted curve. The loss after 100 experiments is displayed in the histogram on the right.

Other impracticalities of Algorithm 6.1 include the need for an informative prior and the poor performance for weakly coupled oscillators. Though Problem 3 requires that the prior is a particular coupling octave, the discussion of Section 5.4 shows that this requirement can be weakened. Furthermore, the problem of weakly coupled oscillators can be eliminated through the use of the techniques introduced in Chapter 5.

Despite these impracticalities, Algorithm 6.1 is clearly robust against two important experimental imperfections: decoherence and measurement error. Although the analysis of these effects has been based on simulation with relatively simple models, there is every indication that Algorithm 6.1 will perform well if implemented in a real experiment. There is a clear argument for attempting such experimental implementation.

MITIGATING DEFECTS

It is better to travel well than to arrive.

SIDDHARTHA GAUTAMA

UNWANTED TLDSes can have a particularly deleterious effect on the quality of control over an SIC-based quantum computer prototype. The purpose of this chapter is to describe a possible mitigation strategy for a particular effect of TLDSes on SIC-based quantum information processors. The goal of the mitigation strategy is to set the qubit frequency ν from an initial frequency ν_0 to a final frequency ν_1 whilst preserving any encoded information.

This is a straightforward task if no TLDS is present. But if a TLDS is present with frequency ω between ν_0 and ν_1, the intermediate value theorem implies that any continuous alteration of ν from ν_0 to ν_1 requires that $\nu = \omega$ at some point. If the qubit is momentarily resonant with the TLDS, the resulting coherent energy exchange could be highly deleterious to the quantum information stored in the qubit. The TLDS mitigation problem is therefore to suppress coherent energy exchange.

Together with Tobias Chasseur, Lukas Theis, Daniel Egger, and Frank Wilhelm, I showed in Chasseur et al. (2015) (reprinted as Appendix A) that coherent energy exchange between a qubit and a TLDS can be suppressed if

$$\nu_0 \ll \omega - g < \omega + g \ll \nu_1, \tag{7.1}$$

where the TLDS parameters are (ω, g). Then the mitigation problem can be solved using a suitably chosen time-dependent qubit frequency $\nu = f(t)$. Crucially, this solution requires extremely accurate knowledge of the TLDS parameters and extremely precise control over the qubit frequency.

This chapter is organized as follows. Section 7.1 describes the relationship of the TLDS mitigation problem with the discussion of Chapter 4. Section 7.2 explains my contributions. Finally, Section 7.3 discusses the possibility of implementing our solution in experiment.

Problem Statement §7.1

This section describes the 'TLDS mitigation' problem, which is the problem of modifying a qubit frequency without altering its state if doing so would bring the qubit momentarily into resonance with a TLDS. To state this problem, it is necessary to consider the exchange Hamiltonian, described in Section 4.1.1, with a qubit frequency ν that varies as a function of time. The function is selected in order to accomplish some task, which in this section is simply to begin at $\nu = \nu_0$, end at $\nu = \nu_1$, and to ensure the initial state of the qubit is approximately equal to the final state of the qubit. In other words, the TLDS mitigation problem is to execute an identity gate with low error rate whilst tuning a qubit across resonance with a TLDS.

The Landau-Zener Hamiltonian described in Section II of Chasseur et al. (2015) is simply the single-oscillator exchange Hamiltonian with tunable $\nu - \omega$. The addition of the control term describes the fact that the qubit frequency need not remain constant between preparation and measurement; in fact, two-qubit gates are implemented in part by tuning distinct qubits into and out of resonance with one another. Therefore, the success of such methods depend on the ability to change the qubit frequency from one value to another whilst preserving any encoded quantum information. Unwanted TLDSes can interfere with this goal. The TLDS mitigation problem is therefore to execute such transformations with low error rates.

Problem 4 (TLDS Mitigation). Consider a swap spectrum containing exactly one oscillator: $\sigma = [(\omega, g)]$. Define ν_0 and ν_1 such that

$$\nu_0 \ll \omega - g < \omega + g \ll \nu_1. \tag{7.2}$$

Choose $f(t)$ such that $f(0) = \nu_0$, $f(1) = \nu_1$, and $\exp(-iH_{\text{exchange}}(t)) \approx \mathbb{1}$.

The solution presented in Chasseur et al. (2015) is found in Section II.B.2. There a linear sweep is set as the input to a pulse optimization algorithm called GRAPE, introduced by Khaneja et al. (2005). GRAPE seeks a choice of control functions that maximize the average fidelity of the resulting time-evolution to some desired unitary operator. In Chasseur et al. (2015), GRAPE proposes a sinusoidal modification to the linear sweep, presented as Fig. 7, that has a time-varying frequency. The modification appears to have the form

$$-\frac{\sin\left(t\sqrt{1+t^2}\right)}{\sqrt{1+t^2}}, \tag{7.3}$$

though this observation has yet to be tested.

The main issue with this solution is that it is not likely to be robust to experimental imperfections in pulse shaping. In particular, investigations of the effect of a constant-frequency, constant-amplitude sinusoidal augmentation to the linear sweep imply that the quality of these pulses for solving the TLDS mitigation problem are very sensitive to the linear sweep velocity and phase offset at resonance. My contribution to this work was to investigate the sensitivity of constant-frequency, constant-amplitude sinusoidally augmented linear sweeps to the sweep velocity. This contribution is described in the following section.

Statement of Contributions §7.2

My contributions are summarized in Figure 7.1. This figure demonstrates the possibility of time-evolutions like that depicted in Figure 5 of Chasseur et al. (2015), reprinted in Appendix A for convenience. Figure 7.1 depicts the tunnelling probability as a function of the parameters v and λ, which is to say it plots the probability of observing the probe in its ground state given that it was initially set to its excited state, the TLDS to its ground state, and the evolution was controlled as per Eq. (2) of Chasseur et al. (2015).

A crucial observation to make about Figure 7.1 is that it indicates the sensitivity of the tunnelling probability to the parameter values for $\epsilon(t)$ from Eq. (2) of Chasseur et al. (2015). The ratio between v and λ needs to be rather finely tuned; moreover, Figure 6 in Chasseur et al. (2015) shows that performance is also highly sensitive to the ratio between v and Ω, and Figure 4 of Chasseur et al. (2015) shows sensitive to the phase φ.

Based on these numerical results, I argued that suppression of tunnelling can be achieved with controls selected by applying a gradient-free optimization algorithm (such as Nelder-Mead) to the tunnelling probability four-surface. The robustness of
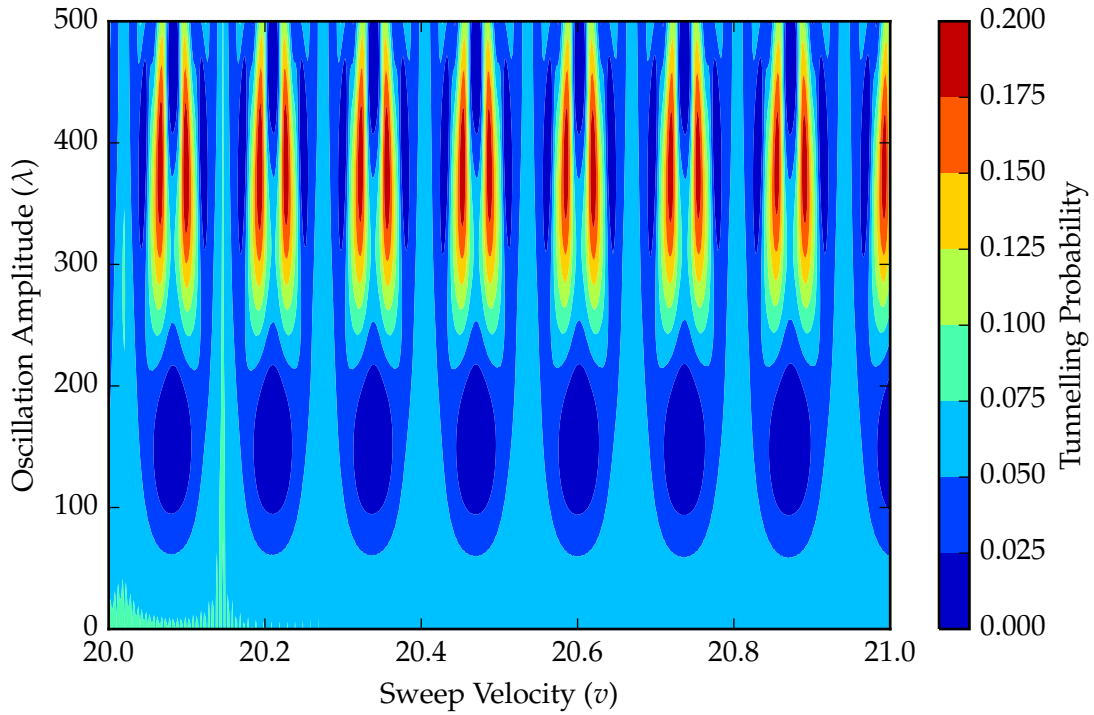
Figure 7.1: An illustration of the final tunnelling probability of the sinusoidally-augmented linear sweep specified in Eq. (2) of Chasseur et al. (2015). Here I have set $\Delta := 1$, $\Omega := 100$, and $\varphi := 0$ and used Mathematica's numerical differential equation solver to generate the time-evolution of the tunnelling probability from $t = -25$ to $t = +25$ for various choices of linear sweep speed v and oscillation amplitude λ. The plotted values represent the probability of observing the probe system in its ground state at $t = +25$. The oscillation frequency of the tunnelling probability for λ fixed at approximately 150 decreases as v is increased.

these control pulses to parameter uncertainty is a key consideration for experimental implementation and therefore constitutes an argument that control methods that are robust to model uncertainty will generally require that some effort be made to efficiently characterize models and re-tune control pulses to account for changes as they occur.

