ERROR MODELS FOR QUANTUM STATE AND PARAMETER ESTIMATION

by

LUCIA SCHWARZ

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2014

DISSERTATION APPROVAL PAGE

Student: Lucia Schwarz

Title: Error Models for Quantum State and Parameter Estimation

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Miriam Deutsch	Chair
Steven van Enk	Advisor
Graham Kribs	Core Member
Jeff Cina	Institutional Representative

and

J. Andrew Berglund	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2014

DISSERTATION ABSTRACT

Lucia Schwarz

Doctor of Philosophy

Department of Physics

September 2014

Title: Error Models for Quantum State and Parameter Estimation

Within the field of Quantum Information Processing, we study two subjects: For quantum state tomography, one common assumption is that the experimentalist possesses a stationary source of identical states. We challenge this assumption and propose a method to detect and characterize the drift of nonstationary quantum sources. We distinguish diffusive and systematic drifts and examine how quickly one can determine that a source is drifting. Finally, we give an implementation of this proposed measurement for single photons.

For quantum computing, fault-tolerant protocols assume that errors are of certain types. But how do we detect errors of the wrong type? The problem is that for large quantum states, a full state description is impossible to analyze, and so one cannot detect all types of errors. We show through a quantum state estimation example (on up to 25 qubits) how to attack this problem using model selection. We use, in particular, the Akaike Information Criterion. Our example indicates that the number of measurements that one has to perform before noticing errors of the wrong type scales polynomially both with the number of qubits and with the error size.

This dissertation includes previously published co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Lucia Schwarz

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Universität Augsburg, Augsburg, Germany

DEGREES AWARDED:

Doctor of Philosophy, Physics, 2014, University of Oregon
Master of Science, Physics, 2009, University of Oregon
Vordiplom, Physics, 2007, Universität Augsburg

AREAS OF SPECIAL INTEREST:

Quantum Information, Statistics, Computational Science, Data Science

PROFESSIONAL EXPERIENCE:

Graduate Research Assistant, Theoretical Quantum Information, University of Oregon 2010-1014

Graduate Teaching Fellow, University of Oregon, 2009-2010

Teaching Assistant, Department of Computer Science, University of Augsburg, Germany, 2007-2008

GRANTS, AWARDS AND HONORS:

Ph.D. Qualifying Exam Award, Department of Physics, University of Oregon 2009

Fellowship from the "Deutscher Akademischer Austauschdienst" (German Academic Exchange Service), Germany, 2008-2009

Scholarship from the "Studienstiftung des Deutschen Volkes" (German National Merit Foundation), Germany, 2006-2008

PUBLICATIONS:

Schwarz, Lucia, & van Enk, S. J. (2013). Error models in quantum computation: An application of model selection. *Phys. Rev. A* **88**, 032318.

Schwarz, Lucia, & van Enk, S. J. (2011). Detecting the Drift of Quantum Sources: Not the de Finetti Theorem. *Phys. Rev. Lett.* **106**, 180501.

ACKNOWLEDGEMENTS

First of all, many thanks to my advisor Steven van Enk for his continued support and guidance throughout my time in graduate school. I also want thank my lab mates Jun Yin, Megan Ray and Imran Mirza for being great colleagues.

A special thanks goes to my friends and fellow graduate students Erin, Shikha, Eryn and Maunta who have stood by me in happy but also in difficult times. Each of them has supported me in their own special way, and I am grateful to have shared so many wonderful experiences with them.

None of this would have been possible without my loving parents who always believed in me and supported me in all my goals, even though I moved halfway across the world and could never explain my work. And my little sister Patricia, who gave me the incentive to be a good role model and who not only listened to all my crazy ideas, but came up with even crazier ones.

Meinen Eltern, die mich in allem bedingungslos unterstützt haben.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

CHAPTER I

INTRODUCTION

1.1. Quantum Computing

The world of quantum mechanics is very different from the one that we can experience with our senses. Quantum mechanical particles have no definite position and momentum at the same time, instead we can only determine the probability for finding the particle in a certain location or with a certain velocity when it is measured. Moreover, any time we observe the particle, we also change its state.

Through these strange properties of quantum mechanical systems, a new kind of computer has been envisioned: the quantum computer. Instead of discrete, classical bits, a quantum computer uses quantum bits called "qubits", which can be in a continuous superposition of two states. Any operation on N qubits will then act on all possible 2^N different states, which means that in principle large calculations can be executed in parallel. The read-out mechanism has to be well thought out to extract exactly the information that we want, since any other information is lost during the measurement.

The potential of a quantum computer is not fully known yet. Its most famous algorithm is Shor's factoring algorithm [50], which can factor large numbers into their prime factors in polynomial time. There is a lot of interest in this algorithm because it will be able to crack the common RSA public key encryption (for more information, see [53]).

Another quantum algorigthm is Grover's search algorithm [25] which can sort an unsorted database in a time that is less than linear in the size of the database. Maybe

the most interesting possibility for scientists is the simulation of quantum systems, which may revolutionize quantum chemistry and condensed matter physics.

Actual implementations of quantum computers are still rudimentary and limited to a small number of qubits. There is a great diversity of physical systems that are proposed for use in quantum computing. One of the first studied systems was nuclear magnetic resonance [63], but the size of such a computer is limited in practice by the size of the molecules that are used. Among the most advanced implementations so far are trapped-ion systems, where the record was set by the group of Rainer Blatt with 14 entangled qubits [43]. This technology allows for complete coherent control of all the qubits.

The other promising system are superconducting qubits, where each qubit is represented by a tiny superconducting circuit that contains Josephson junctions [5, 15].

In general, we can distinguish between three different types of quantum computer. A universal quantum computer is a quantum computer with a universal set of gates, and involves complete coherent control of the qubits. This type of quantum computer will be able to perform any task that a classical Turing machine can do, and quite possibly even more. It is thought that for certain algorithms, a quantum computer will achieve a considerable speed-up.

But for certain problems, we might not need a universal quantum computer. In order to perform quantum simulation, a quantum computer only needs to simulate one particular class of Hamiltonians, and can be engineered to perform this task without being universal. Finally, there are quantum computers that rely on quantum annealing [31], a method for finding the global minimum of a function by adiabatically

cooling quantum fluctuations. These systems are already being built, but can only solve a certain class of problems [8].

As these technologies become more developed, we might eventually see systems that consist of several different parts, a fast interacting system that processes quantum states and a slower system where quantum states can be stored for longer times without too much decoherence. In this case we will also need to develop methods for transfering quantum states from one system to another.

The crucial resource for quantum computing is strongly believed to be entanglement, and all quantum computers use entanglement in some way. First described in the famous paper by Einstein, Podolsky and Rosen [16], entanglement describes correlations between quantum mechanical particles that are stronger than allowed by classical physics. For example, consider a singlet state of two spin-1/2 particles, commonly denoted by $|S\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)$.

If the first system is measured to be in state $|\uparrow\rangle$ in any arbitrarily chosen measurement direction, then the second system will be in state $|\downarrow\rangle$ in the same basis and vice versa, even though quantum mechanics dictates that since the operators for the spin in x-, y- and z-direction do not commute with each other, they cannot have definite values at the same time. This effect cannot be described purely by classical correlations. Moreover, a measurement on one part of this entangled system will affect the state of the other part instantaneously, no matter how much distance is between them. It was discovered by Bell [6] that this effect can not be explained by local hidden variable theories, and in that sense quantum theory is nonlocal.

1.2. Quantum Cryptography

Two entangled particles share these stronger-than-classical correlations even if they are spatially separated over large distances. Even though it is not possible to transfer information faster than light, this principle can still be used in the field of Quantum Cryptography. In order to encrypt and send a message safely between two parties, traditionally called Alice and Bob, they need to share a private key that is unknown to any eavesdropper, called Eve. Up until recently, the classical public key algorithm was considered the most reliable and secure way to generate private keys. It uses the fact that large numbers can not be factored into their prime factors by current classical computers in a short time. If quantum computers become available, this scheme can be broken by the aforementioned Shor algorithm.

In contrast to relying on hard computational problems, quantum key distribution [17] uses the nature of entanglement as a way to transmit a private key between two parties. Light, or photons, have a certain polarization which can be measured in any direction. Specifically, we can choose a basis that defines horizontally and vertically polarized states, and a different basis that is rotated with respect to the first basis by 45 degrees, so its basis states are diagonally and anti-diagonally polarized.

Let's assume that Alice and Bob share N pairs of entangled photons, where each pair is originally in a singlet state. Both Alice and Bob now measure their photons randomly in either the horizontal/vertical basis, or in the diagonal/anti-diagonal basis.

Then they communicate classically about which basis they measured in, and disregard any state where they used different basis states. This will leave them with about $N/2$ pairs, and since the photons were entangled, both Alice and Bob can infer the measurement outcome of each other because they will be perfectly anti-correlated.

This string of bits can then be used as a secure private key to encrypt any message that they want to transmit.

Listening to the classical channel does not give any information about the key, only the basis in which it was measured in. But what happens if there is an eavesdropper that can intercept the quantum channel? If Eve performs a measurement on the transmitted photons, her measurement will project the photon into whichever basis Eve chose for her measurement. This will destroy the entanglement, and Bob's measurement outcome will then not necessarily be correlated with Alice's. By comparing a small part of their private key, Alice and Bob can find out with very high statistical confidence whether their photons were intercepted or not.

This method is secure in principle, since any measurement on the transmitted particles will change their state and thus eavesdroppers can be detected. There are already commercial options of this method of quantum key distribution.

1.3. Challenges of Quantum State Estimation

Even though quantum computers have been implemented in small systems, there are daunting challenges that need to be overcome in order to build a large scale quantum computer that is actually useful. The minimum requirement for such a quantum computer will be the implementation of qubits that can be initialized to a desired states and measured reliably. Furthermore we need to be able to perform computations by applying quantum gates to the qubits. This can be reduced to a universal set of only a few quantum gates, through whose combinations any other unitary interaction can be performed. This universal set of quantum gates needs to include two-qubit gates, so any two qubits need to be able to interact with each

other. In order to build a large-scale quantum computer, all the operations have to be implemented in a fault-tolerant way, i.e. the error probability per qubit has to be small enough so that it can be corrected reliably. If this is the case then we can control decoherence while we scale up the number of qubits, eventually leading to a quantum computer that will outperform any classical computer.

For detecting errors, it is important that the state of a quantum system can be measured reliably. For small systems, this is commonly done by Quantum Tomography, which was first implemented experimentally at the University of Oregon by Raymer in 1993 [54], and since then has become an important tool in the field of quantum information science. The largest system where tomography has been used is a system of 8 qubits [28], which required $65,536$ parameters. In chapter II we will give an introduction into the area of Quantum State Estimation and Quantum Tomography. While the theoretical background is very well understood and Quantum Tomography is widely used in practice, there are still many challenges that need to be overcome in order to perform reliable, large-scale Quantum Tomography. The main part of this thesis will examine some of these problems and provide new methods to solve them.

In chapter III we discuss how tomography is limited if the source of the quantum systems is nonstationary. We develop a method for characterizing quantum sources that can detect whether a source is drifting, and also distinguish between diffusive and systematic drifts. A special focus is placed on the question of how many measurements are needed for reliable estimates. We also show that for single-photon wave packets, our measurement can be implemented by the Hong-Ou-Mandel effect.

In chapter IV, we give an overview of quantum error correction and explain how certain types of errors in a quantum computer can do more harm than others. In

particular, the threshold theorem for fault tolerant quantum computing assumes that all errors that occur are of certain types. But how can we detect that only such benign errors occur? For implementations of tens or even hundreds of qubits, reconstructing the full state becomes impossible, and reconstructing a quantum process that occurs during a computation becomes even more impossible, because as the size of the system grows, the number of variables that are needed to describe it grows exponentially.

In chapter V, we develop a method to tackle this problem, and show its application on quantum state estimation for up to 25 qubits. Our method uses model selection to create and test error models for the state. This will drastically reduce the number of parameters that need to be measured and analyzed. Our model selection tool is the Akaike Information Criterion, a tool borrowed from information theory and statistics that has a very strong theoretical foundation (see appendix), and that has been used for many decades. Our example indicates that the number of measurements that one has to perform before noticing errors of the wrong type scales polynomially both with the number of qubits and with the error size.

Finally, in chapter VI, we conclude the results of this thesis.

Chapters III and V were published and co-authored with S. J. van Enk.

CHAPTER II

QUANTUM STATE ESTIMATION

2.1. Quantum States

A pure quantum state is a normalized vector (called *ket*) in a complex Hilbert space and describes the state of a physical system. This space can be finite- or infinite dimensional.

The simplest finite-dimensional system, which also serves as the basis of quantum computation, is a 2-level system known as a qubit, in analogy to the classical bits. Such a qubit can be represented by many different physical systems, e.g. two discrete energy levels in an atom which are coupled by a laser, or the spin of an electron which has two distinctive states that are coupled by a magnetic field.

The two orthogonal eigenstates of a qubit are labeled as $|0\rangle$ and $|1\rangle$ and form a basis of the two-dimensional Hilbert space. Any pure qubit state then can be written as a linear combination of those basis vectors

$$|\Psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle\,, \tag{2.1}$$

where α and β are complex numbers with $|\alpha|^2 + |\beta|^2 = 1$.

However, since we often deal with ensembles of states, or have incomplete information about our system, it is very useful to introduce a more general type of states called mixed states, which can be described by *density matrices* of the form

$$\rho = \sum_i p_i\,|\Psi_i\rangle\,\langle\Psi_i|\,, \tag{2.2}$$

where $\sum_i p_i = 1$ and $p_i > 0$ for all i. This density matrix contains all the information, however incomplete, that we have about a quantum system [21]. It is hermitian, positive (all eigenvalues are greater or equal to 0) and always obeys $\mathrm{Tr}(\rho) = 1$.

We define the purity of a state as $P = \mathrm{Tr}(\rho^2)$. This quantity is only equal to 1 if the state is pure. For a maximally mixed state of dimension d, which is $\rho_{mixed} = \mathbb{1}/d$, the purity becomes $P = 1/d$.

