FOUNDATIONS AND APPLICATIONS OF SEQUENTIAL MEASUREMENTS

Kieran Flatt

A thesis submitted for the degree of Doctor of Philosophy

School of Physics and Astronomy
College of Science and Engineering
University of Glasgow

June 2019

Abstract

When a system is measured, its state is changed. A mathematical consequence of this statement is that scenarios in which a quantum system is measured repeatedly, or the same system is used to measure many others, require the use of Kraus's formalism. Three projects which fall into this category are discussed in this thesis. One is of foundational interest and two are more oriented towards experiment.

The first piece of work is an analysis of the uniqueness of each of Kraus's formulae for joint and conditional probabilities. Gleason, Busch and others were interested in whether the probability rules of quantum mechanics were constructed ad hoc or whether they had deeper significance. They showed that the Born rule was the only way of calculating quantum probabilities consistent with some basic assumptions about the nature of a physical theory. I extend this work to the sequential measurement case and show that no further assumptions are required for joint, over single, probabilities.

A mathematical technique, the use of operator space, from that work is then developed, in my second reported piece of work, for use as a tool in quantum cryptanalysis. I show that calculations of the best eavesdropping strategies for quantum key distribution protocols can be done in a straightforward manner. I rediscover optimal strategies for BB84 and B92, two of the most commonly discussed protocols, and report a new attack for PBC00.

Multiple-copy state discriminators look for methods of distinguishing states given a number of systems all in that state. An open question is whether a quantum memory, a device which interacts with other systems and does not decohere, aids this problem. In the third piece of work reported here, I compare the ability of two schemes, one which uses a quantum memory and one which does not, for performing multiple-copy state discrimination. One surprising result is that the scheme that uses quantum memory always performs worse than the one which does not. Another is that both schemes tend to the same limit in the case that the resource is an infinite number of copies. This suggests that a quantum memory may not be helpful.

Acknowledgements

Deepest gratitude to Sarah Croke and Steve Barnett, my two supervisors, who suggested interesting topics throughout and helped in so many other ways. Some algebraic gaffes were handled with extreme class. Thank you.

Thanks, of course, to all the other Quantum Theorists at the University of Glasgow: Can Ritboon, Frances Crimin (on whose floor I sit while correcting the thesis), Graeme Weir, Neel Mackinnon, Scarlett Gao, Anette Messinger, Thomas Brougham, Matthias Sonnleitner, Jim Cresser, Fiona Speirits, Jörg Götte, Rob Cameron, Gergely Ferenczi. Wonderful lunchmates, bakers and drinkers, all.

I couldn't have done this research without support from my parents and siblings. Thanks to you as well.

Ale - grazie. Stew, Robbie, Brogues, Hannah, Johnston &c. – thank you so much for your company. David and Sherri – thanks to you and congratulations on your baby. Tilly, all your visits were so fun. Sam, Wardy, Loz, Rick, Liam and Cherrie, who I first learned physics with, I couldn't thank you more.

Author's declaration

I declare that this thesis, presented for the degree of Doctor of Philosophy at the University of Glasgow, was composed by myself, except where indicated in the text by special reference. Parts of this work either have been, or will be, published in:

- K. Flatt, S. M. Barnett, and S. Croke. Gleason-Busch theorem for sequential measurements. *Phys. Rev. A*, 96:062125, 2017
- K. Flatt, S. Croke, and S. M. Barnett. Two-time state formalism for quantum eavesdropping. *Phys. Rev. A*, 98:052339, 2018
- K. Flatt, S. Croke, and S. M. Barnett. Multiple-copy state discrimination of noisy qubits. In preparation

The thesis has not been presented to any other university, either in the United Kingdom or overseas, for examination.

Kieran Flatt
June 2019

These were moments, years,
Solid with reality, faces, namable events, kisses, heroic acts,
But like the friendly beginning of a geometrical progression
Not too reassuring, as though meaning could be cast aside some day
When it had been outgrown. Better, you said, to stay cowering
Like this in the early lessons, since the promise of learning
Is a delusion, and I agreed, adding that
Tomorrow would alter the sense of what had already been learned,
That the learning process is extended in this way, so that from this standpoint
None of us ever graduates from college,
For time is an emulsion, and probably thinking not to grow up
Is the brightest kind of maturity for us, right now at any rate.

John Ashbery, *Soonest Mended*

Contents

Chapter 1

Introduction

That it is impossible to characterise perfectly a system with a single measurement, and that the system's state will be altered by that measurement, begs the question of how much can be known about that system. Answering this leads to the field of quantum information theory. The practical issue of extracting information from a quantum system forms the basis of technological applications and highlights foundational issues. The latter, of course, were well-known to many of the great innovators of the early twentieth century (Einstein, Pauli, Dirac, etc.) but were passed over for several decades. As Olival Freire Jr. has discussed [4], quantum foundations during this period became a clandestine subject, published in unofficial journals and discussed in out-of-hours reading groups. It is probably true that the motives behind this side-lining were practical: it was only in the 1980s that it was possible to perform experiments on single quantum systems. This meant that interpretational issues could be approached somewhat systematically, and also used as the basis for technologies: above all, communication systems and computers. In the few decades since then, measurement theory has been a rich topic for all kinds of researchers. My doctoral work, which is brought together in this thesis, has covered both foundational aspects as well as those more likely to have consequences for experimentalists. I have focused on sequential measurements, processes in which the same system is measured two-or-more times.

In Chapter 2, I present quantum theory, beginning from the standard set of von Neumann postulates to introduce the language of quantum measurement theory: POVMs, Kraus operators and Naimark dilation. Quantum key distribution and state discrimination, two key applications of this framework and those which form the basis of the latter half of the thesis, are also introduced.

In Chapter 3, I examine the Kraus formalism in terms of postulates. Researchers from von Neumann to Gleason to Busch have asked why the Born rule has the form that it does. Over the past decade this kind of questioning has given rise to a field known as quantum reconstructions, in which sets of operational postulates are proposed and used to derive quantum theory. The idea behind this is that questions about the more mysterious aspects (e.g., the measurement problem) can be understood more easily. Towards this goal, I link Gleason and Busch's analysis of the Born rule to Kraus's formalism for joint and conditional probabilities, an analysis which shows that no additional assumptions are

needed to find the state-update rule.

The work in Chapter 4 develops, based on some aspects of the Kraus rule analysis, a tool for developing eavesdropping attacks in quantum key distribution. The work in quantum reconstructions leads to an understanding of sequential measurements in which the space of two-time states is fundamental. In this framework, pre- and post-selection appear as superoperators and Kraus operators appear as states, and this distinction maps naturally onto that in quantum key distribution between Alice and Bob's correlations and Eve's actions. With this framework, I am able to rediscover the optimal and well-known attacks for BB84 and B92 and also find a novel optimal attack for a less-explored protocol, PBC00, which uses the trine states. The surprising result here is that the optimal attack does not give Eve information about the transmitted state, an unexpected result which is counterintuitive but found naturally with the two-time state formalism.

In Chapter 5, I move onto a different topic: multiple-copy state discrimination. How to successfully discriminate two states given a number of copies is still not a deeply explored question, partly due to the difficulty in deriving analytic results. An open question is whether a quantum memory is a useful resource in this problem. My contribution is to calculate the probability of success for two different schemes, one which uses a quantum memory and one which does not. They are both known to be optimal for discriminating two pure states, but I apply them to mixed states representing imperfect preparation. Two surprising results emerge. The first is that both schemes tend towards the same sub-unit probability of success, in the many-copy limit. The second is that the local scheme, the scheme that does not need a quantum memory, is better in all cases. Admittedly this improvement is very small and probably not experimentally detectable, however it still goes against the commonplace that it is always useful to be able to interact coherently.

I conclude with Chapter 6, in which I summarise the contents of the previous chapters and discuss some possible paths towards future work in the fields of quantum reconstructions, eavesdropping strategies and multiple-copy state discrimination.

Chapter 2

Background

One topic of this thesis is that there are many ways to present quantum mechanics. In this thesis, I am mostly concerned with applications in measurement theory and so use the relevant language of density matrices and POVM elements. To develop that framework, I begin from the most common starting point: von Neumann's postulates. The material in §2.1-3 is taken from a variety of standard sources [5, 6, 7, 8, 9, 10, 11].

2.1 Basic theory

Pure states

The fundamental quantities in quantum mechanics are states and observables. The former are represented by vectors, written as kets $|\psi\rangle$, in a complex valued Hilbert space $\mathcal{H}$. The association between states and vectors was formalised by von Neumann as his first postulate. That states are written as vectors is a consequence of the fact that a quantum-mechanical description of reality allows for a continuum of states. In any vector space, the basis can be freely chosen and, because of this, quantum states can exist in superpositions. If the set of vectors $\{|i\rangle\}$ form a basis then the state may be written as

$$|\psi\rangle = \sum_i a_i |i\rangle, \tag{2.1}$$

in which the set of coefficients can be any complex numbers such that the state is normalised. The set of kets also implies a set of bras $\langle\psi|$, which formally speaking are vectors in the space of functionals, and allow inner products to be defined. If a second state is written in the same basis as Eq. 2.1,

$$|\phi\rangle = \sum_j b_j |i\rangle, \tag{2.2}$$

then the inner product is

$$\langle\psi|\phi\rangle = \sum_i a_i^* b_i. \tag{2.3}$$

The probabilistic description of quantum mechanics, which is introduced alongside the idea of measurements, requires that the states are normalised. That is,

$$\langle\psi|\psi\rangle = 1. \tag{2.4}$$

Observables

Alongside states, observables are the other basic quantity in quantum theory. These are associated with operators, which can be defined more generally. An operator is an object which acts upon one vector and outputs another. I am concerned in particular with linear operators, which satisfy

$$\begin{aligned} A\left(|\psi\rangle + |\phi\rangle\right) &= A|\psi\rangle + A|\phi\rangle \\ \left(A + B\right)|\psi\rangle &= A|\psi\rangle + B|\psi\rangle \\ A\left(\alpha|\psi\rangle\right) &= \alpha A|\psi\rangle. \end{aligned} \tag{2.5}$$

Here, A and B are the linear operators and α is any complex number. Linear operators should be defined on the entire vector space so that, in this set of definitions and with a slight abuse of notation, $|\psi\rangle$ and $|\phi\rangle$ need not be states. Out of the whole class of linear operators, quantum theorists finds particular use for those which are Hermitian. This is due to a second postulate of quantum mechanics which associates physically observables quantities with Hermitian operators: every mathematical object of this kind can be experimentally measured and vice versa. Hermitian operators are those which are the same as their complex conjugate transpose, i.e., $A = A^\dagger$. (To be precise, this is the definition of a self-adjoint operator. While these are actually distinct from Hermitian operators, the manner of this distinction is not important in the applications required here.) The eigenvectors of an operator are the set of kets $|\lambda_i\rangle$ satisfying

$$A|\lambda_i\rangle = \lambda_i|\lambda_i\rangle. \tag{2.6}$$

The objects λ_i are called eigenvalues. For Hermitian operators, they are positive, and this can be used to verify whether or not a given operator is Hermitian. Any operator's set of eigenvectors forms a basis, called the eigenbasis, which spans the relevant Hilbert space. That is, any state can be written in terms of the eigenvectors of a given operator. It is often useful to write an operator in terms of outer products of its eigenvectors,

$$A = \sum_i \lambda_i|\lambda_i\rangle\langle\lambda_i|, \tag{2.7}$$

which is called the spectral decomposition of an operator.

Quantum mechanics is an inherently probabilistic theory, and this indeterminacy enters at the level of measurement. When a system in the state $|\psi\rangle$ is measured for the observable A, the possible measurement outcomes are the set of eigenvalues associated with A. Only one of these outcomes will occur, and the probability of that event is given by the Born

rule, written in its most basic form as

$$\mathrm{P}(\lambda_i|\psi) = |\langle\lambda_i|\psi\rangle|^2. \tag{2.8}$$

Alongside the probability of individual outcomes, the measurement of a variable on a system is associated with the average value that would be calculated for that variable after many measurements. This quantity is the expectation value,

$$\langle A\rangle = \langle\psi|A|\psi\rangle, \tag{2.9}$$

which follows from the Born rule and the spectral decomposition of the operator. It is the sum of all possible values of that variable, weighted by the probability that they are measured.

It is well known that a quantum measurement will alter the state of the measured system. For measurements of the type considered by von Neumann, this behaviour is the notorious wave function collapse. The claim is that when an observable A is measured on a pure state with the outcome λ_i, the post-measurement state will be the associated eigenvector, $|\lambda_i\rangle$, suitably normalised. This can be formalised by introducing the projector $\Lambda_i = |\lambda_i\rangle\langle\lambda_i|$, in which case the collapse of the state is

$$|\psi\rangle \to \frac{\Lambda_i|\psi\rangle}{\langle\psi|\Lambda_i|\psi\rangle}. \tag{2.10}$$

This update is sometimes referred to as Lüder's rule. Written in this manner, it generalises readily to the framework of density matrices which I introduce in a later section.

An operator which is found to be very useful for various calculations is I, the identity. This is the operator constructed so that $I|\psi\rangle = |\psi\rangle$ for all possible states. It allows the act of not measuring a system to be represented. In terms of an eigenbasis $|\lambda_i\rangle$ the identity is

$$I = \sum_i |\lambda_i\rangle\langle\lambda_i|. \tag{2.11}$$

Composite Systems

A third postulate concerns the act of bringing together two or more systems. What is meant by a system in this context is a degree of freedom, which is general enough to include both spatially separated particles and two-or-more different variables on the same particle. If one system is in the state $|\psi\rangle_A$, defined by a vector on the Hilbert space $\mathcal{H}_A$, and another system is in the state $|\phi\rangle_B$, similarly defined on $\mathcal{H}_B$, then the object which describes the composite system is the tensor product, $|\psi\rangle_A \otimes |\psi\rangle_B$, of those two states. Composite states of this kind do not exhibit correlated measurement outcomes. However, if $\mathcal{H}_A \otimes \mathcal{H}_B$ is an allowed state space, then by the first postulate it follows that *any* suitably normalised vector on that space is also an allowed state of the composite system. This claim introduces entanglement into quantum theory. My discussion here will be concerned with the set of bipartite states shared between two systems however everything stated can be generalised to multipartite states defined on more than two subsystems.

Entanglement is the idea that measurements on composite systems, under specific conditions, can be correlated despite (in principle, unlimited) spatiotemporal separation of the two systems. This correlation occurs due to the collapse of one system due to the measurement of the other, and not because both quantities were predetermined. This concept rubs up against relativity in various ways which were later explored by Bell, and which are best explained by Maudlin [12]. The most explored entangled states are the Bell states, of which one example is

$$|\Psi^+\rangle = \frac{1}{\sqrt{2}} \left(|0_A 0_B\rangle + |1_A 1_B\rangle \right). \tag{2.12}$$

To see that these two systems are entangled, consider that system B only is measured such that the outcome is associated with the eigenvector $|0\rangle$. By the state-update rule, this will leave the composite system in the state

$$\begin{aligned} |\Psi^+\rangle &\to \frac{(I_A \otimes |0\rangle\langle 0|_B)\,|\Psi^+\rangle}{\langle\Psi^+|\,(I_A \otimes |0\rangle\langle 0|_B)\,|\Psi^+\rangle} \\ &= |0_A 0_B\rangle. \end{aligned} \tag{2.13}$$

It is obvious that no measurement on system A could now be associated with the vector $|1\rangle$, but this was possible before the measurement on B. Hence, the two systems are correlated.

It is natural to ask which composite states are entangled and which are not. The answer is that any non-entangled state is separable: it can be written as $|\psi\rangle_A \otimes |\phi\rangle_B$. The possibility of changing basis (i.e., a state may appear inseparable in one basis but not another) means that it is not always straightforward to determine whether or not a state is entangled. The object which is required is the Schmidt rank. Consider a bipartite state

$$|\psi\rangle = \sum_{ij} c_{ij} |a_i\rangle_A |b_j\rangle_B, \tag{2.14}$$

where $\{|a_i\rangle\}$ and $\{|b_j\rangle\}$ are two arbitrary bases for each space. The coefficients c_{ij} can be considered the elements of a matrix C and the singular valued decomposition theorem (which is introduced in Appendix A) states that this matrix can be decomposed into the form $C = U\Sigma V^\dagger$, where U and V are unitary operators and Σ is a positive semidefinite diagonal matrix. This means that the matrix elements will satisfy $c_{ij} = \sum_k u_{ik}\sigma_k v_{kj}$. With this, the bipartite state can be written as

$$\begin{aligned} |\psi\rangle &= \sum_{ijk} u_{ik}\sigma_k v_{kj} |i\rangle_A |j\rangle_B \\ &= \sum_k \sigma_k \left(\sum_i u_{ik} |a_i\rangle_A \right) \left(\sum_j v_{kj} |b_j\rangle_B \right) \\ &= \sum_k \sigma_k |u_k\rangle_A |v_k\rangle_B. \end{aligned} \tag{2.15}$$

The Schmidt decomposition is the name given to the structure seen in the third line of

this calculation. The objects $\{\sigma_k\}$ are the Schmidt coefficients and the number of them is called the Schmidt rank. Entanglement theory states that if the Schmidt rank is greater than one then the bipartite state cannot be written as a separable state and hence the system is entangled.

Schrödinger's Equation

I have mentioned one way, measurement, by which systems evolve. Such a change is irreversible and transfers information out of the system. If a system does not interact with another then it will instead evolve reversibly, according to Schrödinger's equation:

$$i\hbar\frac{\partial}{\partial t}|\psi\rangle = H|\psi\rangle. \tag{2.16}$$

All of the usual notation is adopted here: $\hbar$ is the reduced Planck constant and H is the Hamiltonian, the operator which governs the total energy in the system. In quantum information theory it is more useful to replace this differential form with a unitary operator, one for which $UU^\dagger = U^\dagger U = I$. If a system is known to be in the state $|\psi_0\rangle$ at the time $t = 0$ then it is assumed that at time t it will be in the state $|\psi_t\rangle = U_t|\psi_0\rangle$. The form of the unitary operator must be found in terms of the Hamiltonian. I begin by substituting $|\psi_t\rangle$ into Eq. 2.16 and then rearrange for

$$\frac{\partial}{\partial t}U_t|\psi_0\rangle = -\frac{iH}{\hbar}U_t|\psi_0\rangle. \tag{2.17}$$

The form of the evolution is the same for all initial states which means that the unitary must satisfy

$$\frac{\partial}{\partial t}U_t = -\frac{iH}{\hbar}U_t. \tag{2.18}$$

The usual solution of this differential equation is used and evolution under the Schrödinger equation can be represented by the operator

$$U_t = \exp\left(-\frac{i}{\hbar}\int_0^t H dt\right), \tag{2.19}$$

where the exponential is defined according to the usual rules for functions of operators. In many situations the Hamiltonian will be time-independent and in this case the operator has an even simpler form

$$U_t = \exp\left(-\frac{i}{\hbar}Ht\right). \tag{2.20}$$

Unitary evolution represents reversible evolution, according to which there is a single state associated with each time t for a given initial state. If a particular unitary is required then it can be constructed by implementing the relevant Hamiltonian according to this equation. Reversible evolution can also be performed in this manner on composite systems, and in general unitary evolution of this kind will tend to increase the level of entanglement between subsystems.

2.2 Mixed states

The formulation, in terms of pure states and projective measurements, of quantum mechanics presented above can be used if three assumptions hold. Firstly, the system does not interact with others, except for a possible measuring device. Secondly, an experimentalist is able to characterise precisely the system at one point in time. Thirdly, all measurements are ideal, so that outcomes correspond to pure states. All of these assumptions can be relaxed and to do so density matrices and positive-operator-valued measurements are required.

Density matrices

To introduce density matrices, the trace function $\mathrm{Tr}(A)$ is required. It is the sum of the diagonal elements of a matrix, i.e.,

$$\mathrm{Tr}(A) = \sum_i \langle i|A|i\rangle. \tag{2.21}$$

A number of the trace's properties will be required. It is a cyclic function: $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. It is also invariant under changes of basis, so that the sum can use any set of orthogonal vectors which span the space. It is linear: $a\mathrm{Tr}(A) = \mathrm{Tr}(aA)$ for a being any number. These properties are used throughout.

The reason for the ubiquity of the trace operation in quantum theory is that it can be used to rewrite the Born rule. I take Eq. 2.8 and multiply it by the identity in the basis $\{|j\rangle\}$:

$$\begin{aligned}
\mathrm{P}(\lambda_i|\psi) &= \langle\psi|\lambda_i\rangle\langle\lambda_i|\psi\rangle \\
&= \sum_j \langle\psi|j\rangle\langle j|\lambda_i\rangle\langle\lambda_i|\psi\rangle \\
&= \sum_j \langle j|\lambda_i\rangle\langle\lambda_i|\psi\rangle\langle\psi|j\rangle \\
&= \mathrm{Tr}\left(P_{\lambda_i}P_\psi\right).
\end{aligned} \tag{2.22}$$

I use the notation $P_\psi = |\psi\rangle\langle\psi|$ for the projector associated with a state $|\psi\rangle$. With the Born rule in this form, I am in a position to introduce the density operator. Mixed states represent systems which are prepared probabilistically, i.e., a system is in the state $|\psi_i\rangle$ with the probability p_i, where $i = 0, 1 \cdots N$. I label this ensemble ρ. The probability of a given measurement outcome is calculated using the Born rule:

$$\begin{aligned}
\mathrm{P}(\lambda_i|\rho) &= \sum_j p_j \mathrm{P}(\lambda_i|\psi_j) \\
&= \mathrm{Tr}\left(P_{\lambda_i}\sum_j p_j P_{\psi_j}\right).
\end{aligned} \tag{2.23}$$

The linearity of the trace has been used to bring the probabilities p_j and the sum into the

argument. It is seen that probabilistic preparation can be mathematically represented by replacing the pure state projectors P_ψ with

$$\rho = \sum_j p_j P_{\psi_j} = \sum_j p_j |\psi_j\rangle\langle\psi_j|. \tag{2.24}$$

This object is the density matrix. Any positive, Hermitian operator with unit trace is a possible density matrix. That it is positive ensures that it has a spectral decomposition and so can be created by a probabilistic mixture of pure states. The requirement that the trace is unity is the usual requirement that a set of probabilities sums to one. With reference to the next chapter, it is interesting to note that the generalisation of pure states would have been different had the Born rule not been equal to the squared amplitude of the inner product of the measurement. It is this that allows the linear trace to represent probabilities. Probabilistic mixtures would have a different structure if the Born rule used, for example, the fourth power of the amplitude. This emphasises the necessary link between the probability rule and the space of states, which link will be explored in much greater detail in the next chapter.

All of the postulates concerning pure states generalise to density matrices. If a von Neumann measurement has outcome $|\lambda_i\rangle$ then the density matrix ρ is updated by

$$\rho \to \frac{P_{\lambda_i}\rho P_{\lambda_i}}{\mathrm{Tr}\left(P_{\lambda_i}\rho P_{\lambda_i}\right)} \tag{2.25}$$

and if instead the state evolves under the unitary U_t then the update will be

$$\rho \to U_t \rho U_t^\dagger. \tag{2.26}$$

Composite systems

Composite systems are included in the obvious manner: if each of a pair of systems is associated with density operators on $\mathcal{H}_A$ and $\mathcal{H}_B$ then any positive semidefinite operator with unit trace on the product space will also be a possible state of the composite system. Mixed states on the product space fall into three categories: uncorrelated, classically correlated and entangled. Uncorrelated states are those that can be written as

$$\rho_{AB} = \rho_A \otimes \rho_B. \tag{2.27}$$

If the overall density operator can be written in this form, then joint probabilities factor and are independent, so there is no correlation between the two measurements. It is also possible to construct density operators that are weighted sums of these:

$$\rho_{AB} = \sum_i p_i \rho_A^i \otimes \rho_B^i. \tag{2.28}$$

In such cases, the measurements will be correlated but in a classical manner, i.e., the correlations are determined before the measurement takes place. The final possibility is, of course, that the state cannot be written in either manner and in such cases the two

systems are entangled. At the time of writing, there is no equivalent tool to the Schmidt rank for diagnosing entanglement of density matrices.

It may be that an experimenter measures only subsystem on $\mathcal{H}_B$ of a bipartite state. The object that describes the measurement statistics in such a case is the reduced density operator, calculated by taking the partial trace,

$$\mathrm{Tr}_A\left(\rho_{AB}\right) = \sum_i \langle i|_A \rho_{AB} |i\rangle_A, \tag{2.29}$$

of the overall system. This objects reproduces all of the measurement statistics which would be found by local measurements. If the joint state is separable, $\rho_{AB} = \rho_A \otimes \rho_B$, then the reduced density operator of system A and B will be ρ_A and ρ_B respectively. That the partial trace is the unique way of ignoring some degrees of freedom in a composite system is because it is the unique map which preserves the Born rule as the probability rule.

Qubits

In this thesis I am always concerned with finite dimensional systems and almost always concerned with qubits, which are ubiquitous throughout quantum information theory. Qubits are the set of two-dimensional states (some common physical examples are the polarisation of a photon or the spin of an electron). They are often written in the computational basis which consists of two orthogonal states labelled $|0\rangle$ and $|1\rangle$. A pure state of the system is written in the bra-ket form

$$|\psi\rangle = a_0|0\rangle + a_1|1\rangle, \tag{2.30}$$

where the coefficients are free to be any two complex numbers such that $|a_0|^2 + |a_1|^2 = 1$. Occasionally it is handy to express the pure states of a qubit in the column vector notation:

$$\psi = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}. \tag{2.31}$$

It is useful to represents the states and operators of qubits with the Pauli operators, a set of four orthogonal operators which, alongside the identity, form a basis in which all two-dimensional operators can be written. In bra-ket notation the whole set is

$$\begin{aligned} I &= |0\rangle\langle 0| + |1\rangle\langle 1| \\ \sigma_x &= |0\rangle\langle 1| + |1\rangle\langle 0| \\ \sigma_y &= i|1\rangle\langle 0| - i|0\rangle\langle 1| \\ \sigma_z &= |0\rangle\langle 0| - |1\rangle\langle 1|. \end{aligned} \tag{2.32}$$

The Pauli operators, ignoring the identity, all have different eigenvectors. For σ_x they are $(|0\rangle \pm |1\rangle)/\sqrt{2}$, for σ_y they are $(|0\rangle \pm i|1\rangle)/\sqrt{2}$ and for σ_z they are the computational basis, $|0\rangle$ and $|1\rangle$. This allows them to be used as the basis for the concept of the Bloch sphere.

This is a constructed used to pictorially represent states, as in Fig. 2.1 . It is a unit-radius sphere in a space with the axes consisting of the eigenvectors of the three Pauli matrices. The north and south poles correspond to the basis states $|0\rangle$ and $|1\rangle$ and all other points which lie on the sphere are other pure state. Because the Pauli matrices form a basis for all matrices in the space of qubits, any density operator can be written as a weighted sum of them, in the form

$$\rho = \frac{1}{2}\left(I + u\sigma_x + v\sigma_y + w\sigma_z\right), \tag{2.33}$$

where u, v, w are three real numbers and must satisfy $u^2 + v^2 + w^2 \leq 1$ to ensure positivity of the density operator. These three parameters are a set of coordinates which are the components of the state's vector on the Bloch sphere. If the equality is satisfied, the state is pure and lies on the surface. Otherwise, if the sum of squares is less than unity, then the density matrix corresponds to a mixed state. Because all qubit states lie on the Bloch sphere, operations can be considered as maps between the related vectors on the Bloch sphere. In particular, any unitary operation can be visualised as rotating a vector around a particular axis by some angle.

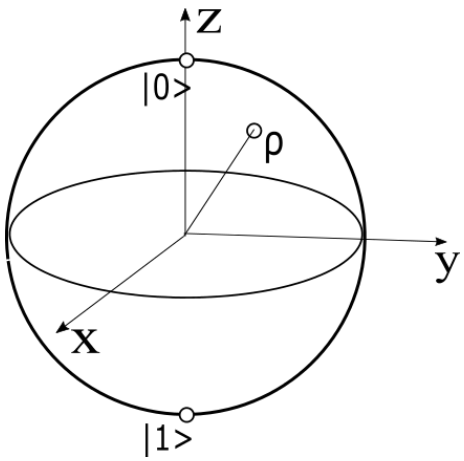


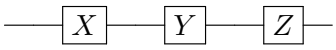
Figure 2.1: The Bloch sphere is a visual way of representing qubits. Pure states are points on the surface and mixed states are inside that surface.

Gate model

In classical information processing, it is common to represent processes by a sequence of logic gates which act on one, two or more bits. By replacing the input bits with qubits, it is also possible to present quantum processes in this manner. Quantum gates are usually defined by their action on the computational basis $|0\rangle, |1\rangle$ and the action on all other states follows from linearity and the superposition principle. I list some commonly used gates here. It is not an exhaustive list. I discuss only the one and two qubit gates which will be needed in what follows.

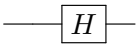
- **Pauli gates.** The most commonly used Pauli gate is the Pauli X, which simply

implements the σ_x operation on the input qubit. This is the quantum equivalent of the classical NOT gate: it transforms the basis states $|0\rangle, |1\rangle$ to $|1\rangle, |0\rangle$ respectively. Of course, all other Pauli gates can implemented in a similar manner and correspond to introducing different phase transformations: Y transforms $|0\rangle$ to $i|1\rangle$ and $|1\rangle$ to $-i|0\rangle$; Z leaves $|0\rangle$ unchanged but maps $|1\rangle$ onto $-|1\rangle$. The notation used for these three gates is:

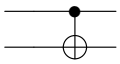


In a quantum circuit diagram, the input is the left-most point and the gates act from left to right, as in classical logic circuits.

- **Hadamard.** Another useful single qubit gate is the Hadamard gate, which maps the basis states $|0\rangle$ and $|1\rangle$ onto $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$ respectively. It is written in the same manner as the Pauli gates:

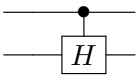


- **CNOT.** Among the possible two-qubit gates, the CNOT is particularly useful. This gate takes in two qubits which are called the control and the target. If the control qubit is in the state $|0\rangle$ then neither qubit is changed however if the control qubit is in the state $|1\rangle$ then a NOT gate is performed on the target. The notation used for this gate is



Here, the upper wire denotes the control and the lower wire denotes the target.

- **Controlled rotations.** The CNOT gate can be thought of as a controlled Pauli-X gate. All other single qubit gates can be implemented in a controlled manner and with the obvious notation:



where I use the Hadamard gate as an example only. Not only the Pauli gates but a rotation by any angle around any axis of the Bloch sphere can be performed in a controlled manner.

An important result of quantum information processing is the existence of *universal* gate-sets: if one can implement particular pairs of gates then any unitary operation can be performed. (The full result states that any unitary can be efficiently approximated in this manner, but this difference is not significant at the level that gates are used here.) A standard set of universal gates for actions on two qubits is the Hadamard gate and the CNOT gate however others are possible. Related to this is the fact that each unitary can be implemented in a number of different ways. One example, which is used in a later section of this thesis, is that a controlled rotation can be implemented by a CNOT gate with single qubit rotations.

2.3 Generalised measurements

One of the von Neumann postulates states that quantum states are associated with vectors, and these were generalised into positive operators. The same set of postulates also link measurement outcomes to vectors: the eigenvectors of the measured variable's Hermitian operator. It should come as no surprise that measurement outcomes are generalised by replacing these eigenvectors with matrices. These objects are the positive operator-valued measurements, or POVMs. Two other names, POM (probability operator measure) and effect, are used throughout the literature to refer to the same quantity. In this thesis I use 'POVM element' and 'effect' both to refer to this quantity, only changing usage for stylistic variation.

POVMs

In quantum information processing, it is often more useful to consider the probability of a given measurement outcome rather than the variable which is being measured. This is precisely analogous to classical information processing, for what is interesting in a Jacquard loom is not the shape and width of the holes but whether or not there is a hole at a given point. For the rest of the thesis I am concerned purely with this aspect of measurement.

The probability of a pure state outcome, according to the objects introduced so far, is given by

$$\mathrm{P}(\lambda_i|\rho) = \mathrm{Tr}\left(P_{\lambda_i}\rho\right). \tag{2.34}$$

As in the previous case, the generalisation is made intuitive by considering a noisy measurement. This example is taken from Ref. [10]. I consider a two-outcome measurement, with possible orthogonal outcomes $|\lambda_0\rangle$ and $|\lambda_1\rangle$, in which a faulty measuring device records with probability p the outcome which did *not* occur. The probability of getting the zero outcome is then

$$\begin{aligned}\mathrm{P}(\lambda_0|\rho) &= (1-p)\mathrm{Tr}\left(P_{\lambda_0}\rho\right) + p\mathrm{Tr}\left(P_{\lambda_1}\rho\right)\\ &= \mathrm{Tr}\left(\left((1-p)P_{\lambda_0} + pP_{\lambda_1}\right)\rho\right)\\ &= \mathrm{Tr}\left(\pi_0\rho\right).\end{aligned} \tag{2.35}$$

In the third line I introduce the object

$$\pi_0 = (1-p)P_{\lambda_0} + pP_{\lambda_1} \tag{2.36}$$

to represent the zeroth measurement outcome. This object is the POVM element introduced above. POVMs find two uses in quantum mechanics. Firstly, they can represent imperfect measurements of the kind shown here. Secondly, it is often the case that one does not want to perform a von Neumann measurement.

In general, a POVM is a set of operators $\{\pi_i\}$ which satisfy the following three condi-

tions:

$$\pi_i = \pi_i^\dagger \tag{2.37}$$

$$\pi_i \geq 0 \tag{2.38}$$

$$\sum_i \pi_i = I. \tag{2.39}$$

The first two of these are interrelated as any positive operator will be Hermitian, although not vice versa. Hermiticity ensures that the POVM outcome is a physical observable; this observable must be positive as it is a probability; a complete set of measurement outcomes must sum to one, hence the third requirement. Any set of operators satisfying these properties can be implemented as a POVM according to Naimark's theorem, which I introduce below.

Kraus operators

A property of von Neumann measurements is that the measured state is left in a pure state. One way to understand this is that such measurements characterise precisely the outcome. This is not true for the generalised measurements which are represented by POVMs and which I consider here. The objects needed to describe the wider range of state-updates are Kraus operators [13], which will play an important role throughout this thesis. Kraus operators can also be called measurement operators or instruments, and as with the other possible names for POVM elements I alternate usage only for stylistic variation. Throughout the literature, one sometimes finds 'effect' used for Kraus operator as well as POVM element but here I avoid that usage. In this section I begin by introducing a general definition for Kraus operators and then present two different methods (one more mathematical and another which is more physical) which justify their usage.

An effect π_i is decomposed into the form

$$\pi_i = \sum_\nu A_i^{\nu\dagger} A_i^\nu. \tag{2.40}$$

A decomposition of this form is not unique and can be done in an infinite number of ways. The objects A_i^ν are the Kraus operators. If π_i corresponds to the outcome of a measurement on the state ρ then the system will be left in the state

$$\rho_i = \frac{\sum_\nu A_i^\nu \rho A_i^{\nu\dagger}}{\mathrm{Tr}\left(\sum_\nu A_i^\nu \rho A_i^{\nu\dagger}\right)}. \tag{2.41}$$

In many of the scenarios that are discussed in this thesis, I am interested in non-degenerate measurement outcomes. In such cases, there is just a single Kraus operator associated with each effect. A decomposition which involves a sum, as above, is indicative of degenerate measurement outcomes, i.e., one could map two measurement outcomes from one set onto a single outcome in another. This idea is referred to as 'coarse graining'.

If two measurements are performed on the same system, the joint measurement is also

represented by a Kraus operator. If two measurement outcomes are represented by the effects $\pi_i^{(1)} = A_i^\dagger A_i$ and $\pi_j^{(2)} = B_j^\dagger B_j$ then the state after both outcomes is

$$\rho_{ij} = \frac{B_j A_i \rho A_i^\dagger B_j^\dagger}{\mathrm{Tr}\left(B_j A_i \rho A_i^\dagger B_j^\dagger\right)}, \tag{2.42}$$

and the joint probability of these two outcomes is

$$\mathrm{P}(i,j|\rho) = \frac{\mathrm{Tr}(\pi_j^{(2)} A_i \rho A_i^\dagger)}{\sum_{ij} \mathrm{Tr}(\pi_j^{(2)} A_i \rho A_i^\dagger)}. \tag{2.43}$$

The state update can be interpreted as a single transformation by introducing $A_{ij} = B_j A_i$, so that the transformation is

$$\rho_{ij} = \frac{A_{ij} \rho A_{ij}^\dagger}{\mathrm{Tr}\left(A_{ij} \rho A_{ij}^\dagger\right)}. \tag{2.44}$$

I have restricted the set of measurements here to require just one Kraus operator but it is, of course, possible to generalise the transformation to include more than one. The POVM element associated with this outcome is $\pi_{ij} = A_i \pi_j^{(2)} A_i^\dagger$, which is positive for the same reason that an updated density operator is positive.

The mathematical foundations of the Kraus operator formalism are in completely-positive maps. Every state is associated with a density operator and so it must be true that any physical transformation of a system will leave it in a state also associated with a density operator, i.e., the map associated with the transformation must preserve the positivity and trace of the density operator. As composite systems are also allowed, the transformation must be completely positive: a map acting on $\mathcal{H}_A$ only must preserve positivity of states on $\mathcal{H}_A \otimes \mathcal{H}_B$. That maps must be completely-positive and trace-preserving (CPTP) enforces that they can be represented by Eq. 2.41 [14].

Naimark Dilation

A useful concept to keep in mind when thinking about Kraus operators is Naimark dilation. System A is entangled with an ancilla, a secondary system labelled B. A measurement of the first system is performed by measuring the ancilla. Naimark's theorem states there is a one-to-one mapping between this way to perform a measurement and the set of POVMs. There are two sides to this claim. It says that every POVM can be implemented in this manner. This is a result which explains why effects are such a powerful construction: Naimark's theorem says that every POVM which can be written down has a physical counterpart. Conversely, it also claims that any measurement of this kind can be represented by a set of effects. I now demonstrate the latter point.

I consider two systems. System A is prepared in the state $\rho_A = \sum_i \lambda_i |\lambda_i\rangle\langle\lambda_i|$, where $\{|\lambda_i\rangle\}$ is the relevant eigenbasis. System B is prepared in the pure state $|\Psi\rangle\langle\Psi|$. The two systems are initially uncorrelated and are entangled by a unitary U. The process is completed by a projective measurement on system B, which has outcome $|i\rangle$. I now show

that this entire process can be described by POVM elements and Kraus operators which act on the Hilbert space $\mathcal{H}_A$ of system A only.

I start with the Kraus operators. The act of measuring system B only can be described by the composite operator $I_A \otimes |i\rangle_B\langle i|_B$. The reduced density operator on A after preparation, unitary interaction and measurement will be

$$\begin{aligned}\rho_A^i &= \mathrm{Tr}_B\left(U\rho_A \otimes \rho_B U^\dagger I_A \otimes |i\rangle_B\langle i|_B\right) \\ &= \sum_{jk} \lambda_k \langle j| \left(U|\lambda_{kA}\Psi_B\rangle\langle\lambda_{kA}\Psi_B|U^\dagger I_A \otimes |i\rangle_B\langle i|_B\right)|j\rangle \\ &= \langle i|_B U|\Psi\rangle_B \rho_A \langle\Psi|_B U^\dagger|i\rangle_B. \end{aligned} \tag{2.45}$$

I identify the Kraus operators here by

$$A_i = \langle i|_B U|\Psi\rangle_B \tag{2.46}$$

and so the state update is

$$\rho_A^i = A_i \rho_A A_i^\dagger. \tag{2.47}$$

Up to a factor of normalisation, this is the state-update rule which is associated with Kraus operators. The construction Eq. 2.46 is not limited to be a positive operator and so, as an instrument, can be any operator. The only further generalisation which is possible is that the state update is over a sum of such terms. It's easy to see that such behaviour is included by coarse-graining over the measurement results on B, i.e., allowing that multiple outcomes occurred on that system.

I now show that this also leads to the POVM description of a measurement. To do this I must show that the probability rule associated with this process is the Born rule in terms of effects. I have

$$\begin{aligned} p_i &= \mathrm{Tr}_{AB}\left(U\rho_A \otimes \rho_B U^\dagger \left(I_A \otimes |i\rangle_B\langle i|_B\right)\right) \\ &= \mathrm{Tr}_A\left(A_i \rho_A A_i^\dagger\right) \\ &= \mathrm{Tr}_A\left(\rho_A A_i^\dagger A_i\right). \end{aligned} \tag{2.48}$$

To get from the first to the second line, I use the result of the previous calculation to get at the Kraus operator description. The third line is then derived from the cyclicity of the the trace function. The definition

$$\pi_i = A_i^\dagger A_i \tag{2.49}$$

relates the effect to the instrument, as was required before. This object is transparently positive, and therefore Hermitian. Summing over all measurement results $|i\rangle$ gives the

identity ($\sum_i |i\rangle\langle i| = I$) which leads to the POVM elements forming a complete set:

$$\begin{aligned}\sum_i \pi_i &= \sum_i \langle\Psi|_B U^\dagger |i\rangle_B \langle i|_B U |\Psi\rangle_B \\ &= \langle\Psi|_B U^\dagger U |\Psi\rangle_B \\ &= \langle\Psi|_B \left(I_A \otimes I_B\right) |\Psi\rangle_B \\ &= I_A \end{aligned} \tag{2.50}$$

All three properties are satisfied which means that Naimark's model of a measurement can be associated with a POVM. I have not proved the converse theorem, that all POVMs can be interpreted in this way, but it is shown elsewhere [11]. The two proofs together give a one-to-one mapping between the two concepts and give a useful model to keep in mind when thinking about quantum measurements.