Outlook §7.3

The work presented in Chasseur et al. (2015) demonstrates that the TLDS mitigation problem can be solved with suitably chosen controls. Sinusoidally-augmented linear sweeps of a probe across resonance with an oscillator are shown to suppress coherent exchange between the probe and oscillator in the swap spectrum model. However, the controls for these pulses need to be very finely tuned in order to suppress that coherent exchange.

This is a more general problem in quantum control. Control pulses are only valuable insofar as they are robust against experimental imperfections and uncertainties. For example, the control pulse needs to respect considerations such as the finite rise time of electronics and random noise introduced to signals. Pulses such as the ones presented in Chasseur et al. (2015) are finely tuned, so the experimental implementation of such pulses is a difficult challenge.

Within the context of this thesis, the most important consideration for the design of a TLDS mitigation strategy such as that of Chasseur et al. (2015) is that the model parameters must be specified to extremely high accuracy and precision. Any uncertainties ensure that the finely-tuned control pulses we propose could misbehave quite badly. The main impact of Chasseur et al. (2015) in this thesis is therefore that solutions to the TLDS characterization problem are needed to correct for the presence of a TLDS.

EIGHT

CONCLUSIONS

Nature is often hidden, sometimes overcome, seldom extinguished.

THE quantum computing enterprise has a long way to go before useful and error-free quantum computations can be carried out in practice. A major outstanding challenge is to characterize the logical errors that affect quantum computer prototypes in a practical yet rigorous way. The project of this thesis is to make progress towards characterizing errors in SIC-based quantum information processors. The results of this thesis are presented primarily in four chapters.

1. Chapter 3 demonstrates the relationship between average gate fidelity, a common performance metric, and the requirements of quantum fault-tolerance threshold theorem. This relationship is demonstrated by showing that

   (a) the upper bound of Wallman and Flammia (2014) is an asymptotically tight approximation to the least upper bound to average fidelity in terms of quantum gate error rate, and

   (b) tighter bounds can be placed on the error rate in terms of average fidelity if knowledge is available about the 'Pauli distance' of the noise process.

2. Chapter 5 explains a data-efficient algorithm for detecting the influence of TLDSes on a SIC-based qubit. This algorithm is presented as the composition of two other algorithms:

    (a) an algorithm for collecting data according to what was called the 'coupling octave sampling heuristic' and

    (b) an algorithm for deciding, based on that data, whether to accept or reject the hypothesis that no TLDSes are visible.

3. Chapter 6 explains a data-efficient algorithm for updating crude estimates of TLDS parameters to highly accurate ones. Unlike the algorithms in Chapter 5, this algorithm is 'online' in the sense that it requires interactive access with the experiment. The algorithm decides, based on a heuristic, what experiments to perform based on the currently available knowledge in the form of a particle representation of the posterior pdf over possible TLDS parameter values.

4. Chapter 7 suggests a method for mitigating one deleterious influence of TLDSes: the possibility of coherent exchange of information between a qubit and a TLDS as the qubit is tuned across resonance with the TLDS. The method is to augment a linear sweep across resonance with a precisely tuned sinusoid.

In presenting these results, I have aimed to problematize clearly enough that the effectiveness of the solutions can be compared to the problem they were designed to solve. Thus the bounds on quantum gate error rate in terms of average gate fidelity that are presented in Chapter 3 are presented as a solution to the problem of comparing gate performance as measured by randomized benchmarking with stated fault-tolerance targets. The algorithms of Chapters 5 and 6 together enable data-efficient solutions to the TLDS characterization problem, under special circumstances at least. The method of tuning qubits across resonance with a TLDS presented in Chapter 7 is aimed at protecting quantum information encoded in the qubit.

The results of this thesis should be seen within the context of the emerging field of characterization, verification, and validation of quantum information processing devices. This field is driven by the twin challenges of the curse of dimensionality that affects any 'brute force' characterization procedure together with the informationally destructive nature of measurement in quantum physics. Until recently, relatively simple and general tools such as quantum process tomography and randomized benchmarking have sufficed to assess progress in quantum information processing prototypes. But devices that are produced today are both well-controlled and highly complex. In my view, further

progress must be assessed based on device-specific methods of characterizing errors in quantum information processors.

In this context, the impact of the results in this thesis are clear. The asymptotic optimality of the upper bound of Wallman and Flammia (2014) demonstrates that average fidelity, the quantity estimated by randomized benchmarking, is not sufficiently informative to justify strong claims regarding the scalability and fault-tolerance properties of experimental devices for which the precise nature of noise remains somewhat murky. The algorithms presented in Chapters 5 and 6 enable an analysis of the influence of unwanted environmental resonances in SIC-based quantum information processors and in any other device for which the swap spectrum model is accurate. Chapter 7 gives a possible approach to mitigating the deleterious influence of such unwanted resonances.

Techniques like these eschew the generality of randomized benchmarking or quantum process tomography and instead achieve more limited objectives. These techniques should form part of an overall strategy to assess just one important source of noise in SIC-based quantum information processors, the influence of TLDSes on the dynamics of an SIC-based qubit. To develop more techniques like these for other noise sources requires not only a deep analysis of the physics of the devices but also advances in the methods of statistical experimental design.

APPENDICES

A

REPRINT OF CHASSEUR ET AL. (2015)

T HE following seven pages are a reprint of Chasseur et al. (2015) and are included to supplement Chapter 7. I contributed to this work at an early stage; the bulk of the work was carried out by the first two authors.

The relevance of this work to my thesis is that the avoided crossing model is identical to the one-oscillator swap spectrum model. One of the aims of the paper is to ensure the diabatic transition of an excited probe between two bias points if the probe comes into resonance with an oscillator during the transition. Diabaticity can be ensured by carefully shaping the time-evolution of the probe frequency with a sinusoidally-augmented linear sweep of the form depicted in Eq. (2).

My work is summarized in Section II.B.1. Frank Wilhelm suggested the augmented linear sweep and argued that it should be possible to suppress population transference between the probe and the oscillation based on a Dyson series expansion. I numerically solved the Schrödinger equation and demonstrated that the first order Dyson series fails to predict the 'tunnelling' probability. I investigated the tunnelling probability as a function of v and λ (keeping Ω constant and $\varphi = 0$) and found that certain values of λ and v would indeed ensure that the tunnelling probability is extremely low. All numerical evidence indicated that the probability could be brought arbitrarily close to zero. I also noted the sensitivity of this effect to perturbations in the parameters v, λ, Ω, and φ.

Engineering adiabaticity at an avoided crossing with optimal control

T. Chasseur,[1] L. S. Theis,[1] Y. R. Sanders,[2] D. J. Egger,[1] and F. K. Wilhelm[1,2]

[1]*Universität des Saarlandes, Saarbrücken, Germany*
[2]*IQC and Dept. of Physics and Astronomy, University of Waterloo, 200 University Avenue W, Waterloo, Ontario N2L 3G1, Canada*

We investigate ways to optimize adiabaticity and diabaticity in the Landau-Zener model with nonuniform sweeps. We show how diabaticity can be engineered with a pulse consisting of a linear sweep augmented by an oscillating term. We show that the oscillation leads to jumps in populations whose value can be accurately modeled using a model of multiple, photon-assisted Landau-Zener transitions, which generalizes work by Wubs *et al.* [New J. Phys. **7**, 218 (2005)]. We extend the study on diabaticity using methods derived from optimal control. We also show how to preserve adiabaticity with optimal pulses at limited time, finding a nonuniform quantum speed limit.

PACS number(s): 32.80.Xx, 02.30.Yy, 03.67.−a

I. INTRODUCTION

The adiabatic theorem, which should be applied with care [1–3], states that if the time evolution of a quantum system is sufficiently slow, transitions between eigenstates can be neglected. It is thus a statement about an approximation rather than a rigorous theorem in the mathematical sense. Adiabatic quantum computing is one paradigmatic example of the usefulness of such time evolutions [4,5]. Another application concerns the control of quantum processor elements by frequency tuning. On the one hand, gate designs often rely on adiabaticity [6,7]. On the other hand, due to spurious couplings—e.g., higher-order interactions beyond nearest neighbors to other parts of the chip [8] and undesired spurious resonators [9–12]—and adiabatic following (i.e., reach a nonadiabatic sweep within the limited bandwidth of a realistic experiment), one desires a detour to diabaticity. Shortcuts to adiabaticity, i.e., arriving at the same final state as the adiabatic evolution but in a shorter time, have been investigated [13]. The study of adiabaticity in periodically driven systems has led to adiabatic Floquet theory [14–16]. The physics of adiabaticity is well captured in the Landau-Zener (LZ) model which analytically describes the behavior of a system when linearly swept through an avoided level crossing [17,18]. The generic nature of this model gives it a wide range of applications, for instance in transitionless quantum driving of spins [19]. Considering more complex, i.e., nonlinear, pulses has useful applications, such as Landau-Zener-Stückelberg interferometry. This allows one to determine the parameters of the avoided crossing by quickly sweeping back and forth through it [20–22]. The LZ model also describes tunneling states in the tunneling model of amorphous solids. Here an atom can move between two adjacent positions separated by a potential barrier. Therefore, nonadiabatic driving of tunneling states affects the dielectric constant in glasses [23,24]. Quantum transients of matter waves in optical lattices can be modeled by multiple LZ transitions [25]. Evidence for interacting defects in glasses has been presented in [26]. Studies of the magnetization of molecular magnets [27] involve a similar situation where a magnetic field is swept over many spins [28]. These dynamics also occur in molecular collisions [29].

Parallel to these developments are those in quantum optimal control where a control pulse is shaped to realize a specific time evolution [30,31]. These methods were originally pioneered for nuclear magnetic resonance and have started to gain popularity in solid-state quantum information devices; for instance, in electron spin qubits to engineer gates [32], as well as in superconducting qubits [33,34] to address leakage [35,36], frequency crowding [37,38] and gate design [39,40]. In such systems the crucial entangling gates are realized with an anticrossing [41]. Within the framework of qubits and optimal control, the LZ model has been used to study quantum speed limits [42–44] and how robust high-fidelity pulses are related to uncertainties in the noncontrollable part of the Hamiltonian [45].