2.2. Measurements

Any observable on a quantum system (described by $|\Psi\rangle$ or ρ) can be described by a hermitian operator $\hat{A}$ that acts on the Hilbert space. The operator has eigenvalues a_i with eigenvectors $|a_i\rangle$, which represent the possible outcomes of the measurement. If $|a_i\rangle$ are normalized, each possible measurement result a_i occurs with probability

$$p_i = |\langle a_i|\Psi\rangle|^2, \tag{2.3}$$

and after the measurement the state has collapsed into the corresponding eigenvector

$$|\Psi\rangle \to |a_i\rangle . \tag{2.4}$$

This can easily be generalized to the density matrix formalism using the projection operators $\hat{A}_i = |a_i\rangle\langle a_i|$. Then the probability of each outcome is

$$p_i = \mathrm{Tr}(\rho \hat{A}_i) \tag{2.5}$$

which is also known as Born's rule, and the state collapses according to

$$\rho \to \frac{\sqrt{\hat{A}_i}\rho\sqrt{\hat{A}_i}}{\mathrm{Tr}\rho\hat{A}_i}. \tag{2.6}$$

This kind of measurement that uses projection operators is known as a *von Neumann* measurement. Since $\hat{A}$ is hermitian, the eigenvectors are orthogonal such that $\mathrm{Tr}(\hat{A}_i\hat{A}_j) = \delta_{ij}$. These operators form a complete set in the space of operators on the Hilbert space, such that $\sum_i A_i = \mathbb{1}$.

It has been found very useful to generalize the notion of measurement operators to not only allow orthogonal projection operators, but any set of positive operators that sum up to unity. Let $\{\Pi_i\}$ be such a set of Positive Operator Valued Measurements (POVM) with $\sum_i \Pi_i = \mathbb{1}$. Note that the Π_i do not necessarily have to be orthogonal, but Born's rule is still valid such that

$$p_i = \mathrm{Tr}(\rho\Pi_i). \tag{2.7}$$

Any POVM can be realized by coupling an auxiliary system to the state and performing a von Neumann measurement on this combined system [47].

2.3. Quantum Tomography

The goal of quantum tomography is to reconstruct the density matrix ρ of a quantum system. However, because of the uncertainty relation that is inherent to quantum mechanics, we can never measure all information about one single copy of the system. Take for instance a spin-qubit, which can be completely characterized by three parameters (e.g. the spin in x-, y- and z direction). Each time we measure any spin component of this qubit, the system collapses to an eigenstate of the measurement

operator and we can therefore not make any subsequent measurements on the original system.

Quantum tomography circumvents this limitation by using not just one, but many identical copies of the same quantum state ρ.

We choose a complete set of measurement operators, in general these can be denoted as POVMs $\{\Pi_i\}$ with $i \in \{1, m\}$ and record the outcome of each measurement for a large number M of such states. If each outcome i occurs x_i times, then the *frequency* $f_i = x_i/M$ of each measurement outcome approaches the *probability* $p_i = \mathrm{Tr}(\rho\Pi_i)$ of each measurement outcome, such that

$$f_i = p_i + \epsilon \tag{2.8}$$

where ϵ is the statistical error due to the finite number of measurements and decreases as $\epsilon = \mathcal{O}(1/\sqrt{M})$ with higher numbers of measurements.

If we had unlimited resources, and access to infinite copies of the system, we could make this error infinitesimally small and the measurement frequencies would be equal to the true probabilities. Then we could reconstruct the exact density matrix as the solution of the set of linear equations that (2.5) gives with $p_i \equiv f_i$ for all i.

Even if we only have a limited amout of resources, we can still use this procedure to reconstruct a density matrix ρ_e that best describes the state of each copy of ρ, by solving the m linear equations given by

$$f_i = \mathrm{Tr}(\rho_e\Pi_i) \tag{2.9}$$

for ρ_e. For more in-depth reviews of quantum tomography, see [46, 41].

2.4. Maximum Likelihood Estimation

Quantum tomography provides us with a reconstructed density matrix, but this density matrix could be unphysical. For instance, we could perform tomography on a qubit using three different measurement settings that measure the spin in x-, y- and z-direction. These measurements are represented by the Pauli matrices σ_x, σ_y and σ_z. Consider the case where we only perform one single measurement per setting, e.g. we use three identical copies of the qubit and measure each spin once. If our qubit is in state $|0\rangle$, it is possible that all the measurements return the result 'up' or '0'. Then the reconstructed density matrix is

$$\rho = \begin{pmatrix} 1 & \frac{1+i}{2} \\ \frac{1-i}{2} & 0 \end{pmatrix}, \tag{2.10}$$

which is clearly not physical because it has one negative eigenvalue. This example is extreme, but it illustrates the following problem: Most states that we are interested in tend to be pure states, which lie at the boundary of the space of physical density matrices, and therefore it is very likely that quantum tomography with a finite number of measurements will return a density matrix that lies outside of this boundary.

Instead of choosing an unphysical density matrix, we would like to choose the one state out of the space of physical density matrices that fits the measurement results best. In order to choose the best density matrix, we introduce the *likelihood* of a density matrix ρ as the probability of the measured outcome (denoted by the set of frequencies $\{f_i\}$), given that a system is in state ρ:

$$\mathcal{L} = P(\{f_i\}|\rho) = \prod_i \mathrm{Tr}(\Pi_i \rho)^{f_i M} \tag{2.11}$$

where M is the total number of measurement records, and f_i is the frequency of each outcome associated with the operator Π_i. If this expression is taken as a function of ρ instead of f_i, it is not a probability distribution since it is not normalized. However, the global maximum of this function over the space of physical density matrices is the density matrix that is most likely to have produced the recorded measurement outcome. For practical reasons, because the likelihood tends to be very large, we usually perform this maximization on the logarithmic likelihood

$$\log \mathcal{L} = \sum_i (f_i M) \log[\mathrm{Tr}(\Pi_i \rho)] = M \sum_i f_i \log[p_i(\rho)] \tag{2.12}$$

where we introduced $p_i(\rho)$ as the probability of measurement outcome i, given that the system is in state ρ. Since the logarithm is monotonous, $\log \mathcal{L}$ is maximized for the same state ρ as $\mathcal{L}$.

Note that if the result of quantum tomography from Eq. (2.9) is physical, it will always agree with the density matrix found by maximum likelihood estimation. This can be easily seen by noticing that the function $\sum_i f_i \log p_i$ is the *relative entropy* (also known as *Kullback-Leibler divergence*) between two probability distributions $\{p_i\}$ and $\{f_i\}$. The relative entropy takes on its maximum when the probability distributions are equal, i. e. $p_i \equiv f_i$ for all $i \in \{1, \ldots, M\}$. But this is just the condition that we solved in Eq. (2.9) for $p_i = \mathrm{Tr}(\rho \Pi_i)$.

For more information on how quantum-state estimation can be improved even further, see [7].

2.5. De Finetti Theorem

One requirement for quantum tomography is to have a source that produces many identical copies of the state that we intend to measure. Realistically, even the most perfect source cannot produce perfect copies of the same state. One may well wonder why quantum tomography assigns just *one* density operator, instead of a distribution of states. A crucial role in this context is played by the de Finetti theorem ([11, 37]).

The de Finetti Theorem states that if one has an extendible permutation-invariant sequence of N quantum systems, then one can assign a quantum state of the form

$$\rho^{(N)} = \int d\rho P(\rho) \rho^{\otimes N} \tag{2.13}$$

to the collection of N systems, with $P(\rho)$ a probability density over density operators. Quantum tomography then succeeds in making the distribution $P(\rho)$ more and more narrow, sharply peaked around some ρ_0. In fact, in the limit of $N \to \infty$, one has $P(\rho) \to \delta(\rho - \rho_0)$. For this to work, we have to ensure that the two prerequisites, extendibility and permutation invariance, hold for the measurement records of the sequence of states that are being used for quantum tomography.

A state is permutation invariant, or symmetric, if any permutation of its subsystems does not change the overall state. Formally, this can be written as

$$\rho_{\Pi(1)} \otimes \rho_{\Pi(2)} \otimes \cdots \otimes \rho_{\Pi(N)} = \rho_1 \otimes \rho_2 \otimes \cdots \otimes \rho_N. \tag{2.14}$$

An example for such a system is a state of N qubits that are either all in state $|0\rangle$ or all in state $|1\rangle$. This state is known as the GHZ state, or Greenberger-Horne-Zeilinger

state [13], and is of the form

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}} \left(|00\ldots0\rangle + |11\ldots1\rangle \right). \tag{2.15}$$

Any permutation of the qubits will always result in the same state.

The other requirement is that the system is extendible, meaning that the sequence of states can be extended to an arbitrarily large symmetric sequence:

$$\mathrm{Tr}_{N+1,\ldots,N+m} \left[\rho_1 \otimes \ldots \rho_N \otimes \rho_{N+1} \ldots \rho_{N+m} \right] = \rho_1 \otimes \cdots \otimes \rho_N \tag{2.16}$$

where Tr_n is the partial trace over subsystem n. The GHZ state does *not* fulfill this requirement: If we choose a GHZ state of $N+1$ qubits, and trace out the last qubit, the resulting state is

$$\rho = \frac{1}{2} \left(|00\ldots0\rangle \langle 00\ldots0| + |11\ldots1\rangle \langle 11\ldots1| \right) \tag{2.17}$$

We see that all the coherence has been lost, and we are not arriving at the original state.

It is easy to see that the usual assumption of a sequence of identical copies

$$\rho = \rho_1 \otimes \rho_1 \otimes \cdots \otimes \rho_1 \tag{2.18}$$

fulfills both of the requirements of the de Finetti theorem. Note that one way to ensure a permutation invariant measurement record is to perform different measurements in a random order.

An important prerequisite for quantum state tomography is therefore that we have a reliable way to check whether a sequence of states fulfills the two requirements

of the de Finetti theorem. In the next chapter we will propose a very simple measurement that can perform such a test.

2.6. Scalability of Quantum State Estimation

Quantum tomography works well for small systems. In the case of single qubits, the density matrix has only three independent variables, so it is sufficient to experimentally determine the expectation values of three different measurement settings in order to reconstruct the entire state. Any useful quantum computer will require a much larger array of qubits though, and the number of elements in the density matrix increases exponentially as $4^N - 1$ in the number of qubits N. Quantum tomography very quickly becomes experimentally and numerically unfeasible.

In practice, most of the time we only consider a very small set of states where many parameters are taken to be zero and only few specific errors are assumed to occur. In that case we can use models of only a few variables to describe an arbitrarily large quantum state. For example, assume that our target state is supposed to be a perfect GHZ state $|\mathrm{GHZ}\rangle$. If we can accurately bring a system of many qubits in such a state, the density matrix will only have 4 non-zero elements, and determining the rest of the density matrix elements will not give us any more information. But of course we need to check if we indeed created such a perfect $|\mathrm{GHZ}\rangle$ state. In order to test if there are any other non-zero parameters in our quantum state, we might try an error model with only one error parameter q and model the state as a mixture of the target state and a maximally mixed state

$$\rho = q\,|\mathrm{GHZ}\rangle\langle\mathrm{GHZ}| + (1-q)\mathbb{1}/D, \tag{2.19}$$

where D is the dimension of the Hilbert space. But we cannot know if our system is indeed described well by such a simple state, or if we could find a better description that could more accurately describe the behaviour of our system. We will consider generalizations of this simple idea in Chapter V, and also provide a method that can compare two or more such models and predict which model will perform better.

CHAPTER III

DETECTING DRIFT OF QUANTUM SOURCES

This chapter was published as *Detecting the Drift of Quantum Sources: Not the de Finetti Theorem*, Phys. Rev. Lett. **106**, 180501 (2011). It was initiated by S. J. van Enk and finished jointly by Lucia Schwarz and S. J. van Enk.

3.1. Drifting Sources

We consider quantum state tomography in the case where the assumption of permutation invariance does not hold, and where, consequently, the de Finetti theorem does not apply. The most relevant case is that of a (slowly) drifting source. For example, it is well known that a laser displays phase diffusion: when one considers two light pulses emitted by the same laser with a short time delay τ between them, there will be a (random) phase difference whose average magnitude increases with τ. Of course, even in this case, one could average over all emitted light pulses, say N instances, to arrive at a single average density matrix. Indeed, if done correctly the averaging procedure restores the permutation invariance, but (i) the average density matrix depends on the number N, and (ii) the averaging procedure throws away potentially useful information. For example, if we are interested in the purity of our quantum states, the single state estimate will be too conservative.

In this section, we set ourselves the task of figuring out how one could detect whether (and how) a source is drifting.

In principle, for detecting drift one could still use a variant of quantum tomography: for example, we split our N quantum systems into two groups of size $N/2$ each: the first half (chronologically) and the second half. For each we estimate a single

density matrix: and if the difference between the two estimates is (not) statistically significant then we conclude our source is (not) likely drifting. This method works to some extent, but is still subject to the same two objections mentioned above. Moreover, it has been known for a few decades that for the detection of given physical quantities (such as a particular matrix element of the density matrix) a targeted method is always superior to performing full tomography [14]. Therefore, we propose and analyze a different method directly targeted at detecting drift. A difference with the above-mentioned method [14] is that we consider a quantity determined by *pairs* of density matrices.

3.2. A Two-state Measurement: The Swap Operator

Consider what one would measure to detect phase diffusion of a (pulsed) laser in the special (but relevant) case where one assumes the laser pulses can be described by coherent states with some fixed (and known) amplitude but a diffusing phase (relative to some phase standard). One would take pairs of the output laser pulses, and split them on a 50/50 beamsplitter in such a way that one particular output would be the vacuum if their phase difference, $\delta\phi$, would be zero. That output's intensity is then

$$I = |\alpha|^2 |1 - \exp(i\delta\phi)|^2/2, \tag{3.1}$$

if $|\alpha|$ is the amplitude of the laser pulses. Thus measuring this intensity determines the phase difference directly.

Now how do we generalize this measurement to arbitrary quantum systems (in particular, qubits)? We first note that the intensity I can also be written in terms of the *overlap* between the two input states, call them $\rho = |\alpha\rangle\langle\alpha|$ and $\rho' = |\alpha'\rangle\langle\alpha'|$,

since $\exp(-2I) = \mathrm{Tr}(\rho\rho')$ in this case. Thus, our choice of generalization will be to measure the overlap between pairs of instances of quantum systems from one and the same source. In other words, we propose to measure the swap operator $\hat{V}$, defined in terms of basis vectors $\{|i\rangle\}$ and $\{|j\rangle\}$ of the two (isomorphic) Hilbert spaces of two instances numbered m and n from our source by

$$\hat{V} = \sum_i \sum_j |i\rangle_m \langle j| \otimes |j\rangle_n \langle i| \,. \tag{3.2}$$

The expectation value of $\hat{V}$ equals the overlap

$$\mathrm{Tr}(\rho_m \otimes \rho_n \hat{V}) = \mathrm{Tr}(\rho_m \rho_n). \tag{3.3}$$

(Note the left-hand side contains the tensor product of two density operators, the right hand-side their matrix product.)