2.4 State discrimination

A common task in quantum information processing is state discrimination [15, 16]. An experimenter has a system of which she knows not the state, but the finite set which that state was drawn from. Her task is to perform a measurement which determines the state. If the possible states are orthogonal pure states then it will be possible for her measurement to leave her certain of the prepared state, by performing a von Neumann measurement with the possible outcomes corresponding to the possible states. As is well-known, if the states are non-orthogonal or mixed then they can never be perfectly distinguished. This follows from the association between measurements and positive operators only. The effects π_0 and π_1 can be used to represent the outcomes of a measurement which seeks to distinguish $|\psi_0\rangle$ from $|\psi_1\rangle$. A perfect measurement will satisfy both $\langle\psi_0|\pi_0|\psi_0\rangle = 1$ and $\langle\psi_1|\pi_0|\psi_1\rangle = 0$. The first condition can be satisfied by requiring $\pi_0 = |\psi_0\rangle\langle\psi_0| + \pi_0'$, where π_0' is a positive operator on the rest of the space. With this definition, the second condition is then $|\langle\psi_1|\psi_0\rangle|^2 + \langle\psi_1|\pi_0'|\psi_1\rangle = 0$. Both terms on the left-hand side must be equal to zero but this can only hold if the two states are orthogonal. This demonstrates that for non-orthogonal states it is impossible to perform perfect discrimination.

Instead, the task is to maximise a chosen figure of merit. There are two main senses in which state discrimination can be optimised. In minimum-error discrimination, which was pioneered by Helstrom [17], every measurement has an outcome although some of them are incorrect. In unambiguous state discrimination, only a subset of measurements are associated with a possible state but for those cases the experimenter knows with certainty which state was prepared. Which type of measurement gives the maximum success probability will depend upon the specific case which is investigated.

Here, in this background, I discuss the basic results of both types of state discrimination and also cases in which there are several copies of the system. This material is used in Chapters 4 and 5 of this thesis. Although state discrimination in general can concern the problem of distinguishing between three-or-more states, I restrict myself to cases involving two states.

Minimum-error measurement

Minimum-error measurement is a method for discriminating states which always returns an answer but without certainty. There are two interesting quantities. The first of these is the highest probability of correctly identifying the state which can in principle be achieved. The second is the measurement which achieves that success rate. Helstrom provided a constructive proof which gives both [17, 18].

To derive the optimal measurement for discriminating two states is straightforward. A system is in one of two states, ρ_0 and ρ_1 (prepared with probability p_0 and p_1 respectively) and measured with outcomes π_0 and π_1 corresponding to the two possible states. The probability that this measurement is successful is

$$\begin{aligned} \mathrm{P}_{succ} &= p_0 \mathrm{P}(\pi_0|\rho_0) + p_1 \mathrm{P}(\pi_1|\rho_1) \\ &= \mathrm{Tr}\left(p_0\rho_0\pi_0 + p_1\rho_1\pi_1\right). \end{aligned} \tag{2.51}$$

The POVM must be complete, which means that $\pi_1 = I - \pi_0$. Using this to eliminate one of the effects gives

$$\mathrm{P}_{succ} = p_1 + \mathrm{Tr}\left[\left(p_0\rho_0 – p_1\rho_1\right)\pi_0\right]. \tag{2.52}$$

The maximum probability occurs when π_0 is a projector onto the eigenvector of $p_0\rho_0 – p_1\rho_1$ with the highest eigenvalue. The best measurement that could be performed is one which includes that as an outcome (similarly, π_1 will be a projector onto the other eigenvector). What value the probability takes in general will depend upon the form of the two density operators however a simple result can be found for the case of two pure states. In this case, $\rho_k = |\psi_k\rangle\langle\psi_k|$ with

$$|\psi_k\rangle = \cos(\theta)|0\rangle + (-1)^k \sin(\theta)|1\rangle, \quad k = 0, 1 \tag{2.53}$$

where $0 \leq \theta \leq \pi/4$. Any pair of states in the Hilbert space can be written in this manner without loss of generality. The two states are non-orthogonal, with an overlap $\langle\psi_0|\psi_1\rangle = \cos(2\theta)$. To find the Helstrom bound and Helstrom measurement, which are the name given to the quantities under consideration, I require the eigendecomposition of

$$p_0\rho_0 – p_1\rho_1 = \begin{bmatrix} \left(p_0 – p_1\right)\cos^2(\theta) & \sin(\theta)\cos(\theta) \\ \sin(\theta)\cos(\theta) & \left(p_0 – p_1\right)\sin^2(\theta) \end{bmatrix}. \tag{2.54}$$

A calculation reveals that the eigenvalues associated with this operator are

$$\lambda_\pm = \frac{1}{2}\left(p_0 – p_1 \pm \sqrt{1 – 4p_0p_1\cos^2(2\theta)}\right), \tag{2.55}$$

which are associated with the eigenvectors

$$|\phi_\pm\rangle = \frac{1}{\sqrt{2}}\left(\sqrt{1 \pm \frac{\left(p_0 – p_1\right)\cos(2\theta)}{1 – 4p_0p_1\cos^2(2\theta)}}|0\rangle \pm \sqrt{1 \mp \frac{\left(p_0 – p_1\right)\cos(2\theta)}{1 – 4p_0p_1\cos^2(2\theta)}}\right)|1\rangle. \tag{2.56}$$

The vector $|\phi_+\rangle$ corresponds to the prepared state $|\psi_0\rangle$. This measurement will succeed

with a probability

$$\mathrm{P}_{succ} = \frac{1}{2}\left(1 + \sqrt{1 + 4p_0 p_1 \cos^2(2\theta)}\right), \tag{2.57}$$

This object is the Helstrom bound; no measurement achieves a higher success probability for discriminating two pure states. If the states are equally likely, $p_0 = p_1 = 1/2$, then the measurement becomes a projector onto the σ_x basis, $|+\rangle, |-\rangle$. This is diagrammed in Fig. 2.2, where it is seen that these are the orthogonal pair which are symmetric around those which are prepared. If the state $|\psi_0\rangle$ is definitely prepared, $p_0 = 1$, then the optimal measurement is a projection onto the pair of states $|\psi_0\rangle$ and $|\psi_{0\perp}\rangle$, the latter of which is the state orthogonal to the former. This is intuitive: one outcome must be the state which was sent, as this ensures the correct outcome, and the other effect needs to complete a basis for the Hilbert space. The POVM in this case can be thought of as hypothesis checking, which seeks to confirm the prior knowledge. In the range $1 > p_0 > 1/2$, the measurement rotates from one basis to the above.

So far I have discussed pure and mixed state discrimination if there are two possible signal states. It is of course possible to discriminate three-or-more, and in this case Helstrom's conditions

$$\left(\sum_i p_i \rho_i \pi_i\right) - p_j \rho_j \geq 0 \ \ \forall j \tag{2.58}$$

$$\pi_i \left(p_i \rho_i - p_j \rho_j\right) = 0 \ \ \forall i, j \tag{2.59}$$

must be satisfied in their full generality. The second condition can be derived from the first, i.e., they are not independent, and are both sufficient and necessary for an optimal measurement. A general solution, in the same sense as for the two-state case, to this set of conditions is not known but has been analysed for a small set of scenarios.

Unambiguous state discrimination

In unambiguous state discrimination schemes, a measurement with $N + 1$ possible outcomes is required to distinguish N states from each other. This is because one of the possible outcomes is inconclusive, i.e., if that outcome occurs the experimenter can only guess which state was prepared according to the prior probabilities. Each of the other outcomes corresponds to a possibly prepared state but, in contrast to the minimum-error measurement, identifies that state with certainty. By incurring the cost of some inconclusive results, one is able to herald the success of the measurement. Unambiguous state discrimination is less well-explored for mixed states, so I present here only the canonical pure state formalism due to Ivanovic, Dieks and Peres [19, 20, 21].

Again, consider that the states $|\psi_0\rangle$ and $|\psi_1\rangle$ are prepared with probability p_0 and p_1 respectively. A POVM with elements π_0, π_1 which satisfy

$$\langle\psi_0|\pi_1|\psi_0\rangle = \langle\psi_1|\pi_0|\psi_1\rangle = 0 \tag{2.60}$$

will have the property that if π_0 is found then $|\psi_1\rangle$ could not have been measured: the

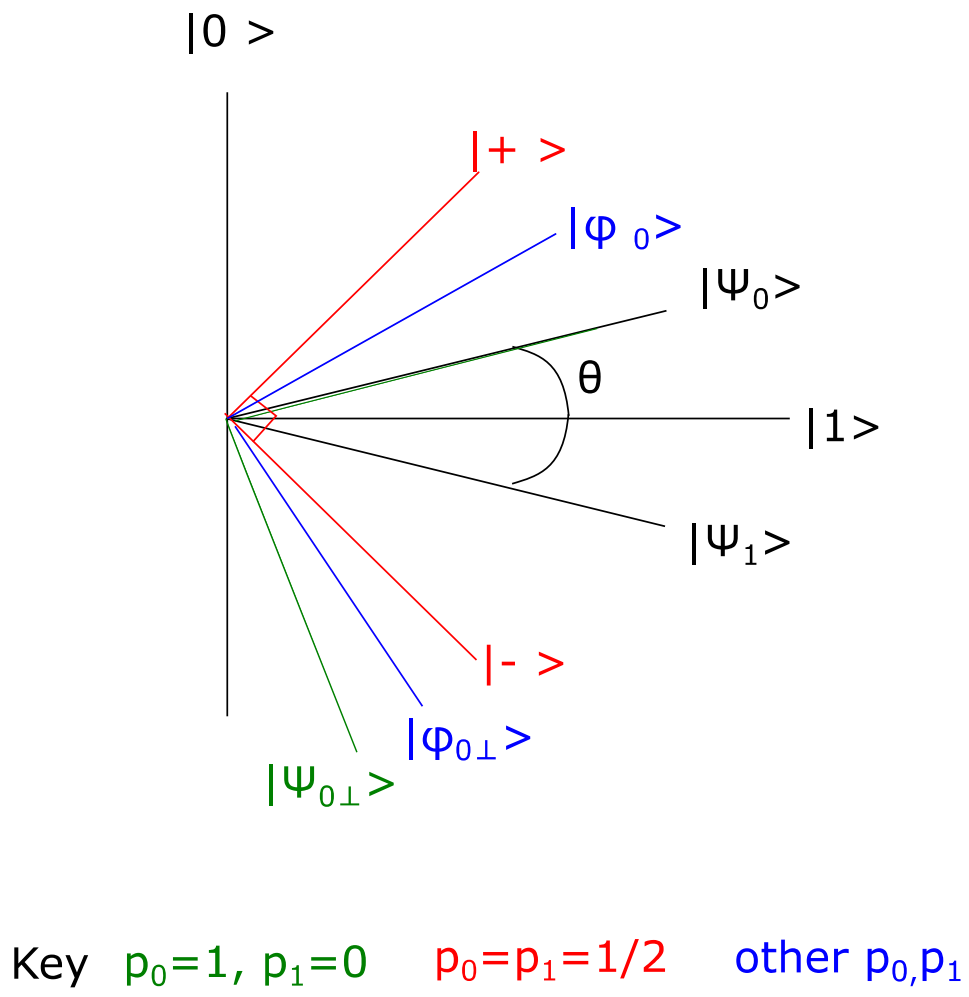


Figure 2.2: Graph displaying the optimal measurements for distinguishing between two states $|\psi_0\rangle$ and $|\psi_1\rangle$ are separated by an angle 2θ and which are prepared with varying prior probabilities. All vectors should be normalised. In red is the basis $|+\rangle, |-\rangle$, which is the optimal discriminating measurement if the two states are equiprobable. In green is the other end of the scale. If one particular state is definitely prepared then the basis must contain that state. In between, a general set of prior probabilities will mean that the best measurement is satisfied by a basis $|\phi_0\rangle, |\phi_1\rangle$, defined according to the Helstrom measurement in Eq. 2.56.

prepared state is guaranteed to be $|\psi_0\rangle$. The requirement can be achieved if π_0 is a projector onto the state orthogonal to $|\psi_1\rangle$, and similarly for the other effect π_1. However, as the two prepared states are in general non-orthogonal, this measurement would not satisfy $\pi_0 + \pi_1 = I$. By themselves the two projectors do not form a POVM. The solution is to weigh each projector by some constant of proportionality and complete the space with a third effect, $\pi_?$. The overall POVM is now

$$\begin{aligned} \pi_0 &= a_0 |\pi_{1\perp}\rangle\langle\pi_{1\perp}| \\ \pi_1 &= a_1 |\pi_{0\perp}\rangle\langle\pi_{0\perp}| \\ \pi_? &= I - \pi_0 – \pi_1. \end{aligned} \tag{2.61}$$

The third outcome, $\pi_?$, is inconclusive as it is equally likely for both possible states: $\langle\psi_0|\pi_?|\psi_0\rangle = \langle\psi_1|\pi_?|\psi_1\rangle$. As that outcome gives no information, the optimal scheme will minimise the probability that it occurs, where the degrees of freedom to optimise over are a_0, a_1. This process is subject to the further constraint that all three POVM elements stay positive, so a_0 and a_1 must both be positive but not so large that $\pi_? < 0$. I use the same parameterisation for the states, Eq. 2.53 as before. A short calculation reveals that the probability of the inconclusive result is

$$\begin{aligned} \mathrm{P}_? &= p_0 \mathrm{P}(?|0) + p_1 \mathrm{P}(?|1) \\ &= 1 – (a_0 p_0 + a_1 p_1) \sin^2(2\theta). \end{aligned} \tag{2.62}$$

Minimising this for general priors is not straightforward but was performed by Jaeger and Shimony [22]. More useful is to focus on the equiprobable case, $p_0 = p_1 = 1/2$. In this case the usual methods of constrained optimisation give $a_0 = a_1 = 1/2\cos^2(2\theta)$. Substitution into the above gives $\mathrm{P}_? = \cos(2\theta)$ and hence

$$\mathrm{P}_{succ} = 1 - \cos(2\theta) = 2\sin^2(2\theta). \tag{2.63}$$

This stays within the expected bound as $0 \leq \theta \leq \pi/4$. This quantity is the Ivanovic-Dieks-Peres limit.

Unambiguous state discrimination becomes more complicated if there are more than two pure states. In the two state case, what allows the scheme to work is that it is possible to perform a measurement which rules out a state, i.e., in some sense it makes more sense to say that π_0 is associated with the state which isn't $|\psi_1\rangle$ than it does to say that it identifies $|\psi_0\rangle$. In the two-state case this is just semantics, but when more than two states are involved then it is not. One requires a measurement which satisfies $\langle\psi_i|\pi_j|\psi_i\rangle = 0$ for all $i \neq j$, and this condition can only be satisfied if the set of states $|\psi_i\rangle$ are linearly independent. This result was first shown by Chefles, who provided one of the only analyses of the three-or-more state case [23].

Multiple-copy discrimination

In some applications of quantum information processing, one needs to discriminate between a set of states given a resource of N copies of the state. In this scenario, it is possible to outperform the single-copy Helstrom bound. If there are two signal states then a global measurement of all N copies will in principle be able to reach the multiple-copy Helstrom bound

$$\mathrm{P}_{succ}^{N} = \frac{1}{2}\left(1 + \sqrt{1 + 4p_0 p_1 \cos^{2N}(2\theta)}\right), \tag{2.64}$$

where all symbols have the same meaning as when used above. At first glance, it is not clear which measurement will reach this bound. The same reasoning as above suggests that it is a POVM on the multiple-system Hilbert space $\mathcal{H}^{\otimes N}$ however this involves finding the eigenvalues of the matrix $p_0\rho_0^{\otimes N} - p_1\rho_1^{\otimes N}$ and it is not straightforward to perform the eigendecomposition of an N dimensional operator, despite the possible symmetries. Furthermore, in the absence of quantum memories, it is still experimentally difficult to interact unitarily with so many systems, which leads to one asking whether it is possible to reach the Helstrom bound with just local measurements. Local measurements are split into two classes. Fixed measurements are those in which the same measurement is performed on each system and the overall result is assigned based on the majority result. Adaptive measurements are those in which each measurement depends upon the previous results.

For the basic case of discriminating two pure states, a local adaptive scheme is known which reaches the Helstrom bound and it turns out to be the simplest form which one might expect. At each stage, one performs the Helstrom measurement for the equivalent single copy case, but updates the prior probabilitis p_0 and p_1 based upon the measurement record. It even turns out that this Bayesian scheme is Markovian, in that it only depends on the directly prior outcome: the best measurement at each stage consists of one of two POVMs, corresponding to the measurement outcome at a previous stage. The signal state predicted by the scheme is the outcome of the final measurement; it depends on no other results in the measurement record. This scheme is discussed in Chapter 5, where I present the POVM at each step and show that it reaches the Helstrom bound. There, it is contrasted with a scheme called 'quantum data gathering', which requires a quantum memory and also reaches the Helstrom bound. Despite its increased experimental complexity, the system generalises straightforwardly to distinguish three-or-more pure states, which problem is known to not be solved by Bayesian updating.

The task of optimally discriminating mixed, rather than pure, states in the multiple-copy case is a much more difficult task, and I am aware of only numerical results in this area. The first point to note is that the distinction between globally optimal and locally optimal measurements breaks down. In the adaptive pure scheme, the measurement which is performed on the first $n < N$ qubits is also the measurement which would best discriminate the two states if the resource was n qubits only in total. That scheme is thus both globally optimal and locally optimal. For mixed states, this does not hold. A counterintuitive result, for example, is that a locally-optimal adaptive scheme performs

worse globally than a fixed, majority-voting scheme.

2.5 Quantum key distribution

Among the many proposed applications of quantum information theory, two uses have received the most attention. Scalable quantum computers are believed to have intrinsic advantages over their classical counterparts, but it seems likely that the technical jumps needed to build a quantum computer will take at least a decade. Quantum communications devices, on the other hand, are provably secure even with current technologies and are commercially available. Quantum key distribution forms the basis of these devices [24]. It consists of a set of protocols which share a key between two parties (who we have come to know as Alice and Bob) and which use the phenomenon of measurement backreaction, which does not occur in classical signal transmission, to alert the legitimate parties to the presence of an eavesdropper (Eve).

The resources required for a typical quantum key distribution protocol are a random number generator, a quantum channel and a classical channel. The latter is authenticated, so that the receiving party has confidence in the identity of the transmitter which is typically achieved by transmitting a pre-shared set of bits. Alice uses the random number generator to select, from a known set, a quantum state which is encoded on a quantum system and sent to Bob. Bob measures the system. Each state that could have been prepared as well as each possible outcome is assigned to one of the two classical bit values, zero or one. After the set of measurements the two parties share information, over the classical channel, which leaves them with correlated strings of bit values.

This correlated data can be used to form a shared secret key by a set of techniques from classical cryptography. They first share a set of their classical bits which allow them to characterise the quantum channel. At this stage they can detect the presence of Eve if the noise is higher than expected. Even if an eavesdropper is present, it is still possible for them to distil a secret key as long as the noise is below the quantum bit error rate, an information-theoretic property of the protocol. Otherwise, they abort at this stage. The next steps are reconciliation, which corrects for the noise in the channel and leaves the two parties with a shared string (which nonetheless will be partially known by Eve), followed by privacy amplification, which decreases the length of the string but leaves Alice and Bob with a private key. The success of a protocol is quantified by the key rate, which is the length of the private key that can be generated per unit time. It is this object which researchers seek to maximise.

I have presented here the prepare-and-measure based scheme, in which Alice prepares a qubit which is measured by Bob. It is common to find quantum key distribution analysed using entanglement-based schemes. In such an approach, the two legitimate parties share a maximally-entangled state, i.e., one of the four Bell states. Each party then measures their qubit - Alice by implementing a POVM equivalent to her set of prepared states, Bob with his original POVM. Formally, this scheme is equivalent to the prepare-and-measure scheme. In entanglement-based quantum key distribution, Eve's range of attacks can be formulated in greater generality, which is why it is more prevalent in the literature.

Eavesdropping and security

I have so far commented only upon the actions of the two legitimate parties. A third, Eve, seeks to learn their private key and cryptanalysis partly involves determining her best attacks. An assumption which underlies this analysis is Kirchoff's principle, which states that one should assume that the eavesdropper has access to all information (i.e., they know the protocol, which states are being sent, all the classical announcements; the correlated pairs of states which are associated with each bit value) apart from the measurement outcomes.

The set of possible attacks are split into three classes, each of differing levels of complexity [25]. The simplest attacks, and those which can be implemented with current technology, are called *individual* attacks. Here, Eve interacts in the same way with each transmitted qubit and is assumed to measure before reconciliation. (She may also measure multiple qubits however each measurement must be the same.) Individual attacks are the topic of Chapter 5 of this document. The most general type of measurement of this kind is that Eve interacts each qubit with an ancilla which is then measured. A measurement of this kind is the most general allowed on individual qubits as shown by Naimark's theorem.

The next-simplest attacks are *collective* attacks. Eve requires one quantum memory (i.e., a qubit which does not decohere) for each transmitted qubit. She interacts all qubits with an individual quantum memory through a unitary, and stores them until reconciliation has occured. Then, she performs a collective measurement (one on the product system of all quantum memories) based on the announcement at this stage. As might be expected, given that Eve has further information, she is able to learn more of the key than in individual attacks.

The most general attacks are those in which Eve can do anything consistent with the laws of quantum mechanics. The assumption which is relaxed, compared with collective attacks, is that the unitary interactions are the same with each transmitted qubit. In *coherent* attacks all N transmitted qubits interact with N quantum memories through a unitary which is able to vary across all subspaces. The measurement, which depends upon the announcement during reconciliation, can also be a global measurement on the product space.

The technological advances required to perform collective or coherent attacks are unlikely to occur for several years but one cannot base a communication system on the principle that the eavesdropper is precisely as powerful as oneself. It is also reasonable to require that any commonly used system will be able to withstand attacks that become available in the next few years, as moving to a new security infrastructure can be highly disruptive. For this reason, information-theoretic security, against arbitrary attacks is required. Privacy amplification allows Alice and Bob to create a secret key from their set of data after reconciliation. The amount of information in the final secret key is the difference between that in the pre-privacy amplification bit string and the upper bound on the information accessible to Eve, i.e., the secret key has length $I(A:B)$–I_E where $I(A:B)$ is the mutual information between Alice and Bob. The measure of information used depends on the type of attack considered; for individual attacks it is simply the Shan-

non information as Eve is left with purely classical bit values, however, this changes for the more general attacks. Security proofs consist of analyses of specific protocols which find an upper bound on how much information Eve can extract, and this gives a level of classical correlation, the quantum bit error rate, between Alice and Bob below which a secret key can be generated. A full treatment of this subject is not required for what follows and is quite involved.

Example: BB84

The discussion is clearer with a concrete example. By far the most widely discussed quantum key distribution protocol is BB84. It was the first to be proposed, forms the basis of most commercial technologies, and has the highest quantum bit error rate.

In BB84, Alice selects her qubits from a set of four possible states, either the computational basis $|0\rangle, |1\rangle$ or the σ_x basis, $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$. The state is then sent to Bob who performs a POVM corresponding to choosing one of the two bases, with equal probability, as his measurement basis. Alice and Bob then announce in which basis they chose to prepare and measure. Only if the two match are the outcomes kept in the record. Bob's outcome $|0\rangle$, for example, is consistent with all possible transmissions except $|1\rangle$, so he requires Alice to say that she prepared in the computational basis in order that he knows which state she did send. If there is no noise, or eavesdropper, then the two parties will know each other's record, and a key can be formed (the classical bit value 0 is assigned to $|0\rangle$ and $|+\rangle$, the classical bit value 1 is assigned to $|1\rangle$ and $|+\rangle$). This step of the protocol is then repeated for every qubit and the post-measurement processes (channel evaluation, reconciliation, privacy amplification) are implemented.

If there is no eavesdropping, Alice and Bob will discard half of all qubits and will be left with a secret key. An eavesdropper will hope to know the bit values in that key and perform a measurement on the transmitted qubits. The fifth chapter of this thesis is dedicated to optimising that attack and so I consider just a limited type of attack here: measure-resend. Eve performs a von Neumann measurement individually on each qubit and needs to calculate the best basis in which to perform it. The sense of optimality I consider here is simply her chance of identifying which classical bit value, rather than the state, Alice transmitted and assume she is not at this point worried about revealing herself through Bob's subsequent measurements. I assign to each of Eve's outcomes the states

$$\begin{aligned} |E_0\rangle &= \cos(\theta)|0\rangle + \sin(\theta)|1\rangle \\ |E_1\rangle &= \sin(\theta)|0\rangle - \cos(\theta)|1\rangle. \end{aligned} \tag{2.65}$$

The task is to find the value of θ which maximises her chance of identifying Alice's transmitted bit value. To do this, I associate with each bit value a density operator

$$\begin{aligned} \rho_0 &= \frac{1}{2}\left(|0\rangle\langle 0| + |+\rangle\langle +|\right) \\ \rho_1 &= \frac{1}{2}\left(|1\rangle\langle 1| + |-\rangle\langle -|\right). \end{aligned} \tag{2.66}$$

The factor of one-half represents the fact that, given the bit value 0, each of the states representing that bit will occur in half of cases. The probability that Eve gets the correct outcome is

$$\begin{aligned} \mathrm{P}_E &= \sum_i \mathrm{P}(E=i|A=i)\mathrm{P}(A=i) \\ &= \frac{1}{2}\left(\langle E_0|\rho_0|E_0\rangle + \langle E_1|\rho_1|E_1\rangle\right). \end{aligned} \tag{2.67}$$

The factor of one-half here is the probability that Alice chose either zero or one as her bit value. Evaluating the expectation values gives

$$\mathrm{P}_E = \frac{1}{4}\left(2+\cos(2\theta)+\sin(2\theta)\right). \tag{2.68}$$

This function maximises when $\theta = \pi/8$ to give $\mathrm{P}_E = (2+\sqrt{2})/4 \approx 0.85$ as the highest probability of Eve's attack being successful. This attack is to measure in the

$$\begin{aligned} |0_B\rangle &= \cos(\pi/8)|0\rangle + \sin(\pi/8)|1\rangle \\ |1_B\rangle &= \sin(\pi/8)|0\rangle - \cos(\pi/8)|1\rangle \end{aligned} \tag{2.69}$$

basis. This pair of vectors is known as the Breidbart basis. Here, I have simply given a flavour as to how eavesdropping is optimised, as a precursor to the more involved calculations of Chapter 5, in which I consider a more general class of attack than projective measurements.

Generalisations

I have restricted this introduction to the set of quantum-key-distribution protocols which have two properties. Firstly, the bit values are encoded on discrete states (e.g., a photon's polarisation). Secondly, the security of the protocols relies on an assumption that Alice and Bob have complete control over the devices they use. Both of these are technological limitations and have been overcome both theoretically and in practice.

Discrete-state quantum key distribution, which is what I discuss above, has a limited key rate due to the difficulty in detecting single photons. This issue is solved by continuous-variable quantum key distribution [26, 25]. Instead of transmitting single photons, photons in a coherent state can be sent and what is measured is the momentum-phase quadrature of the photon. The absence or presence of a photon thus appear as regions of phase space (Gaussian distributions around both the axis and a selected other point). Everything else in a continuous-variable protocol is precisely the same. This modification allows for an increased key-rate as quadrature measurements are implemented by heterodyne and homodyne measurement, which are a more advanced technology than single-photon detection.

The assumption that Alice and Bob can trust their measurement devices is relaxed in device-independent and measurement-device independent protocols [27, 28, 29], in which it can even be assumed that Eve has as much control over the devices as is allowed by

the laws of quantum mechanics. If Alice and Bob share an entangled state, e.g., a Bell state, this can be done by including a Bell test, which verifies that the correlations are intrinsically quantum rather than classical, as one step of the protocol. If this test fails, the legitimate users are alerted to the fact that their equipment is untrustworthy and can abort the protocol.

Chapter 3

Kraus formalism from first principles

A fundamental component of quantum theory is Born's rule, which relates operators with probabilities. In the most simple form, as a representation of an experiment in which a pure state $|\psi\rangle$ is measured projectively, with outcome represented by the state $|\phi\rangle$, it is written as

$$P(\phi|\psi) = |\langle\psi|\phi\rangle|^2. \tag{3.1}$$

It is not obvious why the rule has this particular form. After all, it is well known that Born himself initially associated the wavefunction's amplitude, rather than the squared magnitude of that quantity, with the probability of finding a particle at a given point in space [30]. In order that the theory did not contain negative probabilties, he corrected himself to the accepted form in a footnote however a number of other forms can be constructed on an ad hoc basis that also satisfy that requirement. Why does the formula not use the fourth power of the magnitude, for example? A wide range of authors [31, 32, 33, 34, 35, 36, 37] have analysed this issue. They all find that it is impossible to construct any probabilistic theories which deviate from this form while satisfying some basic physical requirements.

Closely associated with the Born rule is what might be called Kraus's rule [13], which defines the joint probability of two particular consecutive outcomes when measuring the same system, and which is discussed in §2.5. For convenience, I restate the probability rule here. The probability of an experimental outcome associated with the effect $\pi_j^{(2)}$ following an outcome associated with the effect $\pi_i^{(1)} = A_i^{(1)\dagger} A_i^{(1)}$, given that a system has been prepared in the state ρ, is given by

$$\mathrm{P}(i,j|\rho) = \frac{\mathrm{Tr}(\pi_j^{(2)} A_i^{(1)} \rho A_i^{(1)\dagger})}{\sum_{ij} \mathrm{Tr}(\pi_j^{(2)} A_i^{(1)} \rho A_i^{(1)\dagger})}. \tag{3.2}$$

It is natural to ask, in a similar manner as above, which physical assumptions underpin this equation. This is a non-trivial generalisation of the Born rule result(s) in that the first measurement has a back-reaction onto the system's state, given by the Lüders rule

[38]. This begs the question of whether the state update rule is an additional assumption on top of those needed for the probability formula.

I analyse the question of the uniqueness of Kraus's rule in this chapter, which is based on Ref. [1]. I will first contexualise my work with respect to the field of quantum reconstructions and what I refer to as the Gleason-Busch theorem, which pinpoints the sense in which the Born rule is unique. After that, I introduce some results from the theory of operator space and use them to derive firstly the Born rule and then the Kraus rule. Both of these are linked to a set of operational postulates related to the idea that counting individual outcomes constitutes the measurement of a probability. The central point is that measurement back-reaction is not an extra postulate in itself, but is part and parcel of the same set of assumptions from which the Born rule arises. Finally, I provide some short examples which link the mathematical objects used here back to standard quantum theory, and connect this work to some other results, including two-time states and the conditional state theory of Leifer and Spekkens.

3.1 Context

The earliest questions about the uniqueness of the Born rule can be dated back to the publication of von Neumann's *Mathematical Foundations of Quantum Mechanics* [39]. From a historical perspective, this is the first text to formulate quantum mechanics as a self-consistent theory which is derived from a number of postulates, and von Neumann's framing has stayed with us until the present day. In this formulation, the basic assumptions which lead to quantum mechanics are not physical in nature but take the form of mathematical statements, e.g., that observables are associated with Hermitian operators defined upon a Hilbert space. This approach has issues, which I discuss below, but it allows one to ask which aspects of the theory must be postulated and which can be derived.

An important early result in this area was Gleason's theorem [31]. At that point in time generalised measurements had not been formulated, and quantum foundations was focused on the study of the yes-no questions which can be asked by von Neumann measurements. These are associated with probabilties through the Born rule, considered as a map from the set of projectors to the set of real numbers. Because this function also requires a density matrix, the question of the Born rule's uniqueness is also probing the possibility of constructing a larger state space than that which can be represented by density matrices. In particular, some physicists were hopeful that it would be possible to introduce hidden variables through a hypothetical generalisation of Born's rule [5, 6].

Von Neumann's mathematical postulates can be replaced with alternative sets of minimal assumptions. For Gleason, these were that yes-no questions are associated with projectors, that probabilities are positive and form a complete set, and that, if the sum of two different sets of projectors is equal, then the sum of associated probabilities is also equal. Formally, the third postulate is

$$\mathrm{P}(P_{ij}) = \mathrm{P}(P_i) + \mathrm{P}(P_j) \tag{3.3}$$

where $P_{ij} := P_i + P_j$, P is the probability map (which is to be determined) and $P_i = |i\rangle\langle i|$. Gleason's theorem then states that the only map which satisfies these three postulates is $\mathrm{P}(P_i) = \mathrm{Tr}(\rho P_i)$, the Born rule, if P_i lies in a Hilbert space of two or more dimensions. The proof is geometric in nature and relies upon the transformation properties of the spherical harmonics under the rotation group SO(3).

Equation 3.3 can be understood as a claim that instantaneous non-local communication is impossible. Consider a three-level system spanned by the three orthonormal state vectors $|0\rangle$, $|1\rangle$ and $|2\rangle$ as well as the following two possible sets of measurement outcomes: $\{|0\rangle, |1\rangle, |2\rangle\}$ and $\{|0\rangle, |+\rangle = (|1\rangle + |2\rangle)/\sqrt{2}, |-\rangle = (|1\rangle - |2\rangle)/\sqrt{2}\}$. These two measurements correspond to a scenario in which, firstly, the zero outcome is filtered and those outcomes are passed on to one party, and, secondly, the remaining systems are sent to another party who measures in either the $|1\rangle, |2\rangle$ or $|+\rangle, |-\rangle$ basis. The fraction of states which is received by the former party could, if Eq. 3.3 did not hold, then depend upon the latter party's choice of measurement, and this property could then be the basis of an instaneous communication system. This is a stronger form of nonlocality than that represented by the Bell inequalities, which require the two parties to share their measurement outcomes. Thus, while no faster-than-light communication system could be constructed using the latter, it could if the former kind of nonlocality held. Gleason's theorem can then be thought of as an affirmation that quantum mechanics must be local.

As important as Gleason's theorem is, there are reasons to revisit it contemporarily. The first is that a concept of measurement has been developed in the interim period which associates possible outcomes with a greater class of objects than simply projectors [9, 10]. It is not obvious whether this allows for a more general class of probability rules, and thus states. A second issue is the somewhat opaque nature of Gleason's geometric proof. Significant study is required to understand his paper and the symmetry-theoretic argument makes it difficult to connect the physical postulates to the final result [40]. For both reasons, Busch [32] provided a simplified proof that the Born rule, alongside the density matrix, is the only map between effects and the set of real numbers. I refer to this more general result as the Gleason-Busch theorem and prove it below. That Busch's proof is much simpler is related to the larger set of measurements which he considers: Gleason's complicated formalism is designed to get around the issue that, in general, the sum of two projectors is not itself a projector. However, summing two effects does produce a third effect. Caves *et al.* [35] adapted Busch's proof such that it was in a similar form to Gleason's result, and this was built on by Barnett *et al.* [34], who relaxed one of the postulates and emphasised the Bayesian nature of quantum measurements. This allows the probability rule which is associated with quantum retrodiction [41, 42] to be derived from the same axioms.

Closely related to the work of Gleason and Busch is a movement in quantum foundations away from providing interpretations of quantum theory and towards providing sets of axioms from which quantum theory is derived: the field known as 'quantum reconstructions' [33, 43, 44, 45, 46]. A typical work in this field will begin by providing a set of physicial principles, then represent them mathematically, and then derive quantum theory

from these axioms. In such a manner, any mystery about the theorems of quantum mechanics (e.g., problems concerning the distinction between reversible evolution under the Schrödinger equation and irreversible 'collapse') is moved from the results themselves, now clearly defined, to the initial set of axioms which are the only free choice in this form of analysis. Quantum reconstructions are distinguished from previous axiomatisations (e.g., those due to von Neumann and Birkhoff [47] or Mackey [48]) by the physical nature of the initial postulates.

One of the first publications along this line is that due to Hardy [33]. In this work he introduces two parameters: N, defined as the number of states perfectly distinguishable within a system, and K, the number of degrees of freedom, defined as the number of measurements needed to distinguish two different states, within a system. Hardy then provides a relationship between these two variables upon the basis of his 'five reasonable axioms':

H1 Probabilities P_i are understood as the relative frequency of a given outcome i. For a large enough sample, this probability will be constant between different runs of the experiment.

H2 K is a function of N and takes the minimum value allowed by the five axioms, i.e., the number of parameters needed to describe states is minimally linked to the number of measurements needed to distinguish them.

H3 Systems of the same dimension (i.e., systems which can hold the same amount of information) all behave the same.

H4 Composite systems $A \otimes B$ are multiplicative: $K_{AB} = K_A K_B$ and $N_{AB} = N_A N_B$.

H5 There exists a continuous reversible transformation of a system between any two pure states.

These form a basis for deriving all relevant aspects of quantum mechanics: the Born rule, the Hilbert space structure, the state update rule. Hardy starts by showing from H1 that all probabilities can be associated with inner products, and it is the space on which these inner products are defined which is fixed by the remaining axioms. This move is repeated in my own work as in most research into quantum axioms. The fifth axiom is found to be the most important as it is this that introduces the possibility of associating probabilities with superpositions of what he calls 'fiducial measures', which are most straightforwardly translated into the usual langugage as pure states. It is worth noting that this axiom has been criticised for lacking the direct physical meaning of the other axioms [43]. What does an instrumentalist mean when they say that transformations exist between two pure states? Indeed, the continuity implicit here is very similar to a lemma in Gleason's proof that requires him to employ the theory of symmetries. However, this issue does not detract from the main thrust of the paper: that it is indeed possible to provide simple physicial principles for quantum theory.

In the years following Hardy's paper, a number of other reconstructions have appeared which are typically information-theoretic in nature. Among these the most notable is

that due to Chiribella, D'Ariano and Perinotti [44, 46]. The main result of their work is a demonstration that it is, in principle, not possible to construct a physical theory in which state purification is possible without at the same time introducing the uncertainty principle in the form of measurement back-reaction onto the system. This important result emphasises that results which are previously thought of as independent may become entwined upon closer inspection.

Closely related to the work which I am presenting here are two papers in particular. One is an article by Cassinelli and Zanghi [49] which generalises Gleason's theorem to find the Lüders rule for updating states. This is of particular interest as it is an early example of a demonstration that probability measures and state updates are closely linked, however it is distinct from the work presented here for a number of reasons. One is that their work does not proceed from physical postulates. Their work was performed in the context of quantum logic and what might be understood as a linguistics-inspired framework for deriving quantum mechanics, in which measurements are associated with truth statements, and it becomes important to ensure that the implications and negation of that statement are well-defined in the theory. This is very different from the operational arguments that I present below. More importantly, their approach requires the use of a projector which is orthogonal to that being mapped onto probabilties and such an object does not exist for a general positive operator. Hence, their result cannot be generalised to include effects. Further relevant work is that by Shrapnel *et al.* [50]. This axiomatic work begins by assuming that transformations are defined by completely positive maps and use this, alongside a set of postulates, to derive the Born rule and state update rule. This is the opposite approach to that presented here, in which I make some assumptions about what consitutes a measurement and then derive completely positive maps as a consequence of these requirements. While Shrapnel *et al.*'s work is interesting, it lacks the instrumentalist flavour that I seek, as the concept of a completely positive map is abstracted away from the measurement process.

What I have taken from the various derivations of the Gleason-Busch theorem is some specific insights about the mathematical representations required. From quantum reconstructions, I have taken the need to ground any mathematical postulates in physical arguments. I discuss these below, after a mathematical detour in which I introduce operator space.

3.2 Operator space

As in most quantum reconstructions, I begin by establishing a link between probabilities and inner products. The task is then to specify, based on a set of physical assumptions, the vector space on which the inner product is defined. Before discussing my choice of axioms, I introduce operator space. This is also known as Liouville space [51, 52]; both names are used equivalently throughout.

Operator space is defined as the tensor product of a Hilbert space, $\mathcal{H}$, with its dual space, $\mathcal{H}^\dagger$. The former can be considered the space of kets. The latter can be considered the space of bras or, more formally, the space of maps between vectors in $\mathcal{H}$ and the set

of complex numbers. An operator can be written as the outer product of a ket and a bra, hence it should be no surprise that there is a mapping between the set of operators and the just defined space. Typically, one writes

$$A \leftrightarrow |A\rangle\rangle, \tag{3.4}$$

i.e., an operator acting on vectors in $\mathcal{H}$ is itself a vector in operator space, the latter indicated by a doubly angled ket. This mapping is a form of the Choi-Jamiołkowski isomorphism [53, 54], which is between the sets of bipartite states and operations. One interpretation of this is in terms of gate teleportation however, as will be discussed later, this is not a unique reading [55].

I decompose the operator under consideration into an arbitrary basis in order to be more precise about the mapping:

$$A = \sum_{ij} a_{ij}|i\rangle\langle j| \leftrightarrow |A\rangle\rangle = \sum_{ij} a_{ij}|ij^\dagger\rangle\rangle. \tag{3.5}$$

Here, and in what follows, the superscripted dagger indicates a basis vector in the dual space. As a particular example, I consider the identity operation in two dimensions. In the Hilbert space representation this operator is

$$I = \sum_{i=0}^{1} |i\rangle\langle i| = |0\rangle\langle 0| + |1\rangle\langle 1|, \tag{3.6}$$

and, following the above isomorphism, in the Liouville space representation it is

$$|I\rangle\rangle = \sum_{i=0}^{1} |ii^\dagger\rangle\rangle = |00^\dagger\rangle\rangle + |11^\dagger\rangle\rangle. \tag{3.7}$$

The relevance of operator space for the issue of probability rules becomes more clear when the inner product is defined. It is straightforward to show that the natural way to define inner products is to associate them with the trace rule in Hilbert space. This can be seen by defining two operators, firstly A by Eq. 3.5 and secondly B by

$$B = \sum_{ij} b_{ij}|i\rangle\langle j| \leftrightarrow |B\rangle\rangle = \sum_{ij} b_{ij}|ij^\dagger\rangle\rangle. \tag{3.8}$$

The trace over the product of these two operators is

$$\mathrm{Tr}(B^\dagger A) = \sum_{ij} b_{ij}^* a_{ij}.$$

Similarly, in operator space their inner product can be calculated

$$\begin{aligned}\langle\langle B|A\rangle\rangle &= \sum_{ijkl} b_{ij}^* a_{kl} \langle\langle ij^\dagger | kl^\dagger \rangle\rangle \\ &= \sum_{ijkl} b_{ij}^* a_{kl} \delta_{ik} \delta_{jl} = \sum_{ij} b_{ij}^* a_{ij}.\end{aligned}$$

I use the usual notation for the Kronecker delta. Bringing together the above two results gives

$$\langle\langle B|A\rangle\rangle = \mathrm{Tr}(B^\dagger A) \tag{3.9}$$

and there is a natural association between inner products in the space of operators and the trace operation in the space of states, a property known as the Hilbert-Schmidt inner product [56]. The appearance of the dagger inside the trace is a choice; one could also have the inner product as $\langle\langle B|A\rangle\rangle = \mathrm{Tr}(BA)$. If this inner product was used, then the map from operator space to the dual space (i.e., from the space of double-angled kets to double-angled bras) would not need complex conjugation of the coefficients. Throughout, I have chosen to preserve that property from Hilbert space and so use the above form. There is a greater freedom in the choice of inner product in that a linear map may act on one of the two operators so that the inner product would instead be $\langle\langle B|A\rangle\rangle = \mathrm{Tr}(\mathcal{L}(B^\dagger)A)$ [57] however it will be seen later that this freedom has a natural interpretation in terms of the measurement process such that I can use Eq. 3.9 without loss of generality. That the inner product for operator space is the trace operation emphasises the close link between inner products and probabilities.