For the detour to diabaticity we investigate the dynamics of a LZ system under a linear sweep augmented by a fast oscillation. A similar system has been studied in Refs. [46,47]. We investigate how such a pulse can be used to engineer diabaticity. Optimal control methods allow us to deepen this study as well as investigate pulses that keep the evolution adiabatic [13] which are crucial to quantum computing. The work is structured in the following way. Section II discusses the LZ system when the linear sweep is augmented by an oscillating term. The analytics are in Sec. II A while the numerics are presented in Sec. II B. The adiabatic pulses are discussed in Sec. III.

II. ENGINEERING DIABATICITY

The LZ Hamiltonian is defined by $\hat{H}_{LZ} = \varepsilon(t)\hat{\sigma}_z/2 + \Delta\hat{\sigma}_x/2$ where $\hat{\sigma}_x$ and $\hat{\sigma}_y$ are Pauli matrices in the basis states $|0\rangle$ and $|1\rangle$, which also are eigenstates in the absence of the coupling Δ. A finite Δ mixes these states into new energy eigenstates referred to as instantaneous eigenstates and introduces an avoided level crossing localized at $\varepsilon(t) = 0$.

Sweeping through that anticrossing with constant velocity v following $\varepsilon(t) = vt$ results in the celebrated transition probability between $|0\rangle$ at $t \to -\infty$ and $|1\rangle$ and $t \to \infty$

$$P_{LZ} = 1 - \exp\left(-\frac{\pi\Delta^2}{2v}\right). \tag{1}$$

Therefore a linear sweep at high speed, $v \gg \Delta^2$, avoids leakage in our basis of $|0\rangle$ and $|1\rangle$ hence keeping the time evolution *diabatic*, whereas an infinitely slow sweep, $v \ll \Delta^2$,

keeps the system in the instantaneous ground (or excited) state at all times while that state is changing along the Hamiltonian.

A. Analytics of an oscillation-augmented linear sweep

Consider a linear sweep augmented by an oscillation of amplitude λ and frequency Ω

$$\varepsilon(t) = vt + \lambda \cos(\Omega t + \varphi). \qquad (2)$$

The phase φ determines the value of the oscillation when the linear sweep is on resonance with the anticrossing and plays an important role as emphasized below. This form of ε can arise when investigating the dielectric constant of glasses by using an external electric field [23,24]. Wubs *et al.* [46] study a very similar model where ε is linear only and Δ varies sinusoidally in time. We will see that similar physics arises in our model and will connect to their work when appropriate. To study the time evolution produced by Eq. (2) we switch to the interaction picture by the transformation $\hat{U}_0 = \exp\{-i\phi(t)\hat{\sigma}_z/2\}$ resulting in

$$\hat{H}_I = \frac{\Delta}{2} \begin{pmatrix} 0 & e^{i\phi(t)} \\ e^{-i\phi(t)} & 0 \end{pmatrix} \quad \text{with} \quad \phi(t) = \int_0^t d\tau\, \varepsilon(\tau).$$

Using the Taylor expansion of $\sin(x)$ and $\exp(x)$ and the binomial expansion of $(e^{i\Omega t} - e^{-i\Omega t})^n$ these terms are cast into the form

$$e^{\pm i\phi(t)} = \sum_{m=-\infty}^{\infty} (\pm)^m J_m\left(\frac{\lambda}{\Omega}\right) e^{\pm i\frac{1}{2}vt^2 + im(\Omega t + \varphi)}. \qquad (3)$$

J_m is the mth Bessel function of first kind. This expansion is similar to that done in [21,48] and is motivated by coherent destruction of tunneling [49–51]. The expansion in Eq. (3) allows one to identify term m with an $|m|$-photon emission and absorption process [52]. Using the property $J_{-m}(x) = (-1)^m J_m(x)$ the Hamiltonian in the interaction picture is

$$\hat{H}_I = \frac{\Delta}{2} \sum_{m=-\infty}^{\infty} J_m\left(\frac{\lambda}{\Omega}\right)$$
$$\times \left[\cos(m\varphi) \begin{pmatrix} 0 & e^{\frac{i}{2}vt^2 + im\Omega t} \\ e^{-\frac{i}{2}vt^2 - im\Omega t} & 0 \end{pmatrix} \right.$$
$$\left. - i\sin(m\varphi) \begin{pmatrix} 0 & -e^{\frac{i}{2}vt^2 + im\Omega t} \\ e^{-\frac{i}{2}vt^2 - im\Omega t} & 0 \end{pmatrix} \right]. \qquad (4)$$

For $\lambda = 0$ the dynamics under a linear sweep are recovered. Under these conditions, the system undergoes a transition when $t = 0$ and the final population is given by Eq. (1). This jump is a result of the function vt^2 having an extrema at $t = 0$ and around this time the Hamiltonian is no longer rapidly oscillating. Thus if a Magnus expansion were done [53], one would see that the higher-order commutators cannot be neglected; these contributions add up, thereby producing a jump. A Dyson series expansion also fails since as $\varepsilon(t)$ can be zero, there is no small parameter to expand in.

Another way to see this is to note that Eq. (4) can be viewed as a sum of Landau-Zener Hamiltonians taken in the interaction picture where the zero bias point is shifted from $\varepsilon = 0$ to $\varepsilon = m\Omega$. Indeed the terms $\pm i(vt^2/2 + m\Omega t)$ can be
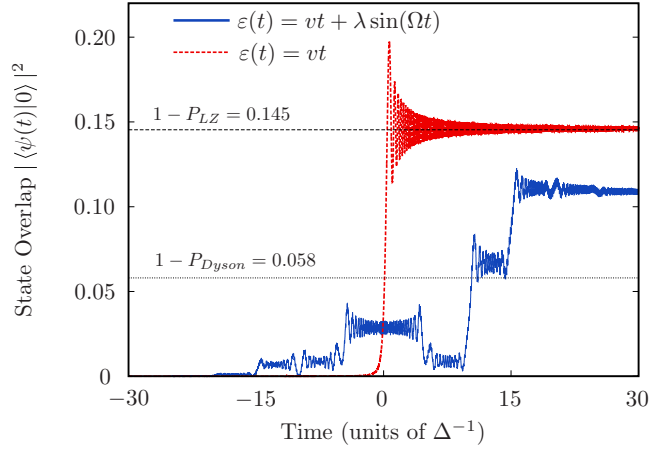

FIG. 1. (Color online) Comparison of a linear sweep with one augmented by an oscillation. The time evolution was computed from $-T/2$ to $T/2$ with $T = 100\Delta^{-1}$ thus cutting off times without significant population changes. The linear speed was $v = 10\Delta^2$, the oscillation frequency and amplitude were $\Omega = 50\Delta$ and $\lambda = 120.24\Delta$ respectively. The population jumps occur at $t = -m\Omega/v$ which here are integer multiples of $5\Delta^{-1}$. The dotted line is the transition probability that would be expected of a first-order Dyson series expansion. As expected, it does not predict the correct probability because the expansion is not valid since ε can be smaller than Δ.

recast into the form

$$\pm i\frac{v}{2}\left(t + \frac{m\Omega}{v}\right)^2 \mp i\frac{m^2\Omega^2}{2v}. \qquad (5)$$

We thus expect many jumps to happen at intervals of $m\Omega/v$ which is illustrated in Fig. 1. This generalizes the results of Ref. [46] where only two separate jumps appear.

1. Multijump model

Building on the observations of the previous section we construct an approximate model that allows us to analytically compute the unitary matrix describing the jumps at $m\Omega/v$. It consists of using the separation into multiple Landau-Zener Hamiltonians and assuming their application is non-overlapping in time, i.e., that the jumps are independent and that the sweep between the jumps only contributes a phase factor, which assumes $\Delta \ll \Omega$. A time evolution that spans up to m_0 photon jumps is thus approximated by

$$\hat{U}_J = \prod_{m=-m_0}^{m_0} \begin{pmatrix} e^{-\frac{\pi\Delta_m^2}{4v}} & e^{i\tilde{\varphi}_m}\sqrt{1 - e^{-\frac{\pi\Delta_m^2}{2v}}} \\ e^{-i\tilde{\varphi}_m}\sqrt{1 - e^{-\frac{\pi\Delta_m^2}{2v}}} & e^{-\frac{\pi\Delta_m^2}{4v}} \end{pmatrix}. \qquad (6)$$

The terms in the matrices are similar to those from the LZ transition probability in Eq. (1) but jump m has $\Delta_m = \Delta J_m(\lambda/\Omega)$ instead of Δ and an associated Stokes phase [21,46] given by

$$\tilde{\varphi}_m \simeq -\frac{\Delta^2}{4v}\ln\left(\frac{T^2 v}{4}\right) - \frac{\pi}{4} - \frac{m^2\Omega^2}{2v} - m\varphi$$
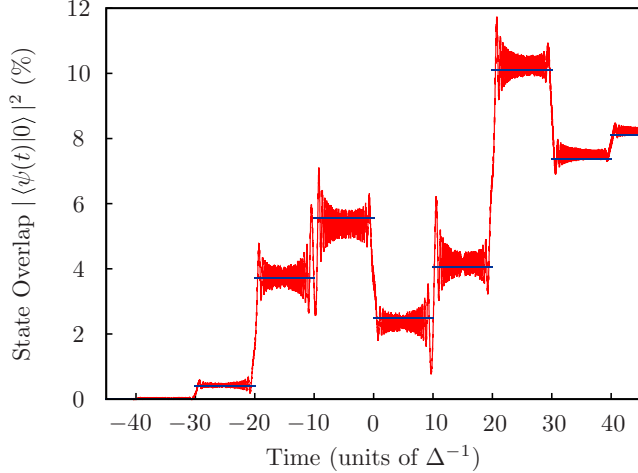
FIG. 2. (Color online) Illustration of the multijump model. The red line shows the full time propagation. The horizontal lines show the state overlap as predicted by the multijump model $\hat{U}_J$. As can be seen this model accurately predicts the photon-assisted LZ transitions.
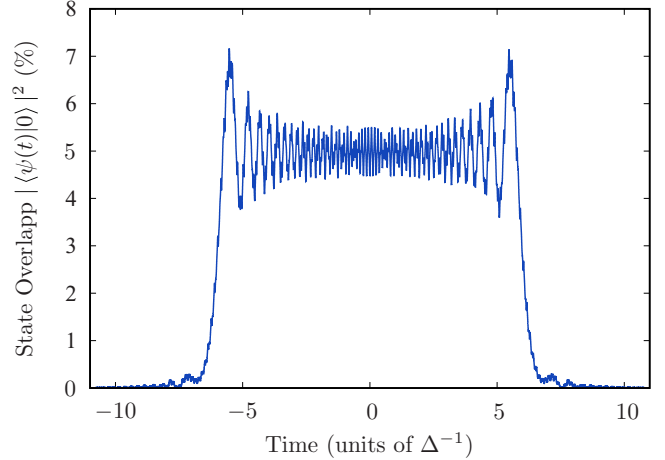