The operator $\hat{V}$ possesses $D(D+1)/2$ eigenvalues $+1$ corresponding to symmetric eigenstates

$$|i\rangle\,|i\rangle\,;\,[|i\rangle\,|j\rangle + |j\rangle\,|i\rangle]/\sqrt{2} \tag{3.4}$$

for $i \neq j$, and $D(D-1)/2$ eigenvalues -1 corresponding to antisymmetric eigenstates

$$[|i\rangle\,|j\rangle - |j\rangle\,|i\rangle]/\sqrt{2} \tag{3.5}$$

for $i \neq j$.

Each measurement of $\hat{V}$ yields one of its two eigenvalues, ± 1, and so only after multiple measurements will one obtain a statistical estimate of the overlap. [And, as a bonus, *if* the two density matrices are identical, then this measurement in fact measures the purity [19, 18], $P = \mathrm{Tr}(\rho_m^2)$.] By comparing the overlap between

adjacent copies, where $|m-n|=1$ with the overlap between outputs that are farther apart, $|m-n|>1$, we obtain information about whether the source is drifting: if the source is not drifting, the overlap is independent of $|m-n|$.

3.3. Characterizing Different Types of Drift

In order to infer more detailed information about the character of the drift or diffusion (beyond the mere statement that the source is or is not stationary), we need some simplifying assumptions about the sequence of states (the space of *all* possible output states of N copies is too large to be either measurable or tractable). Here we make the following two assumptions (one of which has been implicitly used already in the above description): (a) the states are independent, (b) the drifting process is Markovian, such that the overlap between two copies m and n only depends on $|m-n|$ (applicable to, e.g., laser phase diffusion, and testable by our measurement by monitoring V_k, defined in Eq. (3.20), over time). So we write the state of N systems produced by our quantum source as a tensor product [1]

$$\rho^{(N)} = \rho_N \otimes \rho_{N-1} \ldots \otimes \rho_1. \tag{3.6}$$

In this case, we get

$$V_{nm} \equiv \mathrm{Tr}[\rho_n \otimes \rho_m \hat{V}] = \frac{1}{2}[P_n + P_m] - \frac{1}{2}\mathrm{Tr}[\Delta_{mn}^2], \tag{3.7}$$

[1] It is an open question whether this form can be *derived* from *approximate* permutation invariance (plus extendability) of the sequence. The state of laser pulses emitted by a phase-diffusing laser can be written in this form, modulo two subtleties discussed in [62].

where $P_k = \mathrm{Tr}[(\rho_k)^2]$ is the purity of system k, and $\Delta_{mn} = \rho_m - \rho_n$. As special cases of nonstationary sources we consider both diffusion and systematic drift, modeled by

$$\rho_{n+1} = U_r \rho_n U_r^\dagger. \tag{3.8}$$

Diffusive drift occurs when U_r is a *random* unitary matrix, picked from some distribution; a systematic drift occurs when U_r is fixed. In either case, the purity of ρ_n is independent of n: $P_n = P_m \equiv P_1$. We can take stochastic averages over the random distribution of unitaries, which we will indicate by a bar, to get

$$\overline{V}_{nm} = P_1 - \frac{1}{2}\mathrm{Tr}\overline{[\Delta_{nm}^2]}. \tag{3.9}$$

In case the drift process is purely a systematic drift, each unitary U_r is the same, and we get

$$\mathrm{Tr}\overline{[\Delta_{nm}^2]} = |n-m|^2 D_1 \tag{3.10}$$

for some *drift constant* D_1.

If the process that changes the states ρ_n is diffusive (for example, the random distribution of U_r is a Gaussian centered around the identity), then we get a linear relationship between the overlap and the distance $|n-m|$,

$$\mathrm{Tr}\overline{[\Delta_{nm}^2]} = |n-m| D_2, \tag{3.11}$$

for some *diffusion constant* D_2 [2]. In this case, measuring the swap operator between neighboring copies, for which $|n-m| = 1$ and on copies with $|n-m| = 2$ gives us

[2]Equations (3.10) and (3.11) are informative only for states sufficiently far away from the completely mixed state, since the latter is invariant under (3.8), so that D_1 and D_2 are zero for that extreme case.

both the purity

$$P_1 = 2\mathrm{Tr}[\overline{\Delta^2_{n,n+1}}] - \mathrm{Tr}[\overline{\Delta^2_{n,n+2}}] \tag{3.12}$$

and the diffusion constant

$$D_2 = \mathrm{Tr}[\overline{\Delta^2_{n,n+2}}] - \mathrm{Tr}[\overline{\Delta^2_{n,n+1}}]. \tag{3.13}$$

(And similar relations hold when the drift is purely systematic.)

One way to check which sort of drift process one actually has, diffusive, systematic, or a combination thereof, is to measure in addition the quantity $\mathrm{Tr}[\overline{\Delta^2_{n,n+3}}]$, and then calculate the ratio

$$\alpha \equiv \frac{\mathrm{Tr}[\overline{\Delta^2_{n,n+2}}] - \mathrm{Tr}[\overline{\Delta^2_{n,n+1}}]}{\mathrm{Tr}[\overline{\Delta^2_{n,n+3}}] - \mathrm{Tr}[\overline{\Delta^2_{n,n+2}}]}. \tag{3.14}$$

If the ratio is 1, one has a purely diffusive process, if $\alpha = 3/5$ one has a systematic drift, and in all cases in between one has both diffusive and systematic drifts. To see how the number α is determined when there is a combination of systematic and diffusive drifts, let us consider the simplest case of a qubit source. We model the drift process with a unitary matrix $U_r = \exp(i\delta\vec{r}\cdot\vec{\sigma})$, with $\delta \ll 1$, $\vec{\sigma}$ a vector containing the three Pauli matrices, and a random vector $\vec{r}$ that consists of both a diffusive part and a systematic part,

$$\vec{r} = p\,\vec{r}_{\mathrm{const}} + (1-p)\,\vec{r}_{\mathrm{diffusive}} \tag{3.15}$$

with a normally distributed random vector $\vec{r}_{\mathrm{diffusive}}$ and a constant (unit) vector $\vec{r}_{\mathrm{const}}$ (and $0 \le p \le 1$). In this case α depends on p and on the ratio of the two constants

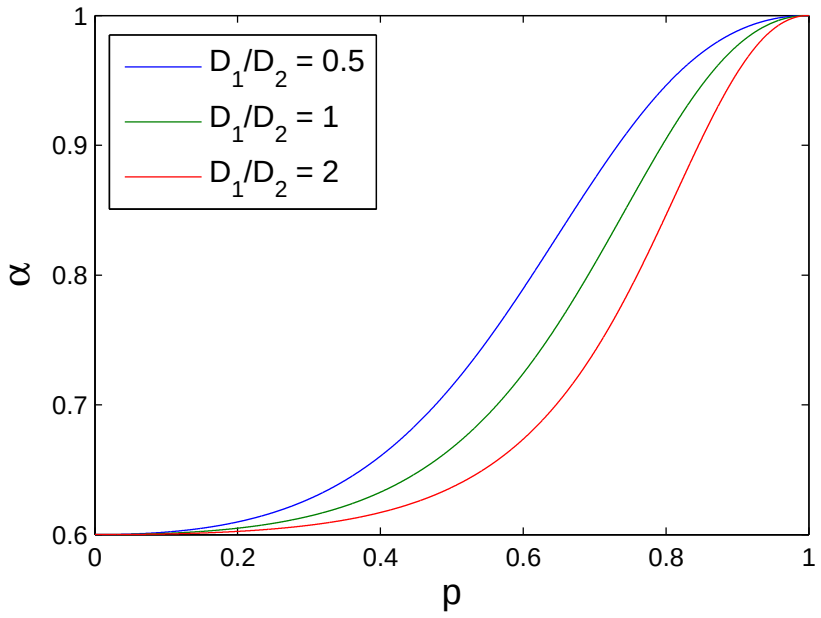

FIGURE 3.1. A simulation of the measureable ratio α (3.16) for a combination of systematic and diffusive drifts, for different drift constants.

D_1 and D_2, with $D_1 = \delta^2 \mathrm{Tr}[\vec{r}_{\mathrm{const}} \cdot \vec{\sigma}, \overline{\rho}]^2$ and $D_2 = \delta^2 \mathrm{Tr}\overline{[\vec{r}_{\mathrm{diffusive}} \cdot \vec{\sigma}, \rho]^2}$:

$$\alpha = \frac{-p^2 + \frac{D_1}{D_2}(-3p^2 + 6p - 3)}{-p^2 + \frac{D_1}{D_2}(-5p^2 + 10p - 5)}. \tag{3.16}$$

In Fig. 3.1, we plot this ratio for different values of D_1 and D_2.

3.4. How Many Measurements?

The next question we consider is, given a source of quantum states, how quickly can we determine (by measuring the swap operator) whether the source is drifting? Let us first consider the case of pure diffusive drift. We could, for example, measure the swap operator between states that are 1 step and 2 steps apart, respectively, and

see if the two numbers are equal or not. Suppose we write the overlaps as

$$\overline{\mathrm{Tr}(\rho_n\rho_{n+1})} = P_1 - D_2/2, \tag{3.17}$$

$$\overline{\mathrm{Tr}(\rho_n\rho_{n+2})} = P_1 - D_2. \tag{3.18}$$

$$\ldots \tag{3.19}$$

Suppose we have N data sets of the measurements of both $\mathrm{Tr}(\rho_n\rho_{n+1})$ and $\mathrm{Tr}(\rho_n\rho_{n+2})$. We get the measured frequencies $f_1^{\pm}$ and $f_2^{\pm}$, respectively, of the measurement outcomes ± 1 in the two cases. The average values are

$$V_k := \overline{\mathrm{Tr}(\rho_n\rho_{n+k})} = f_k^+ - f_k^-, \tag{3.20}$$

for $k = 1, 2$, and the standard error bars are (for large enough N) $\Delta V_k = 2\sqrt{f_k^+ f_k^- / N}$. To decide that the source is drifting, the values of V_1 and V_2 should *not* overlap within their error bars. The necessary condition for that is

$$\frac{1}{2}\Delta V_1 + \frac{1}{2}\Delta V_2 < V_1 - V_2 = D_2/2. \tag{3.21}$$

From assumption (3.17) and Eqns. (3.20), we can write the frequencies in terms of D_1 and P_1, and using all this in Eq. (3.21) we can solve for the minimum necessary number of measurements:

$$N_{\min} = [\frac{1}{D_2}(\quad \sqrt{(1 + P_1 - \tfrac{D_2}{2})(1 - P_1 + \tfrac{D_2}{2})} \tag{3.22}$$

$$+ \quad \sqrt{(1 + P_1 - D_2)(1 - P_1 + D_2)})]^2. \tag{3.23}$$

To detect a systematic drift, a similar calculation gives

$$N_{\min} = [\frac{1}{3D_1}(\quad \sqrt{(1+P_1-\frac{D_1}{2})(1-P_1+\frac{D_1}{2})} \tag{3.24}$$

$$+ \quad \sqrt{(1+P_1-2D_1)(1-P_1+2D_1)})]^2. \tag{3.25}$$

The number of measurements needed, for both diffusive and systematic drifts, is depicted in Fig. 3.2 for various values of P_1.

In principle, one could detect a drifting source a lot faster if one measured the swap operator on states that are $k > 2$ steps apart (in addition to measuring states 1 step apart), simply because $|V_k - V_1|$ will be larger. In an actual experiment, however, the larger the distance between two copies, the longer the earlier copy would have to be stored in memory. We could model the decoherence that the earlier copy undergoes as follows: assume that there is a typical decoherence time scale τ, which, e.g., drives any state towards the totally mixed state. That is, if we keep a system for time t, then

$$\rho \to (\exp(-t/\tau)\rho + (1 - \exp(-t/\tau)\mathbb{1}/D, \tag{3.26}$$

with D the dimension of the Hilbert space of our quantum system. Then assume that the time needed to produce one copy is $\epsilon\tau$ with some (hopefully small) number ϵ. Then we can write the overlap between states n and $n+k$ as

$$P_k \qquad = \mathrm{Tr}(\rho_n \tilde{\rho}_{n+k}) \tag{3.27}$$

$$= e^{-k\epsilon}\mathrm{Tr}(\rho_n \rho_{n+k}) + \tfrac{1}{D}(1 - e^{-k\epsilon})\mathrm{Tr}(\mathbb{1}\rho_{n+k}). \tag{3.28}$$

The inferred overlap between copies n and $n+k$ follows from the measured P_k by multiplying it with $\exp(k\epsilon)$ (and subtracting a known quantity): so the error bar in

the overlap multiplies by the same number. This error thus becomes substantial once $k\epsilon$ becomes of order unity, so that is where we would expect the method to use copies a distance k apart to break down. In Fig. 3.3 we plot the number of measurements needed for various values of ϵ as a function of k, and we can indeed see that for too large values of k, the required number of measurements increases exponentially with k. The optimal k, of course, depends on the specific decoherence process, and also on how fast the source is drifting, but seems to be around $k\epsilon \approx 4$ in our example.