The final obects to define are linear operators in Liouville space, which correspond to superoperators in Hilbert space. As an example, I consider a transformation between two operators A and B:

$$A \to B = L_1 A L_2^\dagger, \tag{3.10}$$

where L_1 and L_2 are two unconstrained operators. In operator space, this pair of operations are represented by single linear operator which can be defined as $L = L_1 \otimes L_2$. The operator space representation of the above transformation is

$$|A\rangle\rangle \to L|B\rangle\rangle = (L_1 \otimes L_2)|B\rangle\rangle. \tag{3.11}$$

In order to make clear the objects that are being used here, it is useful to follow a basic calculation using both standard Hilbert space quantum mechanics and the alternative which I have introduced here. A simple scenario is that a system is prepared as $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and then measured in the $|0\rangle, |1\rangle$ basis. A calculation of the outcome probability and post-measurement state demonstrates the basic physics involved in such a process.

In 'standard' quantum mechanics, one begins by writing down the density operator $\rho = |+\rangle\langle+|$, which represents the prepared state, and the projector $P_0 = |0\rangle\langle 0|$ associated

with the relevant measurement outcome. The probability of this outcome is

$$\mathrm{P}(0|+) = \mathrm{Tr}(|+\rangle\langle+|0\rangle\langle 0|) = \frac{1}{2} \tag{3.12}$$

and the system's post-measurement state will be

$$\rho \to \rho' = \frac{|0\rangle\langle 0|+\rangle\langle+|0\rangle\langle 0|}{\mathrm{Tr}(|0\rangle\langle 0|+\rangle\langle+|0\rangle\langle 0|)} = |0\rangle\langle 0|, \tag{3.13}$$

where the denominator is a factor of normalisation. Both of these solutions are examples of textbook quantum theory.

A different approach uses the framework of operator space. One starts by representing the prepared state by the vector $|\rho\rangle\rangle = |+ +^\dagger\rangle\rangle$ and the projector by the vector $|P_0\rangle\rangle = |00^\dagger\rangle\rangle$. Probabilities are calculcated by writing the Born rule as an inner product, using Eq. 3.9:

$$\mathrm{P}(0|+) = \langle\langle+ +^\dagger |00^\dagger\rangle\rangle = \frac{1}{2}. \tag{3.14}$$

To find the post-measurement state, we can construct a superoperator. Following the discuss above, the required object is $P_0 = |0\rangle\langle 0| \otimes |0^\dagger\rangle\langle 0^\dagger| = |00^\dagger\rangle\rangle\langle\langle 00^\dagger|$. The operator space analogue to the state-update rule is

$$|\rho\rangle\rangle \to |\rho'\rangle\rangle = \frac{|00^\dagger\rangle\rangle\langle\langle 00^\dagger| + +^\dagger\rangle\rangle}{\langle\langle 00^\dagger| + +^\dagger\rangle\rangle} = |00^\dagger\rangle\rangle. \tag{3.15}$$

The results obtained by using objects in operator space are precisely the same as those found previously, demonstrating that it is possible to represent quantum theory in both Hilbert and Liouville spaces. At this point I have given no justification for using the more unfamiliar methods over the tried and tested formalism. Over the rest of this chapter, I hope to demonstrate that operator space provides a natural framework for handling sequences of quantum measurements.

3.3 Operational postulates

The aim of this chapter is to derive the framework of sequential measurement theory from a small set of postulates. In keeping with the spirit of quantum reconstructions, these are inspired by a straightforward interpretation of probabilities as representing relative frequencies of measurement outcomes. However, the principles do not have the same status as those that would be used in a quantum reconstruction. The reason for this is that the postulates are formulated as restrictions upon a map which acts upon the set of effects, without deriving the latter objects themselves as ingredients of measurement theory. Indeed, it is not clear why one would make such a mathematical representation for a physical process without prior knowledge of quantum theory. However, I would argue that the application of this work is not limited by this, and could still find use as part of a work which is wider in scope. The idea would be to derive, by adding further axioms, the idea of using a positive operator to represent measurement outcomes. Here I follow

the works of Gleason and Busch in which proofs start by considering effects to be a basic quantity. For a sequence of measurements, this map will act upon the effect associated with the later occuring outcome, and I will include the first by allowing the map to depend on that measurement outcome. This cannot be the case for a single measurement and so, strictly speaking, two maps are required, and hence two different sets of postulates. The maps are associated with each other by an assumption of the measurement process's causal order.

3.3.1 Single measurements

I begin with the assumption that observables in quantum theory are associated with a positive semi-definite operator, any POVM element π_i. This is represented by the operator space vector $|\pi_i\rangle\rangle$. Probability rules are then understood as consisting of maps ν between objects of this kind and real numbers. The term introduced by Gleason [31] for maps of this kind is 'frame functions'. Frame functions are related to probabilities by

(P0) $\mathrm{P}(i|s,x) = N(s,x)\nu(|\pi_i\rangle\rangle)$,

where s is the preparation procedure and x the measurement procedure (i.e., it represents how the POVM is completed as well as the physical process which occured). The factor $N(s,x)$ is a normalisation constant which is allowed to depend upon the experimental method but not the particular outcome. Postulate P0 is an assumption of *noncontexuality*, here used to mean that the probability of an outcome associated with a particular POVM element is the same however the set is completed, up to a constant of normalisation. I make this requirement in order to prevent the kind of instantaneous communication systems [5, 6] which were discussed in the context of Gleason's theorem. It is worth noting that, if quantum theory was contexual, then effects would no longer be able to consistently define a particular measurement outcome. (This point is discussed by Caves *et al.* [35]. In summary, the point is that, if probabilities could depend upon more than one member of a set, they would no longer be linear functionals of a single effect. Hence the POVM description of a measurement would not be possible.) An element of noncontexuality is being assumed implicitly.

On top of this generic definition, the other postulates are physically motivated requirements for the set of probabilities. For single measurements, I adapt Busch's postulates [32]:

(P1) $0 \leq \mathrm{P}(i|s,x) \leq 1$

(P2) $\sum_i \mathrm{P}(i|s,x) = 1$

(P3) $\mathrm{P}(i|s,x) + \mathrm{P}(j|s,x) + \ldots = \mathrm{P}(i \text{ or } j \text{ or } \ldots |s,x)$.

Each of these can be defended with reference to a simple, idealised, measuring device. When one measures a single quantum system, this device counts the number of occurences of a finite set of outcomes (N_0 for one, N_1 for another, etc.) and probabilities are then assigned simply as the fraction of each outcome out of the total number N of measurements: $P_0 = N_0/N$. If a large enough number of experiments are performed then one

would expect this quantity to tend towards a constant value. The three postulates above are then easily understood. P1 follows from the fact that the ratio of two positive integers will be positive and P2 follows from the fact that one outcome must occur. The additivity postulate, P3, codifies the concept of 'coarse-graining'. Associate with two outcomes in one description just a single outcome in a different description: i.e., outcome A is associated with outcomes zero and one. It must be true that $P_A = P_0 + P_1 = (N_0 + N_1)/N$, which is what the third postulate requires. In the next section, it will be seen that these assumptions alone are enough to ensure that the only probability measure is the Born rule.

3.3.2 Sequential measurements

I derive also the Kraus rule [13], which requires a different set of postulates. The focus is on processes in which a system is measured twice, with outcomes represented by the POVM elements $|\pi_i^{(1)}\rangle\rangle$ and $|\pi_j^{(2)}\rangle\rangle$ respectively. I outline how to extend to cases involving three or more measurements in a later section. In order to include the conditional nature of quantum sequential measurements, I modify the noncontexuality postulate P0 to allow the map to depend upon the first measurement outcome. I use the joint probability distribution

(A0) $\mathrm{P}(i,j|s,x) = N(s,x)\omega_i(|\pi_j^{(2)}\rangle\rangle)$.

The definition of $N(s,x)$ as a representation of the contextual information can be kept as long as one is careful to distinguish between that concept for individual and sequential measurements (if it necessary to distinguish the two, I will use the symbol s' for the sequential case). This formula can be thought of as a manifestation of the idea that the measurement sequence defined above can be thought of as a single measurement if the preparation-first measurement segment is understood as a single preparation procedure. As I keep the constant of proportionality as a representation of the general features of the procedure, conditioning upon the first measurement is a property of the map. This is why a different set of postulates is required: they are requirements upon a different map. It is of course possible to take a different perspective, that the measurement sequence is still a preparation and single measurement process but that both outcomes are associated with a single operator, which can be labelled $|\pi_{ij}\rangle\rangle$. In this case, by P0 one has

$$\begin{aligned}\mathrm{P}(i,j|s,x) &= N(s,x)\nu(|\pi_{ij}\rangle\rangle) \\ &= N(s',x)\omega_i(|\pi_j^{(2)}\rangle\rangle).\end{aligned} \tag{3.16}$$

Thus the newly introduced map ω_i is proportional to my initial construction and the only way that this can hold is if the i-dependence appears as a map acting upon $|\pi_j^{(2)}\rangle\rangle$, i.e. ,

$$\omega_i(|\pi_j^{(2)}\rangle\rangle) = \nu(\mathcal{T}_i(|\pi_j^{(2)}\rangle\rangle)). \tag{3.17}$$

The reader should note that, at this point, I do not introduce $\mathcal{T}_i$ as a *linear* superoperator. In principle there is enough freedom for it to be some other map and this usual requirement

will be seen to emerge from the remaining axioms. When I derive the Kraus form, it is the form of the new superoperator $\mathcal{T}_i$ that is found. I require that the joint probability satisfies

(A1) $0 \leq \mathrm{P}(i,j|s,x) \leq 1$.

(A2) $\sum_j \mathrm{P}(i,j|s,x) = \mathrm{P}(i|s,x)$.

(A3) $\mathrm{P}(i,j|s,x) + \mathrm{P}(i,k|s,x) + \ldots = \mathrm{P}(i,j \text{ or } k \text{ or} \ldots |s,x)$.

These postulates can be thought of as straightforward generalisations of those introduced above and have the same reasoning associated with them. Special attention may be paid to A2, which is underpinned by another assumption: that of a specific causal order in which the results of the first measurement may not depend upon the second. At this point I could still consider a general causal order, i.e., a generalisation in which the results of the first measurement can be influenced by later results. This would be in keeping with the programme of research concerning 'indefinite causual order' [55, 58]. However, the method that I follow is to assume the fixed causal order just discussed. One could also read an assumption of noncontexuality in the same postulate, in the sense that this requires that any post-processing (e.g., discarding) of specific measurement outcomes will not change the probability that they occured in the first place.

The postulates are now used to reconstruct Born's and Kraus's rule, along with the related concept of the state update rule and the idea of two-time states to represent pre- and post-selection.

3.4 Single measurements

In this section I derive the Born rule, adapting Busch's (and Barnett *et al.*'s) proof [32, 34] of Gleason's theorem to the language of operator space. This demonstrates my operational approach and provides useful results for the next step, joint probabilities.

The first step towards the Born rule is to extend the additivity postulate, P3, to allow for linearity, i.e., what is required is to show that $\nu(\alpha|E\rangle\rangle) = \alpha\nu(|E\rangle\rangle)$. I follow the procedure of Barnett *et al.* Once linearity is established, the trace operation follows directly. At this point I consider a set of vectors $|E\rangle\rangle, |F\rangle\rangle \ldots$. These should correspond to positive operators but, for the sense of establishing linearity, need not be effects. (Of course, the interpretation in terms of experimental outcomes does not hold for the more general set.) One must first note that, since all individual outcomes in an experiment are subject to the same preparation and measurement context, the relevant probability formulae will all contain the same $N(s,x)$ and the additivity of probabilities (from P3) extends to additivity over the map $\nu(\cdot)$:

$$\nu(|E\rangle\rangle) + \nu(|F\rangle\rangle) + \ldots = \nu(|E\rangle\rangle + |F\rangle\rangle + \ldots). \tag{3.18}$$

Now, consider an integer n. By this sense of additivity,

$$n\nu(|E\rangle\rangle) = \nu(|E\rangle\rangle) + \nu(|E\rangle\rangle) + \ldots = \nu(n|E\rangle\rangle), \tag{3.19}$$

where n is an integer. The property of additivity enforces linearity over the integers. I introduce a second integer, n' and write

$$\begin{aligned} n'\nu(\frac{n}{n'}|E\rangle\rangle) = \nu(n|E\rangle\rangle) = n\nu(|E\rangle\rangle) \\ \implies \frac{n}{n'}\nu(|E\rangle\rangle) = \nu(\frac{n}{n'}|E\rangle\rangle) \end{aligned} \tag{3.20}$$

and linearity is extended to all nonnegative rational numbers.

Linearity is extended to include the irrational numbers in the continuum limit. Though a formal argument of this is used by Fuchs *et al.* [35], it is essentially a further assumption that

$$\alpha\nu\left(|E\rangle\rangle\right) = \nu\left(\alpha|E\rangle\rangle\right) \tag{3.21}$$

for all $0 \leq \alpha \leq 1$. An instrumentalist might argue that linearity over the rational numbers is enough. If probabilities are to be interpreted as relative occurences, it follows that all relevant quantities are rational numbers.

Combining this result with the original P3, linearity can be formalised in

$$\nu(\sum_i \alpha_i|E_i\rangle\rangle) = \sum_i \alpha_i\nu(|E_i\rangle\rangle), \tag{3.22}$$

where α_i may be any positive numbers. Linearity can be extended also to negative and complex numbers through decomposition of the relevant operator onto the positive/negative or real/complex eigenvalues. However, here I am interested in just the definition of the function for effects and for this set of operators it will always be possible to work in a basis such that all coefficients are positive, hence the negative and complex extensions are not be required.

Now that positive linearity has been established, the Born rule follows in just a few lines. As I assume that the measurement is represented by a positive semidefinite operator then it may be expressed as a vector

$$|\pi_i\rangle\rangle = \sum_\lambda \langle\langle\lambda\lambda^\dagger|\pi_i\rangle\rangle|\lambda\lambda^\dagger\rangle\rangle, \tag{3.23}$$

in which the Hilbert space vectors $|\lambda\rangle$ are simply the eigenbasis of the POVM element π_i. The subscript i indicates that it is part of a set, though no reference will be made to the other elements. From this and Eq. 3.22,

$$\nu(|\pi_i\rangle\rangle) = \sum_\lambda \nu(|\lambda\lambda^\dagger\rangle\rangle)\langle\langle\lambda\lambda^\dagger|\pi_i\rangle\rangle. \tag{3.24}$$

This is as an inner product, seen most clearly by defining $|r\rangle\rangle = \sum_\lambda \nu(|\lambda\lambda^\dagger\rangle\rangle)|\lambda\lambda^\dagger\rangle\rangle$, so that the above result is

$$\nu(|\pi_i\rangle\rangle) = \langle\langle r|\pi_i\rangle\rangle. \tag{3.25}$$

As inner products in Liouville space are associated with the trace operation, this is the Born rule up to the normalisation constant. This should not surprise, as it is a long-

established result that maps between vectors and real numbers are inner products [56]. A few more steps, however, are required before it is explicitly the Born rule. In particular, what remains to be seen is the physical interpretation of the vector $|r\rangle\rangle$. At this point, the only physical postulate that has been used to arrive at the trace operation is P3 and it is the others (P0-2) which will pin down the meaning. I begin by writing the result explicitly as a probability, by P0:

$$\mathrm{P}(i|s,x) = N(s,x)\langle\langle r|\pi_i\rangle\rangle. \tag{3.26}$$

By P2 I have

$$\sum_i \mathrm{P}(i|s,x) = N(s,x)\sum_i \langle\langle r|\pi_i\rangle\rangle = 1 \tag{3.27}$$

and hence

$$N(s,x) = \frac{1}{\sum_i \langle\langle r|\pi_i\rangle\rangle}. \tag{3.28}$$

I anticipate the physical interpretation of this vector by defining $|\rho\rangle\rangle = |r\rangle\rangle/(\sum_i \langle\langle r|\pi_i\rangle\rangle)$ and the probability rule which has been derived is thus

$$\mathrm{P}(i|s,x) = \langle\langle \rho|\pi_i\rangle\rangle. \tag{3.29}$$

I have shown that this formula is the unique way to calculate probabilities from measurement operators, given a small number of principles (additivity, noncontexuality). This argument is a restatement, modified to a vector-space representation, of the Gleason-Busch theorem's proof. All that is left is to demonstrate that this result is consistent with standard quantum mechanics: it must be shown that that the vector $|\rho\rangle\rangle$ is the operator space equivalent of the density operator, i.e., that it has all the same mathematical properties. Firstly, due to postulate P1 the associated operator must have positive eigenvalues. Secondly, by evaluating the inner product $\langle\langle I|\rho\rangle\rangle$ it is established that $\mathrm{Tr}(\rho) \leq 1$. This fact is related to the condition that the sum of POVM operators is equal to or less than the identity, with unit trace occuring if all measurement outcomes are available and no post-selection occurs. Both of these conditions may be relaxed and this would lead to a sub-unit trace. The final step is to verify that the vector $|\rho\rangle\rangle$ is independent of $|\pi_i\rangle\rangle$ as at first glance this does not appear to be the case.

Independence can be demonstrated following another argument due to Barnett *et al.* Consider two effects written in their eigenbases, i.e., $|\pi_0\rangle\rangle = \sum_i \lambda_i |\lambda_i\lambda_i^\dagger\rangle\rangle$ and $|\pi_1\rangle\rangle = \sum_j \eta_j |\eta_j\eta_j^\dagger\rangle\rangle$. The linearity of the function ν means that we must be able to extract probabilities also from the sum of these two $|\pi_{01}\rangle\rangle = |\pi_0\rangle\rangle + |\pi_1\rangle\rangle$, and linearity means that we must have a single operator $|\rho\rangle\rangle$ which acts upon both. This vector can be written in the two different bases as:

$$\begin{aligned} |\rho\rangle\rangle &= \sum_i \nu(|\lambda_i\lambda_i^\dagger\rangle\rangle)|\lambda_i\lambda_i^\dagger\rangle\rangle \\ &= \sum_j \nu(|\eta_j\eta_j^\dagger\rangle\rangle)|\eta_j\eta_j^\dagger\rangle\rangle, \end{aligned} \tag{3.30}$$

where the set of coefficients will always be positive. Any possible effect could also be invoked alongside $|\pi_0\rangle\rangle$ and thus $|\rho\rangle\rangle$ will have positive coefficients in all bases with doubled labels, corresponding to the diagonal elements in the operator representation; $|\rho\rangle\rangle$ must therefore be independent of the particular choice of measurement and hence is independent of the effect that it acts upon. The final step is to note that, by a lemma shown by Barnett *et al.* [34], any operator which has the same diagonal elements in all bases must also have the same off-diagonal elements. For this reason, $|\rho\rangle\rangle$ must be unique, given that the measurement statistics are already defined. The vector must be associated with the overall probability distribution of outcomes.

The vector $|\rho\rangle\rangle$ is associated with a Hilbert space operator; has positive eigenvalues and $\mathrm{Tr}(\rho) = 1$; is independent of the individual effect which it maps onto a set of probabilities. By definition, the normalisation constant $N(s, x)$ depends upon the relevant quantum system's preparation as well as the measurement performed. It contains information about the probabilities associated with the whole set of possible measurement outcomes. To summarise, these are all the properties which one would typically associate with the density operator, and this is how the vector $|\rho\rangle\rangle$ will be interpreted. It is thus safe to refer to Eq. 3.29 as the Born rule. By the Hilbert-Schmidt inner product this is

$$\mathrm{P}(i|s, x) = \mathrm{Tr}(\rho\pi_i). \tag{3.31}$$

It is interesting to note that this equation has a Bayesian flavour to it: we can understand the density operator as a representation of the a priori information one has about the quantum system while the effect π_i represents the probability of a given outcome. This will again be seen in the results of the next section, in which I generalise this result to include sequences of measurements.

3.5 Sequential measurements

I now present the main result of this chapter, a demonstration that Kraus's probability rule follows from the operational postulates presented above. To reiterate, though the experimental validity of this formula is already well established, what is of interest here is the sense in which the mathematical structure enforces that it is unique. The same procedure as above is followed: I exploit the properties of vector spaces to show that an inner product is the required form for the joint probability rule.

This derivation is quite involved so I summarise the proof here and diagram the logical relations in Fig. 3.1. First, it is noted that the noncontexuality postulate, A0, implies that the Kraus rule must be consistent with the Born rule. This statement then leads to the linearity which was derived in that case being extended, with use of A3, to linearity over positive numbers in the joint probability frame function. As in the Born rule proof, I must ensure that probabilities are positive (by A1) and this fixes the form of the Kraus operators. Finally, that the joint probabilities form a complete set (A2) means that the set of Kraus operators forms a valid POVM.

In the Born rule derivation, I started by demonstrating that the map ν must be additive

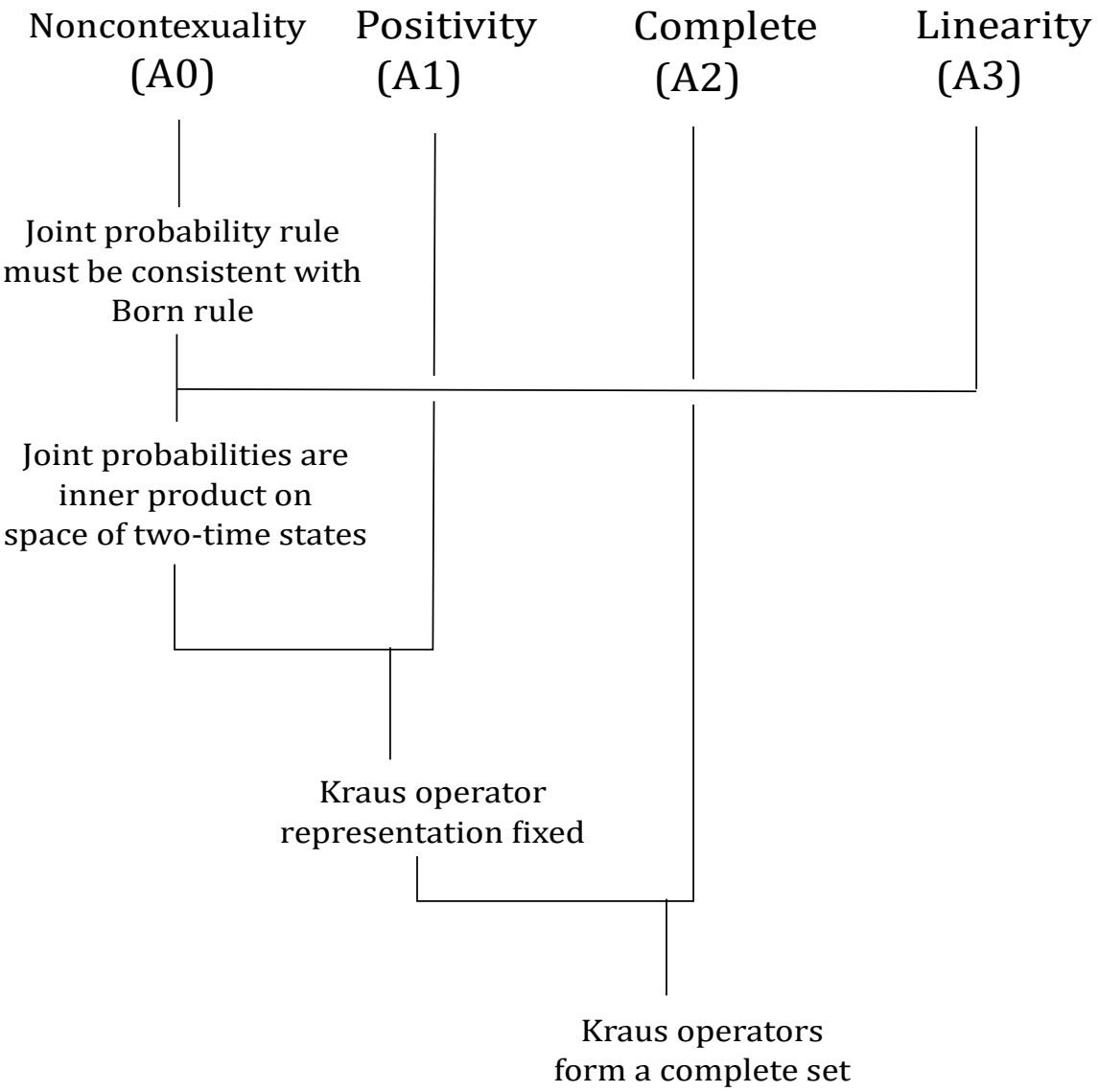


Figure 3.1: A diagram of the derivation of the joint probability rule, which makes explicit how the four axioms fit into the proof. Statements further down the page are implied by those which are further up the page and connected to them by a line.

over all positive operators, as seen in Equation 3.22. The requirement here is that this form of additivity extends to the map $\mathcal{T}_i$, in effect showing that this map must take the form of a linear superoperator. From A3 we have

$$\omega_i(|\pi_j^{(2)}\rangle\rangle) + \omega_i(|\pi_k^{(2)}\rangle\rangle) = \omega_i(|\pi_j^{(2)}\rangle\rangle + |\pi_k^{(2)}\rangle\rangle). \tag{3.32}$$

This is the same form as that for the map ν and so the same arguments can be applied here in order to extend the linearity over positive superoperators. Thus,

$$\alpha_j\omega_i(|\pi_j^{(2)}\rangle\rangle) + \alpha_k\omega_i(|\pi_k^{(2)}\rangle\rangle) = \omega_i(\alpha_j|\pi_j^{(2)}\rangle\rangle + \alpha_k|\pi_k^{(2)}\rangle\rangle), \tag{3.33}$$

with α being positive numbers. I use the relationship between the maps ω_i and ν, given in Eq. 3.17, to introduce the map $\mathcal{T}_i$:

$$\alpha_j\nu\left(\mathcal{T}_i(|\pi_j^{(2)}\rangle\rangle)\right) + \alpha_k\nu\left(\mathcal{T}_i(|\pi_k^{(2)}\rangle\rangle)\right) = \nu\left(\mathcal{T}_i(\alpha_j|\pi_j^{(2)}\rangle\rangle + \alpha_k|\pi_k^{(2)}\rangle\rangle)\right). \tag{3.34}$$

On the left hand side, the linearity of the map ν means that we can bring the coefficients of α into the argument. Finally, as this must hold for the entire set of ν,

$$\alpha_j\mathcal{T}_i\left(|\pi_j^{(2)}\rangle\rangle\right) + \alpha_k\mathcal{T}_i\left(|\pi_k^{(2)}\rangle\rangle\right) = \mathcal{T}_i\left(\alpha_j|\pi_j^{(2)}\rangle\rangle + \alpha_k|\pi_k^{(2)}\rangle\rangle\right). \tag{3.35}$$

This is a statement that $\mathcal{T}_i$ is a linear superoperator. As above, establishing this linearity is the first step in deriving the Kraus rule and at this point only postulate A3 has been used. The positivity and completeness of $\mathcal{T}_i$ is required by the remaining postulates and they fix the map's form. At this point it is helpful to make use of the inner product structure derived from the single measurement case, Eq. 3.31, along with the identification of the map as a superoperator, Eq. 3.35, to write the probability rule as

$$\mathrm{P}(i,j|s,x) = N(s,x)\langle\langle\rho|\mathcal{T}_i|\pi_j^{(2)}\rangle\rangle. \tag{3.36}$$

For future use, I am here explicit about the Hilbert spaces that each of these objects is defined upon: ρ is an operator on $\mathcal{H}_{in}$ and $\pi_j^{(2)}$ is an operator on $\mathcal{H}_{out}$. These two are Hilbert spaces associated with the preparation and second-measurement respectively. The map can then be defined as:

$$\mathcal{T}_i : \mathcal{H}_{out} \otimes \mathcal{H}_{out}^{\dagger} \to \mathcal{H}_{in} \otimes \mathcal{H}_{in}^{\dagger}. \tag{3.37}$$

The subscripts have been adopted in order to avoid confusion throughout the rest of this section, in which operators are defined upon various combinations of the four Hilbert spaces now in use.

While introducing operator space in this chapter I noted that, although the Hilbert-Schmidt inner product is used throughout, there are many ways to define the inner product in operator space, for example by including a linear superoperator. In Eq. 3.36 it can be seen that joint probabilities enter in precisely that manner. In this equation, it is seen that the map can also be thought of as acting upon the density operator $|\rho\rangle\rangle$ and this would

be physically interpreted as a different preparation procedure. This is what was meant when I stated that the more general definitions of inner products do not change how the formulae are physically interpreted.

I now turn to the task of constraining the map, $\mathcal{T}_i$, in such a way as to reconstruct the Kraus form. The condition A1 can be written, using the new form, as

$$0 \geq \langle\langle \rho | \mathcal{T}_i | \pi_j^{(2)} \rangle\rangle \geq 1. \tag{3.38}$$

I focus from now on solely on the bound from below (the requirement that the operator be positive). That the probabilities are less than one is enforced by normalisation at a later point. In fact the requirement of positivity needs to be made stronger. What is actually required is *complete positivity* such that the above requirement holds even in cases where $\mathcal{T}_i$ acts upon only a subsystem. This requirement is a corrolary of the fact that my operational postulates must hold for all possible choices for ρ and $\pi_j^{(2)}$ and, as such, it is not a further postulate. The requirement in full is

$$\langle\langle \rho | \mathcal{T}_{iA} \otimes I_B | \pi_j^{(2)} \rangle\rangle \geq 0, \tag{3.39}$$

where $\mathcal{H}_A$ and $\mathcal{H}_B$ are two Hilbert spaces on which the vectors are to be defined and I_B is the identity operation on the latter. I have derived the requirement that transformations are described by linear, completely positive superoperators. I begin by defining the two vectors $|\rho\rangle\rangle$ and $|\pi_j^{(2)}\rangle\rangle$. As both positive and Hermitian operators they can each be written as a linear superposition of pure states and projectors. For this reason, I restrict my attention to the simpler forms without loss of generality. I write $|\rho\rangle\rangle = |\psi\psi^\dagger\rangle\rangle$ and $|\pi_j^{(2)}\rangle\rangle = |\phi_j\phi_j^\dagger\rangle\rangle$ and also define

$$|\psi\rangle = \sum_i \lambda_i |i_A i_B\rangle \tag{3.40}$$

$$|\phi_j\rangle = \sum_{ik} c_{ik}^{(j)} |i_A k_B\rangle, \tag{3.41}$$

where the Schmidt decomposition [9] has been used to write ψ in a basis such that $\lambda_i \geq 0$. The same basis has also been used to define the projective state although it no longer holds that $c_{ik}^{(j)}$ are always real numbers. From these bases I construct operator space vectors

$$|\rho\rangle\rangle = \sum_{ik} \lambda_i \lambda_k |i_A i_B k_A^\dagger k_B^\dagger\rangle\rangle \tag{3.42}$$

$$|\pi_j^{(2)}\rangle\rangle = \sum_{iklm} c_{ik}^{(j)} c_{lm}^{(j)*} |i_A k_B l_A^\dagger m_B^\dagger\rangle\rangle. \tag{3.43}$$

Bringing all of this together I evaluate the left hand side of Eq. 3.39:

$$\begin{aligned}\langle\langle\rho|\mathcal{T}_{iA}\otimes I_B|\pi_j^{(2)}\rangle\rangle &= \sum_{iklmnp}\lambda_i\lambda_k c_{lm}^{(j)}c_{np}^{(j)*}\langle\langle i_A i_B k_A^\dagger k_B^\dagger|\mathcal{T}_{iA}\otimes I_B|l_A m_B n_A^\dagger p_B^\dagger\rangle\rangle \\ &= \sum_{iklmnp}\lambda_i\lambda_k c_{lm}^{(j)}c_{np}^{(j)*}\langle\langle i_A k_A^\dagger|\mathcal{T}_i|l_A n_A^\dagger\rangle\rangle\langle\langle i_B k_B^\dagger|I_B|m_B p_B^\dagger\rangle\rangle \\ &= \sum_{iklmnp}\delta_{im}\delta_{kp}\lambda_i\lambda_k c_{lm}^{(j)}c_{np}^{(j)*}\langle\langle i_A k_A^\dagger|\mathcal{T}_i|l_A n_A^\dagger\rangle\rangle \\ &= \sum_{ikln}\lambda_i\lambda_k c_{li}^{(j)}c_{nk}^{(j)*}\langle\langle i_A k_A^\dagger|\mathcal{T}_i|l_A n_A^\dagger\rangle\rangle. \end{aligned} \tag{3.44}$$

The right hand side of this equation is an inner product on the subspace A only. The structure of this map can be clarified by replacing the subscripts for the input and output spaces, as introduced earlier, in place of A. Inspection of the calculation up to this point reveals

$$\langle\langle\rho|\mathcal{T}_i\otimes I_B|\pi_j^{(2)}\rangle\rangle = \sum_{ikln}\lambda_i\lambda_k c_{li}^{(j)}c_{nk}^{(j)*}\langle\langle i_{in}k_{in}^\dagger|\mathcal{T}_i|l_{out}n_{out}^\dagger\rangle\rangle. \tag{3.45}$$

The requirement Eq. 3.39, which was derived from the postulates, is that this object is positive for all choices of $|\rho\rangle\rangle$ and $|\pi_j^{(2)}\rangle\rangle$. Writing about the above as an expectation value will constrain $\mathcal{T}_i$ to be a positive superoperator. To bring about this form, I consider the operator $\mathcal{T}_i'$ which is defined on the product space $\mathcal{H}_{in}\otimes\mathcal{H}_{out}^\dagger$ and satisfies

$$\langle\langle i_{in}l_{out}^\dagger|\mathcal{T}_i'|k_{in}n_{out}^\dagger\rangle\rangle = \langle\langle i_{in}k_{in}^\dagger|\mathcal{T}_i|l_{out}n_{out}^\dagger\rangle\rangle. \tag{3.46}$$

Using this definition, the constraint is

$$\begin{aligned}\langle\langle\rho|\mathcal{T}_i\otimes I_B|\pi_j^{(2)}\rangle\rangle &= \left(\sum_{il}\lambda_i c_{li}^{(j)}\langle\langle i_{in}l_{out}^\dagger|\right)\mathcal{T}_i'\left(\sum_{kn}\lambda_k c_{nk}^{(j)*}|k_{in}n_{out}^\dagger\rangle\rangle\right) \\ &= \langle\langle\Psi_j|\mathcal{T}_i'|\Psi_j\rangle\rangle \geq 0, \end{aligned} \tag{3.47}$$

in which I have introduced the notation

$$|\Psi_j\rangle\rangle = \sum_{kn}\lambda_k c_{nk}^{(j)*}|k_{in}n_{out}^\dagger\rangle\rangle. \tag{3.48}$$

Finally, a simple condition for positive probabilities in the operational framework employed has been derived. Eq. 3.47 says that the requirement for an operator to be associated with positive probabilities is that it has positive eigenvalues on the space $\mathcal{H}_{in}\otimes\mathcal{H}_{out}^\dagger$. Vectors on this space contain information about the preparation and second measurement. They can be interpreted as two-time states, which are commonly understood as representing pre- and post-selections [59, 60]. I discuss this point in further detail in the next section. The positivity of $\mathcal{T}_i'$ on the space of two-time states is enough to constrain us to the Kraus form, which follows as the superoperator must have an eigendecomposition on this space:

$$\mathcal{T}_i' = \sum_k |\alpha_{ik}\rangle\rangle\langle\langle\alpha_{ik}|. \tag{3.49}$$

Each vector $|\alpha_{ik}\rangle\rangle = \sum_{lm} \alpha_{lm}^{(ik)} |l_{in} m_{out}^{\dagger}\rangle\rangle$ here is eventually associated with the vector space representation of a Kraus operator. At this point such an identification is only intuitive, as they are not currently normalised and still need to be related to the second measurement's POVM element in the standard way. To return to my original formula of joint probabilities, A0, so far I have shown that joint probabilities can be expressed as

$$\mathrm{P}(i,j|s,x) = N(s,x) \sum_{k} \langle\langle \Psi_j | \alpha_{ik} \rangle\rangle \langle\langle \alpha_{ik} | \Psi_j \rangle\rangle. \tag{3.50}$$

I now demonstrate the identification of the vectors $|\alpha_{ik}\rangle\rangle$, henceforth referred to as Kraus vectors, with the Kraus operators by invoking A2. This task is most straightforward if I use the simplified case that $\rho = |\psi\rangle\langle\psi|$ and $\pi_j^{(2)} = |\phi_j\rangle\langle\phi_j|$, noting again that the positivity of both operators in general means that there is no loss of generality by doing this. After summing over all second outcomes, I have

$$\begin{aligned}
\sum_j \mathrm{P}(i,j|s,x) &= N(s,x) \sum_{jklmnp} \langle\psi| \left(\alpha_{lm}^{(ik)} \alpha_{np}^{(ik)*} |l\rangle\langle n| \langle\phi_j^{\dagger}|m^{\dagger}\rangle\langle p^{\dagger}|\phi_j^{\dagger}\rangle \right) |\psi\rangle \\
&= N(s,x) \sum_{klmnp} \langle\psi| \left(\alpha_{lm}^{(ik)} \alpha_{np}^{(ik)*} |l\rangle\langle n| \langle m| \left(\sum_j |\phi_j\rangle\langle\phi_j| \right) |p\rangle \right) |\psi\rangle \\
&= N(s,x) \langle\psi| \left(\sum_{klmn} \alpha_{lm}^{(ik)} \alpha_{nm}^{(ik)*} |l\rangle\langle n| \right) |\psi\rangle,
\end{aligned} \tag{3.51}$$

Completeness of the second measurement appears in two ways. It has been used above, in the sense that $\sum_j \pi_j^{(2)} = I$, to derive the third line. Completeness was also formalised in A2 which says that $\sum_j \mathrm{P}(i,j|s,x) = \mathrm{P}(i|s,x)$. As $\mathrm{P}(i|s,x) = \langle\psi|\pi_i^{(1)}|\psi\rangle$, the above is consistent with this requirement only if

$$\pi_i^{(1)} = \sum_{klmn} \alpha_{lm}^{(ik)} \alpha_{nm}^{(ik)*} |l\rangle\langle n| = \sum_k A_{ik}^{\dagger} A_{ik}, \tag{3.52}$$

where I use A_{ik} for the Hilbert space operator which maps onto the operator space vector $|\alpha_{ik}\rangle\rangle$. This formula is the usual identification between POVM elements and Kraus operators, and so the result I have derived is seen to be consistent with standard measurement theory. The effect can also be written as

$$\pi_i^{(1)} = \sum_n \langle n^{\dagger}|\alpha_{lm}^{(ik)}\rangle\rangle \langle\langle \alpha_{lm}^{(ik)}|n^{\dagger}\rangle, \tag{3.53}$$

where $\{|n\rangle\}$ is any complete basis for the output space, following the labelling for the daggered/non-daggered spaces given above. This result says that the Kraus vector is associated with the Hilbert space operator of its related POVM element by tracing out the output space. The freedom which is implicit in this is the same as the usual choice one has in decomposing an effect into a set of Kraus operators.

The final piece of the puzzle is to fix the normalisation factor $N(s,x)$, which follows from rewriting Postulate A2 so that it concerns the sum over both outcomes. Again I

begin by summing Eq. 3.50 and look at the simpler case (of projective measurements and pure state preparation) in order to write

$$\sum_{ij} \mathrm{P}(i,j|s,x) = N(s,x)\langle\psi| \left(\sum_{ijk} \langle\phi_j^\dagger|\alpha_{ik}\rangle\rangle\langle\langle\alpha_{ik}|\phi_j^\dagger\rangle \right) |\psi\rangle = 1. \tag{3.54}$$

As the states $|\psi_j\rangle$ are normalised, the above holds only if

$$N(s,x) \sum_{ijk} \langle\phi_j^\dagger|\alpha_{ik}\rangle\rangle\langle\langle\alpha_{ik}|\phi_j^\dagger\rangle = I, \tag{3.55}$$

where the identity I here acts on the input space, according to the labelling given above. It is useful to note that, due to the requirement that the second measurement be complete, I have $\sum_j |\phi_j^\dagger\rangle\langle\phi_j^\dagger| = I$. For this set of projectors to be complete we must have that $\{|\phi_j^\dagger\rangle\}$ forms a complete orthonormal basis of the space and thus the sum over these states in the above equation acts as a trace over the output space. It is also clear from this that $N(s,x) = 1$ as long as the vectors $|\alpha_{ik}\rangle\rangle$ are suitably normalised. To emphasise the distinction, I will henceforth denote the normalised Kraus vectors associated with each of these as $|A_{ik}\rangle\rangle$. This shouldn't be too surprising, as any contexuality associated with the first measurement has already been hidden inside the density matrix as part of the single measurement derivation. From Eq. 3.53 this is simply the usual requirement that the POVM elements associated with the channel are complete.