FIG. 3. (Color online) Example of a pulse with $v = 8\Delta^2$, $\Omega = 50\Delta$ and a time dependent $\lambda(t)$ given by Eq. (7). The ramp and switching times are given by $T + T_s = 3\Omega/v + T_s = 21.55\Delta^{-1}$. The ramp height satisfies $\lambda_r = 2.4048\Omega$ so as to cancel the jump at $m = 0$. The final occupation of $|0\rangle$ is smaller than 10^{-5}.

under the assumption $\Delta^2/4v \leqslant 1$. Note that negative Δ_m occur, leading to an additional phase factor of π. This model properly accounts for the jump heights as shown in Fig. 2 but is completely devoid of the oscillations around these values as is to be expected given that each LZ jump matrix in $\hat{U}_J$ is time independent. Our model might under some conditions be extended to a system with multiple LZ crossings which have recently been investigated [54].

2. Engineering diabaticity

Now the challenge of engineering diabaticity is to sweep through the avoided level crossing such that at the start and end of the evolution we stay in the same uncoupled state, i.e., the time evolution operator up to an irrelevant global phase should be $\hat{U}(-T/2,T/2) = \mathbb{1}$. In Ref. [46] it is shown how time-reversal antisymmetry can lead to a suppression of population transfer yielding $\hat{U} = \mathbb{1}$. This can be done if the time evolution for $t > 0$ reverses the time evolution of $t < 0$ which can be related to the Loschmidt echo [55]. In this section we connect this idea to our model and show how one has to carefully set the ratio λ/Ω to a particular value.

This can be achieved with high accuracy if the time during which the oscillation is present is chosen so that only the photon-assisted jumps at $m = \pm 1$ occur. One has to set the phase φ to zero and choose the ratio λ/Ω to the first zero of the zeroth Bessel function J_0, i.e., $\lambda/\Omega \simeq 2.4048$. This suppresses the transition at $t = 0$ and the photon-assisted jumps at $m = \pm 1$ cancel each other. This follows from the Bessel function property $J_{-1}(x) = -J_1(x)$. To include the finite rise time of any electronics, we choose a linear ramp for the oscillation amplitude

$$\lambda(t) = \begin{cases} \lambda_r & \text{if} \quad |t| < \frac{T-T_s}{2}, \\ \frac{\lambda_r}{T_s}\left(\frac{T+T_s}{2} - |t|\right) & \text{if} \quad \frac{T-T_s}{2} \leqslant |t| \leqslant \frac{T+T_s}{2}, \\ 0 & \text{otherwise}. \end{cases} \quad (7)$$

T_s is the switching time. The height of the ramp is chosen to suppress the jump at $m = 0$, thus $J_0(\lambda_r/\Omega) = 0$ is needed implying $\lambda_r \cong 2.4048\,\Omega$. The duration $T - T_s$ over which the amplitude of the oscillation is held constant has to be long enough to include the $m = \pm 1$ transitions, but the ramp duration plus switching time $T + T_s$ has to be chosen so that there are no longer any oscillations in the pulse for $|t| > 2\Omega/v$ so as to prevent $|m| \geqslant 2$ jumps. This imposes $(T - T_s)/2 > \Omega/v$ and $(T + T_s)/2 < 2\Omega/v$. The time evolution of such a pulse is shown in Fig. 3. There is no jump at $t = 0$ and the $m = \pm 1$ jumps cancel each other. As expected, this results in a pulse which leaves hardly any population in $|0\rangle$, namely, less than 10^{-5}. However, this is phase sensitive. Indeed if $\varphi \neq 0$ then the Stokes phases of $m = \pm 1$ are no longer different by π and the reasoning presented above no longer holds. This is shown in Fig. 4 by the solid red line. At its worst, the phase can produce almost 16% error. This can be understood that only at phase 0, the tunneling events assisted by an odd number of electrons vanish automatically due to destructive interference of time-reversed paths.

If even shorter pulses can be made, the phase variable can be rendered irrelevant. Indeed, for pulses shorter than Ω/v, only the population jump at $m = 0$ could contribute since no photon-assisted processes are resonant. However, it is removed by the right choice of λ/Ω. The error of a shorter pulse with $T + T_s = 13.86\Delta^{-1}$ is shown in Fig. 4 by the dashed blue line.

B. Numerical optimal control

The previous section showed that a linear sweep with an oscillation can reduce population jumps when going through the anticrossing compared to a simple linear sweep. Here we show how a numerical optimization of the four parameters of the pulse in Eq. (2) can reduce an unwanted transition when $T > 2\Omega/v$ and without windowing. Next we show that starting from a linear sweep a time-sliced gradient search provides an
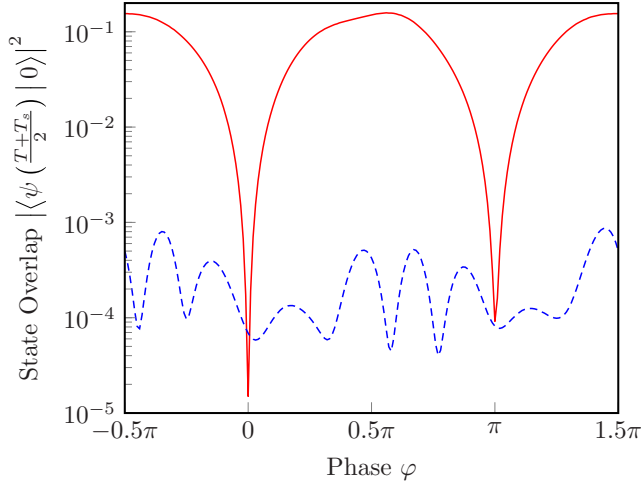
FIG. 4. (Color online) The red solid line shows the phase's effect on the state overlap at $T/2$ of the pulse in Fig. 3. A wrongly tuned phase induces large errors. The dashed blue line corresponds to a shorter pulse where the oscillation is switched off before the $m = \pm 1$ jumps can occur. As expected, the error is much less sensitive to the phase.

FIG. 6. (Color online) Error in logarithmic scale as a function of the error in the parameters (v, Ω) relative to the optimum found by the NM algorithm for the pulse in Fig. 5. As can be seen there is a narrow valley where errors on v and Ω do not affect the fidelity too severely. However, this valley is very steep signifying that the parameters Ω^2 and v need to have the right ratio.

optimal solution with a smaller oscillation but with a chirped frequency.

1. Optimization of a linear sweep with a single frequency oscillation pulse

When the short window of Eq. (7) cannot be created experimentally, diabaticity can still be engineered by optimizing the four parameters $(v, \lambda, \Omega, \varphi)$ of Eq. (2). This is done with the Nelder-Mead (NM) simplex search algorithm [56] which is often used in optimal control [40,57–59]. The metric to be minimized is $|\langle \psi(T)|0\rangle|^2$. Pulses of arbitrarily low population transfer from $|1\rangle$ to $|0\rangle$ can be found with NM. Figure 5
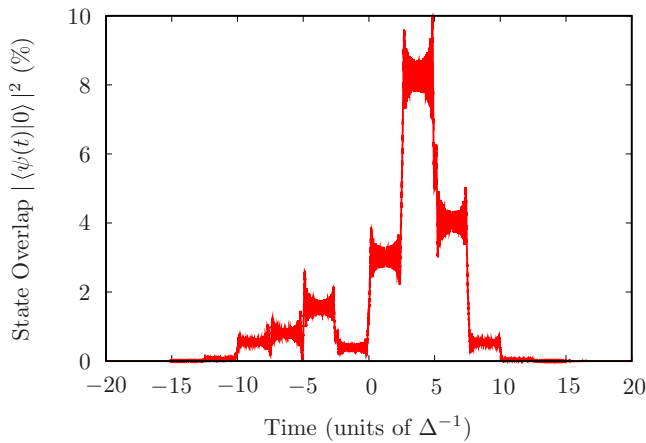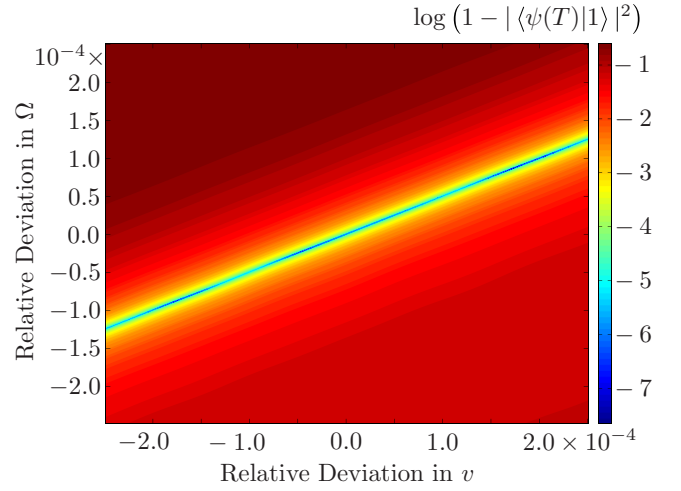

FIG. 5. (Color online) Time evolution under the action of a pulse given by Eq. (2) where the parameters were optimized with the NM algorithm. The pulse duration was set to $T = 200 \, \Delta^{-1}$ and the optimization resulted in $\Omega = 99.9718 \, \Delta$, $\lambda = 311.631 \, \Delta$, $v = 9.99409 \Delta^2$, and $\varphi = 2\pi \, 0.387$. At the end of the pulse, the error is of order 10^{-6}.

shows a pulse with performance perfect down to 10^{-6} where the optimization was stopped; in principle, machine precision can be reached. This pulse does not rely on suppressing the $m = 0$ transition or time-reversal antisymmetry. Instead, the parameters found allow many photon-assisted transitions but they produce a time evolution that leaves no population in $|0\rangle$ at the end of the pulse by tailoring their interference properly. Numerically we found good convergence regardless of the initial parameters, but typically our analytical solutions from the previous sections were used as initial guesses. Many high-quality pulses having different parameters can be found.