3.5. Experimental Implementation of the Swap Operator

Finally, we wish to note that in the case of two independent single photons, when they are viewed as quantum systems with an infinite-dimensional Hilbert space describing polarization, spectral, and transverse spatial degrees of freedom, the swap operator can in fact be measured via the Hong-Ou-Mandel interference effect [30]. This can be shown as follows. We consider two single photon wavepackets impinging on two different input ports (denoted A and B) of a 50/50 beamsplitter. We write the two (mixed) input states in terms of creation and annihilation operators $a^\dagger$ and a (for port A) and $b^\dagger$ and b (for port B) as:

$$\rho_A = \textstyle\sum_{kl} p_{kl} a_k^\dagger \left|0\right\rangle \left\langle 0\right| a_l \tag{3.29}$$

$$\rho_B = \textstyle\sum_{nm} q_{nm} b_n^\dagger \left|0\right\rangle \left\langle 0\right| b_m, \tag{3.30}$$

where the subscripts stand for the mode properties (polarization, frequency etc.) other than their propagation direction. The combined input state is then $\rho_{\text{in}} =$

$\rho_A \otimes \rho_B$. This state gets transformed by the 50/50 beamsplitter in the following way:

$$\rho_{\text{out}} = \sum_{klnm} \frac{p_{kl} q_{nm}}{4} (c_k^\dagger + i d_k^\dagger)(i c_n^\dagger + d_n^\dagger) \left|0\right\rangle \left\langle 0\right| (c_l - i d_l)(-i c_m + d_m), \tag{3.31}$$

where c and d now denote operators of the two output ports C and D. To get the probability P_{cc} of getting a coincidence count, i.e., photo detections at both output ports C and D, we take a partial trace and we get:

$$P_{cc} = \sum_{rs} \left\langle 1_r \right|_c \left\langle 1_s \right|_d \rho_{\text{out}} \left|1_r\right\rangle_c \left|1_s\right\rangle_d = \sum_{rs} \left\langle 0\right| c_r d_s \rho_{\text{out}} c_r^\dagger d_s^\dagger \left|0\right\rangle \tag{3.32}$$

This simplifies to

$$P_{cc} = \frac{1}{2} \sum_k p_{kk} \sum_n q_{nn} - \frac{1}{2} \sum_{kl} p_{kl} q_{lk}. \tag{3.33}$$

The first two sums are the traces of the density matrices and therefore equal 1. It is easy to see that

$$\text{Tr}(\rho_A \rho_B) = \sum_{klnm} p_{kl} q_{nm} \delta_{nl} \delta_{mk} = \sum_{kl} p_{kl} q_{lk}, \tag{3.34}$$

so that we get the simple relation

$$2P_{cc} = 1 - \text{Tr}(\rho_A \rho_B). \tag{3.35}$$

Thus, as announced, the HOM effect measures the overlap between two input states, and hence the swap operator. (And so, if the two single-photon input states are identical, then the HOM interference measurement measures the purity of the input states. Note that this is different from the measurement of single-photon (spectral) purity implemented recently in Ref. [10], which also makes use of the HOM effect, but with a known coherent-state input in the other input port.) Of course, the HOM

effect has been measured many times in the context of characterizing single-photon sources (see, e.g., [49, 39]), but never, as far as we know, systematically on copies more than the minimum distance apart. We also note that for the polarization degree of single photons, the overlap has been measured [29], following ideas from [19].

In conclusion, we proposed the measurement of the swap operator as a means to detect the drifting of a quantum source. This measurement complements quantum tomography, which produces an estimate of a *single* average density matrix, by partially characterizing how this estimate would change over time, for instance, distinguishing between diffusive and systematic drifts. We also analyzed how many measurements are needed to determine that a source is drifting, including the influence of decoherence on the precise measurement strategy. We showed the swap measurement on pairs of single-photon wavepackets is implemented simply by the Hong-Ou-Mandel effect.

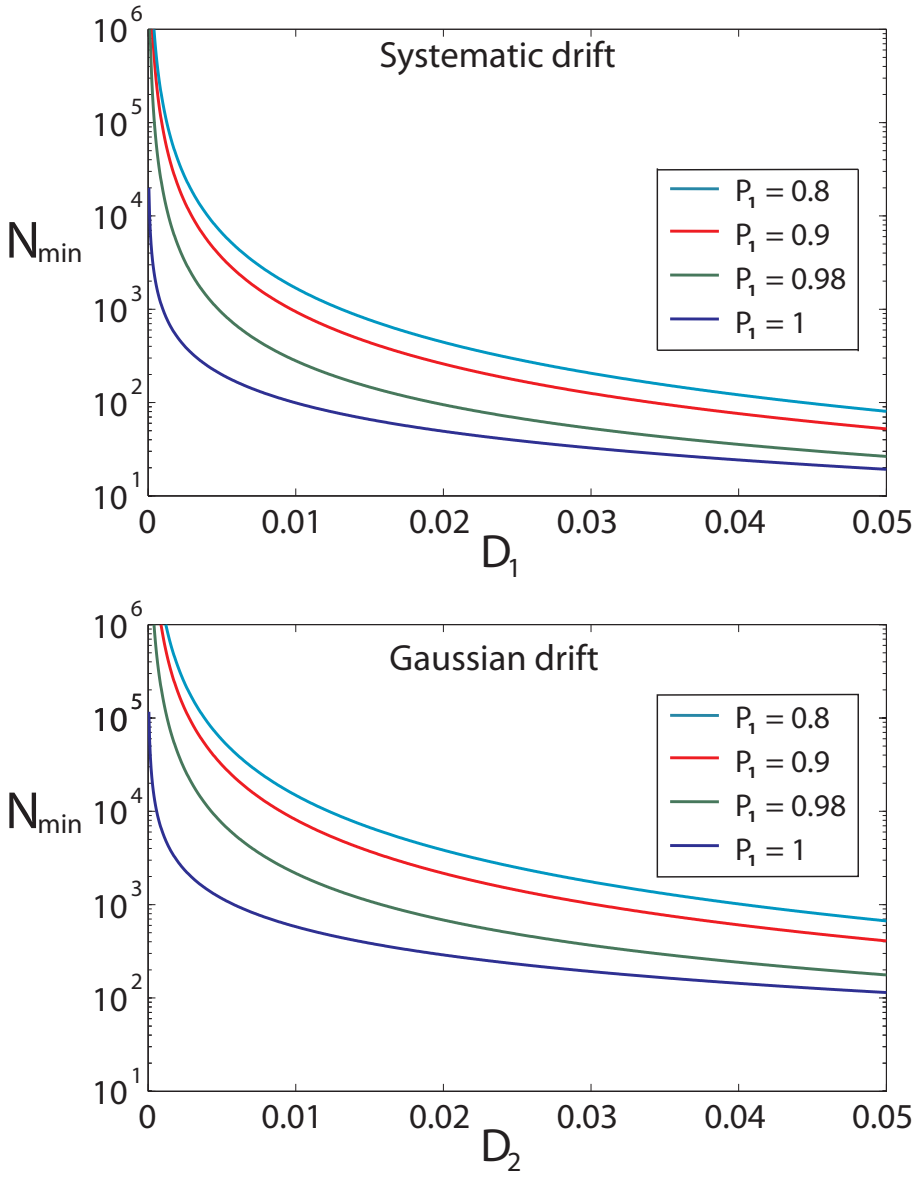

FIGURE 3.2. How many measurements do we need to figure out that a source is drifting? Obviously, the larger the drift is (as measured by the parameter D_2 for diffusive drift or D_1 for systematic drift), the fewer measurements we need. Top: systematic drift, bottom: diffusive drift.

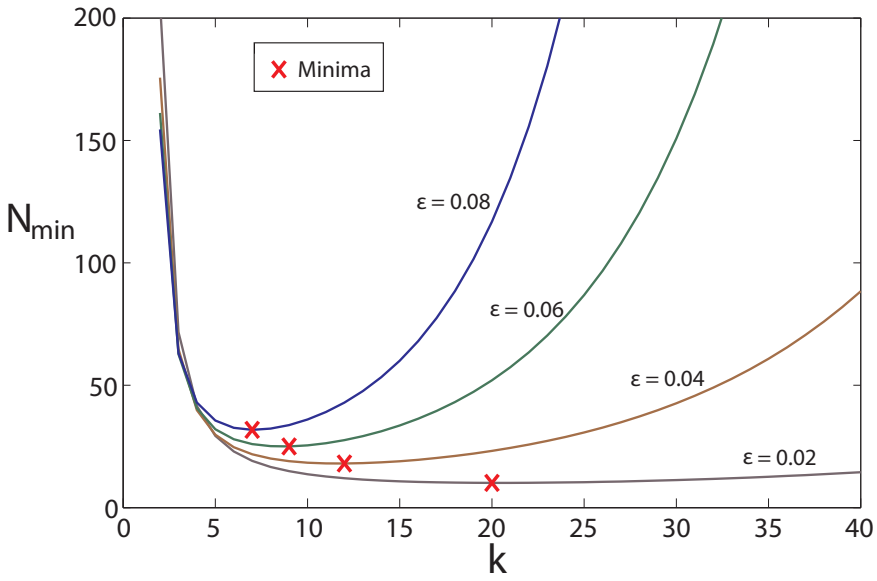

FIGURE 3.3. This plot shows the minimal number of measurements needed to detect drift, as a function of the (temporal) distance between states measured, k, for various values of the decoherence parameter ϵ, for $P_1 = 1$, $D = 2$ and $D_2 = 0.01$ (see main text for definitions). The minimum of each curve determines the optimal distance k between states to be measured (in addition to distance-1 overlaps).

CHAPTER IV

FAULT TOLERANT QUANTUM COMPUTING: WHICH ERRORS ARE OK?

4.1. Errors in Classical versus Quantum Computing

In a classical computer, errors can be corrected using redundancy. The computation can be performed in parallel and the correct result is determined simply by the 'majority vote'. This is a very simple but terribly wasteful procedure, since the bit error rate in our laptops and PCs is typically so low, that error correction is not needed in any standard computer.

This simple sort of error correction is not easily translated to a quantum computer though, since the nature of quantum states makes qubits very different from classical bits. Without correcting errors, a quantum computer cannot perform any nontrivial calculation reliably. The reason for this is that qubits are not digital, and they can occupy a continuum of states, namely all the superpositions of their two basis states $|0\rangle$ and $|1\rangle$. This makes the computation very susceptible to errors, since even small deviations from the original quantum state will change the result, and will build up in subsequent computational steps.

Error correction in quantum computing is not straightforward for several more reasons. Most importantly, the no-cloning theorem [64] prevents us from using a simple repetition code for error correction. Since each qubit is typically represented by just one or a few elementary particles, we cannot make several copies of a state and repeat the calculation multiple times. Additionally, any measurement on a quantum state changes the state itself, so much care has to be taken to not destroy the information that a state carries.

In this section we give a brief overview of the methods that have been developed to not only detect, but also correct an error without destroying the coherence of the state.

4.2. Decoherence

A quantum system in which the qubits are designed to interact with each other to form quantum gates, and with our measurement apparatus to read out the result, invariably also interacts with its environment. This process can be described as the quantum state becoming entangled with its environment through unitary interactions, and then being 'measured', which changes the state. This process is called "Decoherence" and is the main challenge in quantum computing [59].

In general, any error on a quantum state can be described if we introduce some extra qubits from the environment, which interact with our state via a unitary transformation on the combined state, and then discard the extra qubits. This non-unitary evolution can be written in terms of the so-called "Kraus representation" as

$$\rho \to \sum_i E_i \rho E_i^\dagger \tag{4.1}$$

where the Kraus operators E_i sum up to the identity on the Hilbert space of our state such that $\sum_i E_i E_i^\dagger = \mathbb{1}$. For example, for a single qubit state, consider the Kraus operators $E_0 = \sqrt{1-\frac{q}{2}}\mathbb{1}$ (which represents the possibility of no error occuring) and $E_1 = \sqrt{\frac{q}{2}}\sigma_z$ (which represents a phase flip error). If we apply this nonunitary evolution to a pure state of the general form $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the density matrix

will evolve according to

$$\rho = \begin{pmatrix} |\alpha|^2 & \alpha^*\beta \\ \alpha\beta^* & |\beta|^2 \end{pmatrix} \rightarrow \begin{pmatrix} |\alpha|^2 & (1-q)\alpha^*\beta \\ (1-q)\alpha\beta^* & |\beta|^2 \end{pmatrix}. \tag{4.2}$$

This type of decoherence is called dephasing, a process in which the coherence (represented by the off-diagonal elements in the density matrix) between qubits decays and the system transitions from quantum mechanical behavior to classical behavior. Other effects of decoherence include depolarizing, where a pure state evolves towards the fully mixed state, and dissipation, where the populations of quantum states are changing.

4.3. Error Models

We can use the quantum mechanical principle, that any measurement projects a state onto an eigenstate of the measurement operator, to our advantage. If our system has a small (analog) error, we can devise a measurement operator that either projects the state onto its original state with no error, or onto a state where a well-defined error occured. In that way we can turn analog errors into digital (discrete) errors that we can correct.

In a qubit, we can describe these discrete errors as a phase flip, a bit flip, or a combination of both. These actions can be categorized by the Pauli spin operators (see table 4.1). Any errror correction scheme that corrects these errors also corrects any linear combination, and since these three operators span the space of operators on the Hilbert space, we can correct any single-qubit error [23].

If we have N qubits, the state space of our density matrix has $2^N \times 2^N - 1$ real parameters. If we also want to consider quantum processes, that means we

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \text{bit flip} \qquad |0\rangle \Rightarrow |1\rangle,\ |1\rangle \Rightarrow |0\rangle$$

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \text{bit- \& phase flip} \qquad |0\rangle \Rightarrow |1\rangle,\ |1\rangle \Rightarrow -|0\rangle$$

$$\sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{phase flip} \qquad |0\rangle \Rightarrow |0\rangle,\ |1\rangle \Rightarrow -|1\rangle$$

FIGURE 4.1. Three independent 1-qubit errors and their action on the basis states. Any error on 1 qubit can be expressed as a linear combination of these.

would have to correct approximately $4^N \times 4^N$ linearly independent errors. Since it is unlikely (but not impossible) that a large number of errors occur simultaneously, usually not all possible errors are corrected. For example, if we want to consider only one- and two-qubit errors, there would be $\binom{N}{1} + \binom{N}{2}$ possible errors that can occur in the N-qubit density matrix, which is only polynomial in the number of qubits instead of exponential, and consequently is much more practical to implement.

4.4. Error Correction Codes

In classical error correction, we use redundancy to correct errors. For example, in a 3 bit classical code, we use three physical bits to encode one logical bit. The two possible states are encoded as $1_L \equiv 111$ and $0_L \equiv 000$. After each calculation, we can check the state of each physical qubit and determine the state of our logical qubit as the bit state that occurs most often. If only one error has occured, for example we might find our code in state 001, then this code will correct the error and flip the last bit back to 0. Of course not all errors can be caught: If two or three bits are flipped, then the error correction will go wrong. If the probability for each bit error p_I is assumed to be independent, then the probability that this code will produce an

error is

$$p_{error} = 1 - (1 - p_I)^3 - 3p_I(1 - p_I)^2 \simeq 3p_I^2, \tag{4.3}$$

so for small $p_I \ll 1$ the reliability of the code is improved dramatically. This kind of error correction becomes useful only if p_I is low enough so that the error probability will actually decrease if the error correction is applied, here this will be approximately the case when $3p_I^2 < p_I$ and therefore the threshold error rate is given by $p_I < 1/3$.