I have derived the Kraus rule as an extension of the Gleason-Busch theorem. Given that measurements are described by effects (i.e., positive operators on Hilbert spaces), the unique map from a pair of measurement outcomes to the set of real numbers which is consistent with the set of postulates A1-3 is given by

$$P(i,j|s,x) = \sum_k \langle\langle\Psi_j|A_{ik}\rangle\rangle\langle\langle A_{ik}|\Psi_j\rangle\rangle \tag{3.56}$$

in which $|\Psi_j\rangle\rangle$ is a vector containing information about the preparation and second measurement. While it has a different form to the usual Kraus rule as a trace function, it is straightforward to see that the two formulae are equivalent by considering again pure state preparation and a projective second measurement. Writing all three objects in the same basis, $|A_{ik}\rangle\rangle = \sum_{lm} A_{lm}^{(ik)} |lm^\dagger\rangle\rangle$, $|\rho\rangle\rangle = \sum_{qr} \lambda_q \lambda_r^* |qr^\dagger\rangle\rangle$ and $|\pi_j^{(2)}\rangle\rangle = \sum_{st} \beta_s^{(j)} \beta_t^{(j)*} |st^\dagger\rangle\rangle$. From these objects I evaluate first the two-time state vector as $|\Psi_j\rangle\rangle = \sum_{qt} \lambda_q \beta_t^{(j)*} |qt^\dagger\rangle\rangle$. From the above probability rule,

$$\begin{aligned} \mathrm{P}(i,j|s,x) &= \sum_k |\lambda_q^* \beta_t^{(j)} A_{qt}^{(ik)}|^2 \\ &= \mathrm{Tr}\left(\pi_j^{(2)} \sum_k A_{ik} \rho A_{ik}^\dagger \right), \end{aligned} \tag{3.57}$$

where the second line uses the usual isomorphism between operator space and Hilbert space quantities.

State update rule

An auxillary result which follows from the above calculation is the state update rule. I use Bayes's rule to acquire the conditional probability $\mathrm{P}(j|i,s,x)$ and expect that this would behave as a single measurement, with the measured state being that associated with update of the initial state by the first measurement. Again, I use the simplified case in which $|\rho\rangle\rangle = |\psi\psi^\dagger\rangle\rangle$ and $|\pi_j^{(2)}\rangle\rangle = |\phi_j\phi_j^\dagger\rangle\rangle$. Then,

$$\begin{aligned}
\mathrm{P}(j|i,s,x) &= \frac{\mathrm{P}(i,j|s,x)}{\mathrm{P}(i|s,x)} \\
&= \frac{\sum_k \langle\langle \Psi_j|A_{ik}\rangle\rangle\langle\langle A_{ik}|\Psi_j\rangle\rangle}{\langle\langle\rho|\pi_i^{(1)}\rangle\rangle} \\
&= \frac{\langle\phi_j|\left(\sum_k \langle\psi|A_{ik}\rangle\rangle\langle\langle A_{ik}|\psi\rangle\right)|\phi_j\rangle}{\langle\langle\rho|\pi_i^{(1)}\rangle\rangle} \\
&= \mathrm{Tr}\left(\pi_j^{(2)}\frac{\sum_k \langle\psi|A_{ik}\rangle\rangle\langle\langle A_{ik}|\psi\rangle}{\langle\langle\rho|\pi_i^{(1)}\rangle\rangle}\right).
\end{aligned} \tag{3.58}$$

The state update is identified as

$$\begin{aligned}
\rho \to \rho' &= \frac{\sum_k \langle\psi|A_{ik}\rangle\rangle\langle\langle A_{ik}|\psi\rangle}{\langle\langle\rho|\pi_i^{(1)}\rangle\rangle} \\
&= \frac{\sum_k A_{ik}\rho A_{ik}^\dagger}{\mathrm{Tr}(\rho\sum_k A_{ik}A_{ik})}
\end{aligned} \tag{3.59}$$

from the fact that $\mathrm{P}(j|i,s,x) = \mathrm{Tr}(\rho\pi_2)$. Alongside the rule associated with joint probabilities in quantum mechanics, the Lüders rule [38] has been derived. However, none of the postulates concern the measurement dynamics. Importantly, this means that no further assumptions are needed for the dynamics of measurement on top of the probabilistic description: the Born rule and wavefunction collapse are part and parcel of the same mathematical structure, which arises as measurements are associated with positive operators, given some reasonable postulates which a probability rule must satisfy.

3.6 Comments

The work presented here exhibits close links with a number of other areas of research. I discuss a number of these in this section.

Earlier, I discussed Hardy's work in quantum reconstructions and noted that it begins by showing that probabilities can be treated as inner products on a given space, and that once this is shown the task mathematically becomes to pinpoint precisely which space these inner products are defined upon. For single measurements, as has been shown, this is operator space, which has the trace operation as its inner product. While progressing to derive the analogous result for sequential measurements, I skipped over this point and instead focused on treating the channel as a superoperator. This was a choice, made to present the derivation in a greater clarity. Instead, it is possible to treat joint and conditional probabilities also as inner products. The alternative calculation is sketched out

here. Eq. 3.36, which for reference says that $P(i,j|s,x) = \langle\langle\rho|\mathcal{T}_i|\pi_j^{(2)}\rangle\rangle$ (where constants and degenerate channels have been ignored for simplicity), can be written as

$$\begin{aligned} \mathrm{P}(i,j|s,x) &= \mathrm{Tr}\left(\mathcal{T}_i|\pi_j^{(2)}\rangle\rangle\langle\langle\rho|\right) \\ &= \mathrm{Tr}\left(\overline{\mathcal{T}}_i|\Psi_j\rangle\rangle\langle\langle\Psi_j|\right) \end{aligned} \tag{3.60}$$

In the same way that the trace operation in Hilbert space can be represented by an inner product in Liouville space, this trace can also be represented by an inner product in a superoperator space (i.e., $\mathcal{H}_{in} \otimes \mathcal{H}_{in}^\dagger \otimes \mathcal{H}_{out} \otimes \mathcal{H}_{out}^\dagger$). The probability rule is once again an inner product and the task is to find the relevant objects such that this inner product is always positive. The final mathematical structures which result are precisely the same. Such an approach would drive home the link between traces, inner products and probabilities which much work in generalised probabilities and quantum reconstructions relies on.

A more important link is with the two-time state formalism of Aharonov and Vaidman [59, 60]. A two-time state is the product state of a preparation and measurement, as well as superpositions of these objects, and is most commonly used to represent pre- and post-selection. Mathematically, these objects are precisely the vectors $|\Psi_j\rangle\rangle$, defined on $\mathcal{H}_{in} \otimes \mathcal{H}_{out}$ and separated in the same way, that were introduced above. The two-time state interpretation of a sequential measurement process is that vectors on $\mathcal{H}_{in}$ evolve forward in time while those on $\mathcal{H}_{out}^\dagger$ evolve backwards in time, collapsing on any intermediate measurements. Probabilities are then calculated from these objects by taking inner products with another object, associated with any operations between the two, analagous to that found above. One way to understand my work is as an axiomatic foundation for the two-time state formalism, although the interpretation in terms of time evolution does not follow from what has been shown above. Silva *et al* [61] have provided an analysis which shows that such objects are created experimentally by pre- and post-selection. A system is prepared in the state $|\psi\rangle$ and sent to an observer, who performs any measurement before returning the modified system to the preparer. The first party then measures and keep the state only if the desired result $|\phi\rangle$ is the outcome. Measurements on the resulting system can be used to reconstruct the same statistics of the second party's measurement as found for the equivalent two-time state, $|\psi\phi\rangle$. This picture will be useful to keep in mind for the next chapter, when I use the framework in the context of quantum key distribution.

An important tool in quantum information theory is the Choi-Jamiołkowski isomorphism [53, 54]. There are two statements associated with this theorem although the distinction is rarely made. Choi's isomorphism [53] is between the set of operators on a Hilbert space $\mathcal{H}$ and the set of vectors upon the doubled space $\mathcal{H} \otimes \mathcal{H}$. It says that every operator can be represented as a bipartite state and vice versa. This is often understood as a theorem which allows quantum-gate teleportation. In gate teleportation, Alice has access to qubits A and B and Bob has access to qubit C. Alice begins with the state ρ_A and the two parties share a Bell state $|\Psi^+\rangle_{BC}$. The aim is to leave Bob with the state $\mathcal{E}(\rho)$, in which $\mathcal{E}$ is some map. It is a form of state teleportation in which the gate is

implemented during teleportation, and of course one way to do this is to teleport the state and have Bob perform the gate locally. As the steps here are on two different spaces (the first on $\mathcal{H}_A \otimes \mathcal{H}_B$, the second on $\mathcal{H}_C$) they commute and can be implemented in a different order. If Bob first performs the map $\mathcal{E}$ on his qubit before the state teleportation, the state which the two parties share is the Choi state associated with that map. This is the link between Choi's isomorphism and gate teleportation. Jamiołkowski's [54] is similar although the mapping is between operators on $\mathcal{H}$ and vectors on the space $\mathcal{H} \otimes \mathcal{H}^\dagger$, where the dagger indicates the dual space. It is the mathematical theorem upon which operator space is based. There is a subtle distinction between the two in that only one is basis dependent. Choi's map was seen to arise naturally in the derivation above; it is expressed in Eq. 3.46 which defines $\mathcal{T}_i'$ as an operator on $\mathcal{H}_{in} \otimes \mathcal{H}_{out}^\dagger$. Similarly, in Eq. 3.60 we see the superoperator $\overline{\mathcal{T}}_i : \mathcal{H}_{in} \otimes \mathcal{H}_{out} \to \mathcal{H}_{in}^\dagger \otimes \mathcal{H}_{out}^\dagger$ appearing; this is the Jamiołkowski representation of the channel. These maps are visualised in Fig. 3.2. The choice between Kraus operator and Choi-Jamiołkowski representation of sequential measurements is the only freedom available, and the underlying mathematical structure is fixed.

To emphasise the distinction between the superoperator $\mathcal{T}_i$, which was initially introduced and the Choi superoperator $\mathcal{T}_i'$, it is useful to discuss the different representations of the identity superoperator. One is a map of the type $\mathcal{I} : \mathcal{H}_{in} \otimes \mathcal{H}_{in}^\dagger \to \mathcal{H}_{out} \otimes \mathcal{H}_{out}^\dagger$ which leaves states invariant (i.e, $\mathcal{I}|\lambda\rangle\rangle = |\lambda\rangle\rangle \; \forall \; |\lambda\rangle\rangle$). For a two-dimensional Hilbert space, this is

$$\mathcal{I} = |00^\dagger\rangle\rangle\langle\langle 00^\dagger| + |01^\dagger\rangle\rangle\langle\langle 01^\dagger| + |10^\dagger\rangle\rangle\langle\langle 10^\dagger| + |11^\dagger\rangle\rangle\langle\langle 11^\dagger|. \tag{3.61}$$

This object is different to those of interest here, which are the Choi operators on the space $\mathcal{H}_{in} \otimes \mathcal{H}_{out}^\dagger$. These objects are obtained from the 'original' maps by the permutation of indices defined in Eq. 3.46. Doing this for each term individually in $\mathcal{I}$ one finds

$$\begin{aligned} \mathcal{I}' &= |00^\dagger\rangle\rangle\langle\langle 00^\dagger| + |00^\dagger\rangle\rangle\langle\langle 11^\dagger| + |11^\dagger\rangle\rangle\langle\langle 00^\dagger| + |11^\dagger\rangle\rangle\langle\langle 11^\dagger| \\ &= |I\rangle\rangle\langle\langle I|. \end{aligned} \tag{3.62}$$

That there are two different representations of the identity initially seems strange but becomes more intuitive if it is noted that the identity in one case acts upon two-time states rather than states in the more traditional sense. In fact, if one is not careful upon this point then we can quickly get nonsensical results: for example by considering that the probability of measuring the outcome $|1\rangle$ if the state $|0\rangle$ was prepared is given by $\mathrm{P}(0|1) = \langle\langle 01^\dagger|I|01^\dagger\rangle\rangle = 1$.

As discussed earlier, an operator's Choi-Jamiołkowski isomorphism is often interpreted as the channel which teleports it. This channel has also been analysed as a quantum comb and as a conditional state. Quantum combs [62, 63] are the main analytic tool used when analysing quantum networks, which are mathematical objects which combine channels and POVMs. The comb is given by the Choi-Jamiołkowski operator which represents a given quantum network. This is closely linked to what was demonstrated above, in which a Kraus operator is represented by the Choi vector $|A_{ik}\rangle\rangle$ associated with its channel. However, clearly the maps $\mathcal{H}_{in} \to \mathcal{H}_{out}$ which have been used are particularly simple forms

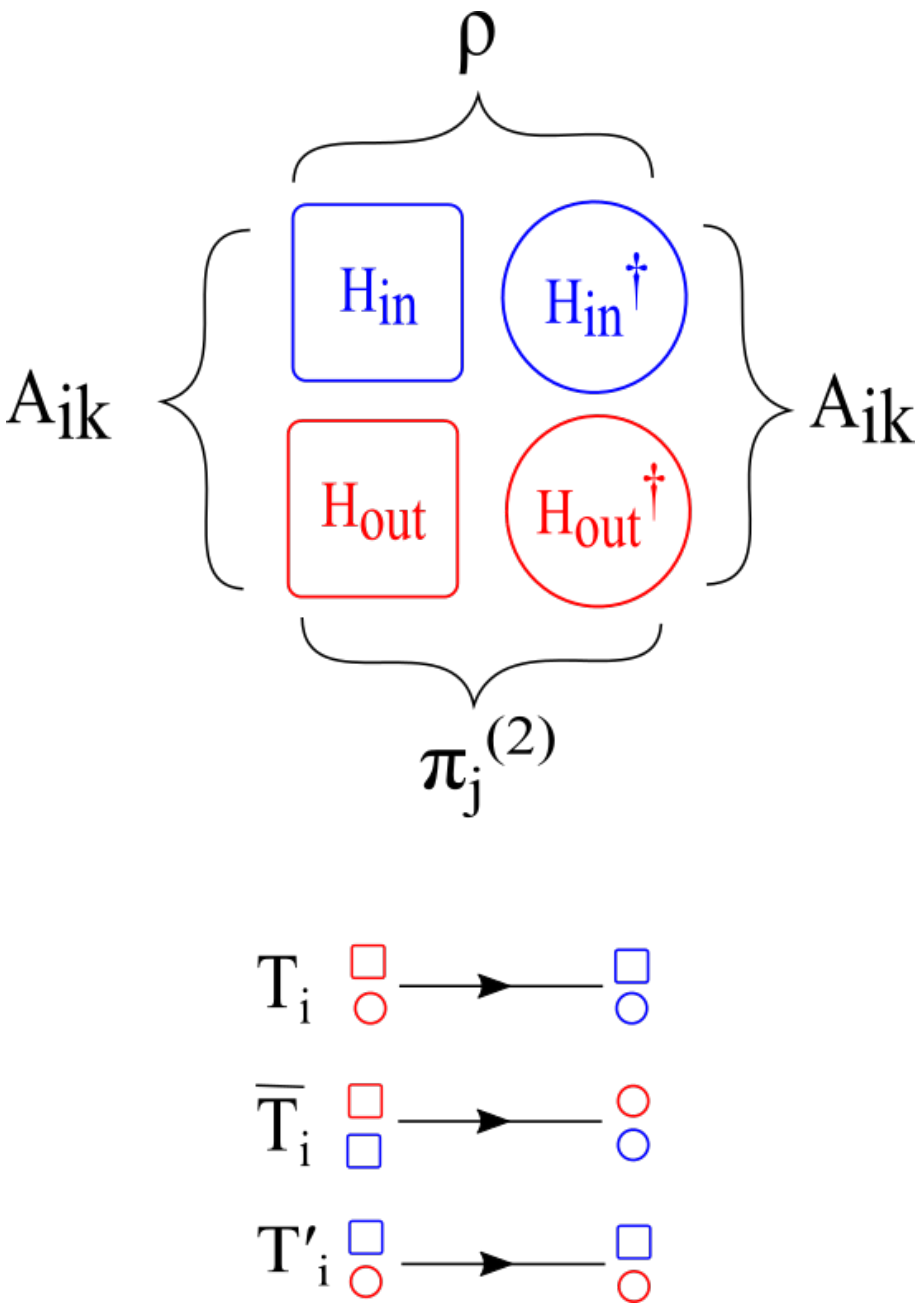


Figure 3.2: A visualisation of the different Hilbert spaces which are used throughout this proof, showing the objects (density matrix ρ, POVM element $\pi_j^{(2)}$ and Kraus operators A_{ik}) which are associated with each pair. Each of the four relevant Hilbert spaces is associated with a different colour and shape according to the upper half of the figure. The maps can then be written as in the lower half.

of quantum network which involve just two Hilbert spaces and hence the more complex cases (involving, for example, concatenations of multiple channels or large numbers of Hilbert spaces) are not included.

Leifer and Spekkens [55] find that the Jamiołkowski operator of a channel is best interpreted as the *conditional state* of the two Hilbert spaces which it relates (and which they associate with spacetime regions rather than the traditional states). This is an operator on a product space defined such that the probability distributions of individual measurements are found by tracing out either space. A similar result is found here: in Eq. 3.53, the effect for the first measurement is derived by tracing out the space associated with the second measurement. Their article uses this isomorphism to put acausal conditional probabilities (on spatiotemporally separated regions) on a similar footing to causally related conditional probabilities. Both of our works use that the space of two-time states and the space of channels are isomorphic, however I consider only a single causal structure and so do not derive their full formalism.

Finally, so far I have provided a derivation for the probability rule for two measurements only. It is natural to ask whether the results generalise to processes involving three or more measurements and the methods developed are easy to generalise. For a non-degenerate measurement, the probability rule is written as

$$\mathrm{P}(i,j|s,x) = \langle\langle \rho | \mathcal{A}_i^{(1)} | \pi_j^{(2)} \rangle\rangle, \tag{3.63}$$

as seen in Eq. 3.46. At this point it is easy to see that the probability rule associated with a potential third measurement would straightforwardly take the form

$$\mathrm{P}(i,j,k|s,x) = \langle\langle \rho | \mathcal{A}_i^{(1)} \mathcal{A}_j^{(2)} | \pi_k^{(3)} \rangle\rangle. \tag{3.64}$$

By considering the preparation-and-first-measurement events are a single preparation, I now introduce $|\rho_i'\rangle\rangle = A_i^{(1)}|\rho\rangle\rangle$, so that the rule is $\langle\langle \rho_i' | \mathcal{A}_j^{(2)} | \pi_k^{(3)} \rangle\rangle$, precisely the same as in the two-measurement case. The rest of the calculations follow the method used earlier, and would result in the three-measurement Kraus rule as well as all the associated state update rules and conditional probabilities. In this and the two measurement case discussed in the main text above I've examined multiple measurements on the same system, however a natural further generalisation would be to look at cases where two or more different systems are measured. This would bring the work presented here closer to the quantum networks and conditional states formalisms.

3.7 Basic examples

In this chapter it is seen that the formalism of two-time states emerges as a natural framework for sequential measurements. As this framework may be unfamiliar, it is useful to explore some short worked examples. In the next chapter, I develop the formalism more fully and show that it can be used to develop attacks for quantum key distribution protocols. Here I look at two shorter examples, demonstrating the impossibility of partial

transposition and analysing an interferometry experiment.

3.7.1 Partial transposition

The transposition, which is positive but not completely positive, is a commonly discussed example of an unphysical map [10]. It finds practical application in entanglement detection. If the density matrix on $\mathcal{H}_B$ only of a bipartite state ρ_{AB} is tranposed, an operation I denote $\mathcal{P}$, then the updated state $\rho_B = \mathrm{Tr}_A(\mathcal{P}(\rho_{AB}))$ is positive, while the overall density matrix can have negative eigenvalues. In the framework used here, this can be seen straightforwardly by writing out the superoperator $\mathcal{P}$ and verifying that it cannot be written in the form required by Eq. 3.49. Transposition is a map that acts as follows:

$$\begin{aligned}
&\mathcal{P}(|0\rangle\langle 0|) \rightarrow |0\rangle\langle 0| \\
&\mathcal{P}(|0\rangle\langle 1|) \rightarrow |1\rangle\langle 0| \\
&\mathcal{P}(|1\rangle\langle 0|) \rightarrow |0\rangle\langle 1| \\
&\mathcal{P}(|1\rangle\langle 1|) \rightarrow |1\rangle\langle 1|,
\end{aligned} \tag{3.65}$$

which is expressed as the following superoperator

$$\mathcal{P} = |00^\dagger\rangle\rangle\langle\langle 00^\dagger| + |10^\dagger\rangle\rangle\langle\langle 01^\dagger| + |01^\dagger\rangle\rangle\langle\langle 10^\dagger| + |11^\dagger\rangle\rangle\langle\langle 11^\dagger|. \tag{3.66}$$

Here I have followed Eq. 3.46 in order that this operator is on the correct Hilbert space however, as the index structure is identical for the alternative configuration, the subscripts are dropped. The condition for allowed operations is that they are positive semi-definite, therefore to show that an operation is disallowed it must be shown that it has negative eigenvalues. In matrix notation the superoperator is

$$\mathcal{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.67}$$

and the eigenvalues λ are found in the usual manner, by solving the equation $|\mathcal{P} - \lambda I| = 0$. This reveals three degenerate eigenvectors with eigenvalue $\lambda = 1$ and one with eigenvalue $\lambda = -1$, a negative value which shows that the superoperator cannot be associated with any physical action. This result has been arrived at by transforming the problem into an eigenvalue calculation, a tool that recurs throughout the next chapter. That negative probabilities occur can be seen by considering the follow series of events. A bipartite system on $\mathcal{H}_A \otimes \mathcal{H}_B$ is prepared in the maximally entangled Bell state $|\Phi_+\rangle = (|0_A 0_B\rangle + |1_A 1_B\rangle)/\sqrt{2}$. System A only is transposed followed by a measurement with the outcome $|\Psi_-\rangle = (|0_A 1_B\rangle - |1_A 0_B\rangle)/\sqrt{2}$. What is the probability of this outcome? I begin by

constructing the two-time state,

$$\begin{aligned}&|\Phi_+\Psi_-^\dagger\rangle\rangle\\&= \frac{1}{2}\left(|00^\dagger\rangle\rangle_A|01^\dagger\rangle\rangle_B - |01^\dagger\rangle\rangle_A|00^\dagger\rangle\rangle_B + |10^\dagger\rangle\rangle_A|11^\dagger\rangle\rangle_B - |11^\dagger\rangle\rangle_A|10^\dagger\rangle\rangle_B\right),\end{aligned} \tag{3.68}$$

which represents the preparation and measurement events. The probability that this occurs, given the discussed operation, is

$$\begin{aligned}&\langle\langle\Phi_+\Psi_-^\dagger|\mathcal{P}_A\otimes I_B|\Phi_+\Psi_-^\dagger\rangle\rangle\\&= \frac{1}{2}\langle\langle\Phi_+\Psi_-^\dagger|\left(-|10^\dagger\rangle\rangle_A|00^\dagger\rangle\rangle_B - |10^\dagger\rangle\rangle_A|11^\dagger\rangle\rangle_B + |01^\dagger\rangle\rangle_A|00^\dagger\rangle\rangle_B + |01^\dagger\rangle\rangle_A|11^\dagger\rangle\rangle_B\right)\\&= -\frac{1}{2},\end{aligned} \tag{3.69}$$

where the identity operation performed on B is again $I_B = |I\rangle\rangle\langle\langle I|$ (as was discussed above). The impossibility of physical transposition is made evident by the appearance of a negative probability.

3.7.2 Interferometry

Interferometers are ubiquitous pieces of equipment in optics. A photon enters a beamsplitter and travels down two arms of different length, such that a phase difference ϕ is introduced and the state becomes $|\rho\rangle = (|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$. The two paths are combined through a second beamsplitter and the photon is measured in the $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$ basis. It is well known that an interference pattern will be displayed in such a case if we post-select on a given outcome, but that this interference pattern will be lost if a which-way measurement is performed to localise the photon within a given arm. This simple experiment can be used to present some of the objects used in my formalism.

We can begin by constructing the two-time state associated with post-selection on a given outcome, for which I choose $|+\rangle$. The two-time state is given by

$$\begin{aligned}|\rho,+\rangle\rangle &= |\rho+^\dagger\rangle\rangle\\&= \frac{1}{2}(|0\rangle + e^{i\phi}|1\rangle)(|0\rangle + |1\rangle)\\&= \frac{1}{2}(|00^\dagger\rangle\rangle + |01^\dagger\rangle\rangle + e^{i\phi}|10^\dagger\rangle\rangle + e^{i\phi}|11^\dagger\rangle\rangle.\end{aligned} \tag{3.70}$$

The act of not measuring the path information can be represented by the identity superoperator, $I = |I\rangle\rangle\langle\langle I|$, so the probability of outcome $|+\rangle$ is

$$\mathrm{P}(+|\rho) = \langle\langle\rho,+|I\rangle\rangle\langle\langle I|\rho,+\rangle\rangle. \tag{3.71}$$

As $\langle\langle I|\rho,+\rangle\rangle = (1+e^{i\phi})/2$, this probability is

$$\mathrm{P}(+|\rho) = \frac{1}{4}(1+e^{-i\phi})(1+e^{i\phi}) = \frac{1}{2}(1+\cos(\phi)). \tag{3.72}$$

The interferometer displays an interference pattern. This is not a surprising result, however

it has allowed me to demonstrate how to construct the two-time states, which are used heavily in the next chapter. The next object to calculate is the equivalent probability in the case that the path degree of freedom is measured interstitially. I consider that the experimenter chooses to measure the arm associated with $|0\rangle$. For this measurement I require a Kraus operator such that the associated effect is $\pi_0 = |0\rangle\langle 0|$. Remembering that POVM elements are found by tracing out the daggered space of the superoperator, there is some freedom in how to complete the channel. I emphasise this point by writing the transformation as $|0\psi^\dagger\rangle\rangle\langle\langle 0\psi^\dagger|$, where $|\psi\rangle$ can be any normalised state. In this case, following the usual probability rule

$$\mathrm{P}(0,+|\rho) = \langle\langle \rho, +|0\psi^\dagger\rangle\rangle\langle\langle 0\psi^\dagger|\rho,+\rangle\rangle, \tag{3.73}$$

for which we can readily evaluate $\langle\langle 0\psi^\dagger|\rho,+\rangle\rangle = (\langle\psi^\dagger|0^\dagger\rangle + \langle\psi^\dagger|1^\dagger\rangle)/2$ and hence find

$$\mathrm{P}(0,+|\rho) = \frac{1}{4}\left(\langle 0^\dagger|\psi^\dagger\rangle + \langle 1^\dagger|\psi^\dagger\rangle\right)\left(\langle\psi^\dagger|0^\dagger\rangle + \langle\psi^\dagger|1^\dagger\rangle\right) \tag{3.74}$$

As there is no dependence upon ϕ here, it is seen that a which-way measurement has destroyed the interference. Again, this is a well-established result however it has allowed me to demonstrate how to use the formalism, in particular highlighting the freedom in assigning Kraus operators from effects.

3.8 Summary

The Kraus formalism has long been the accepted method for handling joint and conditional probabilities in quantum theory. In this chapter, I showed why it plays such an important role: the probability rule is unique in the sense that there is no other linear map from positive operators to probabilities. As a corrollary, the state-update rule associated with the back-reaction of the measurement follows from the same axioms. Furthermore the structure of the two-state vector formalism emerges as the natural way to handle pre- and post-selection. I also provided some simple examples, meant to highlight the properties of the different objects that have arisen in the operator space formalism.

In the next section I provide further examples, demonstrating that the formalism has a role to play in quantum cryptography. This emphasises an important practical application that may be found for foundational work of this kind. It is likely that users of the new quantum technologies, as people unfamiliar with quantum theory, are likely to be sceptical when first introduced to ideas such as the uncertainty principle. The axiomatic approach used here demonstrates that security in fact relies upon some not-so-strange sounding assumptions.

Chapter 4

Two-time states for quantum key distribution

Quantum key distribution (QKD) is a set of protocols which distribute a key, for use in cryptographic exchanges, between two communicating parties and which use the quantum mechanical concept of measurement disturbance to ensure that any eavesdropping is flagged to the legitimate users [64, 65]. In a typical prepare-and-measure scheme [10], one of these parties will produce a quantum state, the signal, and send it through a quantum channel to a second party who measures the state. I follow standard procedure in naming the transmitter Alice and the receiver Bob. After Bob's measurement, one of the two users publically shares information which allows for a pre-determined subset of send-measure correlations to be saved. This process is called sifting. Finally, logical bit values are assigned. According to the principle of measurement disturbance, any interlocutor hoping to know which states were exchanged will leave a measurable trace of their activity: Alice and Bob could in principle uncover them by examining the final set of logical bits. Nonetheless, this illegitimate party (Eve) will attempt to hide behind systemic noise. Quantum cryptanalysis partly involves calculating her best strategy. There is much more to QKD than the outline I've sketched here and the reader is directed to §2.5 for a more wide-ranging discussion.

In designing her eavesdropping strategy, Eve needs to take into account correlations between events in the past (e.g., the signal prepared by Alice) and the future (e.g., Bob's measurement outcomes and the results of the sifting process). There is some friction between this picture and quantum mechanics as it is typically presented, in which there is a preparation procedure followed by a sequence of measurements. In the previous chapter, I introduced a framework which is ideally suited for the task at hand. There, I associate two-time states with the preparation and second measurement in a two-measurement process. In a prepare-and-measure QKD scheme, these two events are the actions of the legitimate users. Also, I associated the quantum channel with the Choi-Jamiołkowski vector of the Kraus operator, an object which I called the Kraus vector. This is the piece of the scheme associated with eavesdropping.

In this chapter, which is adapted from Ref. [2], I show that the formalism of two-time states and Kraus vectors can be used to optimise eavesdropping strategies. Typically,

the problem is transformed into that of finding the eigenvalues of superoperators. After some slight modifications of the framework so that it is more suited for this task, it is applied to three QKD protocols: BB84 [66], B92 [67] and PBC00 [68]. The former two are both well studied in the literature and I acquire the known best schemes. For PBC00, which is less well-explored, a novel result is found: that the best attack does not transfer any information about the signal states to Eve. All of this work is then linked to other aspects of quantum cryptanalysis, for example the recently popular measurement device independent schemes.

4.1 Framework

In the previous chapter I showed, from some basic physical assumptions about quantum theory, that the Kraus rule is the unique joint probability rule. This led me to employ an operator space formalism in which different aspects of a preparation-measurement-measurement scheme are associated with two-time states (for the preparation and second measurement) and Kraus vectors (for the first measurement). This separation maps neatly onto the knowledge which Eve can use to design her eavesdropping strategies. For reference, the probability rule which was derived is

$$\mathrm{P}(i,j|s,x) = \sum_k \langle\langle \Psi_j | A_{ik} \rangle\rangle \langle\langle A_{ik} | \Psi_j \rangle\rangle. \tag{4.1}$$

The object denoted $|\Psi_j\rangle\rangle$ in this equation is the two-time state and is used in the calculations of this chapter to represent correlations between Alice's preparations and Bob's measurements. If their actions are limited to preparing pure states $|\psi\rangle$ and measuring projectively $|\phi_j\rangle$ then the two-time state has the form $|\Psi_j\rangle\rangle = |\psi\phi_j^\dagger\rangle\rangle$. The other object that appears in the probability rule is the vector $|A_{ik}\rangle\rangle$, which is isomorphic to the Kraus operators A_{ik} which act in the usual Hilbert space formalism. This formula can be rewritten as

$$\mathrm{P}(i,j|s,x) = \sum_k \langle\langle A_{ik} | \Psi_j \rangle\rangle \langle\langle \Psi_j | A_{ik} \rangle\rangle. \tag{4.2}$$

It is seen that, whereas the channel was initially seen to act as a superoperator acting upon the two-time state, it is equally valid to think of the two-time state as a superoperator acting upon the intermediate measurement outcome. In terms of information processing, this can be associated with the act of pre- and post-selection. In quantum key distribution schemes, Alice and Bob share classical bits in terms of correlations between particular preparations and measurements; the calculations performed below all begin by constructing a superoperator in terms of the relevant two-time states.

I will use the above probability rules to construct figures of merit which quantify the amount of knowledge Eve can extract from each qubit. Before doing so, I simplify the problem by limiting the range of possible attacks.

I have already made one restriction in setting up the problem by assuming that collective and coherent attacks, which take information from more than one Alice-Bob qubit, are disallowed. The reason for this is that the derivations of eavesdropping strategies

are to be understood partly as a demonstration of the capabilties of the operator space formalism for sequential measurements. I would also argue that investigating particular subsets of attacks can help us to gain insights into the underlying logic of QKD protocols and this will be demonstrated in particular for the three-state protocol PBC00.

A further limit I place on Eve's attacks is that she associates a single Kraus operator with each bit value. There is a good reason to do this, which is that such a measurement can be seen to be minimally disturbing, in the sense that the post-measurement state will be closer to the initial state than if multiple Kraus operators were allowed. This is most easily seen by showing that any two-index Kraus operator can be implemented in two stages. I define $A_i = U_i \pi_i^{\frac{1}{2}}$ and $A'_{ik} = A_{ik} \pi_i^{-\frac{1}{2}} U_i^\dagger$ as the Kraus operators representing this two-stage process, where $\pi_i^{\frac{1}{2}}$ is an effect depending on the required instrument; it is seen that $A_{ik} = A'_{ik} A_i$, so that the operator of interest can be represented in this form, and also that $\sum_i A_i^\dagger A_i = \sum_{ik} A_{ik}'^\dagger A'_{ik} = I$, so that each of these steps is in itself a valid measurement. The two-index attack A_{ik} is implemented by a single index operator A_i followed by a second measurement and, as such, will disturb the incoming state more than A_i by itself would. For this reason, in deriving the eavesdropping strategies I associate a single operator $|E_i\rangle\rangle$ with each classical bit value.

With these assumptions in place, Eq. 4.2 is rewritten as

$$\mathrm{P}(i,j|s,x) = \langle\langle E_i|S_j|E_i\rangle\rangle. \tag{4.3}$$

I have introduced the notation S_j for the superoperator associated with Alice and Bob's correlations, which is to be formed of outer products of two-time states (as well as sums of these). In this sense, the superoperator is similar to density operators, which are also formed of outer products of states. Eve's attack is represented by the Kraus vector $|E_i\rangle\rangle$, with the notation changed to avoid confusion with Alice's preparation. In what follows, I first construct the superoperator relevant to the particular QKD protocol which is analysed. I use this, along with Eq. 4.3, to maximise a set of figures of merit (soon to be introduced) by allowing the set $|E_i\rangle\rangle$ to vary while ensuring simultaneously that they form a complete set,

$$\sum_n \langle n^\dagger|E_i\rangle\rangle\langle\langle E_i|n^\dagger\rangle = I. \tag{4.4}$$

A discussion of this point surrounds Eq. 3.52, in the previous chapter. The set of $|n\rangle$ here is any complete set of basis vectors that span the Hilbert space.

There are a number of senses in which an eavesdropping task might be said to be optimal. In this work I consider two figures of merit. As I am only considering individual attacks, rather than providing a full security analysis, each value seeks to quantify the amount of information which Eve gains from a single qubit. One is the probability that all three parties agree on the bit value, denoted $\mathrm{P}(A = E = B)$ and the other is the probability that all three agree conditioned upon agreement between the two legitimate users, $\mathrm{P}(A = E = B|A = B)$. Of course, these two are related by the usual rule of

conditional probability:

$$\mathrm{P}(A = E = B | A = B) = \frac{\mathrm{P}(A = E = B)}{\mathrm{P}(A = B)}. \tag{4.5}$$

On the denominator of the expression on the right hand side of this equation appears the term for Alice and Bob's agreement. It is this figure which quantifies Eve's ability to avoid detection. After their processes of sifting and privacy amplification, the two legitimate parties will announce a number of their bits and with this information will estimate $\mathrm{P}(A = B)$ in order to decide whether or not an eavesdropper has been present. In an ideal system any errors herald Eve but in reality there will be some inherent noise and Eve will be safe if the induced noise is low enough, as Alice and Bob are forced to assume that all errors are due to Eve. Rigorous security proofs provide quantum bit error rates Q below which the key can be made secure through techniques of privacy amplification. I provide this latter quantity alongside the probability $\mathrm{P}(A \neq B) = 1 - \mathrm{P}(A = B)$ that the two legitimate users disagree with each other as to the bit value, which helps to contextualise the attack.

I take advantage of the principle of *bit symmetry*, by which I mean that, given the high level of symmetry in the protocols under consideration, it should hold that if all three parties relabel which measurement outcomes correspond to which bit values then all probabilities are invariant. A particular example would be all parties agreeing on a particular bit value: the probability that all three bit values are zero should be equal to the probability that all three bit values are one. This is expressed algebraically as $\langle\langle E_0|S_0|E_0\rangle\rangle = \langle\langle E_1|S_1|E_1\rangle\rangle$ and used to simplify some of the expressions that appear. In fact, as shown by Fuchs *et al.* [69], this places no restrictions upon the possible schemes that Eve may consider as there always exists a bit-symmetric attack which can reach the same value for the figures of merit that I consider as one which is not bit symmetric. This property also simplifies the search for optimised Kraus operators in a different way. In the case that the quantum system being represented is a qubit, Eq. 4.4 implies that

$$\sum_i \mathrm{Tr}\left(|E_i\rangle\rangle\langle\langle E_i|\right) = 2. \tag{4.6}$$

As each bit value is associated with a single Kraus operator and these operators are symmetric, this equation implies that $\langle\langle E_0|E_0\rangle\rangle = \langle\langle E_1|E_1\rangle\rangle = 1$ and hence the vectors which I seek below must be normalised.

4.2 General results

The sifting stage in a QKD protocol will remove, from the measurement record, anys outcomes which are irrelevant to the final key. It is directly after this stage that the classical bit values are assigned, and it is these bit values which are represented by superoperators. In particular these superoperators represent post-selection of the three possible outcomes. Two possible outcomes are those in which the parties share a bit value: Alice and Bob both believe the bit value is either 0 or 1. I assign to these cases the superoperators S_0

and S_1. These are constructed of outer products of the two time states, $|\Psi_j\rangle\rangle$, which represent the correlations which leave Alice and Bob with that shared information, as well as sums of those. For example, if a protocol states that Alice sending the state $|0\rangle$ and Bob measuring the state $|0\rangle$ gives them a bit value of 0 then the associated superoperator will be $S_0 = |00^\dagger\rangle\rangle\langle\langle 00^\dagger|$; if the protocol *also* specifies that Alice sending $|1\rangle$ and Bob measuring $|1\rangle$ results in that bit value then the superoperator will include this outer product, i.e., $S_0 = |00^\dagger\rangle\rangle\langle\langle 00^\dagger| + |11^\dagger\rangle\rangle\langle\langle 11^\dagger|$ in this artificial example. In a similar manner the superoperator S_X will represent those outcomes in which Alice and Bob disagree about the value of the key's bit, i.e., Alice believes that it is 0 but Bob believes that it is 1 or vice versa. In general, it is true that Eve can change the sifting rate by her actions and this dynamic is captured by the superoperator $S_S = S_0 + S_1 + S_X$ which represents the chance that a given qubit is *not* sifted from the final key (as all non-sifted bits must be agreed or disagreed upon by the two legitimate parties).

From these superoperators it is possible to arrive at general results which can then be applied in the case of specific eavesdropping protocols to find optimal strategies. This is the route that I take here. It has the advantage of separating out the features which are true of all eavesdropping protocols from those which arise in specific cases.

For each protocol, two strategies are derived which are each optimal in different senses. These are represented by different figures of merit. As I explained, one is the probability that all three parties agree upon the bit value. Such an event is conditional upon a given bit value of the timeslot not being removed during the sifting process, and so the quantity is most generally expressed by

$$\mathrm{P}(A = E = B) = \frac{\sum_i \mathrm{P}(B = E = i | A = i)\mathrm{P}(A = i)}{\mathrm{P}(S)}, \tag{4.7}$$

where i gives the bit value and where I have denoted by $\mathrm{P}(S)$ the probability of a given bit not being sifted, which in full is

$$\mathrm{P}(S) = \sum_i \mathrm{P}(B = E = i | A = i)\mathrm{P}(A = i) + \sum_{ij} \mathrm{P}(B = i, E = i \text{ or } j | A = j)\mathrm{P}(A = j) \tag{4.8}$$

as sifting occurs in the set of cases for which neither party assigns a bit value. All of the protocols considered here are unbiased in the sense that all signal states are sent by Alice with equal probability and so the probabilities $\mathrm{P}(A = i)$ will be factored out of both numerator and denominator. Using that fact, and writing the probabilities in terms of superoperators (Eq. 4.3), gives

$$\mathrm{P}(A = E = B) = \frac{\sum_i \langle\langle E_i | S_i | E_i \rangle\rangle}{\sum_i \langle\langle E_i | (S_0 + S_1 + S_X) | E_i \rangle\rangle}. \tag{4.9}$$

This is further simplified using bit symmetry, as outlined above. This enforces that, for the measurement outcomes which I am searching for, $\langle\langle E_0 | S_0 | E_0 \rangle\rangle = \langle\langle E_1 | S_1 | E_1 \rangle\rangle$ and similarly for the other sets of bit values. With these rules the vector $|E_1\rangle\rangle$ is eliminated

and the figure of merit is expressed as

$$\mathrm{P}(A = E = B) = \frac{\langle\langle E_0|S_0|E_0\rangle\rangle}{\langle\langle E_0|(S_0 + S_1 + S_X)|E_0\rangle\rangle}. \tag{4.10}$$

At first glance this formula suggests that it is possible to ensure agreement between all three parties by enforcing that $\langle\langle E_0|(S_1 + S_X)|E_0\rangle\rangle = 0$, however it is not possible for this condition to hold in general. If only the two outcomes $|E_0\rangle\rangle$ and $|E_1\rangle\rangle$ form a given measurement, this condition is equivalent to the statement that the probability of shared bit value 1 and the probability of disagreement are both equal to zero or, to take a more concrete example, $\mathrm{P}(B = 0, E = 0|A = 1) = \mathrm{P}(B = 1, E = 0|A = 1) = 0$. This can only hold in two cases. One is that $\mathrm{P}(E = 0|A = 1) \neq 0$, so that we have $\mathrm{P}(B = 0|A = 1) = \mathrm{P}(B = 1|A = 1) = 0$ according to Bayes's rule. but this would imply that Bob may believe the bit value to be neither zero nor one, which is obviously absurd. The other possibility is that $\mathrm{P}(E = 0|A = 1) = 0$. This condition can only hold if Eve's POVM element has zero overlap with the density matrix associated with Alice's signal, and this cannot be done if Alice sends a mixed state. It is seen that in general *no* attack can give Eve all of the legitimate user's shared key string. In order to maximise Eq. 4.10 it is necessary to inspect the specific form in each case.