In an experiment the parameters of the pulse can differ from those intended. It is thus important to characterize how robust the pulse is with respect to parameter fluctuations. To study how robust these pulses are we introduce small errors in the optimal values and study the decrease in fidelity. We find that these pulses are very sensitive to small errors in the parameters v and Ω as shown in Fig. 6. The pulse is only sensitive to errors in the ratio between Ω^2 and v, not to these parameters separately. This can be traced back to the observation that the Stokes phase in Eq. (6) strongly depends on Ω^2/v. For parameters φ and λ we find that the optimal pulse is less sensitive and can tolerate roughly an order of magnitude more fluctuations than in v and Ω. Such errors in parameter values could be corrected by using the experiment to close the control loop [59].

2. GRAPE pulses

To go beyond the constraint imposed by the pulse shape of Eq. (2) we use the optimal control algorithm named gradient ascent pulse shape engineering (GRAPE). It tries to optimize a fidelity Φ which is a functional of the pulse $\varepsilon(t)$. The pulse sequence is discretized into N constant pixels u_j of time
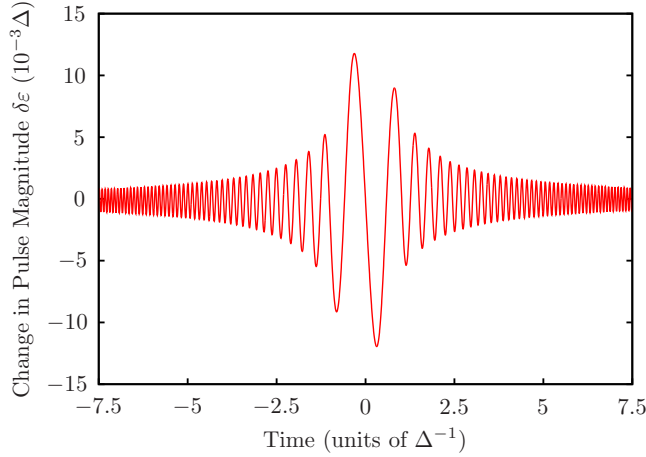
FIG. 7. (Color online) Pulse change due to GRAPE on an initial linear sweep of duration $9.6\Delta^{-1}$ and $v = 40\pi^2\Delta^2$. The figure shows the central region of the pulse where the change is most pronounced. Time and pulse amplitude are respectively given in inverse and proportional units of the gap size Δ.

$\Delta T = N/T$. GRAPE proceeds by iteratively updating the pixel values according to the rule $u_j^{(n)} \mapsto u_j^{(n+1)} = u_j^{(n)} + \epsilon \nabla_j \Phi$ where $\nabla_j \Phi$ is the gradient of the fidelity function with respect to pixel j. Details of the procedure are given in [60]. The gradient is computed analytically [61]. For our situation we use the gate overlap fidelity

$$\Phi = \tfrac{1}{4}\left|\mathrm{Tr}\big\{\hat{\sigma}_0^\dagger \hat{U}[\varepsilon(t)]\big\}\right|^2, \tag{8}$$

where σ_0 is the identity matrix and $\hat{U}[\varepsilon(t)]$ is the time evolution realized by the pulse $\varepsilon(t)$.

When considering the LZ problem we use a linear sweep with speed v without coherent drive as the initial pulse. GRAPE achieved the target error of 10^{-5} by adding to the initial pulse a modulation with a time-dependent frequency. This extra modulation is shown in Fig. 7. Its amplitude is much smaller than the fixed frequency case. The new oscillation is reminiscent of the Stückelberg oscillations with decaying amplitude and increasing frequency. This can be viewed as a rotating frame version of the oscillation-augmented sweep: instead of sweeping the energy splitting, the frequency is swept accordingly.

III. QUANTUM SPEED LIMITS FOR AN ADIABATIC EVOLUTION

This section considers the case when we wish to transfer the population between the two bare states by sweeping through the anticrossing. Equivalently this corresponds to staying in the same energy branch at all times. This can be achieved if the time evolution is adiabatic. Time evolutions of this sort are important in quantum computing where one wishes to exchange a quanta between two quantum elements such as a resonator and a qubit or two qubits [62]. Recently a protocol has been demonstrated that allows fast adiabatic two-qubit gates by making use of only σ_z control and optimal window functions [7]. For quantum computing, the population occupation and the phases can be important.
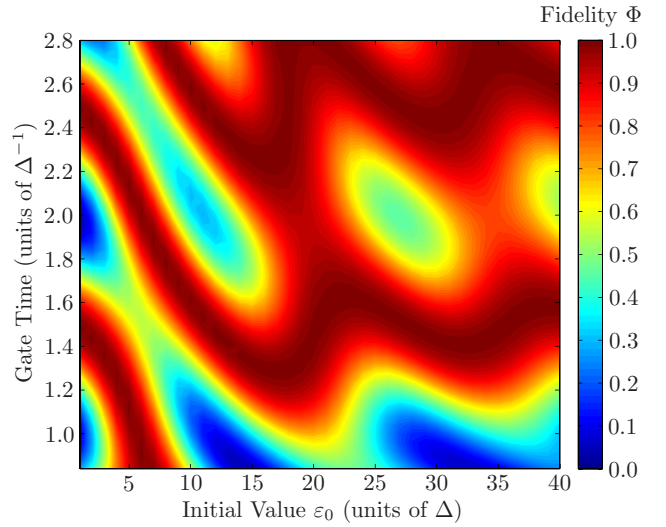

FIG. 8. (Color online) Local phase-insensitive fidelity for the initial pulse given by Eq. (9) for different gate times and start values for $\Delta = 0.04$ GHz in all simulations. This shows that this start pulse fails to keep the population in the energy eigenstate while crossing the anticrossing, i.e., it is not adiabatic.

A. Adiabatic pulses

In this section we wish to realize the gate $\hat{U}_{\mathrm{des}} = \hat{\sigma}_x$ through a pulse with boundary conditions $\varepsilon(\pm T/2) = \pm\varepsilon_0$. The initial pulse for the GRAPE optimization is

$$\varepsilon(t) = -\Delta \tan\left[\frac{\arctan\left(\frac{\Delta}{\varepsilon_0}\right) - \frac{\pi}{2}}{\mathrm{erf}\left(-\lambda\frac{T}{2}\right)}\mathrm{erf}\left(\lambda t\right)\right]. \tag{9}$$

The parameter λT controls the width of the sweep, hence λ alone sets the speed at which the point $\varepsilon = 0$ is crossed. The pulse presented in Eq. (9) is an already studied pulse form designed so that the sweep across the anticrossing is done slowly; but farther away from the anticrossing, the sweep velocity increases [7]. Figure 8 shows the phase-insensitive fidelity, defined by $\Phi_{\mathrm{ins}} = (|\langle 0|\hat{U}|1\rangle|^2 + |\langle 1|\hat{U}|0\rangle|^2)/2$, of the initial pulse in Eq. (9) as a function of (ε_0, T). Φ_{ins} only measures the population transfer. There are regions where the pulse does not perform as well as expected even when the gate time is long.

B. Numerics

To improve on the situation shown in Fig. 8 we use the GRAPE algorithm and study the dependence of the fidelity on gate time T. To preserve the adiabatic nature of the process, the pulse is convoluted with a Gaussian to remove any fast oscillations. The fidelity to be optimized is the same as in Eq. (8) but with the target gate being $\hat{\sigma}_x$ instead of the identity. This guarantees that both phase and population are correct after the gate. Figure 9 depicts the final gate fidelity Φ in the (ε_0, T) parameter space where each pulse corresponds to the result of a GRAPE optimization starting from Eq. (9). As can be seen, all the regions of Fig. 8 that showed a loss of fidelity have been improved upon. This is true as long as the pulse is long enough. Below a certain duration a form of quantum speed limit (QSL)
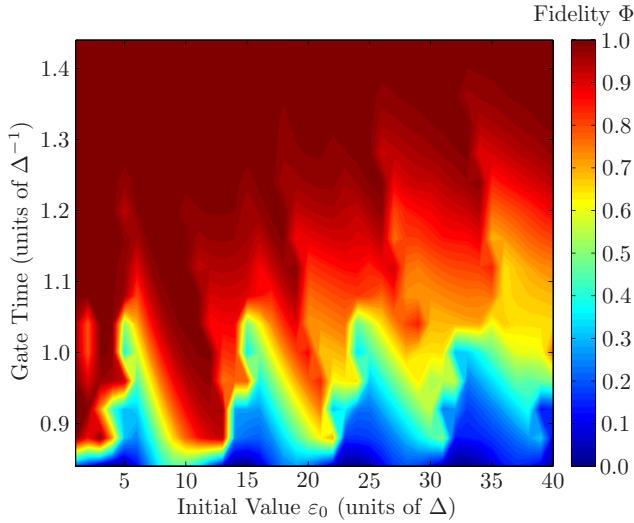
FIG. 9. (Color online) Fidelity of the pulses optimized with GRAPE as a function of duration and sweep range for $\Delta = 0.04$ GHz. Once the gate time is sufficiently large, any fidelity can be reached.
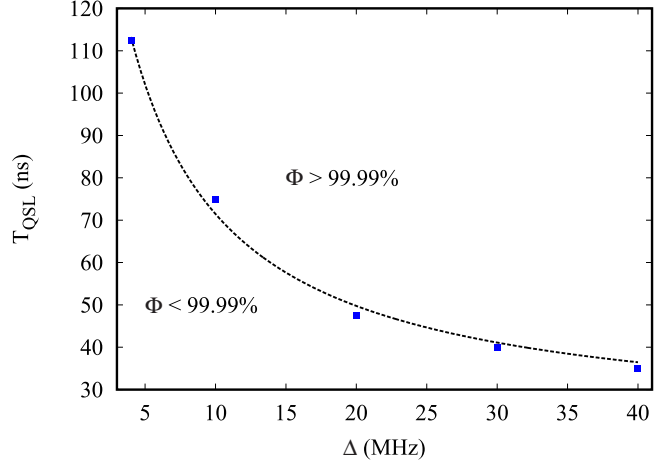