For a quantum system, we might try to use the same principle and encode a state $|\Psi\rangle$ as $|\Psi\rangle \otimes |\Psi\rangle \otimes |\Psi\rangle$. However, this is impossible due to the no-cloning theorem, we can not create copies of a qubit. Peter Shor discovered that it is still possible to spread the information of one qubit into a block of several qubits. For example, for his 9-qubit code [51], the basis states are encoded in three groups of three qubits each:

$$|0\rangle \Rightarrow |\bar{0}\rangle = (|000\rangle + |111\rangle)^{\otimes 3} \tag{4.4}$$

$$|1\rangle \Rightarrow |\bar{1}\rangle = (|000\rangle - |111\rangle)^{\otimes 3} \tag{4.5}$$

Note that this does not violate the no cloning theorem, since for any state $|\Psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$ the encoded state is

$$\left|\overline{\Psi}\right\rangle = \alpha\,|\bar{0}\rangle + \beta\,|\bar{1}\rangle \neq [\alpha(|000\rangle + |111\rangle) + \beta(|000\rangle - |111\rangle)]^{\otimes 3} \tag{4.6}$$

This code has two independent error correction mechanisms. Firstly, any bit flip error can be corrected by taking the majority of each group of three qubits. Secondly, phase errors can be corrected by taking the majority of the three signs in each group.

We can then perform measurements that only extract information about the errors, but not about the state itself. Making use of the projective nature of quantum mechanics, the measurements are chosen such that the state is projected either back into the correct state, or into a state where a definite error has occured, which can then be corrected by applying the inverse operation of the error.

This error correction code can only correct one single qubit error. For example, if we assume that each qubit has an error rate of p_I, then we will get an actual error in our logical qubit only if two or more errors appear in the encoded block, and the error probability will be reduced to $p_{error} \approx C p_I^2$. The factor C depends on the encoding scheme, for example for the nine qubit code we would have $C \approx \binom{9}{2} = 36$, and therefore the threshold error rate below which this code becomes useful is $p_{th} \approx 1/36$.

Many other error correcting codes have been proposed, including codes that can correct more than one errors [34, 55]. There are also lower bounds on the number of qubits that are necessary, which can be derived by counting the number of errors that need to be corrected. For example, to encode a single logical qubit into N physical qubits, each of the 2 basis states ($|0_L\rangle$ and $|1_L\rangle$ can be unchanged, or can be affected by $3N$ linearly independent errors. In total, this adds up to $2 + 2 \cdot 3N$ dimensions which need to fit into the N qubit Hilbert space, and this leads to the *quantum Hamming bound* quantum hamming bound: To encode one qubit, and correct up to one errors, the number of physical qubits needs to fulfill

$$2 + 2 \cdot 3N \leq 2^N. \tag{4.7}$$

In our case we can see that we need at least 5 qubits to encode one logical qubit.

Error correction codes can be classified by the number of qubits that are used to encode one qubit, and by the number of errors they can detect. In order to correct

more errors, it is often useful to concatenate codes, so each encoded qubit is again encoded in a block of qubits. In that case we can reduce the error twice:

$$p_{error} \Rightarrow Cp_I^2 \Rightarrow C(Cp_I^2)^2 = C^3 p_I^4. \tag{4.8}$$

However using more qubits also makes the circuit more susceptible for errors, which poses a limit on the number of layers for encoding.

In general, if we have L layers of encoding, the probability for an error scales like p^{2^L}. Since this is a double exponential in L, for a calculation with T steps typically only $\log(\log(T))$ layers are needed. However, since the number of qubits scales up exponentially with the number of encoding layers, we are still left with $\propto \log(T)$ qubits per block.

Note that this implicitly assumes that the single qubit errors are independent, and do not account for the possiblity that errors on two or more qubits might occur that are spatially correlated.

4.5. Fault Tolerant Quantum Computing and the Threshold Theorem

Implementing error correction schemes does not guarantee that a quantum computation can be performed correctly. The gates used for encoding, detecting and correcting errors may be faulty themselves, and we need to be careful not to correct errors that are falsely detected. For instance, if we detect a bit flip error, we could check again and only correct the error if we find the same bit flip again. Of course as the computation time increases, the probability of errors occuring will also increase.

An error correction scheme is called fault tolerant if it can guarantee that a computation can be performed arbitrarily exact in polylogarithmic time. The first protocol for fault tolerant quantum computing was suggested by Shor [52], since then there have been many improvements [1, 32].

Since any error-correcting code can only detect and correct a certain number of errors, we need to ensure that the total number of errors within a block stays contained. A single error on a qubit can easily propagate and spread to other qubits during a quantum computation. A fault tolerant quantum gate therefore needs to ensure that a single error does not spread. There are many different implementations for state preparation, measurement, error correction, and a universal set of gates, that operate in a fault tolerant way. fault tolerant quantum computing.

Fault tolerant computing assumes a constant error rate per gate, and a (not necessarily identical) constant error per timestep for storing a qubit. As long as this error rate is below a certain threshold p_{th}, the errors will be corrected faster than they appear, and arbitrary long quantum computations can be achieved. Derivations of this error rate tend to vary depending on how conservative their assumptions are. Theoretical proofs have estimated a lower bound for this threshold error rate at $p_{th} \approx 10^{-5}$ to 10^{-3} [4], and the best current experiments typically achieve comparable error rates.

4.6. Which Errors Are Ok?

We have seen that quantum error correction codes are quite efficient in correcting stochastic errors that occur in single qubits, and more advanced codes can even correct independent errors on two or more qubits. In the framework of fault tolerant quantum computation, as long as the error rate is below a certain threshold, any calculation

can be performed safely based on these assumptions. However, we can't be certain that all errors occur independently. For example, we could have spatially correlated errors that affect multiple qubits (or even the entire state) at the same time. If too many qubits are affected by an error, then the error correction will fail and the error can spread uncontrollably. In the next chapter, we will propose a method to model and detect such general errors that affect the entire quantum state of a register.

CHAPTER V

PERMUTATIONALLY INVARIANT STATE RECONSTRUCTION

This chapter was published as *Error models in quantum computation: An application of model selection*, Phys. Rev. A **88**, 032318 (2013). It was initiated by S. J. van Enk and finished jointly by Lucia Schwarz and S. J. van Enk.

5.1. Introduction

In order to develop a quantum computer we need to be able to coherently control and read out a system of many qubits. Verifying how a particular experimental implementation of a quantum computer actually performs will be straightforward once we can run a computation in a fault tolerant manner: we just check whether the answer produced by the computation is correct or not. But before that time arrives we will need to employ other, less conclusive types of tests.

There are two types of generic tests that provide useful information about many-qubit systems: multi-partite entanglement verification tests [26] and randomized benchmarking [36, 42]. However, the information gained is somewhat unspecific: In both cases one may detect that something is wrong, but one will not find out what exactly is wrong. Unfortunately, there is no efficient procedure to figure out what exactly is wrong, simply because we cannot efficiently simulate a generic multi-qubit quantum process on a classical computer. (For smaller systems quantum tomography can be used, and even there one has to be careful with systematic and other errors [45, 38, 61].)

For fault tolerant quantum computation [22] one does need to know not just how large the error probabilities are, but also whether they are of the right type.

This is because threshold theorems [1, 35, 33, 56, 20] need to make explicit use of error models. For example, the calculation of the error threshold may be based on a "local stochastic" error model (for an introduction, see [23]). Errors correlated over a long range may then be disastrous. One mechanism by which such long-range errors might arise is as follows. A laser field's phase and intensity always fluctuate, but, of course, if those fluctuations are always sufficiently small, the errors they cause will be corrected for by quantum error correction. But what if the fluctuations, for just a brief time interval, are large? Then all qubits which happened to have been accessed during that time interval have a much larger probability of error. The problem we consider is how one could notice the presence of such errors.

While there is no systematic and efficient method to solve this problem completely, there is an efficient and well-tested method: model selection [12, 9]. This term refers to a well-developed field of (classical) statistics and inference where the aim is to rank different (statistical) models, each meant to describe some given process. In the present context model selection can be summarized as follows: We design a few-parameter model that describes our predictions of all the processes and errors that occur in our experiment —it may have a few noise parameters with a clear physical meaning, for instance— and compare it with a much larger (but still far from exhaustive) model that includes many (but not all) possible types of errors. As long as the large model contains a number of parameters that scales moderately with the number of qubits, then it still can be analyzed, even for a few dozen qubits. If that large model is ranked higher than our few-parameter model, we conclude that errors occurred that we did not expect.

We are going to discuss an illustrative example of this model selection procedure. We simulate a quantum state estimation experiment on N qubits, which is modeled

after an actual experiment performed on 14 ions in an ion trap in which a 14-qubit GHZ state of high fidelity was generated [43]. We will vary N up to 25 and assume the goal is to generate a perfect GHZ state. We take a 3-parameter model (with three noise parameters describing three different noise processes) as our standard error model and then take a model with $\mathcal{O}(N^3)$ parameters as the much larger error model, which includes many types of errors, although, obviously, not all $\mathcal{O}(4^N)$ possible ones. We assume the data are generated by a "true" state of the form

$$\rho_{true} = (1-q)\rho_{\text{s.e.}} + q\rho_{\text{g.e.}},$$

with the subscript s.e. referring to "standard error model" and "g.e." to the more general error model. We investigate then the following issues: First, does the model selection procedure recognize that the standard error model is indeed correct (i.e., ranked higher than the large general error model) when $q = 0$? Second, in the case that $q \neq 0$, how many measurements does one need to take before one notices that there is in fact an error that lies outside the standard error model? The last question splits naturally into two subquestions, namely, how that number scales with the number of qubits and how it scales with q.

The model selection method we use here is based on the Akaike Information Criterion (AIC) [3]. This method is widely used outside of physics, and by now has been applied on various occasions within quantum information theory as well [60, 40, 65, 27, 61]. Most model selection criteria compare the goodness of fit of each model while penalizing the number of parameters, thus possibly favoring simpler models. The AIC in particular has a clear meaning since it is derived purely from the principles of information (see appendix). It has been found to perform better than

the related Bayesian Information Criterion [9] in quantum state and entanglement estimation [40].

5.2. Preliminaries

5.2.1. Model selection and AIC

Suppose we have taken data and now wish to model the underlying process that generated the data. Our data contains some amount of information about the underlying process, but also statistical fluctuations. How can we determine whether a model is a good description of the underlying process rather than of the statistical fluctuations? In general, models with more parameters will be fitting the data better but are also more likely to fit to the fluctuations, and models with *too many* parameters are *over*fitting. One method to find a compromise between under- and overfitting was proposed by Akaike [3]. He derived an expression for the estimated Kullback-Leibler divergence between one's model and the true underlying process. The Kullback-Leibler divergence is expressed in terms of two probability distributions for the data, the "true distribution" $\{p_i\}$, and the distribution generated by our model, $\{s_i\}$, as follows:

$$KL(p||s) = \sum_i p_i \log \frac{p_i}{s_i}. \tag{5.1}$$

This is a measure for the distance between the two probability distributions $\{p_i\}$ and $\{s_i\}$. It is also called the *relative entropy* and can be understood as the information that is lost if the model $\{s_i\}$ is used instead of the "real" distribution $\{p_i\}$.

Of course, we do not know the true underlying distribution, but, nonetheless, the Kullback-Leibler divergence can be estimated, as was shown by Akaike, using the observed frequencies. Namely, up to a constant that is the same for all models, he

found the divergence to approximately equal

$$AIC = -2\mathcal{L}_{\max} + 2K. \tag{5.2}$$

Here K is the number of parameters of the model, and $\mathcal{L}_{\max}$ is its maximum log-likelihood,

$$\mathcal{L}_{\max} = \max_{\{p_k\}} \sum_k f_k \log p_k, \tag{5.3}$$

with f_k the number of times outcome k was observed and p_k the probability according to the model of obtaining outcome k. Model selection now consists of calculating the AIC for different candidate models, with the lowest score corresponding to the best model. In our context this procedure can distinguish between models that accurately describe the relevant physical (error) processes, and models that spend too many parameters on fitting statistical noise. We thereby gain insights into the actual physical processes that cause errors, and we can tell whether or not errors outside our simple model are significant.

5.2.2. A 3-parameter model for noisy GHZ states

We will simulate an experiment on a noisy GHZ state [24] of N qubits. The ideal GHZ state is a coherent superposition of all qubits in state $|0\rangle$ or all in state $|1\rangle$,

$$|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{2}} \left(|00...0\rangle + |11...1\rangle \right). \tag{5.4}$$

A high fidelity version of this state was created in a trapped ion system for 14 ions [43]. The density matrix $\rho_{\mathrm{GHZ}} = |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$, written in the standard basis $|00...0\rangle, |00...1\rangle, ... |11...1\rangle$, has only four nonzero elements which all equal $\frac{1}{2}$. This

state is maximally entangled and pure. However, a real quantum system in the lab will not be in this perfect state. There might be several effects that act on the qubits during the state preparation and/or storage. As a simple and not unreasonable model we assume just three noise processes, described by three parameters: a small imbalance ε between the populations of $|00..0\rangle$ and $|11..1\rangle$, a systematic phase shift φ of the relative phase between $|00..0\rangle$ and $|11..1\rangle$, and δ which quantifies the loss of coherence between $|00..0\rangle$ and $|11..1\rangle$ due to random phase fluctuations. These three processes will create a mixed state with density matrix

$$\rho_{3P} = \frac{1}{2} \begin{pmatrix} 1+\varepsilon & \dots & \delta\sqrt{1-\varepsilon^2}e^{i\varphi} \\ \vdots & & \vdots \\ \delta\sqrt{1-\varepsilon^2}e^{-i\varphi} & \dots & 1-\varepsilon \end{pmatrix}. \tag{5.5}$$

5.2.3. A large model: permutationally invariant states

For the comparison with our 3-parameter model, we try to find a model that will describe many possible errors and deviations from this simple model. However, we can't model *all* possible errors that may occur. Our goal is therefore to design a model with a fairly large (but still polynomial) number of fitting parameters. If an arbitrary error is affecting our experiment, it will most likely be partially contained in this large model and we will detect it. Of course, this leaves out certain errors that are exactly orthogonal to the large model. To increase the chance of detecting small errors, for an actual experiment several such fairly large models should be considered and compared to each other. There is no systematic way to find good models for this purpose, but any model with a large number of parameters can be used. For simplicity, we only regard one such model in this paper. This suffices for our purpose of determining how many measurements are needed as a function of both N and q.