The second figure of merit which I calculate is the probability that all three parties agree, conditioned upon agreement between the two legitimate users. As the set of outcomes satisfying the latter condition is included in the set of non-sifted results, the superoperator S_X need not be used and this probability can be written as

$$\begin{aligned} &\mathrm{P}(A = E = B|A = B) \\ &= \frac{\sum_i \mathrm{P}(B = E = i|A = i)\mathrm{P}(A = i)}{\sum_i \mathrm{P}(B = E = i|A = i)\mathrm{P}(A = i) + \sum_{ij} \mathrm{P}(B = i, E = j|A = i)\mathrm{P}(A = i)}. \end{aligned} \tag{4.11}$$

Again, that Alice's signal states are equiprobable and that Eve's attacks are bit symmetric allow this expression to be written solely in terms of the vector $|E_0\rangle\rangle$:

$$\mathrm{P}(A = E = B|A = B) = \frac{\langle\langle E_0|S_0|E_0\rangle\rangle}{\langle\langle E_0|(S_0 + S_1)|E_0\rangle\rangle}. \tag{4.12}$$

This result has the suprising implication that, for *any* quantum key distribution protocol, it is possible for Eve to uncover *all* bit values used from the subset in which Alice and Bob have a shared bit value. This will be the case if she constructs Kraus operators which satisfy

$$\langle\langle E_0|S_1|E_0\rangle\rangle = 0. \tag{4.13}$$

If this condition is satisfied then $\mathrm{P}(A = E = B|A = B) = 1$. The reason that this will be possible in all cases for this figure of merit but not the former is that there is no further requirement for Eve to satisfy; she need not ensure simultaneously that $\langle\langle E_0|S_X|E_0\rangle\rangle = 0$ and so, in contrast to the previous figure of merit, can *always* construct a measurement such that she knows the entire bit string for this subset of qubits. Of course, Eve will pay the price of introducing errors which reveal her actions. In general, a whole class of

measurements will satisfy this condition and further considerations must be used to pick out a particular measurement. Below, I am guided by Eve's desire to hide.

4.3 BB84

4.3.1 Scheme

The first and most well-explored QKD scheme is due to Bennett and Brassard and was published in 1984. Hence, it is known as BB84 [66]. This scheme uses four different qubit states: the computational basis states $|0\rangle$ and $|1\rangle$ and the σ_x eigenbasis states $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ and $|-\rangle = (|0\rangle - |1\rangle)/\sqrt{2}$, which can correspond to the horizontal-vertical versus diagonal polarisations of a photon in an optical scheme. Alice has access to a random number generator. She prepares a system in one of these four states with equal probabilities and sends it to Bob. The latter measures with a POVM corresponding to suitably weighted projectors onto the same set of states and records his outcome. After this is repeated for all N resource qubits, one of the two parties will for each qubit announce only the basis from which their state is drawn. In the cases that the two parties agree on the basis then they know unambiguously that they each agree with the other assuming that there is no noise on the channel and no eavesdropping has occured. Otherwise, there is ambiguity in the shared information and those qubits need to be discarded. For example, if Alice sends the state $|0\rangle$ and Bob announces that his result was in the computational basis then each knows the other's bit value. However, Bob's outcomes $|+\rangle$ or $|-\rangle$ are consistent with both $|0\rangle$ and $|1\rangle$ being transmitted and therefore the two parties cannot share information. After the sifting, classical bit values are assigned to each slot. I use the convention in which zero is assigned to $|0\rangle$ and $|+\rangle$ and one is assigned to $|1\rangle$ and $|-\rangle$. In this manner, Alice and Bob are able to share a string of bits which form their secret key.

A large literature exists on the security of and possible attacks upon the BB84 protocol, and it has formed the basis for experimental QKD [70, 71]. The information-theoretic security has been proven under a number of different conditions [72], most famously in an analysis by Shor and Preskill [73]. At the time of writing, the variant of BB84 which performs best is a modification by Gottesman and Lo [74] in which Alice and Bob both make public announcements after the protocol. For that routine, security holds up to a noise level of $Q = 18.9\%$. Alongside this full-blown security analysis a wide range of specific attacks have been considered. The simplest possible is that Eve measures the sent qubit projectively and resends based upon her result. As is shown in §2.5, the pair of states which maximise her chance of correctly identifying the bit are those which form the Breidbart basis. If this attack is performed then the probability that all three parties subsequently share a qubit is $\mathrm{P}(A = E = B) = (3 + 2\sqrt{2})/8 \approx 0.73$, however it is possible to do better, in terms of the figures of merit considered here. Eve's most general possible scheme is to entangle the signal qubit with a probe and then measure the latter. By allowing for this more general set of attacks, researchers have been able to show that Eve's best action is to perform a CNOT gate, acting in the Breidbart basis, in which the

sent qubit is the target and the control is her own suitably prepared probe. This attack is known as the Fuchs-Peres-Brandt attack [69, 75, 76] and is derived below as the attack which maximises the probability of conditional agreement.

4.3.2 Eavesdropping strategy

I now turn to the task of optimising Eve's action and begin by representing the measurements associated with bit value 0 and 1 as $|E_0\rangle\rangle$ and $|E_1\rangle\rangle$ respectively.

The first step is to construct the relevant superoperators which are associated with Alice and Bob's postselection. In the BB84 protocol, the resultant bit value is 0 in two cases: either, Alice sends the state $|0\rangle$ and Bob measures with $|0\rangle$ as the outcome, *or*, Alice sends the state $|+\rangle$ and Bob measures with $|+\rangle$ as the outcome. As discussed above, this information is represented by the superoperator

$$S_0 = |00^\dagger\rangle\rangle\langle\langle 00^\dagger| + |++^\dagger\rangle\rangle\langle\langle ++^\dagger|. \tag{4.14}$$

Similarly, the bit value 1 is assigned to a timeslot in the cases that Alice and Bob agree that the sent state was either $|1\rangle$ or $|-\rangle$. These two cases are represented by the superoperator

$$S_1 = |11^\dagger\rangle\rangle\langle\langle 11^\dagger| + |--^\dagger\rangle\rangle\langle\langle --^\dagger|. \tag{4.15}$$

In terms of these two definitions, the probability that all three parties agree upon the bit value in a given timeslot, conditioned upon the fact that the bit value was not discarded during the sifting stage of the protocol, is given by Eq. 4.10. In fact the BB84 protocol in particular is simpler than others to work with as the rate of sifting is independent of Eve's attack. The reason for this is that sifting depends only upon Bob's choice of measurement basis, and is independent of the state preparation and measurement outcome. Formally this manifests in the fact that $S_S = S_0 + S_1 + S_X = 2I$ (I here is the identity operator which has the property $I|A\rangle\rangle = |A\rangle\rangle$ for all $|A\rangle\rangle$; see previous chapter for discussion of this point) and so the probability rule is simplified to

$$\mathrm{P}(A = E = B) = \frac{\langle\langle E_0|S_0|E_0\rangle\rangle}{2}. \tag{4.16}$$

This figure can be straightforwardly maximised by letting the eavesdropper's Kraus vectors be proportional to the eigenvector of the relevant superoperator which has the largest eigenvalue. The Kraus vector must also be normalised. For the current case, then, the form of $|E_0\rangle\rangle$ that maximises its expectation value upon S_0 is

$$|E_0\rangle\rangle = \frac{1}{N_0}\left(|00^\dagger\rangle\rangle + |++^\dagger\rangle\rangle\right). \tag{4.17}$$

The normalisation constant N_0 here must be such that $\langle\langle E_0|E_0\rangle\rangle = 1$. For this to hold then $N_0 = \sqrt{3}$. The relevant eigenvector of S_1 is evaluated in a similar manner. Bringing

both results together, I represent the overall strategy by the two Kraus vectors

$$\begin{aligned} |E_0\rangle\rangle &= \frac{1}{\sqrt{3}}\left(|00^\dagger\rangle\rangle + |++^\dagger\rangle\rangle\right) \\ |E_1\rangle\rangle &= \frac{1}{\sqrt{3}}\left(|11^\dagger\rangle\rangle + |--^\dagger\rangle\rangle\right). \end{aligned} \tag{4.18}$$

The associated eigenvalue is :

$$\begin{aligned} S_0|E_0\rangle\rangle &= \frac{1}{\sqrt{3}}\left(1 + \langle\langle 00^\dagger| + +^\dagger\rangle\rangle\right)\left(|00^\dagger\rangle\rangle + |++^\dagger\rangle\rangle\right) \\ &= \frac{3}{2}|E_0\rangle\rangle. \end{aligned} \tag{4.19}$$

This result is used to evaluate Eq. 4.16 and I find that

$$\mathrm{P}(A = E = B) = \frac{3}{4} \tag{4.20}$$

is the maximum value that this probabilty can take.

As the reader will be more familiar with the Hilbert space formalism for quantum mechanics, it is helpful to represent the measurement in that form. By the usual isomorphism between the Kraus vectors and operators in Hilbert space, the operation can be expressed using the Pauli operators as

$$\begin{aligned} E_0 &= \frac{1}{\sqrt{3}}\left(|0\rangle\langle 0| + |+\rangle\langle +|\right) \\ &= \frac{1}{2\sqrt{3}}\left(2I + \sigma_x + \sigma_z\right). \\ E_1 &= \frac{1}{\sqrt{3}}\left(|1\rangle\langle 1| + |-\rangle\langle -|\right) \\ &= \frac{1}{2\sqrt{3}}\left(2I - \sigma_x - \sigma_z\right). \end{aligned} \tag{4.21}$$

It is also useful to know the probability that this measurement results in Alice and Bob disagreeing upon the bit value of the key in a given slot. This probability quantifies the possibility that they will discover Eve's action and abort the protocol. In the BB84 strategy, disagreement occurs if Bob's measured state is orthogonal to that sent by Alice. The probability is best calculated using the superoperator S_X introduced earlier which, for this case, takes the form

$$S_X = |01^\dagger\rangle\rangle\langle\langle 01^\dagger| + |10^\dagger\rangle\rangle\langle\langle 10^\dagger| + |+-^\dagger\rangle\rangle\langle\langle +-^\dagger| + |-+^\dagger\rangle\rangle\langle\langle -+^\dagger|. \tag{4.22}$$

In terms of this object, the probability that the two legitimate parties disagree is then

$$\mathrm{P}(A \neq B) = \frac{1}{4}\left(\langle\langle E_0|S_X|E_0\rangle\rangle + \langle\langle E_1|S_X|E_1\rangle\rangle\right). \tag{4.23}$$

The factor of 1/4 is again calculated from the sifting operator. It corresponds to the fact that Alice chooses between four equiprobable states. In order to evaluate this expression,

the expectation of the operator S_X under the pair of Kraus vectors (Eq. 4.18) is required. Taking $|E_0\rangle\rangle$ as a concrete example, one finds first

$$S_X|E_0\rangle\rangle = \frac{1}{2\sqrt{3}}\left(|01^\dagger\rangle\rangle + |10^\dagger\rangle\rangle + |+-^\dagger\rangle\rangle + |-+^\dagger\rangle\rangle\right) \tag{4.24}$$

and, from this, the expectation value is

$$\langle\langle E_0|S_X|E_0\rangle\rangle = \frac{1}{3}. \tag{4.25}$$

Precisely the same value is found for the other measurement outcome, which one might expect given the symmetry of the protocol. Overall the probability that Alice and Bob disagree, given that Eve has used her best attack, is

$$\mathrm{P}(A \neq B) = \frac{1}{6}. \tag{4.26}$$

The next figure of merit to be analysed is the probability that the three parties agree conditioned upon agreement between the two legitimate parties. For the previously derived optimal scheme, Eq. 4.21, this is evaluated from the quantities derived so far:

$$\mathrm{P}(A = E = B|A = B) = \frac{\mathrm{P}(A = E = B)}{1 - \mathrm{P}(A \neq B)} = \frac{3/4}{1 - 1/6} = \frac{9}{10} \tag{4.27}$$

however it may be expected that Eve can do better, at the cost of introducing errors, especially given that she knows more about the correlations between Alice and Bob in the considered scenario. In §4.2, Eq. 4.13, I showed that any measurement such that $|E_0\rangle\rangle$ has an overlap of zero with S_1 will cause all three parties to agree under the conditions of this figure of merit. Any $|E_0\rangle\rangle$ of the form

$$|E_0\rangle\rangle = a|0+^\dagger\rangle\rangle + b|+0^\dagger\rangle\rangle \tag{4.28}$$

satisfies this requirement, as can be seen by inspecting Eq. 4.15. The two coefficients here can be freely chosen, subject to the contraint that the vector $|E_0\rangle\rangle$ is normalised, i.e.,

$$a^2 + ab + b^2 = 1 \tag{4.29}$$

must hold. The other measurement outcome $|E_1\rangle\rangle$ is that which is zero when acted upon by the superoperator S_0 and must take the form

$$|E_1\rangle\rangle = a|1-^\dagger\rangle\rangle + b|-1^\dagger\rangle\rangle, \tag{4.30}$$

again subject to the same constraints. To emphasise: *any* normalised vectors satisfying the above two equations will form a measurment in which $\mathrm{P}(A = E = B|A = B) = 1$. What distinguishes the attacks is the varying levels of noise that they introduce. As an example, I consider the simplest case: $a = 1, b = 0$. This is the pair of vectors $|E_0\rangle\rangle = |0+^\dagger\rangle\rangle, |E_1\rangle\rangle = |1-^\dagger\rangle\rangle$ (or, alternatively the measurement can represented by the

operators $E_0 = |+\rangle\langle 0|, E_1 = |-\rangle\langle 1|$). The first outcome is impossible if Alice sends that state $|1\rangle$ and that result, E_0, will leave all other signal qubits in the state $|+\rangle$. Of these three preparations, $|0\rangle$ and $|+\rangle$ will lead to agreement between Alice and Bob and for $|-\rangle$ they will disagree, however, this latter possibility is not part of the subset of events considered by this figure of merit. Overall, Bob will definitely receive either $|+\rangle$ or $|-\rangle$ and for both of these states there is a fifty percent chance that he disagrees with Alice as to which state was sent. Hence $\mathrm{P}(A \neq B) = 1/2$ for this measurement. It is natural to minimise the noise as a means of selecting from the wider space of possible measurements.

This is a constrained optimisation task: I require the minimum of $\mathrm{P}(A \neq B)$ subject to constraint Eq. 4.29. The first step is to write this quantity in terms of a and b:

$$\mathrm{P}(A \neq B) = \langle\langle E_0|S_X|E_0\rangle\rangle + \langle\langle E_1|S_X|E_1\rangle\rangle = \frac{1}{2}\left(a^2 + b^2\right). \tag{4.31}$$

In order to optimise this subject to the constraint, I introduce the function

$$F(a,b) = \frac{1}{2}\left(a^2 + b^2\right) + \lambda\left(a^2 + b^2 + ab - 1\right), \tag{4.32}$$

where λ is a variable which is to be found. Optimisation will occur when $\partial F(a,b)/\partial a = 0 = \partial F(a,b)/\partial b$ and these two constraints collectively enforce that $a = b$ and fix $\lambda = 2/3$. Substituting $a = b$ into Eq. 4.29 gives $a = \pm 1/\sqrt{3}$, the two possible solutions here being equivalent measurements up to a phase which does not contribute to the measurement process. I choose the uppermost value of a and the measurement which is found to maximise $\mathrm{P}(A = E = B|A = B)$ while simultaneously minimising $\mathrm{P}(A \neq B)$ is represented by the pair of vectors

$$\begin{aligned} |E_0\rangle\rangle &= \frac{1}{\sqrt{3}}\left(|0+^\dagger\rangle\rangle + |+0^\dagger\rangle\rangle\right) \\ |E_1\rangle\rangle &= \frac{1}{\sqrt{3}}\left(|1-^\dagger\rangle\rangle + |-1^\dagger\rangle\rangle\right). \end{aligned} \tag{4.33}$$

All that remains is to make a link between this and the Fuchs-Peres-Brandt attack, which was stated above to be Eve's best attack for the BB84 protocol. This is seen most straightforwardly by writing the two vectors in their Kraus operator representation:

$$\begin{aligned} E_0 &= \frac{1}{\sqrt{3}}\left(|+\rangle\langle 0| + |0\rangle\langle +|\right) \\ &= \frac{1}{\sqrt{6}}\left(\sigma_x + \sigma_z + I\right), \\ E_1 &= \frac{1}{\sqrt{3}}\left(|-\rangle\langle 1| + |1\rangle\langle -|\right) \\ &= \frac{1}{\sqrt{6}}\left(\sigma_x + \sigma_z - I\right). \end{aligned} \tag{4.34}$$

Any set of Kraus operators can be implemented by a CNOT gate which acts in the basis in which they are mutually diagonalised. Here, this is the Breidbart basis therefore I have arrived at the Fuchs-Peres-Brandt attack [69, 75, 76]. While this was already established

as Eve's best attack in the BB84 protocol, it has been found here as the solution to a two step calculation. A general class of measurements was seen to solve and eigenvalue problem and this class was selected from by contrained optimisation.

4.4 B92

4.4.1 Scheme

The second illustration of my method for finding optimal eavesdropping strategies is an analysis of B92, a protocol developed by Bennett who realised that the BB84 protocol could be performed with just two states [67]. In this scheme Alice sends either $|0\rangle$ or $|+\rangle$ with equal probability and Bob measures in the same manner as in the previously discussed strategy, using a POVM consisting of equally weighted projectors onto $|0\rangle, |1\rangle, |+\rangle, |-\rangle$. If the received states are either $|0\rangle$ or $|+\rangle$ then Bob does not know which state was sent, as both $|0\rangle$ and $|+\rangle$ would be consistent with his result. He announces either outcome and those qubits are sifted from the final key. However, in the case that his outcome is $|-\rangle$ then he knows that $|+\rangle$ could not have been sent (as it is orthogonal to the measured state) and so Alice must have prepared $|0\rangle$. These outcomes are recorded and assigned classical bit value 0. Similarly, if he measures $|1\rangle$ then $|+\rangle$ must have been sent and these qubits are given the classical bit value 1. In this manner, a key can be formed. More generally, any two non-orthogonal signal states can be used and a relevant POVM constructed, however the analysis is exactly the same and so I consider only one particular case.

There is a well-known vulnerability to B92 in that Eve can perform unambiguous state discrimination on the incoming bits [15, 19, 20, 21, 77]. The result of this is that she is able to characterise precisely in which state Alice prepared her qubit, at the cost of having no knowledge of the sent state for some subset of results. For this reason, the protocol is only secure up to a noise level of $Q = 3.4\%$ [78, 79].

4.4.2 Eavesdropping strategy

I proceed as in the BB84 analysis, constructing superoperators associated with Alice and Bob's shared bits as well as disagreement between them. There is an added complexity for B92 over BB84 in that, as the post-selected measured states do not form an orthogonal set, it is possible for Eve's measurement to change the amount of sifting which occurs. This is also handled using a superoperator. Using all of these objects, expressions for the various figures of merit will be found and then maximised, again as an eigenvalue problem.

The two shared-bit superoperators are

$$\begin{aligned} S_0 &= |0-^\dagger\rangle\rangle\langle\langle 0\ -^\dagger\ | \\ S_1 &= |+\ 1^\dagger\rangle\rangle\langle\langle +1^\dagger|. \end{aligned} \tag{4.35}$$

Both of these forms can be seen by noting which correlations correspond to which classical bits in the protocol. I also construct a superoperator associated with the sifting that occurs. This includes, as well as the shared bits, those cases in which Bob measures the

state $|1\rangle$ given that Alice has sent the state $|0\rangle$, which cannot occur in the absence of an eavesdropper but which are induced by Eve's measurements. The piece which is associated with cases in which Alice and Bob disagree is

$$S_X = |01^\dagger\rangle\rangle\langle\langle 01^\dagger| + |+-^\dagger\rangle\rangle\langle\langle +-^\dagger|. \tag{4.36}$$

The set of outcomes that are not sifted is associated with the superoperator S_S, which may be written in terms of the above objects as $S_S = S_0 + S_1 + S_X$. As shown in Eq. 4.10, the value $\mathrm{P}(A = E = B)$ is given most simply in terms of the Kraus vector $|E_0\rangle\rangle$ only. At first glance it seems possible to find a measurement in which $\langle\langle E_0|(S_1 + S_X)|E_0\rangle\rangle = 0$, which would give $\mathrm{P}(A = E = B) = 1$. This would be the case if

$$|E_0\rangle\rangle = A|-0^\dagger\rangle\rangle, \tag{4.37}$$

in which A is a variable. The symmetric result associated with the other bit value is

$$|E_1\rangle\rangle = A|1+^\dagger\rangle\rangle. \tag{4.38}$$

However, this cannot be a complete measurement and there must be another outcome associated with each bit value. Why? As discussed following Eq. 4.10, it cannot be true that $\langle\langle E_0|(S_1 + S_X)|E_0\rangle\rangle = 0$, as this implies some nonsensical results (e.g., Bob finding neither bit value for some qubits). Another way to see that this pair of measurement outcomes is unphysical is by checking the normalisation condition, Eq. 4.4, which requires that A satisfies

$$A^2\left(|1\rangle\langle 1| + |-\rangle\langle -|\right) = I. \tag{4.39}$$

On the right hand side is the identity and on the left hand side the only freedom is a constant of proportionality. The bracketed object is clearly not proportional to the identity and so no measurement can be performed which is satisfied by $|E_0\rangle\rangle$ and $|E_1\rangle\rangle$ in their current form. The solution to this problem is to introduce a third possible outcome, $|E_?\rangle$, which allows the set of outcomes to be completed and yet is associated with neither classical bit. If I allow that Eve associates the bit value 0 with the timeslot in half of the cases in which she gets this result, and the bit value 1 in the other half, then it is no longer true that either bit value is encoded in just a single Kraus vector and there is no issue with normalisation. The requirement that the measurement with all three outcomes be trace-preserving is

$$A^2\left(|1\rangle\langle 1| + |-\rangle\langle -|\right) + \sum_i \langle i^\dagger|E_?\rangle\rangle\langle\langle E_?|i^\dagger\rangle = I. \tag{4.40}$$

The third measurement outcome tells Eve nothing about the key which is being shared and so the optimal scheme minimises the chance that this outcome happens. However, that outcome must still correspond to an effect and hence $\pi_? = \sum_i \langle i^\dagger|E_2\rangle\rangle\langle\langle E_2|i^\dagger\rangle$ must

be a positive operator. The operator $\pi_?$ can be expressed, following Eq. 4.40, as

$$\begin{aligned}\pi_? &= I - A^2\left(|1\rangle\langle 1| + |-\rangle\langle -|\right) \\ &= (1 - \frac{A^2}{2})|0\rangle\langle 0| - \frac{A^2}{2}\left(|0\rangle\langle 1| + |1\rangle\langle 0|\right) + (1 - \frac{3A^2}{2})|1\rangle\langle 1|. \end{aligned} \tag{4.41}$$

The task is to maximise A (which parameterises the likelihood of either of the two useful outcomes) while ensuring that $\pi_? \geq 0$. The two eigenvalues of this operator are found to be $\lambda = 1 - (2 \pm \sqrt{2})A^2/2$. In order that $\pi_?$ be postive semi-definite I need to consider only the lowermost of these two eigenvalues. Setting this to zero gives $A^2 = 2 - \sqrt{2}$ and so $\pi_?$ has a single non-zero eigenvector $|\psi_?\rangle$. Any $|E_?\rangle\rangle$ satisfying the pair of equations

$$\begin{aligned}\pi_? &= \sum_i \langle i^\dagger|E_?\rangle\rangle\langle\langle E_?|i^\dagger\rangle = |\psi\rangle\langle\psi| \\ |\psi\rangle &= \frac{\sqrt{2-\sqrt{2}}}{2}\left((-1-\sqrt{2})|0\rangle + |1\rangle\right) \end{aligned} \tag{4.42}$$

is an optimal measurement. This freedom in choosing the Kraus operator is the same as one has in mapping effects onto their associated instruments. What has been shown is that Eve is best-served by performing unambiguous state discrimination: she can know the sent state precisely for some subset of qubits at the cost of losing all information about the rest of them. As mentioned above, this is well established to be the weakness of the B92 protocol.

The probability that all three parties agree upon the bit value, conditioned upon agreement between the legitimate users, is the second figure of merit considered in each of these demonstrations. This object can be maximised in the same manner as in the previously discussed cases. As shown in Section 3.3, one can always satisfy $\mathrm{P}(A = E = B|A = B) = 1$ by finding a pair of Kraus vectors such that $\langle\langle E_0|S_1|E_0\rangle\rangle = 0$, i.e., one ensures that Eve can never get the outcome 0 in the cases which Alice and Bob assign to 1, and vice versa. The pair of outcomes satisfying this condition is

$$\begin{aligned}|E_0\rangle\rangle &= a|-\psi^\dagger\rangle\rangle + b|\phi 0^\dagger\rangle\rangle \\ |E_1\rangle\rangle &= c|1\lambda^\dagger\rangle\rangle + d|\rho +^\dagger\rangle\rangle. \end{aligned} \tag{4.43}$$

The space of measurements of this form is large: there are four complex variables a, b, c, d as well as four states $|\psi\rangle, |\phi\rangle, |\lambda\rangle, |\rho\rangle$, though not all objects may be freely chosen given the requirements of bit symmetry and trace preservation which I am enforcing for Eve's attacks. I begin by using the requirement of bit symmetry, extending the attack so that it is symmetric even in the cases in which Alice and Bob disagree upon the bit value, outcomes outside the regime in which it acts. That the probability of all three parties finding the same outcome should be equal for both bit values, zero and one, gives

$$|a\langle -^\dagger|\psi^\dagger\rangle + b\langle 0|\phi\rangle|^2 = |c\langle 1^\dagger|\lambda^\dagger\rangle + d\langle +|\rho\rangle|^2. \tag{4.44}$$

I now consider the cases in which Alice and Eve agree with each other but not with Bob

(i.e., $A = 0, B = 1, E = 0$ and $A = 1, B = 0, E = 1$). These will occur with equal probability if

$$|a\langle 1^\dagger|\psi^\dagger\rangle|^2 = |c\langle -^\dagger|\lambda^\dagger\rangle|^2. \tag{4.45}$$

Finally, the third condition is arrived at by enforcing bit symmetry between those in which only Eve and Bob agree with each other although not with Alice. I find

$$|b\langle +|\psi\rangle|^2 = |d\langle 0|\rho\rangle|^2. \tag{4.46}$$

There is, of course, a fourth condition: that Alice and Bob agree with each other but not with Eve, but this set is a subset of the cases in which Alice and Bob agree in general. As the figure of merit I consider enforces the latter set, this is not a further requirement. Guided by these formula, I choose a particular subset of measurements which satisfy this set without requiring further optimisation. By inspection, it is seen that Eqs. 4.44, 4.45 and 4.46 are all satisfied if $|\psi\rangle = |-\rangle, |\phi\rangle = |0\rangle$ and $|\rho\rangle = |+\rangle$. Furthermore, in order to contrast with the measurement derived from the previous figure of merit, I choose to look at attacks which preserve the trace with only two outcomes, i.e., I disallow unambiguous state discrimination. This choice of states is $a = c$ and $b = d$, and the two-outcome measurement is complete if

$$\begin{aligned}
&\sum_{ij} \langle i^\dagger|E_j\rangle\rangle\langle\langle E_j|i^\dagger\rangle \\
&= \left(\frac{a^2}{2} + \frac{3b^2}{2} + ab\right)|0\rangle\langle 0| + \frac{b^2 - a^2}{2}\left(|0\rangle\langle 1| + |1\rangle\langle 0|\right) + \left(\frac{3a^2}{2} + \frac{b^2}{2} + ab\right)|1\rangle\langle 1| \\
&= I.
\end{aligned} \tag{4.47}$$

It is easily seen that the parameterisation which satisfies this is $a = b = 1/\sqrt{3}$, so that the Liouville space representation of the measurement is

$$\begin{aligned}
|E_0\rangle\rangle &= \frac{1}{\sqrt{3}}\left(|-\,-^\dagger\rangle\rangle + |00^\dagger\rangle\rangle\right) \\
|E_1\rangle\rangle &= \frac{1}{\sqrt{3}}\left(|+\,+^\dagger\rangle\rangle + |11^\dagger\rangle\rangle\right).
\end{aligned} \tag{4.48}$$

As in the previous case it is useful for the reader to have this written out in the Kraus operator representation:

$$\begin{aligned}
E_0 &= \frac{1}{\sqrt{3}}\left(|-\rangle\langle -| + |0\rangle\langle 0|\right) \\
&= \frac{1}{2\sqrt{3}}\left(2I + \sigma_z - \sigma_x\right) \\
E_1 &= \frac{1}{\sqrt{3}}\left(|1\rangle\langle 1| + |+\rangle\langle +|\right) \\
&= \frac{1}{2\sqrt{3}}\left(2I - \sigma_z + \sigma_x\right).
\end{aligned} \tag{4.49}$$

In order to make clear that this measurement acts as expected, I consider a specific run

in which Alice prepares her qubit in the state $|0\rangle$. In one sixth of cases, Eve incorrectly identifies the bit value as 1, corresponding to the Kraus operator $|E_1\rangle\rangle$ and, if this occurs, the state sent on to Bob will be $|+\rangle$. At this point, it is definite that he disagrees with Alice as the outcome $|-\rangle$, which for him is associated with the bit value 0, cannot occur. Thus, Eve's measurement will ensure that if she has the wrong bit value then so does Bob, which is what I have required of this optimal measurement. Alice and Bob may still disagree in some of the cases in which Eve has correctly identified Alice's bit value and a value of $\mathrm{P}(A \neq B) = 1/5$ is found for this attack on the B92 protocol.

4.5 PBC00

4.5.1 Scheme

The third and final measurement that I consider is PBC00, which uses three trine states to share Alice and Bob's bits [68]. This set of states can be parameterised as

$$|\psi_k\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + e^{i2\pi k/3}|1\rangle\right), \tag{4.50}$$

in which $k = 0, 1, 2$ and which in an optical communication system would correspond to three equiangular linear polarisations of a photon. Alice picks one of these three states with equal probability and transmits it to Bob. The latter measures such that the corresponding POVM

$$\pi_k = \frac{2}{3}|\overline{\psi}_k\rangle\langle\overline{\psi}_k| \tag{4.51}$$

is a set of projectors onto the anti-trine ensemble

$$|\overline{\psi}_k\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - e^{-i2\pi k/3}|1\rangle\right), \tag{4.52}$$

which is the set of states which are orthogonal to the trine states, $\langle\psi_k|\overline{\psi}_k\rangle = 0$. It is impossible for Bob to measure the state orthogonal to that which *was* sent and he has equal probability of measuring each of the two remaining states. Alice now announces one of the states which she did *not* send. There are two possibilities: based upon his measurement, Bob either already knows this, in which case he announces as such and both parties discard the qubit from their key; otherwise, this is new information and both parties now know which state Alice sent without that piece of information being announced publicly. This shared information forms the basis for constructing a classical key. Logical bit values are assigned as such: if Alice announces that she didn't send the state one step clockwise of that which she did, the bit value is 0. If the former is one step anti-clockwise of the latter, the bit value is 1.

As a concrete example, consider that Alice sends a qubit in the state $|\psi_0\rangle$. In the absence of any change of this state due to the quantum channel, Bob will find either the outcome $|\overline{\psi}_1\rangle$ or the outcome $|\overline{\psi}_2\rangle$ with equal probability. If he finds the former then he knows with certainty that Alice sent either $|\psi_0\rangle$ or $|\psi_2\rangle$. If Alice subsequently announces that she did not send $|\psi_1\rangle$ then Bob can still not discover which state she did send. This

result cannot be used as part of the key. On the other hand, Alice may announce that she didn't send $|\psi_2\rangle$. In this case Bob knows that neither $|\psi_1\rangle$ nor $|\psi_2\rangle$ were sent and hence $|\psi_0\rangle$ must be the transmitted state. As the state which Alice announced was one step clockwise of that which she sent, the bit value assigned in this case is 0. An important point is that the classical bit value is decided upon at the point of the announcement, rather than when the state is sent. It is seen that this has interesting implications for the optimal eavesdropping strategy.

Although specific eavesdropping strategies are not explored in the literature for the three-state scheme, security proofs have been performed which show that protocol is secure against intercept-resend attacks up to an error rate of $Q = 9.81\%$ [80, 81]. The protocol has also been experimentally demonstrated [82].

4.5.2 Eavesdropping strategy

In order to find the optimal eavesdropping strategies for this scheme, I again start by constructing the superoperators which represent shared bit values:

$$\begin{aligned} S_0 &= \frac{1}{2}\left(|\psi_0\overline{\psi}_2^\dagger\rangle\rangle\langle\langle\psi_0\overline{\psi}_2^\dagger| + |\psi_1\overline{\psi}_0^\dagger\rangle\rangle\langle\langle\psi_1\overline{\psi}_0^\dagger| + |\psi_2\overline{\psi}_1^\dagger\rangle\rangle\langle\langle\psi_2\overline{\psi}_1^\dagger|\right) \\ &= \frac{3}{8}\left(I - e^{-i2\pi/3}|00^\dagger\rangle\rangle\langle\langle 11^\dagger| - e^{i2\pi/3}|11^\dagger\rangle\rangle\langle\langle 00^\dagger|\right). \end{aligned} \tag{4.53}$$

$$\begin{aligned} S_1 &= \frac{1}{2}\left(|\psi_0\overline{\psi}_1^\dagger\rangle\rangle\langle\langle\psi_0\overline{\psi}_1^\dagger| + |\psi_1\overline{\psi}_2^\dagger\rangle\rangle\langle\langle\psi_1\overline{\psi}_2^\dagger| + |\psi_2\overline{\psi}_0^\dagger\rangle\rangle\langle\langle\psi_2\overline{\psi}_0^\dagger|\right) \\ &= \frac{3}{8}\left(I - e^{i2\pi/3}|00^\dagger\rangle\rangle\langle\langle 11^\dagger| - e^{-i2\pi/3}|11^\dagger\rangle\rangle\langle\langle 00^\dagger|\right). \end{aligned} \tag{4.54}$$

A factor of 1/2 is required in both to bring into the calculation that Alice's announcement will cause fifty percent of outcomes to be sifted. As there is no chance that the announcement can cause sifting if Bob measures the state orthogonal to that which she did send, the disagreement superoperator is

$$\begin{aligned} S_X &= |\psi_0\overline{\psi}_0^\dagger\rangle\rangle\langle\langle\psi_0\overline{\psi}_0^\dagger| + |\psi_1\overline{\psi}_1^\dagger\rangle\rangle\langle\langle\psi_1\overline{\psi}_1^\dagger| + |\psi_2\overline{\psi}_2^\dagger\rangle\rangle\langle\langle\psi_2\overline{\psi}_2^\dagger| \\ &= \frac{3}{4}\left(I - |00^\dagger\rangle\rangle\langle\langle 11^\dagger| - |11^\dagger\rangle\rangle\langle\langle 11^\dagger|\right). \end{aligned} \tag{4.55}$$

The sifting factor could have been moved into the definition of the figures of merit however defining the superoperators in this manner ensures that $S_S = S_0 + S_1 + S_X$ and allows me to use the same general results, Eq. 4.10 and 4.13, as in other cases.

I turn first to the figure of merit $\mathrm{P}(A = E = B)$ and remind the reader that this was shown in Eq. 4.10 to be a ratio of the expectation values of S_0 and S_S both acting upon the vector $|E_0\rangle\rangle$. The important insight here is to note that both of those superoperators differ from the identity only in terms of outer products of $|00^\dagger\rangle\rangle$ and $|11^\dagger\rangle\rangle$. It must be true that the Kraus vector which maximises this ratio is a superposition of these two basis vectors, as adding any further terms would just decrease the constant of normalisation without increasing the overall probability of success. The bit symmetric pair of outcome

vectors is

$$|E_0\rangle\rangle = \frac{1}{\sqrt{2}}\left(|00^\dagger\rangle\rangle + e^{i\phi}|11^\dagger\rangle\rangle\right)$$
$$|E_1\rangle\rangle = \frac{1}{\sqrt{2}}\left(|00^\dagger\rangle\rangle + e^{-i\phi}|11^\dagger\rangle\rangle\right) \tag{4.56}$$

and the optimisation task is simply to find the value of ϕ such that $\mathrm{P}(A = E = B)$ is maximised. It is interesting to note that these correspond to two *unitary* transformations: the former is a rotation by ϕ anti-clockwise around the Bloch sphere and the latter is a rotation by the same angle in the opposite direction. (The corresponding Kraus operators are $E_0 = (|0\rangle\langle 0| + e^{i\phi}|1\rangle\langle 1|)/\sqrt{2}$ and $E_1 = (|0\rangle\langle 0| + e^{-i\phi}|1\rangle\langle 1|)/\sqrt{2}$, both of which can be seen to satisfy the usual condition $UU^\dagger = U^\dagger U = I$ up to a factor of $1/2$, which represents Eve's probability of choosing either.) That this operator is unitary implies the remarkable result that Eve gains *no information* about Alice's state from her intervention! This is an artifact of something I point out above: that the bit value in this strategy is not assigned based upon the choice of signal state, but only upon Alice's later announcement. Eve's best strategy is to change the state which Bob receives and in this manner choose which signal states are subsequently sifted.

The superoperator which represents sifting is found to be

$$S_S = S_0 + S_1 + S_X = \frac{3}{2}\left(I - \frac{1}{4}\left(|00^\dagger\rangle\rangle\langle\langle 11^\dagger| + |11^\dagger\rangle\rangle\langle\langle 00^\dagger|\right)\right) \tag{4.57}$$

which gives, by substitution into Eq. 4.10,

$$\mathrm{P}(A = E = B) = \frac{1 - \cos(\frac{2\pi}{3} - \phi)}{4 - \cos(\phi)}. \tag{4.58}$$

One can straightforwardly maximise this expression, i.e., solve for $d\mathrm{P}(A = E = B)/d\phi = 0$ in the standard manner. This process reveals that the optimal strategy gives $\mathrm{P}(A = E = B) = 3/5$ when the angle satisfies $\sin(\phi) = -5\sqrt{3}/14$, a measurement which has an associated error rate of $\mathrm{P}(A \neq B) = 2/15$. While the particular angle seems odd at first, it can be rationalised to a degree. There is a $\pi/6$ phase difference between the expressions for the probability being either sifted or post-selected. The angle $\phi = \sin^{-1}(-5\sqrt{3}/14)$ lies somewhere between the two.

The other figure of merit considered is the probability that all three parties agree, conditioned upon agreement between Alice and Bob. As was seen, this probability can always be made equal to one by satisfying the requirement $\langle\langle E_0|S_1|E_0\rangle\rangle = 0$, i.e., that there is no chance of Alice and Bob sharing the bit value one if Eve measures zero. As S_1, given by Eq. 4.54, differs from the identity only in the basis vectors $|00^\dagger\rangle\rangle$ and $|11^\dagger\rangle\rangle$ (as was true for the previous calculation), I again consider attacks of the form taken in Eq. 4.56. A quick calculation reveals that the requirement is

$$\langle\langle E_0|S_1|E_0\rangle\rangle = \frac{3}{8}\left(1 - \cos(\frac{2\pi}{3} + \phi)\right) = 0, \tag{4.59}$$

which is trivially solved by $\phi = -2\pi/3$, a more understandable angle of rotation than that calculated above as it corresponds to moving between different choices of trine state. Due to the conjugation introduced by moving between Liouville space and the Hilbert space operator representation, this corresponds to a clockwise rotation of $2\pi/3$ around the Bloch sphere.

This attack corresponds to Eve choosing the bit value which all three parties share, at the cost of of introducing a high error rate. This dynamic can be seen most clearly by considering a particular run of the protocol. I again consider those cases in which Alice sends $|\psi_0\rangle$. If Eve chooses the bit value zero then, without extracting any information from the signal qubit, she rotates the state by $2\pi/3$ and so sends on the state $|\psi_1\rangle$ to the receiver. There is now no chance that Bob's outcome is $|\overline{\psi}_1\rangle$, which would be required in order for him to believe that one is the shared bit value. At this point, the only way that the two legitimate users can share a bit value is if Bob's measurement outcome is $|\overline{\psi}_2\rangle$ and Alice then announces that she did not send $|\psi_1\rangle$; all other cases will either be sifted (Bob measures $|\overline{\psi}_2\rangle$ and Alice announces that she didn't send $|\psi_2\rangle$) or lead to disagreement between the two legitimate users. It follows from this discussion that there is an induced error rate of $\mathrm{P}(A \neq B) = 2/3$ due to Eve's attack. Evaluating $\mathrm{P}(A \neq B)$ confirms this.

4.6 Comments

As a tool for developing eavesdropping strategies, the two-time state formalism is seen to be effective. In all cases, the task of finding optimal strategies in terms of a given figure of merit is transformed into just a few lines of calculation. Furthermore, I have found that interesting insights are still to be gained by investigating specific attacks even in the era of, for example, device independent QKD. One aspect which is highlighted here is that the optimal attack is highly dependent on how bit values are assigned to the signal qubits. For B92, just a single signal state is associated with each classical bit value and Eve is then required to characterise precisely the sent state. This contrasts with PBC00 for which the classical bit value is not associated with any particular state and for which Eve's attack does not extract any quantum information. BB84 lies somewhere in the middle. This insight may help in designing future protocols.

Throughout this chapter, it has been assumed that quantum key distribution schemes are implemented in their 'prepare-and-measure' form, in which a signal qubit is transmitted by one party and then received by a second party. Within the literature on quantum security, an alternative approach is more common in which 'entanglement-based' routines are analysed. The resource in such schemes is that the two legitimate users share a maximally entangled Bell state (e.g., $|\Phi^+\rangle$). Each party then measures their part of the state in an equivalent manner to their preparation or measurement in the associated prepare-and-measure routine. In full security proofs the entanglement based scheme is found to be simpler to analyse although that is not the case for analysing specific attacks, as I have done here. A further reason is more philosophical. Some parties remain sceptical as to the true security advantage allowed by key distribution routines based upon quantum mechanics. Using the prepare-and-measure scheme allows for the principles upon which

security is based to be pinpointed, as in the previous chapter.