FIG. 10. (Color online) Study of the quantum speed limit after optimization of the pulse with the GRAPE algorithm. For each specific Δ a gate time is found for which a high fidelity can be achieved for almost all values of ε_0. The data points in blue are then fitted by the dashed line corresponding to Eq. (10).

is encountered and GRAPE can no longer improve the fidelity. The dependence of this QSL is further studied in Fig. 10. It shows the behavior of the QSL as function of Δ. The figure was created by choosing a gate time for which a good fidelity ($> 99.99\%$) can be found for almost all values of ε_0. The data points (blue squares) are then fitted to

$$T_{\mathrm{QSL}}(\Delta) = t_0 + \frac{c}{\Delta + \Delta_0}, \qquad (10)$$

with c and Δ_0 as fit parameters whereas t_0 is controlled by the fact that we have used "buffer pixels" in GRAPE [39] to prevent steep initial rise and final drop. As seen from the figure this empirical fit describes very well the behavior observed for T_{QSL}.

IV. CONCLUSION

We have analytically discussed the influence of a linear sweep augmented by a fast and strong single frequency oscillation on the dynamics of a two-level system. We showed how such pulses can be used to engineer diabaticity without resorting to large bandwidth control. Optimal control can help go beyond the analytic considerations producing pulses with machine precision level error. In the adiabatic study, improved adiabatic pulses were found with optimal control and their quantum speed limit was discussed.

[1] K.-P. Marzlin and B. C. Sanders, Phys. Rev. Lett. **93**, 160408 (2004).

[2] M. S. Sarandy, L.-A. Wu, and D. A. Lidar, Quantum Inf. Process. **3**, 331 (2004).

[3] D. T. K. Singh, L. Kwek, X. Fan, and C. Oh, Phys. Lett. A **339**, 288 (2005).

[4] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, arXiv:quant-ph/0001106.

[5] A. M. Childs, E. Farhi, and J. Preskill, Phys. Rev. A **65**, 012322 (2001).

[6] L. DiCarlo, J. Chow, J. Gambetta, L. Bishop, B. Johnson, D. Schuster, J. Majer, A. Blais, L. Frunzio, S. Girvin, and R. Schoelkopf, Nature **460**, 240 (2009).

[7] J. M. Martinis and M. R. Geller, Phys. Rev. A **90**, 022307 (2014).

[8] A. Galiautdinov, A. N. Korotkov, and J. M. Martinis, Phys. Rev. A **85**, 042321 (2012).

[9] R. W. Simmonds, K. M. Lang, D. A. Hite, S. Nam, D. P. Pappas, and J. M. Martinis, Phys. Rev. Lett. **93**, 077003 (2004).

[10] J. D. Whittaker, F. C. S. da Silva, M. S. Allman, F. Lecocq, K. Cicak, A. J. Sirois, J. D. Teufel, J. Aumentado, and R. W. Simmonds, Phys. Rev. B **90**, 024513 (2014).

[11] J. H. Cole, C. Müller, P. Bushev, G. J. Grabovskij, J. Lisenfeld, A. Lukashenko, A. V. Ustinov, and A. Shnirman, Appl. Phys. Lett. **97**, 252501 (2010).

[12] R. de Sousa, K. B. Whaley, T. Hecht, J. von Delft, and F. K. Wilhelm, Phys. Rev. B **80**, 094515 (2009).

[13] E. Torrontegui, S. Ibáñez, S. Martínez-Garaot, M. Modugno, A. del Campo, D. Guéry-Odelin, A. Ruschhaupt, X. Chen, and J. G. Muga, Adv. At. Mol. Opt. Phys. **62**, 117 (2013).

[14] H. Breuer and M. Holtaus, Phys. Lett. A **140**, 507 (1989).

[15] S. Guérin and H.-R. Jauslin, Phys. Rev. A **55**, 1262 (1997).

[16] S. Guérin, Phys. Rev. A **56**, 1458 (1997).

[17] L. D. Landau, Phys. Z. Sowjetunion **2**, 46 (1932).

[18] C. Zener, Proc. R. Soc. London A **137**, 696 (1932).

[19] M. V. Berry, J. Phys. A: Math. Theor. **42**, 365303 (2009).

[20] D. M. Berns, M. S. Rudner, S. O. Valenzuela, K. K. Berggren, W. D. Oliver, L. S. Levitov, and T. P. Orlando, Nature **455**, 51 (2008).

[21] S. Shevchenko, S. Ashhab, and F. Nori, Phys. Rep. **492**, 1 (2010).

[22] F. Forster, G. Petersen, S. Manus, P. Hänggi, D. Schuh, W. Wegscheider, S. Kohler, and S. Ludwig, Phys. Rev. Lett. **112**, 116803 (2014).

[23] S. Ludwig, P. Nalbach, D. Rosenberg, and D. Osheroff, Phys. Rev. Lett. **90**, 105501 (2003).

[24] P. Nalbach, D. Osheroff, and S. Ludwig, J. Low Temp. Phys. **137**, 395 (2004).

[25] A. Zenesini, D. Ciampini, O. Morsch, and E. Arimondo, Phys. Rev. A **82**, 065601 (2010).

[26] H. M. Carruzzo, E. R. Grannan, and C. C. Yu, Phys. Rev. B **50**, 6685 (1994).

[27] L. Bogani and W. Wernsdorfer, Nat. Mater. **7**, 179 (2008).

[28] I. Chiorescu, W. Wernsdorfer, A. Muller, H. Bogge, and B. Barbara, J. Magn. Magn. Mater. **221**, 103 (2000).

[29] H. Nakamura, *Nonadiabatic Transition* (World Scientific, Singapore, 2002).

[30] S. Rice and M. Zhao, *Optical Control of Molecular Dynamics* (Wiley, New York, 2000).

[31] P. Brumer and M. Shapiro, *Principles of the Quantum Control of Molecular Processes* (Wiley, New York, 2003).

[32] P. Cerfontaine, T. Botzem, D. P. DiVincenzo, and H. Bluhm, Phys. Rev. Lett. **113**, 150501 (2014).

[33] J. Clarke and F. K. Wilhelm, Nature **453**, 1031 (2008).

[34] M. Devoret and R. Schoelkopf, Science **339**, 1169 (2013).

[35] F. Motzoi, J. M. Gambetta, P. Rebentrost, and F. K. Wilhelm, Phys. Rev. Lett. **103**, 110501 (2009).

[36] E. Lucero, J. Kelly, R. C. Bialczak, M. Lenander, M. Mariantoni, M. Neeley, A. D. O'Connell, D. Sank, H. Wang, M. Weides, J. Wenner, T. Yamamoto, A. N. Cleland, and J. M. Martinis, Phys. Rev. A **82**, 042339 (2010).

[37] R. Schutjens, F. A. Dagga, D. J. Egger, and F. K. Wilhelm, Phys. Rev. A **88**, 052330 (2013).

[38] V. Vesterinen, O.-P. Saira, A. Bruno, and L. DiCarlo, arXiv:1405.0450.

[39] D. Egger and F. K. Wilhelm, Supercond. Sci. Technol. **27**, 014001 (2014).

[40] J. Kelly, R. Barends, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, I.-C. Hoi, E. Jeffrey, A. Megrant, J. Mutus, C. Neill, P. J. J. O'Malley, C. Quintana, P. Roushan, D. Sank, A. Vainsencher, J. Wenner, T. C. White, A. N. Cleland, and J. M. Martinis, Phys. Rev. Lett. **112**, 240504 (2014).

[41] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, C. Neill, P. O'Malley, P. Roushan, A. Vainsencher, J. Wenner, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Nature **508**, 500 (2014).

[42] L. S. Schulman, Lecture Notes Phys. **734**, 107 (2008).

[43] L. Vaidman, Am. J. Phys. **60**, 182 (1992).

[44] T. Caneva, M. Murphy, T. Calarco, R. Fazio, S. Montangero, V. Giovannetti, and G. E. Santoro, Phys. Rev. Lett. **103**, 240501 (2009).

[45] M. D. Grace, J. M. Dominy, W. M. Witzel, and M. S. Carroll, Phys. Rev. A **85**, 052313 (2012).

[46] M. Wubs, K. Saito, S. Kohler, Y. Kayanuma, and P. Hänggi, New J. Phys. **7**, 218 (2005).

[47] H. Zhong, Q. Xie, J. Huang, X. Qin, H. Deng, J. Xu, and C. Lee, Phys. Rev. A **90**, 023635 (2014).

[48] M. C. Goorden and F. K. Wilhelm, Phys. Rev. B **68**, 012508 (2003).

[49] F. Grossmann, T. Dittrich, P. Jung, and P. Hänggi, Phys. Rev. Lett. **67**, 516 (1991).