The ideal GHZ state is permutationally invariant, in the sense that any permutation of its subsystems leaves the overall state unchanged. Mathematically, this can be expressed as

$$\rho_{\mathrm{GHZ}} = \frac{1}{N!} \sum_{\pi_k \in S_N} V(\pi_k) \rho_{\mathrm{GHZ}} V(\pi_k)^\dagger, \tag{5.6}$$

where the sum is over all $N!$ permutations π_k of the N qubits, and $V(\pi_k)$ is the unitary representation of the operator that permutes the subsystems according to the permutation π_k. Since our simple 3-parameter model is also permutationally invariant, it makes sense to use as the large model the set of *all* permutationally invariant (PI) states. This set has been shown to be very convenient for quantum state reconstruction and entanglement detection [58, 57, 44]. Many experiments aim at generating GHZ states, W states or Dicke states, all of which are PI. Moreover, if the PI part of a state ρ is entangled, then so is ρ.

(Note that this choice does not imply that we think the actual state is permutationally invariant, and nor does it imply that we think the error process is permutationally invariant. All that matters is that our model will include the permutationally invariant part of the actual error process. As long as that part does not vanish, we will detect it. Recall that we *cannot* analyze all possible error models!)

As shown in [44], any permutationally invariant state can be represented as a block-diagonal matrix

$$\rho_{\mathrm{PI}} = \bigoplus_{j=j_{min}}^{N/2} P_j \rho_j \otimes \frac{\mathbb{1}}{K_j}, \tag{5.7}$$

where ρ_j is a spin-j matrix of dimension $2j+1$ and $\{P_j\}$ is a probability distribution over the spin values j, and K_j is the dimension of the non-PI part of the spin-j state,

given by

$$K_j = \binom{N}{N/2 - j} - \binom{N}{N/2 - j - 1}. \tag{5.8}$$

The dimension of the permutationally invariant subspace grows as $\propto N^3$ with the number of qubits N. This model fits our purposes very well. For a dozen or more qubits the model contains a substantial number of parameters, but not so many that we cannot analyze it.

It was shown in Refs. [57, 44] that the necessary and sufficient number of different measurements needed to gain full information about a permutationally invariant state is

$$D_N = \binom{N+2}{N}. \tag{5.9}$$

In particular, we can choose to measure observables of the form $\hat{A}^{\otimes N}$, that is, we can measure the same single-qubit observable on each qubit. We just have to pick D_N different single-qubit observables, the outcomes of which ought to be more or less uniformly distributed on the Bloch sphere [57, 44]. In principle, we can choose any set of random, linearly independent projective measurement operators, but it is advantageous to use an evenly distributed set of measurement operators, analogous to SIC-POVMs [48]. We achieved this by minimizing the frame potential

$$F = \sum_{j,k} |\langle \Psi_j | \Psi_k \rangle|^2, \tag{5.10}$$

where each measurement is a projective measurement such that $\hat{A}_i = |\Psi_i\rangle \langle \Psi_i|$.

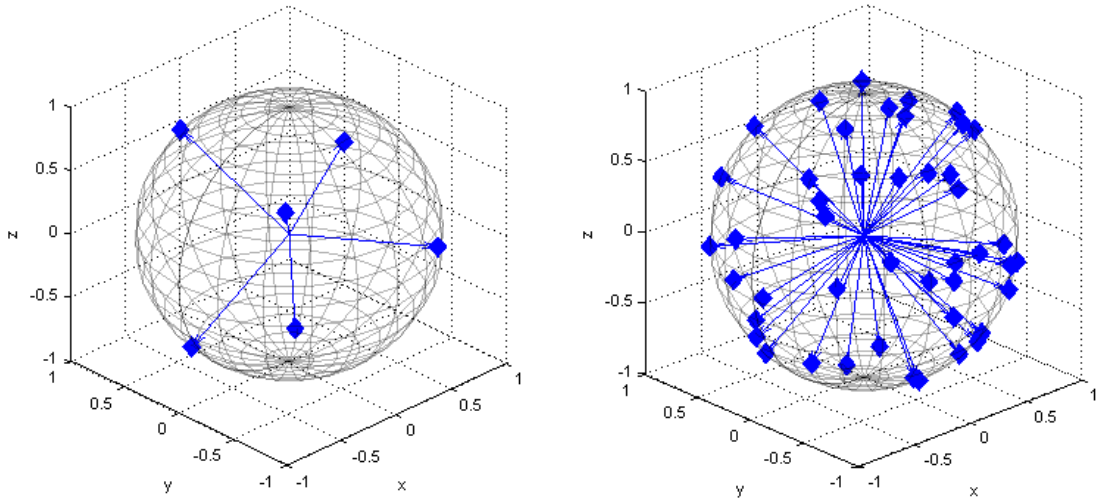

FIGURE 5.1. Uniformly distributed measurement settings on the Bloch sphere for 2- and 8-qubit states

It was shown [57] that the necessary and sufficient number of different measurements to gain full information about a permutationally invariant state is

$$D_N = \binom{N+2}{N} = \mathcal{O}(N^2). \tag{5.11}$$

In Fig. 5.1, we show two examples of a set of D_N measurement operators, represented by their Bloch vectors.

5.3. Numerical Results

We simulate an experiment on N qubits. The "true" state that generates the data is chosen to be an unequal mixture of a noisy GHZ state ρ_{3P} (contained in the 3-parameter model) and a randomly picked permutationally invariant state ρ_{PI} orthogonal to the 3-parameter states (just to make sure the overlap of the actual state

with the 3-parameter subspace does not vary with N). We write

$$\rho_{true} = (1-q)\,\rho_{3P} + q\,\rho_{\mathrm{PI}}. \tag{5.12}$$

The parameter q determines the probability of "wrong" types of errors, namely, those outside our standard error model. We simulate a certain number of measurements,

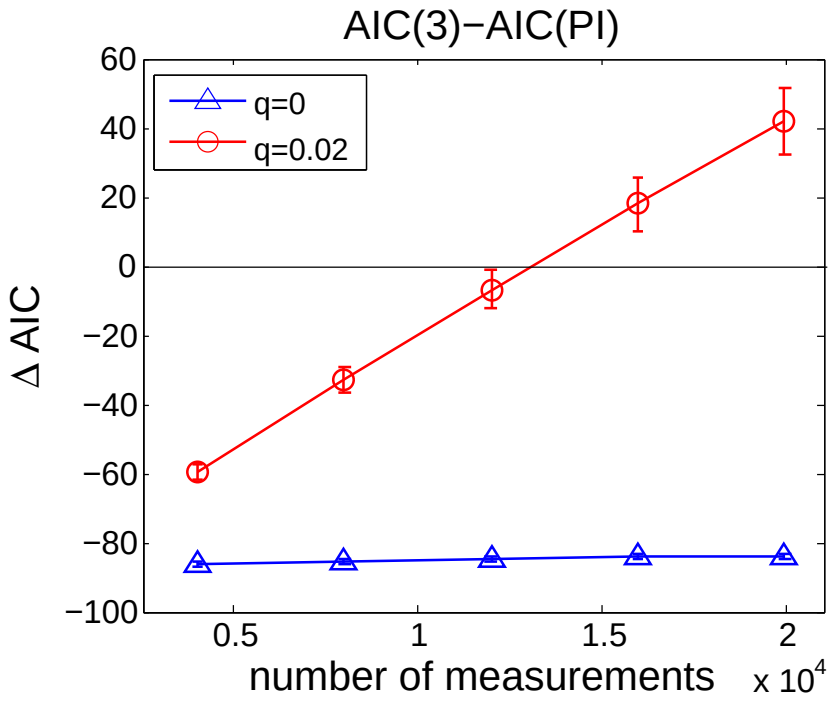

FIGURE 5.2. The differences in AIC values ΔAIC for a state of $N = 5$ qubits plotted against the total number of measurements. There are 21 measurements settings in this case, and the PI model contains 55 parameters. The simulation was run 100 times and the average ΔAIC is plotted. Error bars refer to the spread of ΔAIC over the 100 runs.

where each single measurement consists of measuring N times the same single-qubit observable, where the latter is chosen from the set of D_N single-qubit observables. So a single measurement yields N outcomes 0 or 1. We assume for simplicity that each of the D_N observables is measured the same number of times. We then find numerically the maximum likelihood state for the three-parameter model as well as for the large

PI model. This is easy for the three-parameter model since the minimization is over just 3 parameters. For the PI model we apply an iterative algorithm described in [44] for which the required computation time increases only polynomially in the number of qubits. Using the two maximum likelihoods thus obtained, we can calculate the respective AIC values for the 3P and PI models and plot the difference, which we denote by ΔAIC. Negative values of ΔAIC correspond to the 3-parameter model being favored, whereas positive values indicate that the PI model is better.

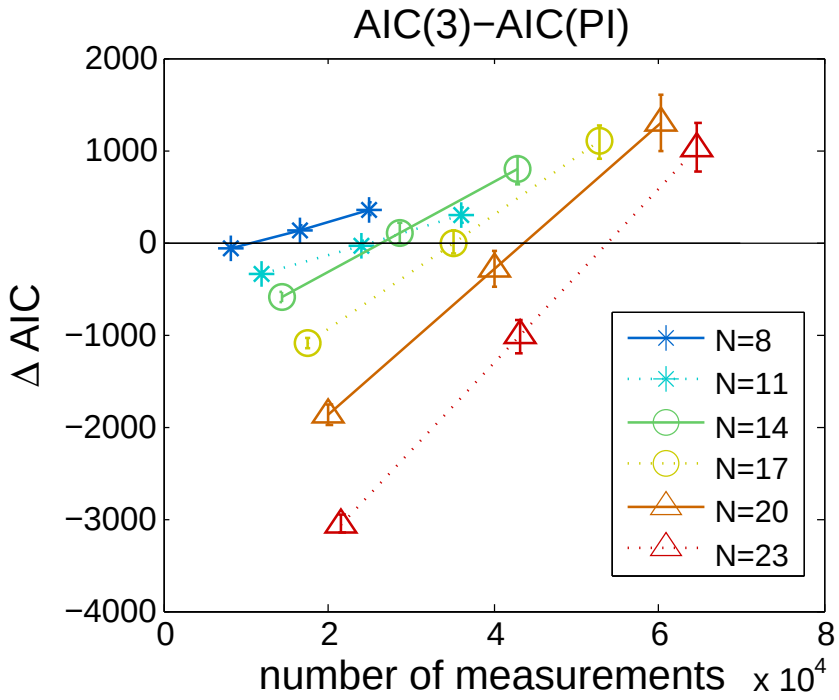

FIGURE 5.3. The differences in AIC for several different numbers of qubits, plotted against the total number of measurements. Note that a single measurement on N qubits yields N binary outcomes. For very small numbers of measurements ΔAIC approaches twice the difference in the number of parameters of the two models ($\approx N^3/3$). In this plot we used $q = 0.02$.

Fig. 5.2 shows ΔAIC for two different "true" states, one with $q = 0$, the other with $q = 0.02$. For $q = 0$ the Akaike Information Criterion correctly always prefers the 3-parameter model. This is not as trivial (since the data are generated from a 3-

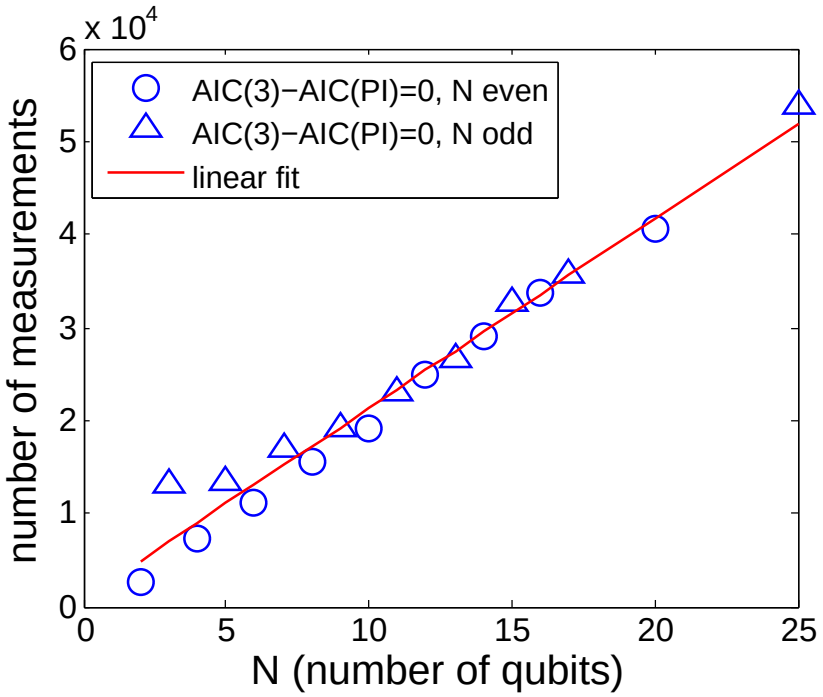

FIGURE 5.4. The average number of measurements required to reach the point where both models are rated equally. Small even and odd numbers of qubits behave slightly differently. ($q = 0.02$)

parameter state!) as it may seem, because the statistical fluctuations are substantial (note that each observable is measured just a few dozen times for the smallest total number of measurements in the plot). For $q = 0.02$ we see that a relatively small number of measurements suffices to start favoring the PI model over the 3-parameter model, and the more measurements one performs, the firmer that conclusion gets. For very small numbers of measurements, a nonzero q cannot be detected yet, and we may interpret the point where ΔAIC crosses zero as the point where sufficiently many measurements have been taken to detect the presence of errors outside our standard (three-parameter) error model.

Let us consider how that crossing point changes with the number of qubits. A range of results for different N is plotted in Fig. 5.3. With increasing N the crossing

point clearly moves towards larger numbers of measurements. We plot the crossing point as a function of N in Fig. 5.4. We see that the necessary total number of measurements to detect a fixed perturbation q increases only linearly in the number of qubits N, which shows that this can be measured very efficiently. (The number of single-qubit measurements needed grows as N^2.)