To some extent, choosing which framework (either measurement- or entanglement-based) to analyse in is a matter of taste, as the two are formally equivalent. In any QKD routine Alice and Bob's actions must determine pure states, otherwise it would be impossible for the receiving party to know with confidence that their resultant bit value is the same as the transmitter. I denote by $|\psi_A\rangle$ and $|\psi_B\rangle$ two particular outcomes associated with Alice and Bob's respective measurements in the entanglement based scheme. The probability distribution of these outcomes is $P(A,B) = |\langle\psi_A|\langle\psi_B|\Phi^+\rangle|^2$. It is seen that Alice and Bob's measurements are already represented by a bipartite state in this framework. Indeed, it is the two-time state $|\Psi_B\rangle \leftrightarrow |\psi_A\psi_B\rangle$ which I have used throughout (to be precise, it is the Jamiołkowski isomorphism of the two time state rather than the Choi isomorphism). Eve's attacks in the entanglement-based scheme will act as a channel on the bipartite state therefore, if her attack is associated with a Kraus operator $E_i^\dagger$, the state is written as $E_i^\dagger \otimes I|\Phi^+\rangle$. This form is the same as the objects $|E_i\rangle\rangle$ which I have used throughout. The overall probability rule which includes Eve's attacks is

$$\begin{aligned} P(A,B,E) &= |\langle\psi_A|\langle\psi_B|E_i \otimes I|\Phi^+\rangle|^2 \\ &= \langle\Phi^+|E_i \otimes I|\psi_A\psi_B\rangle\langle\psi_A\psi_B|E_i^\dagger \otimes I|\Phi^+\rangle. \end{aligned} \tag{4.60}$$

It can be seen that the expression for probability on the second line here is the same to Eq. 4.3, up to a choice of either Choi or Jamiołkowski isomorphism. Thus, the two frameworks are formally equivalent.

Another area of quantum cryptography is closely related to the two-time state formalism which I have used. This is the set of device independent (DI) and measurement-device independent (MDI) schemes [27, 28, 29]. The development of research into these areas is required by the necessity that a typical quantum-communications user will not be able to fully control all devices in their system. For example, components may be purchased from a third party who may or may not take advantage of the buyer's trust. To get around this possibility, DI and MDI protocols allow for full security even if an eavesdropper is in full control of the quantum channel. This is done by including Bell tests upon entangled states as part of the routine in such a way as to alert Alice and Bob if the equipment is behaving in a predictable (and hence, not fully quantum) manner. On a formal level the reason that these schemes are interesting here is that they typically require Alice and Bob to send qubits to a central untrusted server. For a similar reason as in the entanglement-based schemes, a probability rule emerges which is similar to that from two-time states. As security proofs in MDI analyses are typically simpler than those for measurement device *dependent* schemes, this suggests that it may be possible to map results from the former field onto the latter. This would be a job for future research.

4.7 Summary

In the previous chapter I derived several aspects of sequential measurement theory based upon some assumptions that we would want the probability rule to have, such as noncon-

texuality and completeness. This analysis suggested a framework of two-time states and Kraus vectors for the natural way to handle series of measurements. As a demonstration of this point, this chapter includes analyses of three different quantum key distribution protocols: the well-explored BB84 and B92 protocols and the lesser-examined PBC00. I gain some general insights into the possiblity of maximising my two figures of merit.

Some natural extensions of this work present themselves. One could modify the scheme so that it more closely represents practical implementations, for example, by allowing for loss in the quantum channel. I have also pointed out above that the method of analysis is similar in spirit to that used in MDI and DI quantum key distribution, suggesting that there may be fruitful applications in using the two-time state formalism to results in that field. However, for the remainder of this thesis I turn my attention to a different problem entirely: multiple-copy state discrimination.

Chapter 5

Multiple-copy state discrimination with noisy preparation

Throughout this thesis I have been writing, sometimes quite abstractly, on the topic of quantum measurement. In this chapter I consider a problem, multiple-copy state discrimination, which is more concrete than the axiomatic analysis of Chapter 3. To begin, I recap some basic concepts. If two states are orthogonal, then it is always possible to distinguish them with certainty. More generally, if the two states are non-orthogonal, then this cannot be done. The topic concerned with understanding this issue is called state discrimination [10, 15, 17, 83] and much research on this topic is concerned with finding the optimal measurement, i.e., minimising the probability of either incorrectly, or inconclusively, identifying the state. Unambiguous state discrimination [15, 19, 20, 21], definitively identifies the state in some measurements but in the rest is inconclusive. This chapter focuses on the other approach, minimum-error discrimination, in which every outcome gives an answer but at the cost of finding the incorrect outcome in some cases. One aims to minimise the error probability and the best-case value is given by the Helstrom bound [10] which, for two equiprobable pure states, is

$$\mathrm{P}_H = \frac{1}{2}\left(1 + \sqrt{1 - \cos(2\theta)}\right), \tag{5.1}$$

in which 2θ is the angle between the states to be discriminated. This was introduced in §2.4, where a more detailed discussion can be found, but is repeated here for convenience.

Here, I am concerned with multiple-copy state discrimination [84, 85, 86, 87], in which there are two-or-more systems in a given state. State-discrimination schemes for multiple copies are roughly grouped into two categories: local, in which each system is independently measured (allowing that the measurement which is performed can depend on the previous outcomes), and collective, in which a single measurement is performed on the overall product system. An example of each kind is discussed below. At the most general level, allowing that any number of initial states could be prepared, and that these states are either pure or mixed, some broad features emerge. In general, that is for discriminating two or more states which may or may not be pure, one can do better by measuring collectively rather than locally. This turns out not to hold for two-pure-state discrimina-

tion. As I show below, there exists a local scheme that is able to do as well as the best collective protocol [88, 89, 90]. Both schemes are able to reach the Helstrom bound, which for N copies becomes

$$\mathrm{P}_H^N = \frac{1}{2}\left(1 + \sqrt{1 - \cos^N(2\theta)}\right). \tag{5.2}$$

A further distinction can be made between global and local properties in terms of what is optimal, as well as measurement, and this leads to some counterintuitive results [86], particularly when applied to mixed-state discrimination. In the case of local measurements, the distinction between local and global optimality is the idea that there exists a scheme in which one identifies the state at each stage with suboptimal probability but overall, by taking the transmitted state to be the most commonly found outcome, does reach the related Helstrom bound. Strangely, this holds even if one performs the same measurement on each system. Such odd results highlight one useful property of multiple-copy state discrimination: it allows researchers to test their understanding of quantum measurement theory. It also has practical applications and one example is metrology: environmental details are imprinted onto a quantum state and measuring the probe corresponds to measuring that property, e.g., the direction of a magnetic field.

In this chapter, the question that I look to answer is: does having access to a quantum memory improve the resilience of the measurement scheme to noise? In particular, I take one example of local measurement (Acín *et al.*'s local adaptive scheme [88]) and one example of collective measurement (Blume-Kohout *et al.*'s quantum data gathering [91]) and calculate the probability that the scheme, as designed for distinguishing pure states, correctly identifies the transmitted state in the prescence of preparation noise. In any actual protocol, it will be impossible to perfectly prepare the resource systems. Instead of pure states, it is instead mixed states which must be discriminated. Two surprising results are found. In the many-copy limit, I show that both schemes tend to the same probability of success, which is less than one. I also find that the local adaptive measurement scheme consistently outperforms quantum-data-gathering, which goes against the accepted wisdom that 'nonlocality without entanglement' is always useful [85, 92, 93]. In the first section, I define some basic quantities that are used throughout. Following that, I present detailed algebraic derivations for the success probability of both schemes in the prescence of noise. The final section compares the performance of the two schemes and considers how they might be improved.

5.1 Basic model

In this chapter, I derive the probability of success for two multiple-copy state-discrimination schemes (local adaptive measurements and coherent measurements). In principle either of these schemes could be tailored to discriminate in a large range of situations. The most general case would be that any number of copies of any number of pure or mixed states can be prepared in a Hilbert space of any dimension. However, I restrict myself to the simplest non-trivial case: two non-orthogonal pure states, of which there are N copies, defined on a two-dimensional space. This allows me to find analytic solutions for the probability in

a way that is not possible in general.

Any pair of non-orthogonal states can be parameterised by

$$|\psi_k\rangle = \cos(\theta)|0\rangle + (-1)^k \sin(\theta)|1\rangle, \tag{5.3}$$

with $k = 0, 1$. The overlap of the two possible states defined in this way is $\cos(2\theta)$. In an ideal multiple-copy state discrimination scheme one has N copies of $|\psi_k\rangle$, i.e., the multipartite state $|\psi_k\rangle^{\otimes N}$, however I assume here that there is some noise in the preparation such that this does not hold. In practice, it is never possible to prepare a state perfectly in a chosen pure state. Even if tomography is performed for a large number of copies then θ will only be restricted to some probability distribution of non-zero width. To model this, I associate with each system S_i a level of noise, represented by a change $\delta\theta_i$ in the angle which parameterises the state. The actual state of the system will then be

$$|\tilde{\psi}_k^{(i)}\rangle = \cos(\theta + \delta\theta_i)|0\rangle + (-1)^k \sin(\theta + \delta\theta_i)|1\rangle. \tag{5.4}$$

The values of the noise parameter $\delta\theta_i$ on each system are uncorrelated in this model and are related to the preparation fidelity F of the experimental apparatus. I assume that the probability distribution of the noise is symmetric, i.e., that $\mathrm{P}(\delta\theta_i) = \mathrm{P}(-\delta\theta_i)$. One could be more specific about the type of noise (i.e., usually one would assume that it is Gaussian) however this level of detail is not needed to relate the fidelity to the probability of success. By definition, the fidelity is given by the average overlap of the pure state Eq. 5.3 and that in Eq. 5.4. The latter object will be $|\langle\tilde{\psi}_k^{(i)}|\psi_k^{(i)}\rangle|^2 = \cos^2(\delta\theta)$ and therefore

$$\langle\cos^2(\delta\theta_i)\rangle = F. \tag{5.5}$$

Here, $\langle\cdot\rangle$ has its usual meaning as the average, here taken over the noise distribution. From this it is easy to see that

$$\langle\sin^2(\delta\theta_i)\rangle = 1 - F \tag{5.6}$$

and hence

$$\langle\cos(2\delta\theta_i)\rangle = 2F - 1. \tag{5.7}$$

The above three results all follow simply from the definition of fidelity. I can furthermore use the assumption that the noise is symmetric about $\delta\theta = 0$ for

$$\langle\sin(2\delta\theta_i)\rangle = 0. \tag{5.8}$$

It turns out that these four basic results are all that is needed to calculate the probabilty of success for each of the schemes which are considered here. With these results, it is possible to characterise, with a single parameter F, the noise on each of the individual N systems. Because the noise on each qubit is independent, one may average over the noise on each state individually.

5.2 Local adaptive measurement

The main aim of this work is to ask whether collective, rather than individual, measurements are more successful at discriminating imperfectly prepared states. As an example of an individual measurement scheme, I consider the local adaptive scheme.

5.2.1 Scheme

I follow here the scheme of Acín *et al.* [88] although similar results have been acquired by others [89, 90]. The main result of their work is a demonstration that it is possible to reach the Helstrom bound for the two bipartite states $|\psi_k\rangle^{\otimes N}$ by performing just individual measurements on each of the copies and allowing for each measurement to depend upon the measurement record. The scheme which does this and which I develop in more detail shortly turns out to be Bayesian updating, by which is meant that after each outcome one simply updates their prior probabilities associated with the preparation of each of the two states and then performs the associated Helstrom measurement. What is particularly surprising is that the measurement which should be performed at the nth step is dependent only upon the single result at the $(n-1)$th step, i.e., if the outcome associated with $|\psi_0\rangle$ was found at the prior stage then one measurement is performed; if instead $|\psi_1\rangle$ was found then a different measurement is optimal. In this sense, the scheme is Markovian.

Acín *et al.*'s scheme acts in the following manner. A projective measurement is performed upon each of the N copies and the user's final guess is that the prepared state is that associated with the final measurement *only*. The measurement which is performed at each step is represented by a projector onto the state

$$|\omega(i_N x_{N-1})\rangle = \cos(\phi_x - i_N \frac{\pi}{2})|0\rangle + \sin(\phi_x - i_N \frac{\pi}{2})|1\rangle. \tag{5.9}$$

In this formula $i_N = 0, 1$ represents the outcome of the measurement of the Nth qubit; x_{N-1} is the measurement record which is thought of as a bit string of all previous outcomes. For maximum notational clarity, a short example is one in which zero followed by one were the outcomes of two consecutive measurements. This would be represented in the notation introduced here by $i_1 = 0, i_2 = 1, x_1 = 0$ and $x_2 = 01$, i.e., x are bit strings and i are individual outcomes. The angle ϕ_x depends on the measurement record (I omit the subscript $N-1$ on x here for neatness; different notation is introduced below for other bit strings) and is to be found such that the overall probability of success is optimal. I consider just the case in which $|\psi_0\rangle$ and $|\psi_1\rangle$ occur with equal probability and then the success probability is

$$\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\sum_x \left(|\langle\omega(0x_{N-1})|\psi_0\rangle|^2 \mathrm{P}(x_{N-1}|0) + |\langle\omega(1x_{N-1})|\psi_1\rangle|^2 \mathrm{P}(x_{N-1}|1)\right), \tag{5.10}$$

in which $\mathrm{P}(x_{N-1}|0)$ is the probability that the bit string x_{N-1} occured given that the state $|\psi_0\rangle$ was prepared. The sum is over all possible bit strings of length $(N-1)$ as the probability of an overall outcome for the scheme is dependent upon the final outcome only.

The authors then show that this figure is maximised if the angle ϕ_x satisfies

$$\cos(2\phi_x) = (-1)^{i_{N-1}} \cos(2\theta) \sqrt{\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^{2N}(2\theta)}}, \tag{5.11}$$

and for reference I derive this result in Appendix B of this thesis. On the right hand side the only appearance of the measurement record is in the value i_{N-1}, which is the outcome of the measurement directly previous to the one being considered. This is the sense in which the scheme is Markovian. The authors are able to show that this scheme reaches the Helstrom bound for the measurement, i.e., if there is no noise in the preparation

$$\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\sqrt{1-\cos^{2N}(2\theta)}. \tag{5.12}$$

The method which they use to show this does not generalise to the noisy case and so, below, I present an alternate calculation which, though it is somewhat more involved, can be extended to cases with imperfect preparation fidelity.

5.2.2 Success probability

The central value which is calculated here is the probability $\mathrm{P}_N^{\mathrm{ad}}$ of identifying the correct state. This calculation is fairly technical and so I summarise the structure here. Firstly, I perform the calculation in the perfect-fidelity case: I show that Acín *et al.*'s local adaptive scheme satisfies the Helstrom bound. This is done by deriving an inductive relationship, one which relates $\mathrm{P}_N^{\mathrm{ad}}$ to $\mathrm{P}_{N-1}^{\mathrm{ad}}$, which is then solved. Secondly, I introduce noise into the model which results in a different form for the inductive expression. This is again solved analytically. Finally, I discuss the behaviour of the protocol in the case of a large number of copies. In this regime it is possible to derive the success probability differently and this is found to agree with the many-copy limit of the general formula.

I begin with the clean case. Eq. 5.10 is to be evaluated as a function of the overlap $\cos(2\theta)$ of the two possible states. This calculation requires the probability that one finds either 0 or 1 on the final measurement:

$$\begin{aligned}
\mathrm{P}(0|x,0) &= |\langle\omega(0x_{N-1})|\psi_0\rangle|^2 = \frac{1}{2}\left(1+\cos(2\theta)\cos(2\phi_x)+\sin(2\theta)\sin(2\phi_x)\right) \\
\mathrm{P}(1|x,1) &= |\langle\omega(1x_{N-1})|\psi_1\rangle|^2 = \frac{1}{2}\left(1-\cos(2\theta)\cos(2\phi_x)+\sin(2\theta)\sin(2\phi_x)\right).
\end{aligned} \tag{5.13}$$

Substituting these formulae into Eq. 5.10 gives

$$\begin{aligned}
\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\Bigg[1 + \frac{1}{2}\sum_x \Big(&\sin(2\theta)\sin(2\phi_x)\left(\mathrm{P}(x|0)+\mathrm{P}(x|1)\right) \\
&+\cos(2\theta)\cos(2\phi_x)\left(\mathrm{P}(x|0)-\mathrm{P}(x|1)\right)\Big)\Bigg].
\end{aligned} \tag{5.14}$$

At this point the two trigonometric functions of ϕ_x, both of which are evaluated from Eq.

5.11, are used in the function, which becomes

$$\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\Bigg[\, 1 + \frac{1}{2}\frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}}\sum_x (\mathrm{P}(x|0)+\mathrm{P}(x|1)) \\ + \frac{1}{2}\cos^2(2\theta)\sqrt{\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^{2N}(2\theta)}}\sum_x (-1)^{i_{N-1}}(\mathrm{P}(x|0)-\mathrm{P}(x|1))\Bigg]. \tag{5.15}$$

Given that a bit value is prepared, some entry must appear in the measurement record. This means that the first sums that occur in this expression are over complete sets of outcomes, and $\sum_x (P(x|0)+P(x|1)) = 2$. The second sum can be evaluated by using the rules of conditional probability. In the above expression, x is the list of possible outcomes of the first $(N-1)$ measurements. I adopt $\dot{x}$ to denote the series of the first $(N-2)$ results, so that

$$\mathrm{P}(x|a) = \mathrm{P}(i_{N-1}\dot{x}|a) = \mathrm{P}(i_{N-1}|\dot{x},a)\mathrm{P}(\dot{x}|a). \tag{5.16}$$

The relevant sum is hence

$$\sum_x (-1)^{i_{N-1}}(\mathrm{P}(x|0)-\mathrm{P}(x|1)) \\ = \sum_{\dot{x}}\sum_{i_{N-1}} (-1)^{i_{N-1}}(\mathrm{P}(i_{N-1}|\dot{x},0)\mathrm{P}(\dot{x}|0) - \mathrm{P}(i_{N-1}|\dot{x},1)\mathrm{P}(\dot{x}|1)), \tag{5.17}$$

where I have also separated out the sum into contributions to x from $\dot{x}$ and from the penultimate outcome. From Eq. 5.13 each of the sums involved in the right hand side can be evaluated. A general expression is found:

$$\sum_{i_{N-1}} (-1)^{i_{N-1}}\mathrm{P}(i_{N-1}|\dot{x},a) \\ = \cos(2\theta)\cos(2\phi_{\dot{x}}) + (-1)^a \sin(2\theta)\sin(2\phi_{\dot{x}}) \tag{5.18}$$

in which $a = 0$ or 1. The above two results combine for

$$\sum_x (-1)^{i_{N-1}}(\mathrm{P}(x|0)-\mathrm{P}(x|1)) = \sum_{\dot{x}} (\sin(2\theta)\sin(2\phi_{\dot{x}})(\mathrm{P}(\dot{x}|0)+\mathrm{P}(\dot{x}|1)) \\ + \cos(2\theta)\cos(2\phi_{\dot{x}})(\mathrm{P}(\dot{x}|0)-\mathrm{P}(\dot{x}|1))\,. \tag{5.19}$$

This expression is then compared with Eq. 5.14, the initial probability formula. It is seen that the right-hand side of the above expression is precisely the same as the series in Eq. 5.14 except that it is over the series $\dot{x}$ (the first $N-2$ results) instead of x (the first $N-1$ results). Hence,

$$\sum_x (-1)^{i_{N-1}}(\mathrm{P}(x|0)-\mathrm{P}(x|1)) = 2\left(2\mathrm{P}_{N-1}^{\mathrm{ad}} - 1\right) \tag{5.20}$$

and an inductive expression for the probability of success has been found:

$$\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\left[1 + \frac{\sin^2(2\theta) + \cos^2(2\theta)\sqrt{1-\cos^{2N-2}(2\theta)}(2\mathrm{P}_{N-1}^{\mathrm{ad}}-1)}{\sqrt{1-\cos^{2N}(2\theta)}}\right]. \tag{5.21}$$

The solution to this equation is the Helstrom bound, Eq. 5.2, as claimed above. To verify this, substitute $\mathrm{P}^{\mathrm{ad}}_{N-1} = (1 + \sqrt{1 - \cos^{2N-2}(2\theta)})/2$ into the right hand side:

$$\begin{aligned}\mathrm{P}^{\mathrm{ad}}_{N} &= \frac{1}{2}\left[1 + \frac{\sin^2(2\theta) + \cos^2(2\theta)(1 - \cos^{2N-2}(2\theta))}{\sqrt{1 - \cos^{2N}(2\theta)}}\right] \\ &= \frac{1}{2}\left[1 + \sqrt{1 - \cos^{2N}(2\theta)}\right]. \end{aligned} \tag{5.22}$$

The final piece is to verify that this satisfies also $N = 1$, so that the chain of induction holds. In this scheme, one is required to perform the single-copy Helstrom measurement on the first qubit, and so the proof that $\mathrm{P}^{\mathrm{ad}}_{1}$ has the above form is the same as the original derivation of the single-copy pure-state Helstrom bound, which I provided in §2.4. Interested readers should consult that discussion for more details. That the inductive measurement has this structure is due to the Markovianity of the scheme, i.e., it depends upon just the previous outcome rather than the entire measurement record. I have followed Acín *et al.* by demonstrating that this measurement is as successful as any possible measurement for distinguishing pure states. In the case that the preparation is imperfect, the discrimination is instead between mixed states. To emphasise, what I want to know is how well the pure-state scheme does in this context. I apply the Markovian scheme used here, rather than true Bayesian updating, and seek an analogue of Eq. 5.22. The solution to this problem is what I now present.

As stated above, the first step that this calculation requires is to derive an expression which is equivalent to Eq. 5.21 but is valid in the case that there is preparation noise. The objects which change if the system is noisy are the expressions for the probability of measuring a set of outcomes x given that the state $|\psi_a\rangle$ was sent. The generalisation of this result is what leads to a different success probability and is found by replacing the state overlap in Eq. 5.13 such that it is instead between the measured state and the noisy state, Eq. 5.4. The general result which I find is

$$\begin{aligned}&\mathrm{P}(i_{N-1}|x,a) = |\langle\omega(i_{N-1}x)|\tilde{\psi}^{(i)}_{a}\rangle|^2 \\ &= \frac{1}{2}\left[1 + (-1)^{i_{N-1}}\left(\cos(2\theta + 2\delta\theta_N)\cos(2\phi_x) + (-1)^a \sin(2\theta + 2\delta\theta_N)\sin(2\phi_x)\right)\right] \\ &= \frac{1}{2}\left[1 + \cos(2\delta\theta)(-1)^{i_{N-1}}\left(\cos(2\theta)\cos(\phi_x) + (-1)^a\sin(2\theta)\sin(2\phi_x)\right)\right. \\ &\left. + \sin(2\delta\theta)(-1)^{i_{N-1}}\left((-1)^a\cos(2\theta)\sin(2\phi_x) - \sin(2\theta)\cos(2\phi_x)\right)\right]. \end{aligned} \tag{5.23}$$

Here, $a = 0, 1$ signifies the state which was transmitted. At this point I can apply the results from averaging over the noise's probability distribution, Eqs. 5.7 and 5.6. The expectation value of this probability is found to be

$$\begin{aligned}&\mathrm{P}(i_{N-1}|x,a) = \\ &\frac{1}{2}\left[1 + (2F-1)(-1)^{i_{N-1}}\left(\cos(2\theta)\cos(2\phi_x) + (-1)^a\sin(2\theta)\sin(2\phi_x)\right)\right]. \end{aligned} \tag{5.24}$$

I start from the same position as before, Eq. 5.10, and substitute instead this result for

the probabilities of the final outcome. The result of this process is almost precisely the same although, as might be expected from the form of the conditional probability just above this sentence, an extra factor of $(2F-1)$ appears before the sum in Eq. 5.14. I have

$$\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\left[1 + \frac{1}{2}(2F-1)\sum_x \left(\sin(2\theta)\sin(2\phi_x)\left(P(x|0)+P(x|1)\right)\right.\right. \\ \left.\left. + \cos(2\theta)\sin(2\phi_x)\left(P(x|0)-P(x|1)\right)\right)\right]. \tag{5.25}$$

The substitution of the measurement angle ϕ_x and the algebraic manipulation of the resultant expression proceed exactly as in the noiseless case, and the resultant expression for the probability of success with a resource of N qubits, in terms of the equivalent probability of success with $(N-1)$ qubits, is given by

$$\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\left[1 + (2F-1)\frac{\sin^2(2\theta) + \cos^2(2\theta)\sqrt{1-\cos^{2N-2}(2\theta)}(2\mathrm{P}_{N-1}^{\mathrm{ad}}-1)}{\sqrt{1-\cos^{2N}(2\theta)}}\right]. \tag{5.26}$$

Again, this is an inductive formula which connects $\mathrm{P}_N^{\mathrm{ad}}$ with $\mathrm{P}_{N-1}^{\mathrm{ad}}$. Despite the simplicity of the generalisation, the analytic solution to this result is much more complicated than in the noiseless case. After some playing around, I find the solution in terms of a series $\mathcal{S}_N$:

$$\mathrm{P}_N^{\mathrm{ad}} = \frac{1}{2}\left[1 + (2F-1)^N\sqrt{1-\cos^{2N}(2\theta)} + \frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}}\mathcal{S}_N\right]. \tag{5.27}$$

I have introduced the notation

$$\mathcal{S}_N = \sum_{i=1}^{N}(2F-1)^{N+1-i}(1-(2F-1)^{i-1})\cos^{2N-2i}(2\theta), \tag{5.28}$$

which helps to condense some of the calculations which follow. I now verify that this solves the above equation and then evaluate the geometric summations which appear in the series term. This will give the overall probability of success. As before, that this is the solution can be most easily seen by first writing out the $(N-1)$ term:

$$\mathrm{P}_{N-1}^{\mathrm{ad}} = \frac{1}{2}\left[1 + (2F-1)^{N-1}\sqrt{1-\cos^{2N-2}(2\theta)} + \frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N-2}(2\theta)}}\mathcal{S}_{N-1}\right]. \tag{5.29}$$

This must then be substituted into the right hand side of Eq. 5.26. If this gives the general

formula for the solution, Eq. 5.27, that solution is valid. I find

$$\begin{aligned}
\mathrm{P}_N^{\mathrm{ad}} &= \frac{1}{2}\Bigg[1+(2F-1)\frac{\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}} \\
&\quad +(2F-1)^N\cos^2(2\theta)\frac{1-\cos^{2N-2}(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}}+(2F-1)\frac{\cos^2(2\theta)\sin^2(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}}\mathcal{S}_{N-1}\Bigg] \\
&= \frac{1}{2}\Bigg[1+(2F-1)^N\frac{1-\cos^{2N}(2\theta)}{\sqrt{1-\cos^{2N}(2\theta)}} \\
&\quad +\frac{\sin^2(2\theta)(2F-1)}{\sqrt{1-\cos^{2N}(2\theta)}}\left(1-(2F-1)^{N-1}+\cos^2(2\theta)\mathcal{S}_{N-1}\right)\Bigg]. \qquad (5.30)
\end{aligned}$$

It is seen, by examining the definition of the series, that $\mathcal{S}_N = (2F-1)(1-(2F-1)^{N-1}+(2F-1)\cos^2(2\theta)\mathcal{S}_{N-1})$ and so the above equation demonstrates the validity of the solution which I have provided. The easiest way to see that the second line of this equation follows from the first is by multiplying the second term by the factor $1+(2F-1)^{N-1}-(2F-1)^{N-1}$ and grouping relevant terms. All that remains is to evaluate the series, Eq. 5.28, which consists of two geometric summations and which are evaluated in the usual manner. After some simplfication, I am left with

$$\mathcal{S}_N = (2F-1)\frac{1-(2F-1)^N\cos^{2N}(2\theta)}{1-(2F-1)\cos^2(2\theta)}-(2F-1)^N\frac{1-\cos^{2N}(2\theta)}{1-\cos^2(2\theta)}. \qquad (5.31)$$

Between this and Eq. 5.27, I have an expression for the probability of success for the local adaptive measurement scheme which is applied to noisy qubits. This is to be compared with the probability of success for an equivalent scheme which uses a quantum memory. After deriving the equivalent expression to Eq. 5.30, I compare and contrast the behaviour of these two functions. Before doing so, I hope to persuade the reader that Eq. 5.30 is indeed the correct result.

5.2.3 Many-copy limit

It is natural, for a number of reasons, to investigate the many-copy limit. Understanding the behaviour of Acín *et al.*'s scheme in this regime gives a better feel for how it works in general. Also, as the protocol behaves differently, it is possible to use an alternative method to calculate the probability of success for a large number of copies. It can thus be used as a check of the final result, Eq. 5.30, found in the previous section. The $N\to\infty$ limit of that equation is

$$\begin{aligned}
\lim_{N\to\infty}\mathrm{P}_N^{\mathrm{ad}} &= \frac{1}{2}\left[1+\frac{(2F-1)\sin^2(2\theta)}{1-(2F-1)\cos^2(2\theta)}\right] \\
&= 1-\frac{1-F}{1-(2F-1)\cos^2(2\theta)}. \qquad (5.32)
\end{aligned}$$

With this expression alone some basic checks can be done. If the fidelity is perfect, $F=1$, the above expression is unity: a user of the local adaptive scheme would know for sure

which state had been prepared if they could measure an infinite amount of copies. This seems intuitive and is backed up by the Helstrom bound. Another parameter of interest is the angle between the two states. If the two states are precisely the same, $|\psi_0\rangle = |\psi_1\rangle$, then it must be impossible to distinguish them and so there is a 50 % probability of success. It cannot be the case that identical mixed states can be distinguished and so the probability must again be one-half. (However, it might be noted that these two limits do not commute. I discuss point in more detail in §5.4.) Again, this is confirmed by looking at the above equation: if $\theta = 0$ then the probability of success is given by one-half. It can also be seen that, as $1/2 \geq F \geq 1$, the fraction on the right-hand side will also lie in the same range and hence the probability is never negative or greater than one, two results which would suggest that this expression is invalid.

A final check can be performed by returning to the local adaptive scheme, in particular the measurement angles given by Eq. 5.11. In the limit of an infinite number of copies that equation becomes

$$\lim_{N\to\infty} \cos(2\phi_x) = (-1)^{i_{N-1}} \cos(2\theta). \tag{5.33}$$

This equation tells us that, as N gets large, the scheme tends towards hypothesis checking: if one measurement identifies that the state $|\psi_0\rangle$ was sent, the next measurement consists of a von Neumann measurement of that state paired with its orthogonal partner. In the perfect-fidelity case, if one repeated this measurement an infinite number of times then every subsequent outcome would be $|\psi_0\rangle$ however, in the presence of noise, sometimes the outcome will be $|\psi_1\rangle$ and then the measurement switches to checking that $|\psi_1\rangle$ was prepared. Thus, if there is any noise in the system then a user can never have complete confidence in their state identification.

The probability of success for the scheme of repeated hypothesis checking can be evaluated straightforwardly. There are two probabilities of interest. One is the probability $\mathrm{P}(a|i_{N-1} = a, a)$ that the Nth measurement is correct given that the $(N-1)$th was also correct. The other is the probability $\mathrm{P}(a|i_{N-1} = \overline{a}, a)$, the probability that the Nth measurement was correct even though the measurement directly previous to that was incorrect. By definition, the former is simply the fidelity:

$$\mathrm{P}(a|i_{N-1} = a, a) = F. \tag{5.34}$$

The other quantity is given by

$$\begin{aligned} \mathrm{P}(a|i_{N-1} = \overline{a}, a) &= |\langle\overline{\psi}_{\overline{a}}|\tilde{\psi}_a\rangle|^2 \\ &= (\cos(2\delta\theta)\cos(2\theta) - \sin(2\delta\theta)\sin(2\theta))^2 \\ &= F - \cos^2(2\theta)(2F-1). \end{aligned} \tag{5.35}$$

The probability $\mathrm{P}^{\mathrm{ad}}_{N+1}$ of success when measuring qubit S_{N+1} can then be expressed in

terms of the probability of success on the previous qubit

$$\begin{aligned} \mathrm{P}^{\mathrm{ad}}_{N+1} &= \mathrm{P}(a|i_{N-1}=a,a)\mathrm{P}^{\mathrm{ad}}_{N} + \mathrm{P}(a|i_{N-1}=\overline{a},a)(1-\mathrm{P}^{\mathrm{ad}}_{N}) \\ &= F\mathrm{P}^{\mathrm{ad}}_{N} + (F-\cos^2(2\theta)(2F-))(1-\mathrm{P}^{\mathrm{ad}}_{N}) \\ &= F-\cos^2(2\theta)(2F-1)+\cos^2(2\theta)(2F-1)\mathrm{P}^{\mathrm{ad}}_{N}. \end{aligned} \tag{5.36}$$

Repeated application of this formula gives a general expression for the probability of success after N' more measurements are made.

$$\mathrm{P}^{\mathrm{ad}}_{N+N'} = (\cos^2(2\theta)(2F-1))^{N'}P^{\mathrm{ad}}_{N}+(F-\cos^2(2\theta)(2F-1))\sum_{i=0}^{N'-1}(\cos^2(2\theta)(2F-1))^{i}. \tag{5.37}$$

I am interested in the many-copy limit. If I take the limit $N'\to\infty$ in the above, the first term vanishes, as $0\leq(2F-1)\cos^2(2\theta)\leq 1$, and I am left with

$$\begin{aligned} \lim_{N'\to\infty}\mathrm{P}^{\mathrm{ad}}_{N'} &= (F-\cos^2(2\theta)(2F-1))\sum_{i=0}^{\infty}\left(\cos^2(2\theta)(2F-1)\right)^{i} \\ &= \frac{F-\cos^2(2F-1)}{1-\cos^2(2F-1)} \\ &= 1-\frac{1-F}{1-\cos^2(2\theta)(2F-1)}. \end{aligned} \tag{5.38}$$

This is the same result as derived above for the general form of the local adaptive measurement scheme. Here it has been arrived at in terms of the scheme's limiting form, in which one checks their guess by performing the relevant projective measurement. This backs up the general result.

5.3 Quantum data gathering

There is some debate in the literature as to whether state discrimination can usefully take advantage of a quantum memory, i.e., a resource qubit which is not allowed to decohere throughout the experiment. One scheme which uses such a device is quantum data gathering. In this chapter I introduce the scheme and again find the probability of success in the perfect and imperfect preparation scenarios.

5.3.1 Scheme

Quantum data gathering, which was introduced by Blume-Kohout *et al.* [91], is an alternative multiple-copy state discrimination scheme. Given N qubits which are all prepared in one of two states, the protocol will predict the state with the maximum possible probability, that given by the Helstrom bound. It is a form of collective measurement and uses a probe qubit on top of the N resource qubits. This probe qubit is initialised in a fiducial state and then interacts unitarily with each of the N qubits, after which it is itself measured using the Helstrom measurement for the two possible final states. This scheme can be generalised to distinguish between a greater number of possible states.

I first describe the behaviour of the scheme if the measured qubits have all been prepared perfectly. A probe qubit is initialised the state $|0\rangle_A$, where the subscript A identifies the probe space. For only the first qubit, the interaction is a SWAP. This, of course, leaves the sample qubit S_1 in the state $|0\rangle_{S_1}$ and the probe in the state $|\psi_k\rangle_A$. An interaction then occurs between the probe and the second qubit, labelled S_2. This interaction is such as to leave the probe in the state

$$|\psi_k^{(2)}\rangle_A = \sqrt{\frac{1}{2}\left(1+\cos^2(2\theta)\right)}|0\rangle_A + (-1)^k\sqrt{\frac{1}{2}\left(1-\cos^2(2\theta)\right)}|1\rangle_A, \tag{5.39}$$

while leaving the sample qubit in the state $|0\rangle_{S_2}$. The probability of successfully distinguishing $|\psi_0^{(2)}\rangle$ from $|\psi_1^{(2)}\rangle$ is now identical to the probability of successfully distinguishing $|\psi_0\rangle^{\otimes 2}$ from $|\psi_1\rangle^{\otimes 2}$, as the overlap of both pairs of states is $\cos^2(2\theta)$. The subsequent interactions all follow this pattern. After interacting with N qubits, the state of the probe will be

$$|\psi_k^{(N)}\rangle_A = \cos(\theta_N)|0\rangle_A + (-1)^k\sin(\theta_N)|1\rangle_A, \tag{5.40}$$

where I have introduced the notation

$$\cos(\theta_N) = \sqrt{\frac{1}{2}\left(1+\cos^N(2\theta)\right)}. \tag{5.41}$$

All interactions are such as to leave the sample qubit in the state $|0\rangle_{S_N}$. Thus, the action of each unitary can be written as $U_n|\psi_k\rangle_{S_n}|\psi_k^{(N-1)}\rangle_A = |0\rangle_{S_n}|\psi_k^{(N)}\rangle_A$. A full description of the behaviour must also include the states which are orthogonal to this basis and, though there is some freedom, it seems natural to choose $U_n|\psi_{k\perp}\rangle_{S_n}|\psi_{k\perp}^{(N-1)}\rangle_A = |1\rangle_{S_n}|\psi_{k\perp}^{(N)}\rangle_A$. This extension can be taken advantage of as a diagnostic for the success of the protocol, as I discuss below. In terms of the computational basis of each object, the unitary which performs in such a manner is

$$\begin{aligned}
U_n|0_{S_n}0_A\rangle &= \frac{\cos(\theta)\cos(\theta_{n-1})}{\cos(\theta_n)}|0_{S_n}0_A\rangle + \frac{\sin(\theta)\sin(\theta_{n-1})}{\cos(\theta_n)}|1_{S_n}0_A\rangle \\
U_n|1_{S_n}1_A\rangle &= \frac{\sin(\theta)\sin(\theta_{n-1})}{\cos(\theta_n)}|0_{S_n}0_A\rangle - \frac{\cos(\theta)\cos(\theta_{n-1})}{\cos(\theta_n)}|1_{S_n}0_A\rangle \\
U_n|1_{S_n}0_A\rangle &= \frac{\sin(\theta)\cos(\theta_{n-1})}{\sin(\theta_n)}|0_{S_n}1_A\rangle + \frac{\cos(\theta)\sin(\theta_{n-1})}{\sin(\theta_n)}|1_{S_n}1_A\rangle \\
U_n|0_{S_n}1_A\rangle &= \frac{\cos(\theta)\sin(\theta_{n-1})}{\sin(\theta_n)}|0_{S_n}1_A\rangle - \frac{\sin(\theta)\cos(\theta_{n-1})}{\sin(\theta_n)}|1_{S_n}1_A\rangle.
\end{aligned} \tag{5.42}$$

The total space of the product state $|\psi_k\rangle^{\otimes N}$ has 2^N dimensions, however not all of the space is required to construct an optimal measurement. Product states lie in a subspace of the overall Hilbert space of the sample qubits, and this subspace has been mapped onto the two dimensions of the probe's Hilbert space.

After information is extracted from all N sample qubits, the probe is measured using the Helstrom measurement, i.e., projectively onto the eigenvalues of $|\psi_0^{(N)}\rangle\langle\psi_0^{(N)}| - |\psi_1^{(N)}\rangle\langle\psi_1^{(N)}|$. In the perfect fidelity case this measurement succeeds with the optimal probability. Otherwise, it has a smaller chance of success. I calculate the relevant Helstrom

measurement at the point that it is required.

5.3.2 Gate implementation

It is useful to know how to perform the quantum-data-gathering routine physically for a number of reasons. From the perspective of understanding the scheme's resilience to noise, knowing how many gates are required can place an upper-bound on the level of gate noise. Though I am in this thesis primarily focused on the preparation noise on the measured qubits, the gate noise will also limit the possibility of successfully performing any coherent measurement. In particular, and as discussed in the §2.2, it is two-qubit gates which cause the most trouble when processing quantum information. In any physical system (e.g., ion traps, NV centres) there is typically a particular two-qubit gate which can be performed most reliably and which allow a full set of quantum operations to be performed. The most commonly used is the CNOT gate and I express the required unitary in terms of these and single-qubit rotations.

To summarise, the gate sequence which satisfies the above unitary is a CNOT with the sample as the control and the probe as the target, followed by a controlled rotation which has the sample as its target. As a controlled rotation can be implemented by single qubit rotations and one CNOT, two CNOTs are required at each stage in the protocol.

I now show this in more detail. This can most easily be seen by introducing the defintions

$$\begin{aligned}\cos(\phi_N) &= \frac{\cos(\theta)\cos(\theta_{N-1})}{\cos(\theta_N)} \quad \sin(\phi_N) = \frac{\sin(\theta)\sin(\theta_{N-1})}{\cos(\theta_N)} \\ \cos(\xi_N) &= \frac{\cos(\theta)\sin(\theta_{N-1})}{\sin(\theta_N)} \quad \sin(\xi_N) = -\frac{\sin(\theta)\cos(\theta_{N-1})}{\sin(\theta_N)}.\end{aligned} \tag{5.43}$$

In terms of this parameterisation, the unitary can be succinctly written as

$$\begin{aligned}U_N|0_{S_N}0_A\rangle &= (\cos(\phi_N)|0\rangle_{S_N} + \sin(\phi_N)|1\rangle_{S_N})\,|0\rangle_A \\ U_N|1_{S_N}1_A\rangle &= (\sin(\phi_N)|0\rangle_{S_N} - \cos(\phi_N)|1\rangle_{S_N})\,|0\rangle_A \\ U_N|1_{S_N}0_A\rangle &= (-\sin(\xi_N)|0\rangle_{S_N} + \cos(\xi_N)|1\rangle_{S_N})\,|1\rangle_A \\ U_N|0_{S_N}1_A\rangle &= (\cos(\xi_N)|0\rangle_{S_N} + \sin(\xi_N)|1\rangle_{S_N})\,|1\rangle_A.\end{aligned} \tag{5.44}$$

The first step is to note that what is required is a rotation which is controlled upon addition, in the computational basis, of the bit values of the probe and sample respectively; if S_N and A agree then one rotation is performed; if they disagree then another is performed. This function can be implemented by first using a CNOT gate, with the probe as its target, which will act to register the parity of the two qubits on the targeted system.