[50] M. Grifoni and P. Hänggi, Phys. Rep. **304**, 229 (1998).

[51] M. Wubs, Chem. Phys. **375**, 163 (2010).

[52] S. Ashhab, J. R. Johansson, A. M. Zagoskin, and F. Nori, Phys. Rev. A **75**, 063414 (2007).

[53] W. Warren, J. Chem. Phys. **81**, 5437 (1984).

[54] N. Sinitsyn, arXiv:1501.06083.

[55] R. A. Jalabert and H. M. Pastawski, Phys. Rev. Lett. **86**, 2490 (2001).

[56] J. A. Nelder and R. Mead, Comput. J. **7**, 308 (1965).

[57] M. J. Biercuk, H. Uys, A. P. VanDevender, N. Shiga, W. M. Itano, and J. J. Bollinger, Nature **458**, 996 (2009).

[58] P. Doria, T. Calarco, and S. Montangero, Phys. Rev. Lett. **106**, 190501 (2011).

[59] D. J. Egger and F. K. Wilhelm, Phys. Rev. Lett. **112**, 240503 (2014).

[60] N. Khaneja, T. Reiss, C. Kehlet, T. Schulte-Herbrüggen, and S. J. Glaser, J. Magn. Reson. **172**, 296 (2005).

[61] S. Machnes, U. Sander, S. J. Glaser, P. de Fouquières, A. Gruslys, S. Schirmer, and T. Schulte-Herbrüggen, Phys. Rev. A **84**, 022305 (2011).

[62] M. Mariantoni, H. Wang, T. Yamamoto, M. Neeley, R. C. Bialczak, Y. Chen, M. Lenander, E. Lucero, A. D. O'Connell, D. Sank, M. Weides, J. Wenner, Y. Yin, J. Zhao, A. N. Korotkov, A. N. Cleland, and J. M. Martinis, Science **334**, 61 (2011).

REFERENCES

Aharonov, D. and Ben-Or, M. (2008). Fault-tolerant quantum computation with constant error rate. *SIAM Journal on Computing*, 38(4):1207–1282.

Aharonov, D., Kitaev, A., and Preskill, J. (2006). Fault-tolerant quantum computation with long-range correlated noise. *Physical Review Letters*, 96:050504.

Aliferis, P. (2007). *Level reduction and the quantum threshold theorem*. PhD thesis, California Institute of Technology.

Aliferis, P., Gottesman, D., and Preskill, J. (2008). Accuracy threshold for postselected quantum computation. *Quantum Information and Computation*, 8:181–244.

Bardeen, J., Cooper, L. N., and Schrieffer, J. R. (1957). Theory of superconductivity. *Physical Review*, 108:1175–1204.

Barends, R., Kelly, J., Megrant, A., Sank, D., Jeffrey, E., Chen, Y., Yin, Y., Chiaro, B., Mutus, J., Neill, C., O'Malley, P., Roushan, P., Wenner, J., White, T. C., Cleland, A. N., and Martinis, J. M. (2013). Coherent Josephson qubit suitable for scalable quantum integrated circuits. *Physical Review Letters*, 111:080502.

Barends, R., Kelly, J., Megrant, A., Veitia, A., Sank, D., Jeffrey, E., White, T. C., Mutus, J., Fowler, A. G., Campbell, B., Chen, Y., Chen, Z., Chiaro, B., Dunsworth, A., Neill, C., O'Malley, P., Roushan, P., Vainsencher, A., Wenner, J., Korotkov, A. N., Cleland, A. N., and Martinis, J. M. (2014). Superconducting quantum circuits at the surface code threshold for fault tolerance. *Nature*, 508(7497):500–503.

Beigi, S. and König, R. (2011). Simplified instantaneous non-local quantum computation with applications to position-based cryptography. *New J. Phys.*, 13(9):093036.

Bhattacharya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Mathematical Society*, 5:99–109.

Black, J. and Halperin, B. (1977). Spectral diffusion, phonon echoes, and saturation recovery in glasses at low temperatures. *Physical Review B*, 16:2879–2895.

Brown, K. R., Wilson, A. C., Colombe, Y., Ospelkaus, C., Meier, A. M., Knill, E., Leibfried, D., and Wineland, D. J. (2011). Single-qubit-gate error below $\mathbf{10^{-4}}$ in a trapped ion. *Physical Review A*, 84:030303.

Chang, J. B., Vissers, M. R., Córcoles, A. D., Sandberg, M., Gao, J., Abraham, D. W., Chow, J. M., Gambetta, J. M., Beth Rothwell, M., Keefe, G. A., Steffen, M., and Pappas, D. P. (2013). Improved superconducting qubit coherence using titanium nitride. *Applied Physics Letters*, 103:012602.

Chasseur, T., Theis, L. S., Sanders, Y. R., Egger, D. J., and Wilhelm, F. K. (2015). Engineering adiabaticity at an avoided crossing with optimal control. *Physical Review A*, 91:043421.

Chen, Y., Neill, C., Roushan, P., Leung, N., Fang, M., Barends, R., Kelly, J., Campbell, B., Chen, Z., Chiaro, B., Dunsworth, A., Jeffrey, E., Megrant, A., Mutus, J. Y., O'Malley, P. J. J., Quintana, C. M., Sank, D., Vainsencher, A., Wenner, J., White, T. C., Geller, M. R., Cleland, A. N., and Martinis, J. M. (2014). Qubit architecture with high coherence and fast tunable coupling. *Physical Review Letters*, 113:220502.

Chow, J. M., Gambetta, J. M., Córcoles, A. D., Merkel, S. T., Smolin, J. A., Rigetti, C., Poletto, S., Keefe, G. A., Rothwell, M. B., Rozen, J. R., Ketchen, M. B., and Steffen, M. (2012). Universal quantum gate set approaching fault-tolerant thresholds with superconducting qubits. *Physical Review Letters*, 109:060501.

Chow, J. M., Gambetta, J. M., Magesan, E., Abraham, D. W., Cross, A. W., Johnson, B. R., Masluk, N. A., Ryan, C. A., Smolin, J. A., Srinivasan, S. J., and Steffen, M. (2014). Implementing a strand of a scalable fault-tolerant quantum computing fabric. *Nature Communications*, 5.

Chuang, I. L. and Nielsen, M. A. (1997). Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467.

Clarke, J. and Wilhelm, F. K. (2008). Superconducting quantum bits. *Nature*, 453(7198):1031–1042.

Cole, J. H., Schirmer, S. G., Greentree, A. D., Wellard, C. J., Oi, D. K. L., and Hollenberg, L. C. L. (2005). Identifying an experimental two-state Hamiltonian to arbitrary accuracy. *Physical Review A*, 71:062312.

Dankert, C., Cleve, R., Emerson, J., and Livine, E. (2009). Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80:012304.

Dennis, E., Kitaev, A., Landahl, A., and Preskill, J. (2002). Topological quantum memory. *J. Math. Phys.*, 43(9):4452–4505.

DiVincenzo, D. P. (2000). The physical implementation of quantum computation. *Fortschritte der Physik*, 48:771–783.

Emerson, J., Alicki, R., and Życzkowski, K. (2005). Scalable noise estimation with random unitary operators. *Journal of Optics B: Quantum and Semiclassical Optics*, 7(10):S347.

Epstein, J. M., Cross, A. W., Magesan, E., and Gambetta, J. M. (2014). Investigating the limits of randomized benchmarking protocols. *Physical Review A*, 89:062321.

Fedorov, A., Steffen, L., Baur, M., da Silva, M. P., and Wallraff, A. (2012). Implementation of a Toffoli gate with superconducting circuits. *Nature*, 481(7380):170–172.

Ferrie, C., Granade, C. E., and Cory, D. G. (2013). How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies. *Quantum Information Processing*, 12(1):611–623.

Flammia, S. T. and Liu, Y.-K. (2011). Direct fidelity estimation from few pauli measurements. *Physical Review Letters*, 106:230501.

Fowler, A. G. (2012). Proof of finite surface code threshold for matching. *Physical Review Letters*, 109:180502.

Fowler, A. G. (2013). Optimal complexity correction of correlated errors in the surface code. arXiv:1310.0863.

Fowler, A. G., Mariantoni, M., Martinis, J. M., and Cleland, A. N. (2012). Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86:032324.

Fredkin, E. and Toffoli, T. (1982). Conservative logic. *Int. J. Theor. Phys.*, 21(3–4):219–253.

Fuchs, C. and van de Graaf, J. (1999). Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227.

Gilchrist, A., Langford, N. K., and Nielsen, M. A. (2005). Distance measures to compare real and ideal quantum processes. *Physical Review A*, 71:062310.

Granade, C., Ferrie, C., Hincks, I., and Sanders, Y. (2016). QInfer: Library for Statistical Inference in Quantum Information. http://dx.doi.org/10.5281/zenodo.45087.

Granade, C. E., Ferrie, C., Wiebe, N., and Cory, D. G. (2012). Robust online Hamiltonian learning. *New Journal of Physics*, 14(10):103013.

Haah, J., Harrow, A., Ji, Z., Wu, X., and Yu, N. (2015). Sample-optimal tomography of quantum states. arXiv:1508.01797.

Hájek, A. (2012). Interpretations of probability. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Stanford University.

Hayes, D., Clark, S. M., Debnath, S., Hucul, D., Inlek, I. V., Lee, K. W., Quraishi, Q., and Monroe, C. (2012). Coherent error suppression in multiqubit entangling gates. *Physical Review Letters*, 109:020503.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 186(1007):453–461.

Johnston, N. (2015). QETLAB: A MATLAB toolbox for quantum entanglement, version 0.8. http://qetlab.com.

Johnston, N., Kribs, D. W., and Paulsen, V. I. (2009). Computing stabilized norms for quantum operations via the theory of completely bounded maps. *Quantum Information and Computation*, 9(1):16–35.

Josephson, B. (1962). Possible new effects in superconductive tunnelling. *Physics Letters*, 1(7):251 – 253.