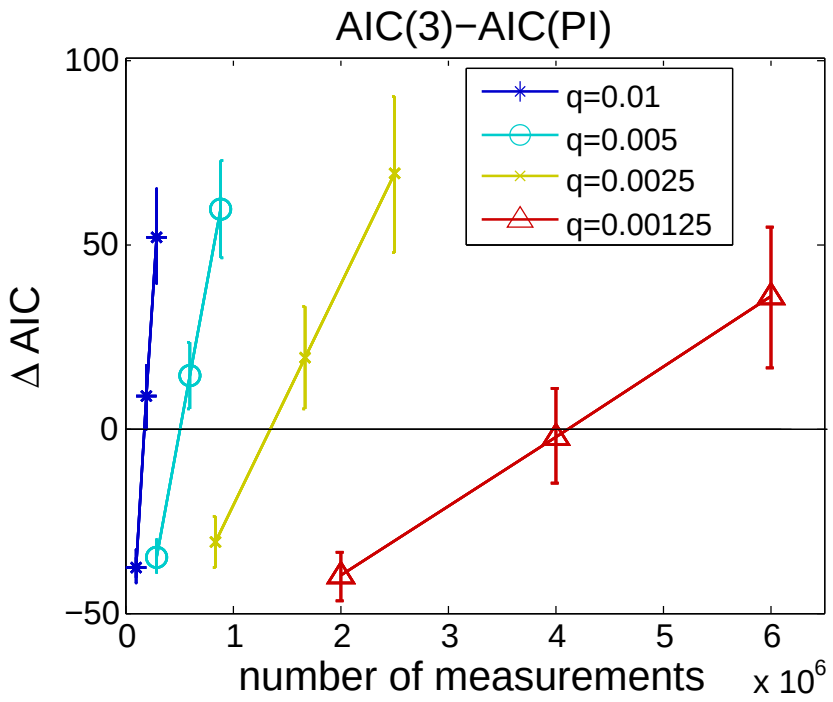

FIGURE 5.5. ΔAIC as a function of the number of measurements performed, for four different values of q and $N = 5$ qubits.

It is also useful to investigate how the number of measurements needed to detect a nonzero value of q depends on that value.

The plots of Fig. 5.5 and Fig. 5.6 show that the total number of measurements needed increases only moderately with $1/q$. This dependence becomes more favorable with increasing N, presumably because there are more ways to detect errors that occur with a given probability.

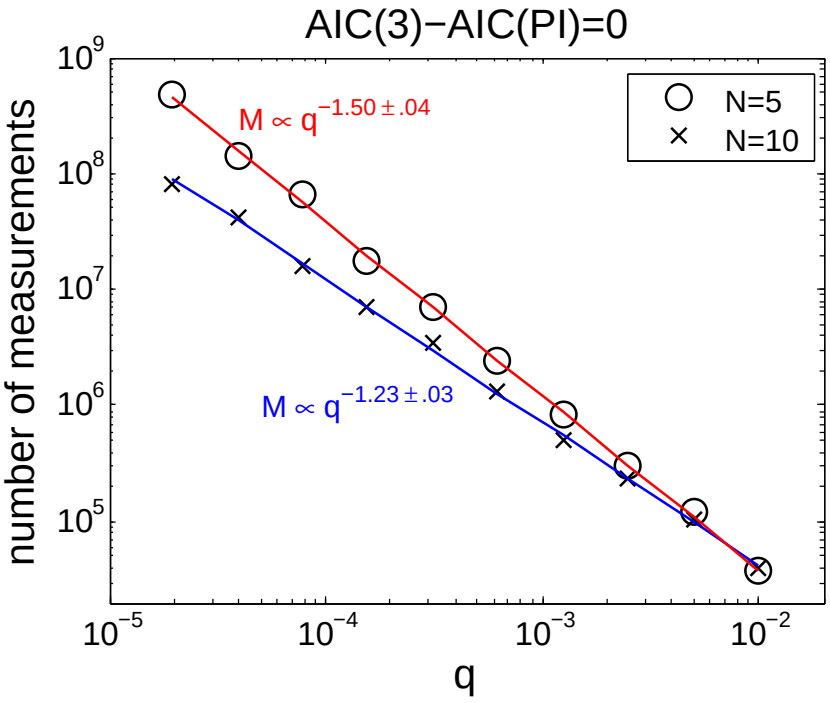

FIGURE 5.6. The minimum number of measurements M needed to detect a perturbation of strength q, both for $N = 5$ and for $N = 10$ qubits.

5.4. Conclusions

We showed by example how to use the Akaike Information Criterion (AIC) to select between different error models in the context of quantum computing. Thanks to the AIC one does not need exponentially many parameters to describe an experiment on multiple qubits. Instead, we compared a small model (with 3 parameters) with an intermediate-sized model ($\mathcal{O}(N^3)$ parameters). The former stands for a standard error model in the context of fault tolerant quantum computing, the larger model stands for other (undesired) types of errors. Since it is crucial to know whether one's implementation satisfies the condition for the fault tolerance error threshold theorems to apply, our method, which works for dozens of qubits, should be quite useful here. In our specific example the number of (unentangled) N-qubit measurements needed to detect errors of the wrong type turned out to scale linearly with the number of

qubits and less than quadratically with the inverse of the wrong error probability. The latter scaling even improves with increasing number of qubits.

CHAPTER VI

CONCLUSIONS

Inspired by recent advances in experimental quantum computation, in this thesis we have asked the question of how to detect errors in our quantum processor, especially the types of errors that have so far been ignored.

We considered two cases. In the first part of this work, we challenged the basis of quantum state tomography: The assumption that the source of quantum states that we are analyzing, produces a sequence of identical states. In more technical terms, the so-called De Finetti theorem requires the sequence of states to be extendible and permutation-invariant in order for us to be able to describe the sequence as a product of identical density matrices. This should not at all be taken for granted, since any experiment will suffer from fluctuations. Therefore we investigated the question of how noise and fluctuations will influence quantum tomography, and we found relations between the amplitudes of the fluctuations and the number of measurements it will take to detect their presence. Additionally we also looked into how different types of fluctuations could affect an experiment differently, and described a method to distinguish between systematic and diffusive drift. The measurement that we proposed is described by the swap operator, which acts on pairs of states and provides a measure of how much difference there is between those two state. In general, the implementation of this operator depends on the physical representation of states. We provided an implementation for the case of single photons, in which case this measurement would simply be a Hong-Ou-Mandel interference on a beam splitter. This measurement should be taken as an addition to standard quantum tomography and as a systematic test of whether the results of tomography are reliable.

In the second part of this work, we focused on quantum state estimation of large systems. In quantum computing experiments, the largest systems so far consist of on the order of 15 qubits but this number is soon expected to increase. Typically, error models for quantum computers assume that only uncorrelated single-qubit and two-qubit errors occur, which we know can be corrected reliably. We instead assume that *any* possible error should be considered. Since the state space grows exponentially in the number of qubits, a complete analysis is not feasible for any but the smallest systems. As a solution, we propose the use of a statistical tool, the Akaike Information Criterion, to select between different error models and choose the model that is probably closest to the truth. This still includes making educated guesses about the errors that can occur, but is a reliable and scalable way to compare different models with varying numbers of parameters.

We numerically simulated experiments of up to 25 qubits and compared two different models, including one that includes many errors that go beyond the standard error model in the context of fault tolerant quantum computing. Our method provides a good indication of whether error correction can be implemented successfully, which is crucial to know for any type of quantum computing. We found in our example that the number of measurements that are necessary to detect errors of the wrong type scaled only linearly with the number of qubits, and less than quadratically with the inverse of the wrong error probability. This scaling is very favorable and indicates that this type of test can be used during quantum state estimation of large systems, or even for reliable quantum process tomography which requires an even larger parameter space.

APPENDIX

DERIVATION OF THE AKAIKE INFORMATION CRITERION

A.1. Prerequisites

A.1.1. Likelihood

The task of model selection is to select a model for a system that is closest to the truth. The only way we can gather information about this system is by making experimental observations, and therefore model selection relies heavily on measurement data.

Assume we have a set of measurement data $\mathbf{x} = \{x_i\}$, $i = 1..M$ out of M observations, where each x_i denotes one single measurement outcome, and we want to use a specific model to explain the outcomes. This model might have one or more parameters (e.g. if the model is a gaussian distribution, the parameters could be the mean and variance) that we denote by the vector $\boldsymbol{\theta}$. We can calculate the probability to get our measurement outcome $\mathbf{x}$ under the assumption that our model is in fact correct, as $g(\mathbf{x}|\boldsymbol{\theta})$. Often times the parameters $\boldsymbol{\theta}$ of the model are unknown. The *likelihood function* is defined as the probability to get a certain measurement outcome $\mathbf{x}$, given that the model parameters $\boldsymbol{\theta}$ are true:

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{x}) = g(\mathbf{x}|\boldsymbol{\theta}). \tag{A.1}$$

Note that this is as a function of $\boldsymbol{\theta}$ and the measurement outcomes are fixed, and therefore the likelihood function is not a probability distribution. In many cases it is

convenient to consider the logarithm of the likelihood

$$\ell(\boldsymbol{\theta}|\mathbf{x}) = \log \mathcal{L}(\boldsymbol{\theta}|\mathbf{x}) = \log g(\mathbf{x}|\boldsymbol{\theta}) \tag{A.2}$$

which is a strictly negative function. The empirical value of $\boldsymbol{\theta}$ that maximizes $\mathcal{L}$ (or, equivalently, ℓ, since the logarithm is a monotonous function) for a given set of measurements $\mathbf{x}$ is denoted by $\hat{\boldsymbol{\theta}}_M$:

$$\ell(\hat{\boldsymbol{\theta}}_M|\mathbf{x}) = \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta}|\mathbf{x}) \tag{A.3}$$

A.1.2. Models

In the calculation of the likelihood, it is assumed that the underlying model for the data is known and only the parameters of the model need to be estimated. How can we know though if a model is the true model, or even a reasonably good approximation to the true model? In some cases, we can't even know if there exists a mathematical model that completely describes the process.

In general, models with more parameters will be fitting the data better and thus have a higher maximum likelihood. However, there is always the danger of overfitting, especially if the number of parameters of a model is close to the order of magnitude of the sample size.

Most importantly, we have to ask ourselves the question as to what we expect from a "good" model. Is it the model that is closest to the "true" model? Is it the model that gives us the best understanding of the underlying processes? Or is it the model that best predicts future observations best?

A different way to look at it is the concept that the measurement data contain some amount of *information* about the underlying process, but also some *noise*. The task of the model is to extract this information and translate it into a more compact or understandable form. [9]

A.1.3. The Kullback-Leibler divergence

The *Kullback-Leibler divergence* provides a way to compare two probability distributions. It is defined as

$$\mathrm{KL}(p||g) = \sum_i p_i \log \frac{p_i}{q_i} \tag{A.4}$$

for discrete random variables and can be extended to

$$\mathrm{KL}(p||g) = \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x \tag{A.5}$$

for continuous distributions.

Although the Kullback-Leibler divergence is not symmetric, and therefore not a metric, it measures a distance between two probability distributions. It is strictly non-negative and only equal to zero if the two distributions are identical. Akaike [2] proposed to use this information-theoretical quantity to estimate the goodness of models, i.e. the distance between a model $g(x|\boldsymbol{\theta})$ and the underlying *true* distribution of the data $p(x)$

$$\mathrm{KL}(p||g) = \int p(x) \log \frac{p(x)}{g(x|\boldsymbol{\theta})} \mathrm{d}x = \mathbb{E}_p \log \frac{p(x)}{g(x|\boldsymbol{\theta})}, \tag{A.6}$$

where in the last step we wrote the integral as an expectation value over the true distribution p.

A.1.4. Estimating the Kullback-Leibler divergence

Since the true distribution $p(x)$ is not known, we can only find an estimate for the Kullback-Leibler divergence of a certain model $g(x|\boldsymbol{\theta})$. We do not assume that $p(x)$ is contained in $g(x|\boldsymbol{\theta})$, but we denote the (unknown) model that is closest to $p(x)$ (in the sense that it minimizes KL$(p||g)$)by it's parameter $\boldsymbol{\theta}_0$. Writing the KL-divergence as

$$\mathrm{KL}(p||g) = \int p(x) \log p(x) \mathrm{d}x - \int p(x) \log g(\mathbf{x}|\boldsymbol{\theta}_0) \mathrm{d}x \tag{A.7}$$

we can see that the first term is just a constant, so it is sufficient to only calculate the second term. Using the data $\mathbf{x} = \{x_i\}$, and the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_M(\mathbf{x})$, we can estimate the second term as

$$\int p(x) \log g(\mathbf{x}|\boldsymbol{\theta}_0) dx \approx \int p(x) \log g(\mathbf{x}|\hat{\boldsymbol{\theta}}_M) \mathrm{d}x = \mathbb{E}_x \log g(\mathbf{x}|\hat{\boldsymbol{\theta}}_M(\mathbf{x})), \tag{A.8}$$

where the expectation value over the true distribution is approximated by an expectation value over the measurement data. However, this estimate will be biased because we use the same set of data for both determining the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ and taking the expectation value. To get an unbiased estimate, we really should have two different, independent sets of data $\mathbf{x}$ and $\mathbf{y}$. Then we could find the unbiased estimate

$$T = \mathbb{E}_x \mathbb{E}_y \log g(\mathbf{x}|\hat{\boldsymbol{\theta}}_M(\mathbf{y})) \tag{A.9}$$

In the next section, we will outline how this estimate can be found.

A.2. Derivation of the AIC

A.2.1. Preliminaries

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_M$, which is found by using the data to maximize the likelihood function, is an estimator for the unique, but unknown parameter $\boldsymbol{\theta}_0$ which minimizes the KL-divergence for this particular model w.r.t. the truth. $\hat{\boldsymbol{\theta}}$ will converge to $\boldsymbol{\theta}_0$ for large M. One can show that

$$\sqrt{M}(\hat{\boldsymbol{\theta}}_M - \boldsymbol{\theta}_0) \to \mathcal{N}(0, \Sigma) \tag{A.10}$$

where the variance-covariance matrix Σ is defined by

$$\Sigma = \mathbb{E}_y(\hat{\boldsymbol{\theta}}(y) - \boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}(y) - \boldsymbol{\theta}_0)^T \tag{A.11}$$

Note that, if the true model p is contained within $g(\mathbf{x}|\boldsymbol{\theta})$, i.e. $p(x) = g(\mathbf{x}|\boldsymbol{\theta}_0)$, then this matrix is equal to the inverse of the Fischer Information Matrix

$$\mathcal{I}(\boldsymbol{\theta}_0) = \mathbb{E}_g \left[\frac{\partial^2 \log g(x|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \tag{A.12}$$

We get another useful relation by noting that $\boldsymbol{\theta}_0$ minimizes the (unknown) KL-divergence, and therefore

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left[\int p(x) \log \left(\frac{p(x)}{g(x|\boldsymbol{\theta})} \right) dx \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0 \tag{A.13}$$

Taking into account that $p(x)$ doesn't depend on $\boldsymbol{\theta}$, and interchanging the derivative and integral, we get

$$\frac{\partial}{\partial \boldsymbol{\theta}} \left[\int p(x) \log g(x|\boldsymbol{\theta}) dx \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbb{E}_x \left[\frac{\partial}{\partial \boldsymbol{\theta}} \log g(x|\boldsymbol{\theta}) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0 \tag{A.14}$$

A.2.2. Estimation of the bias

We want to maximize the *expected* log-likelihood (A.9) where the two expectation values are independent and both with respect to the truth. In the following we will calculate two Taylor-expansions in $\boldsymbol{\theta}$, one around the true value $\boldsymbol{\theta}_0$ and another one around the maximum-likelihood value $\hat{\boldsymbol{\theta}}$.