This gate updates the basis vectors to

$$\begin{aligned}
|0_{S_N}0_A\rangle &\rightarrow |0_{S_N}0_A\rangle \\
|1_{S_N}1_A\rangle &\rightarrow |1_{S_N}0_A\rangle \\
|1_{S_N}0_A\rangle &\rightarrow |1_{S_N}1_A\rangle \\
|0_{S_N}1_A\rangle &\rightarrow |0_{S_N}1_A\rangle.
\end{aligned} \tag{5.45}$$

One sees that the following action completes the unitary. If the probe is now in the state $|0\rangle$, then the unitary operation

$$W_0(\phi_N) = \begin{bmatrix} \cos(\phi_N) & \sin(\phi_N) \\ \sin(\phi_N) & -\cos(\phi_N) \end{bmatrix} \tag{5.46}$$

should act on the sample state. If instead the probe is in the state $|1\rangle$, then

$$W_1(\xi_N) = \begin{bmatrix} \cos(\xi_N) & \sin(\xi_N) \\ -\sin(\xi_N) & \cos(\xi_N) \end{bmatrix} \tag{5.47}$$

is the sample-operation which should occur. The easiest way to achieve this action is to act firstly and unconditionally upon the sample state with $W_0(\phi_N)$ and then, conditioned upon the probe state, to perform a third unitary, $W_0(\phi_N - \xi_N) = W_1(\xi_N)W_0^\dagger(\phi_N)$ on the sample state. A controlled rotation can be implemented by performing a CNOT sandwiched between two rotations. In this case the latter operations must both be $W_0((\phi_N - \xi_N)/2)$, which can be seen most easily by noting that a controlled operation can be written as a four-by-four matrix with the identity in the upper-left quadrant, the desired operation in the lower-right quadrant and zeroes elsewhere. Bringing all of this together, the sequence of operators gives

$$\begin{aligned}
&\begin{bmatrix} W_0(\frac{\phi_N-\xi_N}{2}) & 0 \\ 0 & W_0(\frac{\phi_N-\xi_N}{2}) \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \sigma_x \end{bmatrix} \begin{bmatrix} W_0(\frac{\phi_N-\xi_N}{2}) & 0 \\ 0 & W_0(\frac{\phi_N-\xi_N}{2}) \end{bmatrix} \\
&= \begin{bmatrix} W_0(\frac{\phi_N-\xi_N}{2})W_0(\frac{\phi_N-\xi_N}{2}) & 0 \\ 0 & W_0(\frac{\phi_N-\xi_N}{2})\sigma_x W_0(\frac{\phi_N-\xi_N}{2}) \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ 0 & W_0(\phi_N - \xi_N) \end{bmatrix},
\end{aligned} \tag{5.48}$$

which is the desired matrix. I have arrived at a gate scheme which, using just CNOTs, implements the unitary. The quantum circuit corresponding to this appear in Fig. 5.1.

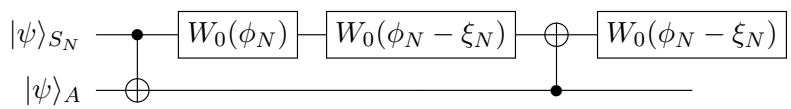


Figure 5.1: Quantum circuit which performs the unitary operation required for quantum data gathering. Gate W_0 defined in text.

It is beyond the scope of this work to provide a bound on the contribution to the

success probability due to gate noise, however a high-level argument is quick to introduce. Typically, gate fidelity is measured by the diamond norm [94, 95, 96]. This quantity is a measure defined on the space of matrices. It is not straightforward to work with, hence the limited scope of this argument. The only property of the diamond norm relevant here is that it satisfies the triangle inequality, in terms of distances between gate sequences. That is, if a state is acted on by either the channel E followed by E' or the channel F followed by F', the distance between the state after these processes is characterised by the distance $\|E \cdot E' - F \cdot F'\|_\diamond$, which can be bounded by

$$\|E \cdot E' - F \cdot F'\|_\diamond \leq \|E - F\|_\diamond + \|E' - F'\|_\diamond \tag{5.49}$$

as was shown by Aharanov *et al.* [97]. Thus the distance between the ideal gate sequence $E^{\otimes N}$ and the noisy implementation $\tilde{E}^{\otimes N}$ will be bounded by

$$\|E^{\otimes N} - \tilde{E}^{\otimes N}\|_\diamond \leq N\|E - \tilde{E}\|_\diamond. \tag{5.50}$$

I have shown that $2N$ CNOT gates will be needed to implement the routine and so the diamond norm of the overall gate sequence is at most $2N$ times the diamond norm of the distance between the individual CNOT gate and the noisy equivalent. In the final step of the quantum data gathering protocol, the actor seeks to distinguish between the two possible final states of the probe qubit. These two possible states will be shifted by the same amount, proportional to the diamond norm of the overall gate sequence. So, based on the argument here, I can say that the probability of success will decrease by a quantity proportional to $2N$. However, finding the relevant constant of proportionality is not straightforward. Sanders *et al.* find that the relation between the gate fidelity (the typically quoted quantity) can depend on the specific form of noise [98] and so a more-detailed model is needed to take this calculation further. I have considered only the two-qubit gates here as they will dominate the gate error. Single qubit gates have a much greater fidelity.

5.3.3 Success probability

My strategy for calculating the overall probability of success is to write each interaction as a Kraus operator acting on the probe. This is done by considering that the sample qubits are subsequently measured in the computational basis. If such a measurement is not performed then the density operator which they measure is represented by summing over both outcomes. This is the same as tracing out the interaction. I calculate the density operator of the probe after interacting with N of the qubits and the success probability is then calculated as the expectation value of the relevant projector. This method also allows one to calculate the probability that specific measurement records occur, if the samples are subsequently measured. Such a tool is useful in examining possible modifications of the scheme, as will be seen later.

The two Kraus operators, written in the computational basis $|0\rangle, |1\rangle$, are:

$$\begin{aligned} M_{0,k}^{(n)} &= \langle 0|_{\mathcal{S}_n} U_n |\tilde{\psi}_k\rangle_{\mathcal{S}_n} \\ &= \begin{bmatrix} \frac{\cos(\theta+\delta\theta_n)\cos(\theta)\cos(\theta_{n-1})}{\cos(\theta_n)} & \frac{(-1)^k \sin(\theta+\delta\theta_n)\sin(\theta)\sin(\theta_{n-1})}{\cos(\theta_n)} \\ \frac{(-1)^k \sin(\theta+\delta\theta_n)\sin(\theta)\cos(\theta_{n-1})}{\sin(\theta_n)} & \frac{\cos(\theta+\delta\theta_n)\cos(\theta)\sin(\theta_{n-1})}{\sin(\theta_n)} \end{bmatrix}, \\ M_{1,k}^{(n)} &= \langle 1|_{\mathcal{S}_n} U_n |\tilde{\psi}_k\rangle_{\mathcal{S}_n} \\ &= \begin{bmatrix} \frac{\cos(\theta+\delta\theta_n)\sin(\theta)\sin(\theta_{n-1})}{\cos(\theta_n)} & \frac{(-1)^{k+1} \sin(\theta+\delta\theta_n)\cos(\theta)\cos(\theta_{n-1})}{\cos(\theta_n)} \\ \frac{(-1)^k \sin(\theta+\delta\theta_n)\cos(\theta)\sin(\theta_{n-1})}{\sin(\theta_n)} & \frac{-\cos(\theta+\delta\theta_n)\sin(\theta)\cos(\theta_{n-1})}{\sin(\theta_n)} \end{bmatrix}. \end{aligned} \tag{5.51}$$

These act upon the Hilbert space $\mathcal{H}_A$ only. The first of these outcomes can be considered as indicating success, in the sense that if the fidelity is perfect, all information about the sample is transferred onto the probe. The second Kraus operator, which is not designed with this process in mind, will thus transfer information about the perpendicular state onto the probe. It can be considered as a failure of the protocol. In this sense, one can think about the subsequent sample measurement as a diagnostic of the scheme's performance.

It is useful at this point to rewrite these operators in terms of the probe basis at each state, i.e., the behaviour of interest is how the basis vectors $|\psi_k^{(n-1)}\rangle$ and $|\psi_{k\perp}^{(n-1)}\rangle$ are mapped on to those vectors $|\psi_k^{(n)}\rangle$ and $|\psi_{k\perp}^{(n)}\rangle$ which form a natural basis for the next stage of the protocol. This can be done if I first introduce the rotation matrix

$$\begin{bmatrix} |0\rangle \\ |1\rangle \end{bmatrix} = \begin{bmatrix} \cos(\theta_n) & \sin(\theta_n) \\ (-1)^k \sin(\theta_n) & (-1)^{k+1}\cos(\theta_n) \end{bmatrix} \begin{bmatrix} |\psi_k^{(n)}\rangle \\ |\overline{\psi}_k^{(n)}\rangle \end{bmatrix}. \tag{5.52}$$

The Kraus operators can be written in a form more useful for calculation, one in terms of rotations between the two relevant bases. I have

$$\begin{aligned} M_{0,k}^{(n)} &= \begin{bmatrix} \cos(\theta_n) & (-1)^k \sin(\theta_n) \\ \sin(\theta_n) & (-1)^{k+1}\cos(\theta_n) \end{bmatrix} \\ &\times \begin{bmatrix} \frac{\cos(\theta+\delta\theta_n)\cos(\theta)\cos(\theta_{n-1})}{\cos(\theta_n)} & \frac{(-1)^k \sin(\theta+\delta\theta_n)\sin(\theta)\sin(\theta_{n-1})}{\cos(\theta_n)} \\ \frac{(-1)^k \sin(\theta+\delta\theta_n)\sin(\theta)\cos(\theta_{n-1})}{\sin(\theta_n)} & \frac{\cos(\theta+\delta\theta_n)\cos(\theta)\sin(\theta_{n-1})}{\sin(\theta_n)} \end{bmatrix} \\ &\times \begin{bmatrix} \cos(\theta_{n-1}) & \sin(\theta_{n-1}) \\ (-1)^k \sin(\theta_{n-1}) & (-1)^{k+1}\cos(\theta_{n-1}) \end{bmatrix} \\ &= \begin{bmatrix} \cos(\delta\theta_n) & 0 \\ \frac{-\sin(2\theta)\sin(\delta\theta_n)\cos(2\theta_{n-1})}{\sin(2\theta_n)} & \frac{\cos(2\theta+\delta\theta_n)\sin(2\theta_{n-1})}{\sin(2\theta_n)} \end{bmatrix}. \end{aligned} \tag{5.53}$$

Similarly, the other Kraus operator is rewritten as

$$M_{1,k}^{(n)} = \begin{bmatrix} 0 & \sin(2\theta+\delta\theta_n) \\ \frac{-\sin(\delta\theta_n)\sin(2\theta_{n-1})}{\sin(2\theta_n)} & \frac{\sin(\delta\theta_n)\cos(2\theta_{n-1})-\sin(2\theta+\delta\theta_n)\cos(2\theta_n)}{\sin(2\theta_n)} \end{bmatrix} \tag{5.54}$$

To avoid any confusion, it must be emphasised that these two matrices include a basis rotation and hide the index of the transmitted state in such a matter. For example, the top

left element in each case corresponds to the coefficient of the object $|\psi_k^{(n)}\rangle\langle\psi_k^{(n-1)}|$. Written in this manner, the forms for the two update matrices at each stage can be understood more clearly as successes and failures. In particular, that the upper-right element of $M_{0,k}^{(n)}$ is zero means that a series of previous failures will not increase the probability of succeeding at a later point in the protocol. Conversely, the zero-valued upper-left element of $M_{1,k}^{(n)}$ says that a failure at any point in the protocol will delete all of the previously acquired information, leaving the probe uncorrelated to the state which it is hoping to identify. This suggests that there is potential for modifying the scheme: after each interaction, the sample can be measured in the computational basis. If the measurement outcome is $|1\rangle$ then the user would reinitialise the probe and the next two interactions would be a SWAP followed by U_2. As more and more interactions occur, the greater the chance of losing all information, i.e., the outcome $|1\rangle$, and the user is forced to consider how the increased probability of overall success at each point plays off against the probability that all will be lost. These dynamics are discussed in greater detail in a later section. Here, I assume that the sample is not measured, and sum over all outcomes.

I use these objects to calculate the overall probability that, given a resource of N sample qubits, the quantum data gathering protocol correctly identifies the relevant state. I first aquire, by successive use of the above Kraus operators in the usual state update rule, a general formula for the density matrix $\rho_k^{(n)}$ of the probe after interacting with n of the qubits. With $\rho_k^{(n)}$ in place then one must simply use the Born rule to acquire the probability of the state associated with $|\psi_0\rangle$ being the measurement outcome given that state $|\psi_0\rangle$ was sent, for example.

The first probe-sample interaction is a SWAP gate. At all points I assume that there is no noise in the gates themselves but on the preparation only. It is easy to see that the probe's state after this gate is

$$\begin{aligned}\rho_k^{(1)} &= \left(\cos(\delta\theta_1)|\psi_k\rangle + \sin(\delta\theta_1)|\psi_{k\perp}\rangle\right)\left(\cos(\delta\theta_1)\langle\psi_k| + \sin(\delta\theta_1)\langle\psi_{k\perp}|\right) \\ &= \begin{bmatrix} \cos^2(\delta\theta_1) & \cos(\delta\theta_1)\sin(\delta\theta_1) \\ \cos(\delta\theta_1)\sin(\delta\theta_1) & \sin^2(\delta\theta_1) \end{bmatrix}. \end{aligned} \tag{5.55}$$

The basis in which each $\rho_k^{(n)}$ is written is that of the ideal (noiseless) probe state at that point (i.e., $|\psi_k^{(n)}\rangle$) as well as the state orthogonal to that. In this manner, the index k is hidden inside the matrix's basis. As I assume that the noise on each individual qubit is independent, I can average over the noise parameter at each stage. The density matrix after the first step is

$$\rho_k^{(1)} = \begin{bmatrix} F & 0 \\ 0 & 1-F \end{bmatrix}. \tag{5.56}$$

This is a result which could have been constructed without calculation. It follows from the definition of the fidelity. In the case that $F = 1$ then the probe is in the state $|\psi_k^{(n)}\rangle$. The lowest possible value of F is $1/2$, and at this point the state becomes maximally mixed.

I can use this form to find $\rho_k^{(2)}$, which allows me to guess at the structure of the more

general result $\rho_k^{(N)}$. I have

$$\rho_k^{(1)} \to \rho_k^{(2)} = M_{0,k}^{(2)} \rho_k^{(1)} M_{0,k}^{(2)\dagger} + M_{1,k}^{(2)} \rho_k^{(1)} M_{1,k}^{(2)\dagger}. \tag{5.57}$$

The calculations are much more concise if each term on the right hand side here is evaluated individually and then both results brought together. Firstly,

$$\begin{aligned}
&M_{0,k}^{(2)} \rho_k^{(1)} M_{0,k}^{(2)\dagger} \\
&= \begin{bmatrix} F\cos^2(\delta\theta_2) & \frac{-F\sin(2\theta)\cos(\delta\theta_2)\sin(\delta\theta_2)\cos(2\theta)}{\sin(2\theta_2)} \\ \frac{-F\sin(2\theta)\cos(\delta\theta_2)\sin(\delta\theta_2)\cos(2\theta)}{\sin(2\theta_2)} & \frac{F\sin^2(2\theta)\sin^2(\delta\theta_2)\cos^2(2\theta)+(1-F)\cos^2(2\theta+\delta\theta_2)\sin^2(2\theta)}{\sin^2(2\theta_2)} \end{bmatrix} \\
&= \begin{bmatrix} F^2 & 0 \\ 0 & \frac{(1-F)(1-F+2(2F-1)\cos^2(2\theta)-(3F-1)\cos^4(2\theta))}{\sin^2(2\theta_2)} \end{bmatrix}.
\end{aligned} \tag{5.58}$$

In the second line of this equation I have averaged over the noise variable $\delta\theta_2$ in keeping with what has been presented above (strictly speaking the two lines of this equations are not equal (the second should include as at this point I have not averaged over $\delta\theta_i$) however I have chosen to keep the notation as simple as possible and hope that the meaning at each point is clear). This choice of basis hides the digit k and shows the advantage of folding the state to be discriminated into the basis. The second piece of interest is

$$\begin{aligned}
&M_{1,k}^{(2)} \rho_k^{(1)} M_{1,k}^{(2)\dagger} \\
&= \begin{bmatrix} (1-F)\sin^2(2\theta+\delta\theta_2) & 0 \\ 0 & \frac{F\sin^2(2\delta\theta_2)\sin^2(2\theta)+(1-F)(\sin(\delta\theta_2)\cos(2\theta)-\sin(2\theta+\delta\theta_2)\cos(2\theta_2))^2}{\sin^2(2\theta_2)} \end{bmatrix} \\
&+ \frac{(1-F)\sin(2\theta+\delta\theta_2)(\sin(\delta\theta_2)\cos(2\theta)-\sin(2\theta+\delta\theta_2)\cos(2\theta_2))}{\sin(2\theta_2)} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\
&= (1-F) \begin{bmatrix} F+(2F-1)\cos^2(2\theta) & 0 \\ 0 & \frac{F+(1-2F)\cos^2(2\theta)+(3F-2)\cos^4(2\theta)+(1-2F)\cos^6(2\theta)}{\sin^2(2\theta_2)} \end{bmatrix} \\
&+ (1-F)\frac{\cos^2(2\theta_2)\sin^2(2\theta)}{\sin^2(2\theta_2)}\sigma_x^{(2)}.
\end{aligned} \tag{5.59}$$

Here, σ_x is the Pauli matrix and the superscript attached to it indicates that the basis is the same as for the other piece. I have expressed both of these forms in terms of orders of $\cos^2(2\theta)$ only for ease of reading. It is useful to note here that $\sin^2(2\theta_2) = 1 - \cos^4(2\theta)$. Bringing both results together gives the density matrix of the probe qubit after interacting with just two samples

$$\begin{aligned}
\rho_k^{(2)} &= M_{0,k}^{(2)} \rho_k^{(1)} M_{0,k}^{(2)\dagger} + M_{1,k}^{(2)} \rho_k^{(1)} M_{1,k}^{(2)\dagger} \\
&= \begin{bmatrix} F-(1-F)(2F-1)\cos^2(2\theta) & 0 \\ 0 & (1-F)(1+(2F-1)\cos^2(2\theta)) \end{bmatrix} \\
&+ (1-F)\frac{\cos^2(2\theta)\sin^2(2\theta)}{\sin^2(2\theta_2)}\sigma_x^{(2)}.
\end{aligned} \tag{5.60}$$

This object is the density matrix of the probe once it has interacted with the first two sample qubits, both of which are in a mixed state associated with noise governed by the parameter F. One can convince themself further that this is the correct form by considering two properties which we would expect of a density matrix. Firstly, it satisfies $\mathrm{Tr}(\rho_k^{(2)}) = 1$, the usual normalisation condition. Secondly, in the perfect-fidelity limit $F = 1$ only the upper-left element of the matrix remains. This corresponds to the probe being found in the state $|\psi_k^{(2)}\rangle$, i.e., that which would be found in the clean case. Here, it is found as a specific point of a larger space.

That the state takes this form - a matrix with diagonal elements plus one which is proportional to σ_x - suggests a possible route towards my main target, the density matrix after interactions with any number of the sample qubits. This route is to calculate what happens to each of those pieces when the two updating matrices act upon them. What is found is that the dynamics of the probe are governed by the following two forms:

$$\begin{bmatrix} a_n & 0 \\ 0 & 1-a_n \end{bmatrix} \to \begin{bmatrix} a_{n+1} & 0 \\ 0 & 1-a_{n+1} \end{bmatrix} + b_{n+1}\sigma_x^{(n+1)}$$
$$\sigma_x^{(n)} \to c_{n+1}\sigma_x^{(n+1)}. \tag{5.61}$$

I have used a_n, b_n, c_n to represent various pieces of the density matrix after i interactions. Using the forms of $M_{0,k}^{(n)}$ and $M_{1,k}^{(n)}$ which I derived above it is possible to find those parameters as functions of F and θ. The result is, once again, a set of geometric summations which can be evaluated to give an analytic expression for the overall probability of success when measured by the minimum-error measurement. The rest of the calculation is a purely algebraic exercise, of which the next step is to evaluate the parameters a_n, b_n and c_n.

The first piece to analyse is the form of the update for a normalised matrix which has only diagonal elements. That is, I consider a generic matrix in which the upper-left element is a_n and the lower-right element is $1 - a_n$. How is this object updated by the two Kraus operators? I begin by calculating the form of each term independently and then sum. Each of the following is the result of performing the relevant matrix multiplication using the basic form of the two matrices, and then averaging over $\delta\theta_n$. These calculations follow the method used above for the two qubit case and the algebra procedes largely the same way.

The first piece of the calculation is

$$M_0^{(n)} \begin{bmatrix} a_{n-1} & 0 \\ 0 & 1-a_{n-1} \end{bmatrix} M_0^{(n)\dagger}$$
$$= \begin{bmatrix} a_{n-1}F & 0 \\ 0 & a_{n-1}\frac{(1-F)\sin^2(2\theta)\cos^2(2\theta_{n-1})}{\sin^2(2\theta_n)} + (1-a_{n-1})\frac{\sin^2(2\theta_{n-1})(F\cos^2(2\theta)+(1-F)\sin^2(2\theta))}{\sin^2(2\theta_n)} \end{bmatrix} \tag{5.62}$$

and the second piece is

$$
\begin{aligned}
&M_1^{(n)} \begin{bmatrix} a_{n-1} & 0 \\ 0 & 1-a_{n-1} \end{bmatrix} M_1^{(n)\dagger} \\
&= \begin{bmatrix} (1-a_{n-1}) & 0 \\ \times \left[F\sin^2(2\theta) + (1-F)\cos^2(2\theta)\right] & \\ & a_{n-1}\frac{(1-F)\sin^2(2\theta_{n-1})}{\sin^2(2\theta_n)} \\ 0 & +(1-a_{n-1})\frac{(1-F)\cos(2\theta_{n-1})+(2-F)\cos(2\theta_n)+(1-2F)\cos(2\theta_{n+1})}{\sin^2(2\theta_n)} \end{bmatrix} \\
&+ (1-a_{n-1})\frac{(1-2F)\cos(2\theta_n)\sin^2(2\theta)}{\sin(2\theta_n)}\sigma_x^{(n)}. \qquad (5.63)
\end{aligned}
$$

The sum of these two pieces gives the update for that part of the density matrix. After some algebraic manipulation it is possible to simplify the bottom right term greatly and the resulting object is

$$
\begin{aligned}
&\sum_i M_i^{(n)} \begin{bmatrix} a_{n-1} & 0 \\ 0 & 1-a_{n-1} \end{bmatrix} M_i^{(n)\dagger} \\
&= \begin{bmatrix} a_{n-1}\cos^2(2\theta)(2F-1) & 0 \\ +F\sin^2(2\theta) + (1-F)\cos^2(2\theta) & \\ 0 & 1-a_{n-1}\cos^2(2\theta)(2F-1) \\ & -F\sin^2(2\theta) - (1-F)\cos^2(2\theta) \end{bmatrix} \\
&- (1-a_{n-1})\frac{(2F-1)\cos(2\theta_n)\sin^2(2\theta)}{\sin(2\theta_n)}\sigma_x^{(n)}. \qquad (5.64)
\end{aligned}
$$

In this equation, the upper-left and lower-right elements explicitly sum to one so it clear that the update is trace preserving.

I also require the equivalent update for the σ_x piece. For $M_{0,k}^{(n)}$ the result (after matrix-multiplication and subsequent averaging over $\delta\theta_n$) is

$$
\begin{aligned}
M_{0,k}^{(n)}\sigma_x^{(n-1)}M_{0,k}^{(n)\dagger} &= \frac{F\cos(2\theta)\sin(2\theta_{n-1})}{\sin(2\theta_n)}\sigma_x^{(n)} \\
&+ \frac{2(1-F)\sin^2(2\theta)\cos(2\theta_{n-1})\sin(2\theta_{n-1})}{\sin^2(2\theta_n)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \qquad (5.65)
\end{aligned}
$$

In a similar manner I find that the other piece is

$$
\begin{aligned}
M_{1,k}^{(n)}\sigma_x^{(n-1)}M_{1,k}^{(n)\dagger} &= -\frac{(1-F)\cos(2\theta)\sin(2\theta_{n-1})}{\sin(2\theta_n)}\sigma_x^{(n)} \\
&- \frac{2(1-F)\sin^2(2\theta)\cos(2\theta_{n-1})\sin(2\theta_{n-1})}{\sin^2(2\theta_n)} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}. \qquad (5.66)
\end{aligned}
$$

Hence, the sum of these two pieces is

$$
\sum_i M_{i,k}^{(n)}\sigma_x^{(n-1)}M_{i,k}^{(n)\dagger} = \frac{(2F-1)\sin(2\theta_{n-1})\cos(2\theta)}{\sin(2\theta_n)}\sigma_x^{(n)} \qquad (5.67)
$$

and the probe's updates always follow the structure which I presented above. There are two pieces to the density matrix at each point: a part with elements only along the diagonal and a part which is proportional to the σ_x matrix when it is written in the probe basis relevant to that stage of the measurement. With Eqs. 5.64 and 5.67 I am in a position to calculate the density matrix for a given generic case. For convenience it is useful to introduce the following notation

$$x = \cos^2(2\theta)(2F-1) \tag{5.68}$$

$$y = F\sin^2(2\theta) + (1-F)\cos^2(2\theta) \tag{5.69}$$

$$r_n = \frac{\cos(2\theta_n)\sin^2(2\theta)(2F-1)}{\sin(2\theta_n)} \tag{5.70}$$

$$s_n = \frac{\sin(2\theta_{n-1})(2F-1)\cos(2\theta)}{\sin(2\theta_n)}. \tag{5.71}$$

In terms of these parameters the two expressions involved in the state update can be expressed as

$$\begin{aligned}
&\sum_i M_{i,k}^{(n)} \begin{bmatrix} a_{n-1} & 0 \\ 0 & 1-a_{n-1} \end{bmatrix} M_{i,k}^{(n)\dagger} \\
&= \begin{bmatrix} xa_{n-1}+y & 0 \\ 0 & 1-xa_{n-1}-y \end{bmatrix} - (1-a_{n-1})r_n\sigma_x^{(n)} \qquad (5.72)\\
&\sum_i M_{i,k}^{(n)}\sigma_x^{(n-1)}M_{i,k}^{(n)\dagger} = s_n\sigma_x^{(n)}. \qquad (5.73)
\end{aligned}$$

Repeated applications of these formulae to the initial density matrix, which I have written above, result in a density matrix

$$\rho_k^{(N)} = \begin{bmatrix} A_N & 0 \\ 0 & 1-A_N \end{bmatrix} - B_N\sigma_x^{(N)}, \tag{5.74}$$

in which

$$A_N = Fx^{N-1} + y\sum_{i=0}^{N-2} x^i \tag{5.75}$$

$$B_N = \sum_{i=1}^{N-1}(1-A_i)r_{i+1}\prod_{j=i+2}^{N} s_j. \tag{5.76}$$

I have a formula for the density matrix of the probe after interacting with N qubits. The first piece to evaluate is the series in A_N. This is the straightforward expression of geometric summation:

$$\begin{aligned}
\sum_{i=0}^{N-2} x^i &= \frac{1-x^{N-1}}{1-x} \\
&= \frac{1-\cos^{2N-2}(2\theta)(2F-1)^{N-1}}{1-\cos^2(2\theta)(2F-1)}. \qquad (5.77)
\end{aligned}$$

The first term is

$$\begin{aligned} A_N &= F\cos^{2N-2}(2\theta)(2F-1)^{N-1} \\ &\quad + \left(F\sin^2(2\theta) + (1-F)\cos^2(2\theta)\right)\frac{1-\cos^{2N-2}(2\theta)(2F-1)^{N-1}}{1-\cos^2(2\theta)(2F-1)} \\ &= 1-(1-F)\frac{1-\cos^{2N}(2\theta)(2F-1)^N}{1-\cos^2(2\theta)(2F-1)}. \end{aligned} \tag{5.78}$$

The next step is to use this to evaluate the other coefficient, B_N. The expression Eq. 5.76 for this gives

$$\begin{aligned} B_N &= \sum_{j=1}^{N-1}(1-A_j)r_{j+1}\prod_{i=j+2}^{N}s_i \\ &= \sum_{j=1}^{N-1}\left[(1-F)\frac{1-\cos^{2j}(2\theta)(2F-1)^j}{1-\cos^2(2\theta)(2F-1)}\right. \\ &\quad \left.\times\frac{\cos(2\theta_{j+1})\sin^2(2\theta)(2F-1)}{\sin(2\theta_N)}\left(\cos(2\theta)(2F-1)\right)^{N-j-2}\right]. \end{aligned} \tag{5.79}$$

After a few lines of rearrangment one finds that this can be written as

$$B_N = (1-F)\cos^{N-1}(2\theta)(2F-1)^{N-1}\frac{\sin^2(2\theta)}{\sin(2\theta_N)}\sum_{j=1}^{N-1}\frac{1-\cos^{2j}(2\theta)(2F-1)^j}{(2F-1)^j}. \tag{5.80}$$

Once again, the task has come down to evaluating a pair of geometrical progressions for objects containing $\cos(2\theta)$ and $(2F-1)$. After evaluating these terms using the standard formula and then simplifying as much as possible, I have

$$\begin{aligned} B_N &= (1-F)\frac{\sin^2(2\theta)\cos^{N-1}(2\theta)}{\sin(2\theta_N)} \\ &\quad\times\left(\frac{1-(2F-1)^{N-1}}{1-(2F-1)} - \cos^{N+1}(2\theta)(2F-1)^{N-1}\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^2(2\theta)}\right). \end{aligned} \tag{5.81}$$

The denominators of the two fractions inside the brackets on the right hand side can both be simplified further, to $1-F$ and $\sin^2(2\theta)$ respectively. I have chosen to leave them in their current form as it makes clear that there are no issues when taking the limits $F\to 1$ and $\theta\to 0$. At this point it worth pausing with the calculation to summarise what has been achieved so far. The quantum data gathering routine uses N copies of one of two possible quantum states $|\psi_k\rangle$. The information on each of these is transferred to a probe which is initialised in the fiducial state $|0\rangle$. After interacting with all N copies, and assuming that the experimenter does not measure the sample qubits subsequently, the probe will be found in a state $\rho_k^{(N)}$. This state can be defined in terms of two coefficients, A_N and B_N using equations 5.74, 5.78 and 5.81. These objects are calculated from the original Kraus operators which were constructed to represent the update of the probe state if the unitary used at each stage either succeeds or fails. To find the probe's state both outcomes are summed. Once the probe is in this state it is measured with a minimum-

error measurement corresponding to the two possible probe states which would occur in the $F = 1$ (noisless) cases. This part of the scheme is introduced to the analysis now.

Minimum-error measurement of the probe seeks to distinguish the two states which the probe can be in for a given preparation. In the noiseless case this will be the pair of non-orthogonal states parameterised by

$$|\psi_k^{(N)}\rangle = \cos(\theta_N)|0\rangle + (-1)^k \sin(\theta_N)|1\rangle, \quad k = 0, 1 \tag{5.82}$$

which is repeated here for convenience. As is well known, the minimum-error measurement for two states ρ_0 and ρ_1 consists of projecting onto the eigenvectors of the difference between them, $\rho_0 - \rho_1$, the so-called Helstrom measurement. In the calculations so far I have worked in the basis $|\psi_k^{(N)}\rangle, |\psi_{k\perp}^{(N)}\rangle$. This is the natural basis to work in as it hides the label of the particular state k. I introduce also $|\overline{\psi}_k^{(N)}\rangle$ to represent the state that has *not* been transmitted. In terms of the calculational basis, that state can be written as

$$|\overline{\psi}_k^{(N)}\rangle = \cos(2\theta_N)|\psi_k^{(N)}\rangle + \sin(2\theta_N)|\psi_{k\perp}^{(N)}\rangle. \tag{5.83}$$

Using this, one can evaluate the required matrix as

$$|\psi_k^{(N)}\rangle\langle\psi_k^{(N)}| - |\overline{\psi}_k^{(N)}\rangle\langle\overline{\psi_k^{(N)}}| = \sin(2\theta_N) \begin{bmatrix} \sin(2\theta_N) & -\cos(2\theta_N) \\ -\cos(2\theta_N) & -\sin(2\theta_N) \end{bmatrix}. \tag{5.84}$$

As stated above, the measurement which maximises the chance of correctly identifying the transmitted state is that which consists of a projection onto the eigenvalues of this matrix [17, 18] (correct identification corresponding to the positive eigenvalue; incorrect identification corresponding to the negative eigenvalue, due to how the task has been set up). One finds in the usual way that the two eigenvectors (with eigenvalues ± 1) are

$$|\psi_+^{(N)}\rangle = \sqrt{\frac{1+\sin(2\theta_N)}{2}}|\psi_k^{(N)}\rangle + \sqrt{\frac{1-\sin(2\theta_N)}{2}}|\psi_{k\perp}^{(N)}\rangle \tag{5.85}$$

$$|\psi_-^{(N)}\rangle = \sqrt{\frac{1-\sin(2\theta_N)}{2}}|\psi_k^{(N)}\rangle - \sqrt{\frac{1+\sin(2\theta_N)}{2}}|\psi_{k\perp}^{(N)}\rangle. \tag{5.86}$$

The probability of successfully identifying the transmitted state (to reiterate: given that there is noise in the sample's state but that the state discrimination is identical to that in the noiseless case) will be $\mathrm{P}_N^{QDG} = \langle\psi_+^{(N)}|\rho_k^{(N)}|\psi_+^{(N)}\rangle$. The probe's density matrix is split into two pieces and I evaluate the expectation value which is associated with each of those seperately. One finds

$$\begin{aligned} &\langle\psi_+^{(N)}| \begin{bmatrix} A_N & 0 \\ 0 & 1-A_N \end{bmatrix} |\psi_+^{(N)}\rangle = \frac{1}{2}\left(1 - \sin(2\theta_N)\right) + \sin(2\theta_N)A_N \\ &\langle\psi_+^{(N)}|\sigma_x^{(N)}|\psi_+^{(N)}\rangle = \cos(2\theta_N). \end{aligned} \tag{5.87}$$

Finally, I am in a position to evaluate the probability that the quantum data gathering routine correctly discriminates two quantum states. In terms of A_N and B_N, which appear

above in equations 5.78 and 5.81 respectively, one finds

$$\begin{aligned} \mathrm{P}_N^{QDG} &= \langle\psi_+^{(N)}|\rho_k^{(N)}|\psi_+^{(N)}\rangle \\ &= \langle\psi_+^{(N)}|\begin{bmatrix} A_N & 0 \\ 0 & 1-A_N \end{bmatrix}|\psi_+^{(N)}\rangle - B_n\langle\psi_+^{(N)}|\sigma_x^{(N)}|\psi_+^{(N)}\rangle \\ &= \frac{1}{2}\left(1-\sin(2\theta_N)\right) + \sin(2\theta_N)A_N - \cos(2\theta_N)B_N. \end{aligned} \tag{5.88}$$

Putting all pieces into one expression, this result is

$$\begin{aligned} &\mathrm{P}_N^{QDG} = \\ &= \frac{1}{2}\left(1+\sqrt{1-\cos^{2N}(2\theta)}\right) - (1-F)\frac{1-\cos^{2N}(2\theta)(2F-1)^N}{1-\cos^2(2\theta)(2F-1)}\sqrt{1-\cos^{2N}(2\theta)} \\ &- (1-F)\sin^2(2\theta)\cos^{2N-2}(2\theta)\frac{1}{\sqrt{1-\cos^{2N}(2\theta)}} \\ &\times\left(\frac{1-(2F-1)^{N-1}}{1-(2F-1)} - \cos^2(2\theta)(2F-1)^{N-1}\frac{1-\cos^{2N-2}(2\theta)}{1-\cos^2(2\theta)}\right) \end{aligned} \tag{5.89}$$

I have, finally, arrived at an expression for the probability that the quantum data gathering routine successfully identifies the transmitted state if that state is prepared imperfectly. For reference, the Helstrom bound is $\mathrm{P}_H^N = (1+\sqrt{1-\cos^{2N}(2\theta)})/2$, the leading order term here. (I have substituted objects written in terms of $\cos^N(2\theta)$ for those such as $\sin(2\theta_N)$ in order to make the relation to the Helstrom bound explicit.) This expression can be compared with that which was derived for an equivalent local scheme and which is written in Eq. 5.27. Both expressions have a roughly similar structure: a leading term which is the Helstrom bound follow by two terms which are proportional to $2\mathrm{P}_H^N - 1$ and the reciprocal of that quantity respectively. The two cases have different coefficients in each case. I discuss how the two schemes behave numerically in more detail below. For now it is useful to provide some basic checks that the current result, Eq. 5.89, behaves healthily. One thing that would be expected is that the success-probability would become equal to the Helstrom bound in the limit that there is no noise. This can be seen easily as the two terms in which the quantum data gathering probability differs from the Helstrom bound both contain factors of $1-F$ and as such they go to zero in the perfect-fidelity limit. Another test-case is letting the two states to be discriminated be the same state, represented by $\theta = 0$. If the two states are the same then no measurement can distinguish them and, as noise effects both states equally, this should still be true irrespective of F's value. In this limit I consider in turn each of the three terms in Eq. 5.89: the first, which is the Helstrom bound, goes to $1/2$; the second term disappears as it contains $1-\cos^{2N}(2\theta)$; and the third goes to zero as it is proportional to $\sin^2(2\theta)$. (One must be careful with the latter limit due to a denominator which also goes to zero, however there is no issue.) Thus I have recovered from the general formula that two identical states are indistinguishable. In both of the cases that I have presented, I find that the calculated results agree with what must have been true a priori, and have further confidence that the general form is correct. In the next section, I present this function graphically for a selection of parameters

and compare the function to the equivalent result for local adaptive measurement. Before doing so, I look at the limit of a large number of copies as another test of the formula.

5.3.4 Many-copy limit

Just as in the local adaptive scheme for quantum state discrimination, it is instructive to look at the limiting behaviour of the protocol: how does the unitary act after interacting with a large number of copies? The probability of success can be calculated for this measurement, and compared with the many-copy limit which is derived from the general form of the probability of success, Eq. 5.89. Finding that both are equal encourages one that the original calculation was performed correctly. The many-copy limit for Eq. 5.89 is

$$\lim_{N\to\infty} \mathrm{P}_N^{QDG} = 1 - \frac{1-F}{1-\cos^2(2\theta)(2F-1)}. \tag{5.90}$$

A quick look at the equivalent result for local adaptive measurements, Eq. 5.32, might surprise the reader: the two schemes have precisely the same many-copy limit! Despite the great differences between the two protocols, both of them tend towards having the same resilience to preparation noise. This seems to indicate that the above expression is fundamental in some way. Unfortunately I do not currently have an explanation for this result, by which I mean an interperation of what that physical formula represents. It should be noted that, in both cases, the scheme aims to distinguish between two mixed states (the noisy case) but uses a scheme designed for pure states. This suggests that the many-copy limit which is derived in both cases is a systematic limit which could be bettered by altering the scheme.

For local adaptive measurement, I showed how the behaviour of that scheme in the many-copy limit had some intuitive properties, as a hypothesis-checking measurement, and was able to derive the relevant probability using that picture. Here, I do the same for quantum data gathering. In this limit, one finds that $\cos(\theta_N) = \sin(\theta_N) = 1/\sqrt{2}$. The 'natural basis' (as I have been referring to it) of the probe in this limit thus becomes

$$|\psi_0^{(N)}\rangle = |+\rangle \quad |\psi_1^{(N)}\rangle = |-\rangle \tag{5.91}$$

and the unitary (see Eq. 5.44) becomes

$$\begin{aligned}
U_N|0_{S_N}0_A\rangle &= \cos(\theta)|0_{S_N}0_A\rangle + \sin(\theta)|1_{S_N}0_A\rangle = |\psi_{0S_N}0_A\rangle \\
U_N|1_{S_N}1_A\rangle &= \sin(\theta)|0_{S_N}0_A\rangle - \cos(\theta)|1_{S_N}0_A\rangle = |\psi_{0\perp S_N}0_A\rangle \\
U_N|1_{S_N}0_A\rangle &= \sin(\theta)|0_{S_N}1_A\rangle + \cos(\theta)|1_{S_N}1_A\rangle = |\psi_{1\perp S_N}1_A\rangle \\
U_N|0_{S_N}1_A\rangle &= \cos(\theta)|0_{S_N}1_A\rangle - \sin(\theta)|1_{S_N}1_A\rangle = |\psi_{1S_N}1_A\rangle.
\end{aligned} \tag{5.92}$$

Due to the form of the probe in this limit, it is useful to write the action of the unitary so that the probe term is written in terms of the σ_x basis and the sample is written in terms of the transmitted states (and those which are orthogonal to them). The most intelligible

way to express this unitary is in the form

$$\begin{aligned} U_N|\psi_{0S_N}+_A\rangle &= |0_{S_N}+_A\rangle \\ U_N|\psi_{0\perp S_N}+_A\rangle &= |1_{S_N}-_A\rangle \\ U_N|\psi_{1S_N}-_A\rangle &= |0_{S_N}-_A\rangle \\ U_N|\psi_{1\perp S_N}-_A\rangle &= |1_{S_N}+_A\rangle. \end{aligned} \tag{5.93}$$

This shows that the unitary can be interpreted in a similar manner to the limiting case: as hypothesis checking. I associate the probe state $|+\rangle$ with the belief that the transmitted state was $|\psi_0\rangle$. If the probe starts off in the former state and interacts with a sample state in the latter, it will stay in the state $|+\rangle$ and leave the sample in the state $|0\rangle$, such as the protocol is designed for. If, however, the probe is in the state $|+\rangle$ and then interacts with a sample in the state $|\psi_{0\perp}\rangle$, the former will be left in the state $|-\rangle$ (corresponding to guessing that the transmitted state was $|\psi_1\rangle$) and the former left in the state $|1\rangle$, which signals a 'failed' protocol. Of course, this cannot occur unless there is noise in the state preparation; in the noiseless case the probe will stay in the state $|+\rangle$ after all interactions. The scenario which occurs if the most likely transmitted state was instead $|\psi_1\rangle$ is the same but with the roles of $|+\rangle$ and $|-\rangle$ reversed.