Kelly, J., Barends, R., Fowler, A. G., Megrant, A., Jeffrey, E., White, T. C., Sank, D., Mutus, J. Y., Campbell, B., Chen, Y., Chen, Z., Chiaro, B., Dunsworth, A., Hoi, I. C., Neill, C., O'Malley, P. J. J., Quintana, C., Roushan, P., Vainsencher, A., Wenner, J., Cleland, A. N., and Martinis, J. M. (2015). State preservation by repetitive error detection in a superconducting quantum circuit. *Nature*, 519(7541):66–69.

Khaneja, N., Reiss, T., Kehlet, C., Schulte-Herbrüggen, T., and Glaser, S. J. (2005). Optimal control of coupled spin dynamics: design of NMR pulse sequences by gradient ascent algorithms. *Journal of Magnetic Resonance*, 172(2):296 – 305.

Kimmel, S., da Silva, M. P., Ryan, C. A., Johnson, B. R., and Ohki, T. (2014). Robust extraction of tomographic information via randomized benchmarking. *Physical Review X*, 4:011050.

Kitaev, A. Y. (1997). Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191.

Knill, E. (2005). Quantum computing with realistically noisy devices. *Nature*, 434(7029):39–44.

Knill, E., Laflamme, R., and Zurek, W. H. (1998a). Resilient quantum computation. *Science*, 279(5349):342–345.

Knill, E., Laflamme, R., and Zurek, W. H. (1998b). Resilient quantum computation. *Science*, 279(5349):342–345.

Knill, E., Leibfried, D., Reichle, R., Britton, J., Blakestad, R. B., Jost, J. D., Langer, C., Ozeri, R., Seidelin, S., and Wineland, D. J. (2008). Randomized benchmarking of quantum gates. *Physical Review A*, 77:012307.

Kueng, R., Long, D., Doherty, A., and Flammia, S. (2015). Comparing experiments to the fault-tolerance threshold. arXiv:1510.05653.

Landau, L. (1932). Zur theorie der energieubertragung. ii. *Physics of the Soviet Union*, 2(2):46–51.

Leggett, A. J. and Vural, D. C. (2013). "tunneling two-level systems" model of the low-temperature properties of glasses: Are "smoking-gun" tests possible? *The Journal of Physical Chemistry B*, 117(42):12966–12971. PMID: 23924397.

Levin, D. A., Peres, Y., and Wilmer, E. L. (2009). *Markov Chains and Mixing Times*. American Mathematical Society, Providence.

Lindblad, G. (1976). On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, 48(2):119–130.

Lisenfeld, J., Grabovskij, G. J., Müller, C., Cole, J. H., Weiss, G., and Ustinov, A. V. (2015). Observation of directly interacting coherent two-level systems in an amorphous material. *Nature Communications*, 6:6182.

Liu, J. and West, M. (2001). Combined parameter and state estimation in simulation-based filtering. In Doucet, A., de Freitas, N., and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*, Statistics for Engineering and Information Science, pages 197–223. Springer New York.

Lucero, E., Barends, R., Chen, Y., Kelly, J., Mariantoni, M., Megrant, A., O'Malley, P., Sank, D., Vainsencher, A., Wenner, J., White, T., Yin, Y., Cleland, A. N., and Martinis, J. M. (2012). Computing prime factors with a Josephson phase qubit quantum processor. *Nature Physics*, 8(10):719–723.

Magesan, E. (2012). *Characterizing Noise in Quantum Systems*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada.

Magesan, E., Gambetta, J. M., and Emerson, J. (2011). Scalable and robust randomized benchmarking of quantum processes. *Physical Review Letters*, 106:180504.

Magesan, E., Gambetta, J. M., and Emerson, J. (2012a). Characterizing quantum gates via randomized benchmarking. *Physical Review A*, 85:042311.

Magesan, E., Gambetta, J. M., Johnson, B. R., Ryan, C. A., Chow, J. M., Merkel, S. T., da Silva, M. P., Keefe, G. A., Rothwell, M. B., Ohki, T. A., Ketchen, M. B., and Steffen, M. (2012b). Efficient measurement of quantum gate error by interleaved randomized benchmarking. *Physical Review Letters*, 109:080505.

Magesan, E., Puzzuoli, D., Granade, C. E., and Cory, D. G. (2013). Modeling quantum noise for efficient testing of fault-tolerant circuits. *Physical Review A*, 87:012324.

Majorana, E. (1932). Atomi orientati in campo magnetico variabile. *Il Nuovo Cimento*, 9(2):43–50.

Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419.

Mariantoni, M., Wang, H., Bialczak, R. C., Lenander, M., Lucero, E., Neeley, M., O'Connell, A. D., Sank, D., Weides, M., Wenner, J., Yamamoto, T., Yin, Y., Zhao, J., Martinis, J. M., and Cleland, A. N. (2011). Photon shell game in three-resonator circuit quantum electrodynamics. *Nature Physics*, 7(4):287–293.

Martinis, J., Cooper, K., McDermott, R., Steffen, M., Ansmann, M., Osborn, K., Cicak, K., Oh, S., Pappas, D., Simmonds, R., and Yu, C. (2005). Decoherence in Josephson qubits from dielectric loss. *Physical Review Letters*, 95:210503.

Neeley, M., Ansmann, M., Bialczak, R. C., Hofheinz, M., Katz, N., Lucero, E., O'Connell, A., Wang, H., Cleland, A. N., and Martinis, J. M. (2008). Process tomography of quantum memory in a Josephson-phase qubit coupled to a two-level state. *Nature Physics*, 4(7):523–526.

Nielsen, M. A. (2002). A simple formula for the average gate fidelity of a quantum dynamical operation. *Physics Letters A*, 303(4):249 – 252.

Nielsen, M. A. and Chuang, I. L. (2000). *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge.

O'Donnell, R. and Wright, J. (2015). Efficient quantum tomography. arXiv:1508.01907.

O'Malley, P. J. J., Kelly, J., Barends, R., Campbell, B., Chen, Y., Chen, Z., Chiaro, B., Dunsworth, A., Fowler, A. G., Hoi, I.-C., Jeffrey, E., Megrant, A., Mutus, J., Neill, C., Quintana, C., Roushan, P., Sank, D., Vainsencher, A., Wenner, J., White, T. C., Korotkov, A. N., Cleland, A. N., and Martinis, J. M. (2015). Qubit metrology of ultralow phase noise using randomized benchmarking. *Physical Review Applied*, 3:044009.

Preskill, J. (1998). Reliable quantum computers. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 454(1969):385–410.

Puzzuoli, D., Granade, C., Haas, H., Criger, B., Magesan, E., and Cory, D. G. (2014). Tractable simulation of error correction with honest approximations to realistic fault models. *Physical Review A*, 89:022306.

Raussendorf, R. and Harrington, J. (2007). Fault-tolerant quantum computation with high threshold in two dimensions. *Physical Review Letters*, 98:190504.

Reichardt, B. (2006). *Error-detection-based quantum fault tolerance against discrete Pauli noise*. PhD thesis, University of California, Berkely, Berkeley, California, United States of America.

Ryan, C. A., Laforest, M., and Laflamme, R. (2009). Randomized benchmarking of single- and multi-qubit control in liquid-state NMR quantum information processing. *New Journal of Physics*, 11(1):013034.

Sanders, Y. R., Wallman, J. J., and Sanders, B. C. (2016). Bounding quantum gate error rate based on reported average fidelity. *New Journal of Physics*, 18(1):012002.

Savage, J. E. (1997). *Models of Computation: Exploring the Power of Computing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition.

Shalibo, Y., Rofe, Y., Shwa, D., Zeides, F., Neeley, M., Martinis, J. M., and Katz, N. (2010). Lifetime and coherence of two-level defects in a Josephson junction. *Physical Review Letters*, 105:177001.

Shor, P. (1999). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41(2):303–332.

Shor, P. W. (1995). Scheme for reducing decoherence in quantum computer memory. *Physical Review A*, 52:R2493–R2496.

Siemwiorek, D. (1991). Architecture of fault-tolerant computers: an historical perspective. *Proceedings of the IEEE*, 79(12):1710–1734.

Stenberg, M. P., Sanders, Y. R., and Wilhelm, F. K. (2014). Efficient estimation of resonant coupling between quantum systems. *Physical Review Letters*, 113(21):210404.

Stueckelberg, E. (1932). Theorie der unelastischen stösse zwischen atomen. *Helvetica Physica Acta*, 5:369.

Svore, K. M., Cross, A. W., Chuang, I. L., and Aho, A. V. (2006). A flow-map model for analyzing pseudothresholds in fault-tolerant quantum computing. *Quantum Information & Computation*, 6(3):193–212.

Svore, K. M., Terhal, B. M., and DiVincenzo, D. P. (2005). Local fault-tolerant quantum computation. *Physical Review A*, 72:022317.

Toffoli, T. (1980). *Reversible Computing/Automata, Languages and Programming*, volume 85 of *Lecture Notes in Computer Science*, pages 632–644. Springer, Berlin.

Unruh, W. G. (1995). Maintaining coherence in quantum computers. *Phys. Rev. A*, 51:992–997.

Wallman, J., Granade, C., Harper, R., and Flammia, S. T. (2015). Estimating the coherence of noise. *New Journal of Physics*, 17(11):113020.

Wallman, J. J. and Flammia, S. T. (2014). Randomized benchmarking with confidence. *New Journal of Physics*, 16(10):103032.

Watrous, J. (2009). Semidefinite programs for completely bounded norms. *Theory Comp.*, 5(11):217–238.

Watrous, J. (2011). Channel distinguishability and the completely bounded trace norm. Lecture Notes.

Whittle, P. (2000). *Probability via Expectation*. Springer-Verlag New York.

Zagoskin, A. and Blais, A. (2007). Superconducting qubits. *Physics in Canada*, 63(4):215–227.

Zener, C. (1932). Non-adiabatic crossing of energy levels. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 137(833):696–702.