We start with the integrand in (A.9), and expand it to second order:

$$\log g(x|\hat{\boldsymbol{\theta}}) \approx \log g(x|\boldsymbol{\theta}_0) + \left[\frac{\partial \log g(x|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right]^T [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] + \frac{1}{2} [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0]^T \left[\frac{\partial^2 \log g(x|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^2} \right] [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] \tag{A.15}$$

Taking the expectation value w.r.t. x, the linear term goes to zero because of (A.14), and we write the matrix in the quadratic term as

$$-\mathbb{E}_x \left[\frac{\partial^2 \log g(x|\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^2} \right] = I(\boldsymbol{\theta}_0) \tag{A.16}$$

We then take the second expectation value and get

$$E_y E_x \log g(x|\hat{\boldsymbol{\theta}}) \approx E_x \log g(x|\boldsymbol{\theta}_0) - \mathbb{E}_y \left[\frac{1}{2} [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0]^T I(\boldsymbol{\theta}_0) [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0] \right] \tag{A.17}$$

and we can rewrite the second term as

$$\mathbb{E}_y\left[\frac{1}{2}[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0]^T I(\boldsymbol{\theta}_0)[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0]\right]=\frac{1}{2}\mathrm{Tr}\left[I(\boldsymbol{\theta}_0)\mathbb{E}_y[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0]^T\right]=\frac{1}{2}\mathrm{Tr}\left[I(\boldsymbol{\theta}_0)\Sigma\right] \tag{A.18}$$

Next, we expand the first of (A.17) term as a Taylor series:

$$\log g(x|\boldsymbol{\theta}_0)\approx\log g(x|\hat{\boldsymbol{\theta}}(x))+\left[\frac{\partial\log g(x|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right]^T[\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}]+\frac{1}{2}[\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}]^T\left[\frac{\partial^2\log g(x|\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}^2}\right][\boldsymbol{\theta}_0-\hat{\boldsymbol{\theta}}] \tag{A.19}$$

. Note that here, the expected value for $\boldsymbol{\theta}$ is taken with respect to the distribution of x. The linear term vanishes, because $\hat{\boldsymbol{\theta}}$ is by definition the value that minimizes $\log g(x|\boldsymbol{\theta})$. Taking the needed expectation value, we get

$$\mathbb{E}_x\log g(x|\boldsymbol{\theta}_0)\approx\mathbb{E}_x\log g(x|\hat{\boldsymbol{\theta}})-\frac{1}{2}\mathrm{Tr}\left[\mathbb{E}_x(I(\hat{\boldsymbol{\theta}})\Sigma\right] \tag{A.20}$$

Using $I(\boldsymbol{\theta}_0)\approx I(\hat{\boldsymbol{\theta}})$, our final estimate for the expected KL-divergence is

$$T=\mathbb{E}_x\log g(x|\hat{\boldsymbol{\theta}}(x))-\mathrm{Tr}[I(\boldsymbol{\theta}_0)\Sigma] \tag{A.21}$$

If p is a subset of g, then it can be shown that $\mathrm{Tr}[I(\boldsymbol{\theta}_0)\Sigma]=K$, where K is the number of parameters, i.e. the length of $\boldsymbol{\theta}$. Even if p is not contained, but close to g, this is still a good approximation.

Akaike multiplied this estimator by -2, in order to be comparable to the χ^2 fit, defining his AIC as

$$AIC=-2\log g(x|\hat{\boldsymbol{\theta}}(x))+2K \tag{A.22}$$

REFERENCES CITED

[1] Aharonov, D. and M. Ben-Or (1997). Fault-tolerant quantum computation with constant error. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, STOC '97, New York, NY, USA, pp. 176–188. ACM.

[2] Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control 19*(6), 716–723.

[3] Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*, pp. 199–213. Springer.

[4] Aliferis, P., D. Gottesman, and J. Preskill (2006, March). Quantum accuracy threshold for concatenated distance-3 codes. *Quantum Info. Comput. 6*(2), 97–165.

[5] Barends, R., J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, and B. Campbell (2014). Superconducting quantum circuits at the surface code threshold for fault tolerance.

[6] Bell, J. S. (1964). On the Einstein Podolsky Rosen paradox. *Physics 1*(3), 195–200.

[7] Blume-Kohout, R. (2010). Optimal, reliable estimation of quantum states. *New J. Phys. 12*, 043034.

[8] Boixo, S., T. F. Rnnow, S. V. Isakov, Z. Wang, D. Wecker, D. A. Lidar, J. M. Martinis, and M. Troyer (2014, February). Evidence for quantum annealing with more than one hundred qubits. *Nature Physics 10*(3), 218–224.

[9] Burnham, K. P. and D. R. Anderson (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Verlag.

[10] Cassemiro, K. N., K. Laiho, and C. Silberhorn (2010). Accessing the purity of a single photon by the width of the hong-ou-mandel interference. *arXiv:1007.25999*.

[11] Caves, C., C. Fuchs, and R. Schack (2002). Unknown quantum states: The quantum de finetti representation. *J. Math. Phys. 43*(9), 4537–4559.

[12] Claeskens, G. and N. L. Hjort (1993). *Model selection and model averaging*. Cambridge University Press.

[13] D. M. Greenberger, M. A. Horne, A. Z. (1989). Going beyond bell's theorem. In K. M. (Ed.), *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, pp. 69–72. Kluwer.

[14] D'Ariano, G. M., U. Leonhardt, and H. Paul (1995, Sep). Homodyne detection of the density matrix of the radiation field. *Phys. Rev. A 52*(3), R1801–R1804.

[15] Devoret, M. H. and R. J. Schoelkopf (2013). Superconducting circuits for quantum information: An outlook. *Science 339*(6124), 1169–1174.

[16] Einstein, A., B. Podolsky, and N. Rosen (1935, May). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev. 47*, 777–780.

[17] Ekert, A. K. (1991, Aug). Quantum cryptography based on bell's theorem. *Phys. Rev. Lett. 67*, 661–663.

[18] Ekert, A. K., C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek (2002, May). Direct estimations of linear and nonlinear functionals of a quantum state. *Phys. Rev. Lett. 88*(21), 217901.

[19] Filip, R. (2002). Overlap and entanglement-witness measurements. *Phys. Rev. A 65*(6), 62320.

[20] Fowler, A. G. (2012). Proof of finite surface code threshold for matching. *Phys. Rev. Lett. 109*(18), 180502.

[21] Fuchs, C. A. (2002, May). Quantum Mechanics as Quantum Information (and only a little more). *arXiv:quant-ph/0205039*.

[22] Gottesman, D. (1998). Theory of fault-tolerant quantum computation. *Physical Review A 57*(1), 127.

[23] Gottesman, D. (2009). An introduction to quantum error correction and fault-tolerant quantum computation. In *Quantum Information Science and Its Contributions to Mathematics, Proceedings of Symposia in Applied Mathematics*, Volume 68, Providence, Rhode Island, pp. 13. Amer. Math. Soc.

[24] Greenberger, D. M., M. A. Horne, A. Shimony, and A. Zeilinger (1990). Bell's theorem without inequalities. *American Journal of Physics 58*(12), 1131–1143.

[25] Grover, L. K. (1996). A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, STOC '96, New York, NY, USA, pp. 212–219. ACM.

[26] Gühne, O. and G. Tóth (2009). Entanglement detection. *Physics Reports 474*(1), 1–75.

[27] Guţă, M., T. Kypraios, and I. Dryden (2012). Rank-based model selection for multiple ions quantum tomography. *New Journal of Physics 14*(10), 105002.

[28] Häffner, H., W. Hänsel, C. F. Roos, J. Benhelm, D. C. al kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, and P. O. Schmidt (2005). Scalable multiparticle entanglement of trapped ions.

[29] Hendrych, M., M. Duek, R. Filip, and J. Fiurek (2003). Simple optical measurement of the overlap and fidelity of quantum states. *Phys. Lett. A 310*(2-3), 95–100.

[30] Hong, C. K., Z. Y. Ou, and L. Mandel (1987, Nov). Measurement of subpicosecond time intervals between two photons by interference. *Phys. Rev. Lett. 59*(18), 2044–2046.

[31] Kadowaki, T. and H. Nishimori (1998, Nov). Quantum annealing in the transverse ising model. *Phys. Rev. E 58*, 5355–5363.

[32] Kitaev, A. (2003). Fault-tolerant quantum computation by anyons. *Annals of Physics 303*(1), 2 – 30.

[33] Kitaev, A. Y. (1997). Quantum error correction with imperfect gates. In *Quantum Communication, Computing, and Measurement*, pp. 181–188. Springer.

[34] Knill, E., R. Laflamme, and L. Viola (2000, Mar). Theory of quantum error correction for general noise. *Phys. Rev. Lett. 84*, 2525–2528.

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

[36] Knill, E., D. Leibfried, R. Reichle, J. Britton, R. Blakestad, J. Jost, C. Langer, R. Ozeri, S. Seidelin, and D. Wineland (2008). Randomized benchmarking of quantum gates. *Physical Review A 77*(1), 012307.

[37] König, R. and R. Renner (2005). A de Finetti representation for finite symmetric quantum states. *J. Math. Phys 46*(122108).

[38] Langford, N. K. (2013). Errors in quantum tomography: diagnosing systematic versus statistical errors. *New Journal of Physics 15*(3), 035003.

[39] Legero, T., T. Wilk, A. Kuhn, and G. Rempe (2006). Characterization of single photons using two-photon interference. *Advances In Atomic, Molecular, and Optical Physics 53*, 253–289.

[40] Lougovski, P. and S. J. van Enk (2009). Characterizing entanglement sources. *Physical Review A 80*(5), 052324.

[41] Lvovsky, A. and M. Raymer (2009). Continuous-variable optical quantum-state tomography. *Rev. Mod. Phys. 81*(1), 299–332.

[42] Magesan, E., J. Gambetta, and J. Emerson (2011). Scalable and robust randomized benchmarking of quantum processes. *Physical review letters 106*(18), 180504.

[43] Monz, T., P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt (2011). 14-qubit entanglement: Creation and coherence. *Physical Review Letters 106*(13), 130506.

[44] Moroder, T., P. Hyllus, G. Toth, C. Schwemmer, A. Niggebaum, S. Gaile, O. Gühne, and H. Weinfurter (2012). Permutationally invariant state reconstruction. *New Journal of Physics 14*(10), 105001.

[45] Moroder, T., M. Kleinmann, P. Schindler, T. Monz, O. Gühne, and R. Blatt (2013, Apr). Certifying systematic errors in quantum experiments. *Phys. Rev. Lett. 110*, 180401.

[46] Paris, M. and J. Rehácek (2004). *Quantum state estimation*. Springer Verlag.

[47] Peres, A. (1995). Quantum Theory: Concepts and Methods. *American Journal of Physics 63*(3), 285+.

[48] Renes, J. M., R. Blume-Kohout, A. J. Scott, and C. M. Caves (2004). Symmetric informationally complete quantum measurements. *Journal of Mathematical Physics 45*, 2171.

[49] Santori, C., D. Fattal, J. Vuckovic, G. S. Solomon, and Y. Yamamoto (2002). Indistinguishable photons from a single-photon device. *Nature 419*(6907), 594–597.

[50] Shor, P. (1999). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review 41*(2), 303–332.

[51] Shor, P. W. (1995, Oct). Scheme for reducing decoherence in quantum computer memory. *Phys. Rev. A 52*, R2493–R2496.

[52] Shor, P. W. (1996). Fault-tolerant quantum computation. In *Proceedings of the 37th Annual Symposium on Foundations of Computer Science*, FOCS '96, Washington, DC, USA, pp. 56–. IEEE Computer Society.

[53] Singh, S. (1999). *The Code Book: The Evolution of Secrecy from Mary, Queen of Scots, to Quantum Cryptography* (1st ed.). New York, NY, USA: Doubleday.

[54] Smithey, D., M. Beck, M. Raymer, and A. Faridani (1993). Measurement of the wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Phys. Rev. Lett. 70*(9), 1244–1247.

[55] Steane, A. (1996). Multiple-particle interference and quantum error correction. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences 452*(1954), 2551–2577.

[56] Steane, A. M. (2003). Overhead and noise threshold of fault-tolerant quantum error correction. *Physical Review A 68*(4), 042322.

[57] Toth, G., W. Wieczorek, D. Gross, R. Krischek, C. Schwemmer, and H. Weinfurter (2010). Permutationally invariant quantum tomography. *Physical review letters 105*(25), 250403.

[58] Tóth, G., W. Wieczorek, R. Krischek, N. Kiesel, P. Michelberger, and H. Weinfurter (2009). Practical methods for witnessing genuine multi-qubit entanglement in the vicinity of symmetric states. *New Journal of Physics 11*(8), 083002.

[59] Unruh, W. G. (1995, Feb). Maintaining coherence in quantum computers. *Phys. Rev. A 51*, 992–997.

[60] Usami, K., Y. Nambu, Y. Tsuda, K. Matsumoto, and K. Nakamura (2003). Accuracy of quantum-state estimation utilizing akaike's information criterion. *Physical Review A 68*(2), 022314.

[61] van Enk, S. J. and R. Blume-Kohout (2013). When quantum tomography goes wrong: drift of quantum sources and other errors. *New Journal of Physics 15*(2), 025024.

[62] van Enk, S. J. and C. A. Fuchs (2002). The quantum state of a propagating laser field. *Quantum Information and Computation 2*(2), 151.

[63] Vandersypen, L. M. K., M. Steffen, G. Breyta, C. S. Yannoni, M. H. Sherwood, and I. L. Chuang (2001, December). Experimental realization of Shor's quantum factoring algorithm using nuclear magnetic resonance. *Nature 414*, 883–887.

[64] Wootters, W. K. and W. H. Zurek (1982, October). A single quantum cannot be cloned. *Nature 299*(5886), 802–803.

[65] Yin, J. and S. J. van Enk (2011). Information criteria for efficient quantum state estimation. *Physical Review A 83*(6), 062110.