Unfortunately, there is no obvious method of calculating the many-copy limit of the quantum data gathering in the same way as is possible for the local adaptive scheme. However, a similar calculation is possible which uses the limit form of the Kraus operators, which I label $M^{\infty}_{i,k}$ and which have a much simpler form than the general case:

$$\begin{aligned} M^{\infty}_{0,k} &= \begin{bmatrix} \cos(\delta\theta_N) & 0 \\ 0 & \cos(2\theta+\delta\theta_N) \end{bmatrix} \\ M^{\infty}_{1,k} &= \begin{bmatrix} 0 & \sin(2\theta+\delta\theta_N) \\ -\sin(\delta\theta_N) & 0 \end{bmatrix}. \end{aligned} \tag{5.94}$$

These operators are written in the computational basis, as are all of the operators in this limit. As I showed earlier, the density matrix when written in the natural basis will consist of two pieces: a trace-one operator with only diagonal terms and an operator proportional to σ_x. I only need to find out how these pieces update in the many copy limit and, beginning with the latter, I consider each in term. A quick calculation reveals that

$$\sigma_x \rightarrow \sum_i M^{\infty\dagger}_{i,k}\sigma_x M^{\infty}_{i,k} = (2F-1)\cos(2\theta)\sigma_x \tag{5.95}$$

is how the σ_x piece of the density operator will be updated after each interaction. The effect is only to multiply the piece by a factor which is less than one. I am here interested in the limit of an infinite number of copies. It is clear that this term will be suppressed, tending to zero in that limit, and the only contribution to the many-copy density matrix will come from the other term. After one interaction the state update of a trace-one,

diagonal matrix is

$$\begin{aligned}\begin{bmatrix} A & 0 \\ 0 & (1-A) \end{bmatrix} \rightarrow & \sum_i M_{i,k}^{\infty\dagger} \begin{bmatrix} A & 0 \\ 0 & (1-A) \end{bmatrix} M_{i,k}^{\infty} \\ &= \begin{bmatrix} (2F-1)\cos^2(2\theta)A & 0 \\ 0 & -(2F-1)\cos^2(2\theta)A \end{bmatrix} \\ &+ \begin{bmatrix} F-(2F-1)\cos^2(2\theta) & 0 \\ 0 & 1-F+(2F-1)\cos^2(2\theta) \end{bmatrix}. \end{aligned} \tag{5.96}$$

I have broken the resulting operator into two terms. The first of these contains the parameter A, which is between one and zero, and characterises the probe's density operator at a given point in the protocol. It is seen that, as for the σ_x term, this term has been multiplied by a positive factor which is less than one and so, if the state is updated many times, that term's contribution will tend to zero. What remains is a different, diagonal trace one matrix. The upper-left term will be the probability of success (as the two post-interaction states are orthogonal in this regime) and after a little consideration it is seen that this has the form

$$\begin{aligned} \mathrm{P}_{\infty}^{QDG} &= \big(F-(2F-1)\cos^2(2\theta)\big)\sum_{i=0}^{\infty}\cos^{2i}(2\theta)(2F-1)^i \\ &= 1-\frac{1-F}{1-\cos^2(2\theta)(2F-1)}. \end{aligned} \tag{5.97}$$

I have once again arrived at the same value for the many-copy limit of multiple-copy data gathering schemes as was arrived at in previous calculations (the local adaptive scheme, and above as the limit of another function). The calculation here should be easier to follow, and the many-copy limit of the interaction operator's structure shows that the two methods have a similar underlying logic.

5.4 Comments

In the previous two sections, I derived analytic expressions for the probability that the local adaptive and quantum-data-gathering schemes correctly identify the noisy equivalents of the pure states which they were designed for. In this section, I compare the two schemes in more detail and discuss how each can be improved.

In Figures 5.2 and 5.3 are plots showing the behaviour of the two functions, Eqs. 5.30 and 5.89, for two different fidelities ($F = 0.95$ and 0.99) and two different angles ($\theta = \pi/6$ and $\pi/8$) so that there are four plots overall. Each graph also displays the Helstrom bound for the perfect fidelity case of that angle. Despite the range of parameters given, the same basic behaviour reoccurs. In all cases, the two schemes approach the same asymptote, as was seen earlier, and differ relatively little before that. There is a slight advantage to using the local adaptive scheme, as it more quickly approaches the asymptote in all cases. Admittedly that improvement is small (for example, in the $F = 0.99$ cases I find numerically that the difference is in the fourth or fifth decimal place), but that there

is any improvement is interesting in itself. It is typical to think that the global optimum of all measurements will be one which requires coherence, i.e., corresponds to measuring components of the product state which are not in the two dimensional subspace of the relevant states. However, here this is not the case. This result is in keeping with previous results in which local measurement is also generally better and it also holds in those cases that the improvement is very small [99]. As would be expected, for both schemes it is still seen the performance gets worse, in terms of the number of copies needed to get close to the asymptotic value, if either the preparation is more noisy (F is smaller) or the possible states are closer together (θ is smaller).

It must be emphasised that in both cases here I have taken the schemes *as they are optimised for pure states but applied them to mixed states*. It is obvious that, if one is to take into account the noisiness, this cannot be the best scheme as I have shown that the infinite-copy limit has an upper bound of the probability which is less than one. It must be true that there exists a scheme which discriminates perfectly in this limit. A straightforward argument leads to this conclusion: if one ignored the set of possibly transmitted states and performed tomography, the mixed state could be characterised completely and this would definitively identify the index of the transmission. It cannot be true that having some initial information causes one to perform worse, ergo, a discrimination scheme must be able to also reach the same bound. For both local adaptive measurement and quantum data gathering, an obvious method for improving the scheme exists.

In the local adaptive protocol, an important feature of Acín *et al.*'s scheme is that it is Markovian, by which is meant that the measurement performed on the Nth copy of the sample depends only upon the result of the $(N-1)$th measurement. Thus, their result does not utilise the entire measurement record. A true Bayesian scheme would be to update the probabilities which are used in deriving the Helstrom measurement after each sample, based upon all previous outcomes. One would expect that this improves the probability of success.

In the quantum data gathering protocol, the scheme can be made more flexible and thus perform better by measuring the samples after performing the protocol. To recap, in that scheme the sample qubit is left in the state $|0\rangle$ after the unitary if there is no noise but if the preparation is imperfect, then the sample qubit's post-interaction state can be something else. I took this into account by associating the state $|1\rangle$ with a failure. Importantly, it can be seen from the construction of the relevant Kraus operator in Eq. 5.54 that a failure in the scheme is irreversible: one loses all information about the state before measurement. In deriving Eq. 5.89, I sum over all possible measurement records, i.e., the experimentalist is assumed to ignore the possible outcomes. A method that might be expected to improve the scheme is to measure instead the sample qubits after the interaction and, if one finds that it is the state $|1\rangle$, to start the protocol again. An issue with this approach is that the probability that an interaction fails at some point increases linearly with the number of interactions. The question one is lead to is: when to stop? Presumably if one has access to one hundred copies of a state and has successfully interacted with ninety-nine of those, the small increase in the probability of success is not worth the risk of losing all the data

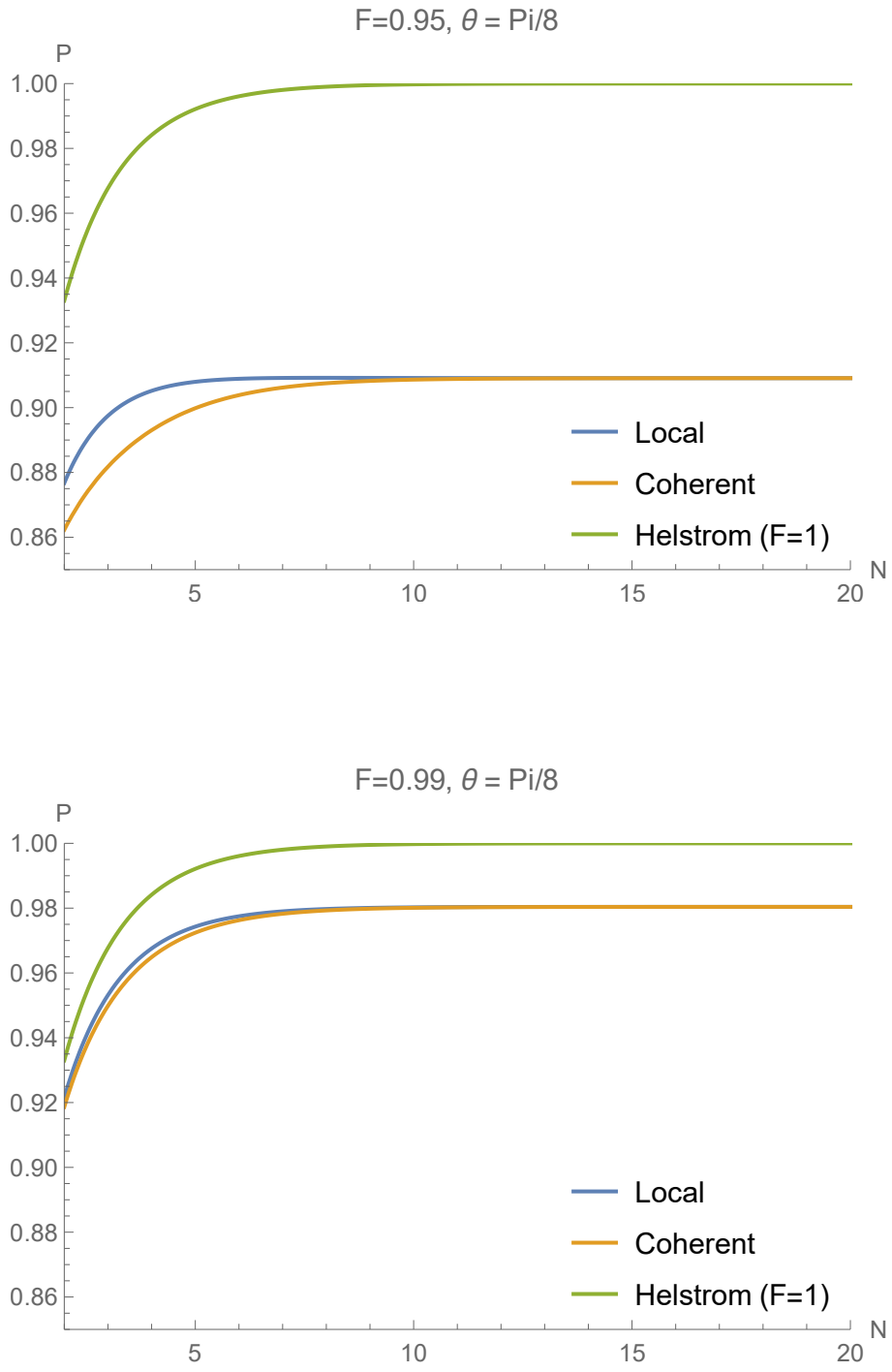


Figure 5.2: These plots display the three functions governing the probability of success for mutliple-copy state discrimination as a function of increasing number of resource qubits. Here the two plots display the success rate for discriminating two states, separated by an angle $\theta = \pi/8$ as defined in the text, which have been prepared with fidelities $F = 0.95$ and $F = 0.99$. The Helstrom bound which is plotted is that for distinguishing between the two relevant pure states.

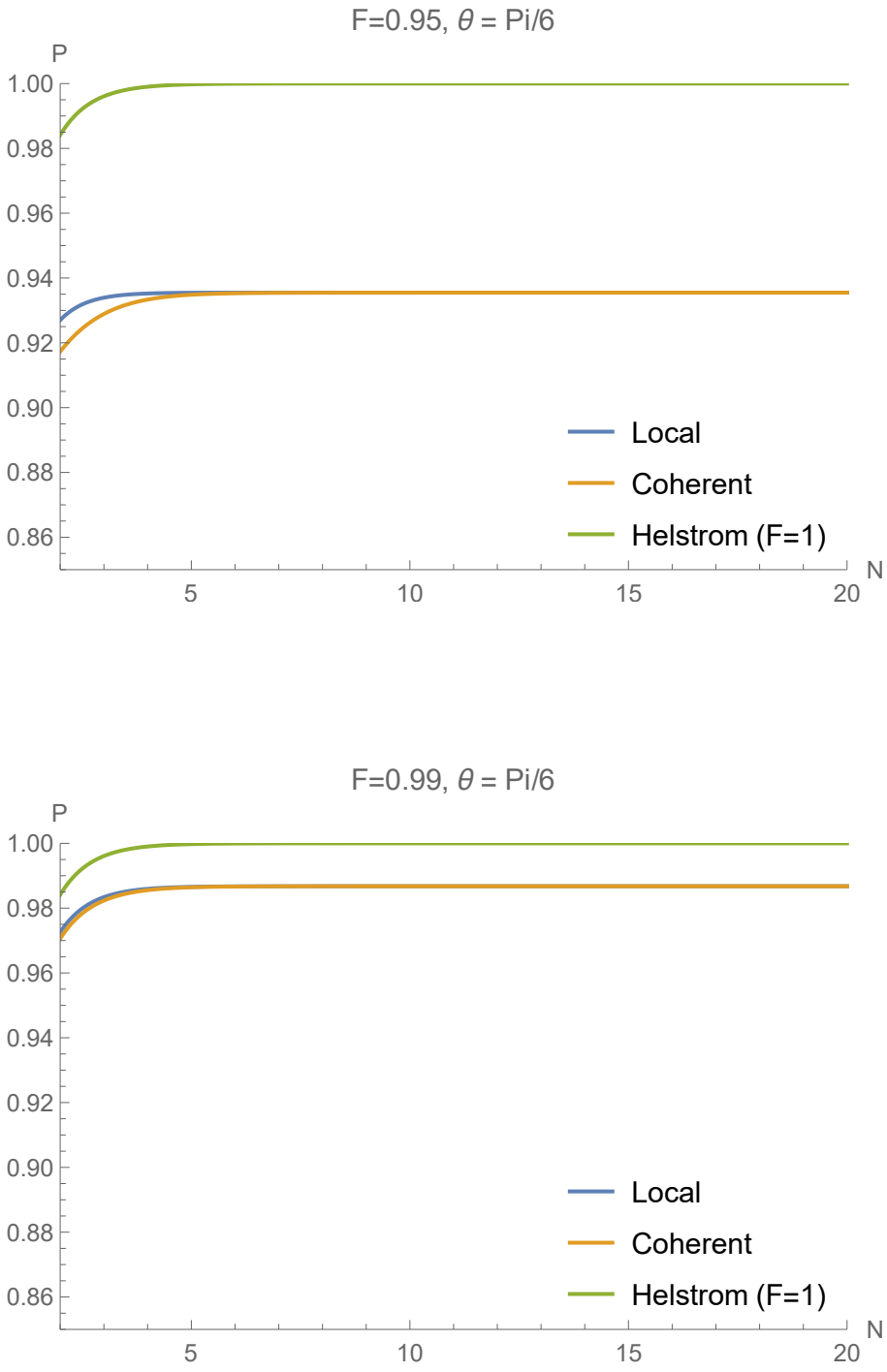


Figure 5.3: These plots display the three functions governing the probability of success for mutliple-copy state discrimination as a function of increasing number of resource qubits. Here the two plots display the success rate for discriminating two states, separated by an angle $\theta = \pi/6$ as defined in the text, which have been prepared with fidelities $F = 0.95$ and $F = 0.99$. The Helstrom bound which is plotted is that for distinguishing between the two relevant pure states.

so far gathered. To answer this question requires a better understanding of how quickly the success-probability function approaches its asymptote.

To approximate the asymptotic behaviour of the function, I write the success probability from Eq. 5.89 as

$$\mathrm{P}_N^{QDG} = \left(\lim_{N\to\infty} \mathrm{P}_N^{QDG}\right) - \epsilon_N^{QDG}. \tag{5.98}$$

The quantity ϵ_N thus quantifies how close the function is to the asymptote at a given value of N. As I am interested in the point at which it is very close to that limit, I wish ϵ_N to be small and as such can approximate the function. I Taylor-expand the various expressions in that function so that it is written as a series in terms of $\cos^2(2\theta)$ and $(2F-1)$ to the power of N and ignore all but the first order. The result can then be rearranged so that it gives the number of measurements required for a given proxmity to the limit:

$$N \approx \frac{\log\left(\frac{\cos^2(2\theta)\epsilon_N^{QDG}}{2F(1-F)}\right)}{2\log\left(\cos(2\theta)\right)}. \tag{5.99}$$

The quality of this approximation can be checked by evaluating a specific case. I consider that I wish to achieve a distance of $\epsilon_N = 0.01$ from the best possible accuracy for an experiment in which $\theta = \pi/6$ and in which the fidelity reaches $F = 0.99$. The choice of $\epsilon = 0.01$ here is somewhat arbitrary although it seems reasonable to only require that the experiment only runs as well as the fidelity allows, i.e., $\epsilon = 1 - F$, and this also simplifies the above expression (which contains $\epsilon/(1-F)$). Evaluating that formula with these parameters gives $N \approx 1.49$, so we would need only two copies to come within one-percent of the many-copy limit. If both the many-copy limit and the success-probability are evaluated with the relevant choices of the various parameters then I find $\mathrm{P}_N^{QDG} - \left(\lim_{N\to\infty} \mathrm{P}_2^{QDG}\right) = 0.016$, so the approximation is good within half-a-percent. Direct evaluation then finds that in fact three, rather than two, copies are required to achieve that level of success. A similar result is found if the same angle is used with a fidelity of $F = 0.95$, with evaluation of the approximate expression giving $N = 4$ for the required number of copies but calculating directly from Eq. 5.89 telling us that $N = 5$ is the correct result. The approximation is fairly good. The main takeaway from these calculations should be that, in general, only a small number of copies are needed to perform the quantum data gathering routine as well as possible.

Based on this analysis, one might suggest a modification of the routine which takes into account the possibility for post-selection. An experimentalist would then: calculate the number N from the above equation; interact her probe with each sample-qubit as specified in the above unitary operations while measuring the samples in the computational basis; when she has N successes, measure the probe; when a failure occurs, start from the beginning. However, a calculation reveals that there are some subtleties with this approach, the reason being that the probability of getting a succession of successes in even a very fidelitous, yet still imperfect, protocol is low enough to cancel out the advantage due to those successes. A short calculation demonstrates this property. I distinguish between two probabilities: P(succ), which is the overall probability of success, and $\mathrm{P}(\mathrm{succ}|\rho_0^N)$, the

probability that the correct transmission is correctly identified, conditioned upon successful interactions with N qubits in a row. As before, the calculation must begin by calculating the probe's density matrix after N such interactions. That is, I must calculate

$$\rho_{0,k}^{(N)} = \frac{M_{0,k}^{(N)} M_{0,k}^{(N-1)} \cdots \rho^{(1)} \cdots M_{0,k}^{(N-1)\dagger} M_{0,k}^{(N)\dagger}}{\mathrm{Tr}(M_{0,k}^{(N)} M_{0,k}^{(N-1)} \cdots \rho^{(1)} \cdots M_{0,k}^{(N-1)\dagger} M_{0,k}^{(N)\dagger})}. \tag{5.100}$$

I begin by evaluating the numerator of this equation, in the same manner as I have done previously, by deriving an inductive expression and evaluating the resultant terms for general N. One finds

$$M_{0,k}^{(N)} M_{0,k}^{(N-1)} \cdots \rho^{(1)} \cdots M_{0,k}^{(N-1)\dagger} M_{0,k}^{(N)\dagger} = \tag{5.101}$$
$$\begin{bmatrix} F^N & 0 \\ 0 & (1-F)\frac{\sin^2(2\theta)}{\sin^2(2\theta_N)} \left(F^{N-1}\cos^2(2\theta_{N-1}) + (1 - F + (2F-1)\cos^2(2\theta))^{N-1} \right) \end{bmatrix}.$$

In terms of this quantity, all the relevant quantitites - $\mathrm{P}(\rho_0^N)$, $\mathrm{P}(\mathrm{succ}|\rho_0^N)$, $\mathrm{P}(\mathrm{succ})$ - can be found directly. These are:

$$\begin{aligned} &\mathrm{P}(\rho_0^N) = \mathrm{Tr}(M_{0,k}^{(N)} M_{0,k}^{(N-1)} \cdots \rho^{(1)} \cdots M_{0,k}^{(N-1)\dagger} M_{0,k}^{(N)\dagger}) \\ &\mathrm{P}(\mathrm{succ}|\rho_0^N) = \langle \psi_+^{(N)} | \rho_{0,k}^{(N)} | \psi_+^{(N)} \rangle \\ &\mathrm{P}(\mathrm{succ}) = \mathrm{P}(\mathrm{succ}|\rho_0^N)\mathrm{P}(\rho_0^N) = \\ &= \langle \psi_+^{(N)} | M_{0,k}^{(N)} M_{0,k}^{(N-1)} \cdots \rho^{(1)} \cdots M_{0,k}^{(N-1)\dagger} M_{0,k}^{(N)\dagger} | \psi_+^{(N)} \rangle. \end{aligned} \tag{5.102}$$

I now evaluate the relevant probabilities. To make the discussion concrete, and in keeping with the above numerical calculations, I use the case $F = 0.99, \theta = \pi/6$, however the broad picture (i.e., how the different quantities play off against each other) holds in general. I showed above that one needs to interact successively with three qubits in order to get within .01 of the asymptotic limit, which suggests that one might want to use a modified scheme of stopping the process once three successive interactions have been performed without failing. In this case, if the scheme is run without checking the sample qubits the asymptotic value is $\lim_{N\to\infty} \mathrm{P}_N^{QDG} \approx 0.9868$. This can be constrasted with the probability that the probe is measured successfully, conditioned upon three successful probe-sample interactions, for which one finds $\mathrm{P}(\mathrm{succ}|\rho_0^N) = 0.9948$ through numerical evaluation, which is a clear improvement. However, this is not the whole picture as one also needs to know the probability that the probe is left in the state ρ_0^N, which turns out to be $\mathrm{P}(\rho_0^N) = 0.9713$. In this modification of the scheme, the overall success-rate is $\mathrm{P}(\mathrm{succ}) = 0.9713 \times 0.9948 = 0.9662$, so that inspecting the qubits has made the protocol perform worse. There are some caveats to the conclusion the modified scheme is worse. One is that success of the scheme is now signalled. As in unambiguous state discrimination, a minimum error overall is sacrificed in order to improve the probability of correct identification in a subset of cases. For some experiments, it may be that this scheme provides a more useful characterisation of the state. Another point is that the overall probability of success depends upon how large the reserve of sample qubits is. In many applications one will have access to many more

than the three qubits required here. The question then becomes: what is the probability of getting three, for example, good outcomes in a row across the total measurement record? The combinatorics involved in this case are too complicated to generate analytic results but it seems likely that taking this account gives favour to the modified scheme, especially as one stops gaining much information after the third or fourth measurement in the standard case.

Finally, there is a subtle point concerning the asymptotic behaviour. Eq. 5.32 provides an example of non-commuting limits for the two cases $F = 1$ and $\theta = 0$. If the limit is first taken to $F = 1$, the fraction term of this formula becomes equal to zero and thus the probability of success becomes equal to unity, regardless of the angle between the two states. This is, of course, the expected behaviour for almost all possible states, apart from one point: that $\theta = 0$, i.e., that the two states become identical. In that case the best guess at the transmitted state will always have a fifty-percent probability of success, however many times it is measured. That result is seen when the many-copy limit is evaluated at $\theta = 0$ directly: irrespective of F, the fraction becomes $1/2$ and hence so does the overall probability of success. The order in which those two cases is evaluated must be done angle-first and this is an example of non-commuting limits. However, for all other cases this issue does not exist. The reason for this odd behaviour is that the original measurements are ill-defined in such a limit. In the quantum-data-gathering protocol this is because the two-dimensional subspace which is occupied by the product states $|\psi\rangle^{\otimes N}$, which information is copied onto the two dimensions of the probe qubit, becomes a one-dimensional subspace. As long as one is careful about this limit, though, the issue can be avoided.

5.5 Summary

In this chapter, I have analysed the noise-resilience of two multiple-copy state-discrimination schemes. Local adaptive measurement is an individual scheme, in which each qubit is measured. The quantum data gathering scheme interacts all qubits coherently with a probe and extracts a single measurement datum. If the qubits are prepared perfectly then both schemes reach the Helstrom bound. I set out to find out if there is a difference in the response of each protocol if the measured qubits are imperfectly prepared. This lead me to two surprising conclusions. Firstly, I showed that the local scheme *always* outperforms the collective scheme, a result which goes against the commonly held notion that a quantum memory is always useful [85, 93].

The other unexpected result is that both probabilities converge upon the same asymptotic limit. This is rather tantalising, as it suggests that the systematic error (applying the best measurement for $F = 1$ generally) has some generic features. An obvious starting point for future work on multiple-copy state discrimination is to explore this behaviour. Another natural extension of the work presented here is to derive analytic results for true Bayesian updating for the local adaptive scheme applied to the noisy states, one that uses the whole measurement record rather than the Markovian scheme relevant for the case of pure states. Though most analyses of mixed-state discrimination are numerical, it is likely

that this simplified case is tractable. In general, more analytic results in the area would help. Investigating Bayesian updating for a limited set of mixed states, and understanding when it is no longer optimal, would be a good starting place.

Chapter 6

Conclusion

A quantum measurement will change the state of the measured system. In this thesis, I have explored a few implications (some rather abstract, others more practical) of this statement.

Kraus gave us a calculational framework for sequences of measurements. Why is this framework the only game in town? This is the first question that I answer in this thesis. Following work by Hardy, Busch, Gleason and others, my answer was that Kraus's probability rule is the unique map from pairs of positive operators to real numbers which is consistent with some (hopefully reasonable sounding) propositions about the nature of quantum measurement. My propositions are based on a counting argument, with which I contest that probability measurements must be thought of as relative frequencies of different outcomes. The mathematical argument uses a trick common to many works in the field of quantum reconstructions: I first show that probabilities can be represented as inner products, and then find the space on which these inner products take place. This space turns out to be the space of two-time states. I find the joint probability rule and from it the state-update rule follows.

Quantum reconstructions continues to draw attention for physicists with foundational questions. Between my initial work and the writing of this thesis, the most significant contribution has come from Masanes *et al.* who provide a new derivation of the Born rule [37]. Essentially, it is a statement that the Born rule is the only associative map from rays to real numbers. On the face of it, this is in keeping with many of the works that I've discussed throughout. However, unlike that work, the more recent results uses the Schrödinger evolution of the quantum state as part of the proof. An obvious implication is that unitary and non-unitary quantum evolutions can be considered two sides of some, deeper, process. However, the mechanism by which this occurs has not yet been mapped out. Also, like many works in the field (a categeory in which I include my own work) they *assume* that a measurement is a special type of quantum evolution which maps a state onto a real number, rather than deriving this behaviour from an underlying framework. My feeling is that, going forward, quantum reconstructions need to provide sets of axioms which are unrelated to measurements and yet which imply both Schrödinger's and Lüders' equations for updating quantum states. This would help to clarify the measurement problem in more depth but, for that reason, is obviously a difficult task.

I followed this foundational discussion with an eavesdropping analysis for some quantum key distribution protocols. The former work led me to see that the natural space to represent sequences of measurement outcomes is the two-time state space. I wanted a physical problem which used the relevant objects in this space and noted that they map onto the actions of the different parties that take part in quantum key distribution; the problem that I chose was to optimise eavesdropping attacks in the individual measurement regime. BB84 and B92, the two protocols which are most often discussed, were re-analysed alongside PBC00, which is not as well-explored. I found that BB84 is best attacked by the Brandt-Peres-Fuchs attack; that B92 is best attack with unambiguous state discrimination; and that PBC00 is weak against qubit-rotation attacks. The final of these is particularly surprising as it means that an eavesdropper does not need to know which state is sent in order to known which bit was set. This highlights that the map between signal states and bits is key to the security of quantum key distribution routines.

If this work has further applications, they are likely to be in the field of device-independent key distribution, which I showed has a similar probability rule. It may be that general security proofs can be developed for these schemes using the tools that I developed. This point should be explored further.

The third strand of work presented here is an analysis of multiply-copy state discrimination in the noisy preparation regime. A variety of protocols are able to reach the Helstrom bound for situations in which one needs to distinguish between two pure states and has access to multiple copies of those states but it is not obvious how well those protocols perform if those pure states are replaced with mixed states. I showed that the Bayesian local adaptive measurement scheme is better than the collective quantum data gathering at distinguishing between two pure states if they have been prepared imperfectly, and also showed that both schemes have the same many-copy limit of their success probability.

While the two-pure-state regime of multiple-copy state discrimination is more-or-less solved, all other possibilities are almost unexplored, and almost no analytic results have been found. While this is mostly due to the difficulty of calculation in this area, some work should be possible. I expect that, for the range of mixed states which I look at in the work presented here, i.e., those of the form $F|\psi_i\rangle\langle\psi_i| + (1-F)|\psi_{i\perp}\rangle\langle\psi_{i\perp}|$ with $i = 0, 1$, the optimal measurement scheme should be tractable. A natural starting point would be an analysis of the true Bayesian updating scheme, one in which each measurement uses the entire measurement record, for this set of states. It should be possible to find the probability of success for this scheme, and even to verify whether or not it is optimal. Even if it is sub-optimal, such an analysis might provide some intuition for some of the odder results which appear in multiple-copy mixed-state discrimination. A move into the problem of discriminating three-or-more states is probably more difficult and will likely have to wait until experimental accuracy has improved.

Appendix A

Singular-value decomposition

The Schmidt decomposition of a bipartite state is a powerful tool in quantum information as it can be used to diagnose entanglement in a bipartite state. The possibility of performing the Schmidt decomposition follows from the singular value decomposition of a matrix. In this appendix I introduce the concept of singular value decompositions, prove that they will always be possible, and introduce a worked example.

The basic theorem of singular value decompositions is that there will always exist a diagonal matrix Σ and unitary matrices U and V such that *any* matrix A can be decomposed into the form

$$A = U\Sigma V^\dagger. \tag{A.1}$$

This construct is used in a number of places in this thesis in its application in the Schmidt decomposition, which can be used to diagnose entanglement and which forms the basis of the method of quantum data gathering in Chapter 5. In quantum mechanics one is always concerned with square matrices however this theorem holds more generally.

I begin by demonstrating that it is always possible to decompose A in this manner. The proof uses the object $A^\dagger A$, which is clearly positive semi-definite (as is $AA^\dagger$). For this reason it has a unique spectral decomposition in terms of eigenvectors λ_i and eigenvectors $|\lambda_i\rangle$, such that

$$A^\dagger A|\lambda_i\rangle = \lambda_i|\lambda_i\rangle. \tag{A.2}$$

I act on each side of this equation with A to produce

$$AA^\dagger A|\lambda_i\rangle = \lambda_i A|\lambda_i\rangle. \tag{A.3}$$

This demonstrates that $A|\lambda_i\rangle$ is an eigenvalue of the positive operator $AA^\dagger$. Thus, the set of objects $|\psi_i\rangle = A|\lambda_i\rangle/\sqrt{\lambda_i}$ form a normalised, orthogonal basis for the space. I evaluate

$$\langle\psi_i|A|\lambda_j\rangle = \sqrt{\lambda_i}\delta_{ij} \tag{A.4}$$

which verifies that the operator can be 'diagonalised' as a map between the two bases I am using. The final step is to write this operator in a particular basis, which I label $\{|i\rangle\}$.

I define the unitaries U and V such that $U|i\rangle = |\psi_i\rangle$ and $V|i\rangle = |\lambda_i\rangle$. Then, from Eq. A.4,

$$\langle i|U^\dagger AV|j\rangle = \sqrt{\lambda_i}\delta_{ij} \tag{A.5}$$

and it is clear that $\Sigma = U^\dagger AV$ is a positive semi-definite operator. This can be rearranged for $A = U\Sigma V^\dagger$. This is the main content of the singular value decomposition theorem.

An example will make this concept clearer. I consider the matrix

$$A = \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix}. \tag{A.6}$$

According to the proof of the theorem above, the object that I need to evaluate is $A^\dagger A$. The eigenvectors of this matrix form the elements of the unitary operator V. I find

$$A^\dagger A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix} \tag{A.7}$$

which has normalised eigenvectors $v_1 = [1/\sqrt{2}, -1/\sqrt{2}]^T$ and $v_2 = [1/\sqrt{2}, 1/\sqrt{2}]^T$ so that

$$V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}. \tag{A.8}$$

I can use this to find the elements of Σ and U given that I know $AV = U\Sigma$. The left-hand side of this equation is

$$AV = \begin{bmatrix} 0 & 2\sqrt{2} \\ -\sqrt{2} & 0 \end{bmatrix}. \tag{A.9}$$

The unitary operation U can be written so that its elements are two normalised column vectors u_1, u_2 and the operator Σ is diagonal, with upper-left element σ_1 and lower-right element σ_2. The product of these two matrices is $U\Sigma = [\sigma_1 u_1, \sigma_2 u_2]$ Hence, from the above I have

$$\sigma_1 u_1 = \begin{bmatrix} 0 \\ -\sqrt{2} \end{bmatrix}. \tag{A.10}$$

Which implies that $\sigma_1 = \sqrt{2}$ and $u_1 = [0, -1]^T$. Similarly, evaluating the other piece gives $\sigma_2 = 2\sqrt{2}$ and $u_2 = [1, 0]^T$. Bringing everything together I have

$$\Sigma = \begin{bmatrix} \sqrt{2} & 0 \\ 0 & 2\sqrt{2} \end{bmatrix}, \; U = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \; V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}. \tag{A.11}$$

Which can be seen to satisfy $A = U\Sigma V^\dagger$ upon evaluation. U and V are two unitaries and Σ is a diagonal matrix. Writing A in this form is the singular value decomposition.

Appendix B

Local-adaptive measurement

In this appendix I derive the best measurement strategy for the local-adaptive measurement scheme. The measurement at each stage is written as

$$|\omega(i_N x_{N-1})\rangle = \cos(\phi_x – i_N \frac{\pi}{2})|0\rangle + \sin(\phi_x – i_N \frac{\pi}{2})|1\rangle, \tag{B.1}$$

in which i is the individual measurement outcome at that stage; x_N is the measurement record up to that point; and ϕ_x is a parameter derived from that measurement record. This notation is all introduced in the main body of the text, following Eq. 5.9. The free parameter, ϕ_x, can be varied to reach the highest possible probability of success and the result, as mentioned in the text, is that this satisfies

$$\cos(2\phi_x) = (-1)^{i_{N-1}} \cos(2\theta) \sqrt{\frac{1–4p_0p_1\cos^{2N-2}(2\theta)}{1–4p_0p_1\cos^{2N}(2\theta)}}. \tag{B.2}$$

The surprising point is that the only piece of the measurement record which appears in this formula is i_{N-1}. This is the result directly prior to the measurement, and so the scheme is said to be Markovian. This is the behaviour that I derive here.

The probability that the scheme succeeds is given by

$$\mathrm{P}_N^{ad} = \sum_x \left(p_0 \mathrm{P}(0|x,0)\mathrm{P}(x|0) + p_1 \mathrm{P}(1|x,1)\mathrm{P}(x|1)\right). \tag{B.3}$$

The pieces $\mathrm{P}(a|x,a)$ can be evaluated in terms of the two angles ϕ_x and θ and substituted into this equation. The first part of this requires the formula

$$\mathrm{P}(i_N|x,a) = \frac{1}{2}\left(1 + (-1)^{i_N}\cos(2\theta)\cos(2\phi_x) + (-1)^{i_N+a}\sin(2\theta)\sin(2\phi_x)\right), \tag{B.4}$$

for the probability of the outcome i_N on the final qubit given that the state ψ_a was sent and that the first $(N-1)$ outcomes have resulted in a bit string x, which formula I use repeatedly throughout this derivation. Substitution of this into the above probability

formula gives

$$\mathrm{P}_N^{ad} = \frac{1}{2}\sum_x \left(p_0\mathrm{P}(x|0)\left(1+\cos(2\theta)\cos(2\phi_x)+\sin(2\theta)\sin(2\phi_x)\right)\right.$$
$$\left. +p_1\mathrm{P}(x|1)\left(1-\cos(2\theta)\cos(2\phi_x)+\sin(2\theta)\sin(2\phi_x)\right)\right), \tag{B.5}$$

The maximum point occurs when the differential of this function is equal to zero, which means that the requirement is

$$\frac{d\mathrm{P}_N^{ad}}{d\phi_x} = \sum_x \left(\left(p_0\mathrm{P}(x|0)+p_1\mathrm{P}(x|1)\right)\sin(2\theta)\cos(2\phi_x)\right.$$
$$\left. -\left(p_0\mathrm{P}(x|0)-p_1\mathrm{P}(x|1)\right)\cos(2\theta)\sin(2\phi_x)\right) = 0. \tag{B.6}$$

In the local-adaptive scheme, there is no distinction between local and global optimisation; it is assumed in advance that the best scheme will be locally optimal and only later shown that this also leads to the globally optimal result, the Helstrom bound. If local optimality is assumed, then the above equation must hold for all x and so each term in the sum will be equal to zero. The condition is

$$\sin(2\theta)\cos(2\phi_x)\left(p_0\mathrm{P}(x|0)+p_1\mathrm{P}(x|1)\right) = \cos(2\theta)\sin(2\phi_x)\left(p_0\mathrm{P}(x|0)–p_1\mathrm{P}(x|1)\right). \tag{B.7}$$

After a small amount of rearrangement this can be written solely in terms of $\cos(2\phi_x)$ only:

$$\cos(2\phi_x) = \frac{p_0\mathrm{P}(x|0)–p_1\mathrm{P}(x|1)}{\sqrt{(p_0\mathrm{P}(x|0)+p_1\mathrm{P}(x|1))^2–4p_0p_1\mathrm{P}(x|0)\mathrm{P}(x|1)\cos^2(2\theta)}}\cos(2\theta). \tag{B.8}$$

The next step is to simplify this object, which can be done by showing that the product $\mathrm{P}(x|0)\mathrm{P}(x|1)$ is proportional to the squared-sum $(p_0\mathrm{P}(x|0)+p_1\mathrm{P}(x|1))^2$. This is done with the usual rules of conditional probability to expand each $(N-1)$-length bit string x in terms of the final value, i_{N-1} and the previous $(N-2)$-length bit string which I label $\dot{x}$. I use again Eq. B.4, replacing x by $\dot{x}$ and i_N by i_{N-1}. I consider first the product of probabilities, which is

$$\mathrm{P}(x|0)\mathrm{P}(x|1) = \mathrm{P}(\dot{x}|0)\mathrm{P}(\dot{x}|1)\mathrm{P}(i_{N-1}|\dot{x},0)\mathrm{P}(i_{N-1}|\dot{x},1). \tag{B.9}$$

I can then evaluate the probabilities of i_{N-1} on the right hand side using the previously written formula. After several lines of manipulation, this simplifies to

$$\mathrm{P}(x|0)\mathrm{P}(x|1) = \frac{1}{4}\left(\cos(2\theta)+(-1)^{i_{N-1}}\cos(2\phi_{\dot{x}})\right)^2\mathrm{P}(\dot{x}|0)\mathrm{P}(\dot{x}|1). \tag{B.10}$$

I also require the squared-sum mentioned above. This is acquired in a similar fashion, by writing the bit-string probability in terms of its final value, substituting in for that

probability and then simplifying. The result of this process is the relation

$$\begin{aligned}(p_0\mathrm{P}(x|0)+p_1\mathrm{P}(x|1))^2 &= (p_0\mathrm{P}(i_{N-1}|\dot{x},0)\mathrm{P}(\dot{x}|0)+p_1\mathrm{P}(i_{N-1}|\dot{x},1)\mathrm{P}(\dot{x}|1))\\ &= \frac{1}{\cos^2(2\theta)}\frac{1}{4}\left(\cos(2\theta)+(-1)^{i_{N-1}}\cos(2\phi_{\dot{x}})\right)^2\\ &\times (p_0\mathrm{P}(\dot{x}|0)+p_1\mathrm{P}(\dot{x}|1))^2. \end{aligned} \tag{B.11}$$

With these two results, I find the expression

$$\cos^2(2\theta)\frac{(p_0\mathrm{P}(x|0)+p_1\mathrm{P}(x|1))^2}{\mathrm{P}(x|0)\mathrm{P}(x|1)} = \frac{(p_0\mathrm{P}(\dot{x}|0)+p_1\mathrm{P}(\dot{x}|1))^2}{\mathrm{P}(\dot{x}|0)\mathrm{P}(\dot{x}|1)}. \tag{B.12}$$

This can be used iteratively to find an expression for the left-hand-side fraction in terms of the equivalent expression for the zero-length bit strings only. In that case, one can use the single-qubit Helstrom bound to satisfy the expression and this leads to the result

$$\mathrm{P}(x|0)\mathrm{P}(x|1) = \cos^{2N-2}(2\theta)(p_0\mathrm{P}(x|0)+p_1\mathrm{P}(x|1))^2. \tag{B.13}$$

With this expression B.8 can be simplified and I am left with

$$\cos(2\phi_x) = \mathrm{sgn}\left(p_0\mathrm{P}(x|0)-p_1\mathrm{P}(x|1)\right)\cos(2\theta)\frac{1\text{–}4p_0p_1\cos^{2N-2}(2\theta)}{1\text{–}4p_0p_1\cos^{2N}(2\theta)}. \tag{B.14}$$

The final step is to introduce the Markovianity; this is done by simplifying the expression $\mathrm{sgn}\left(p_0\mathrm{P}(x|0)-p_1\mathrm{P}(x|1)\right)$. This is done by expanding the argument of sgn in terms of the final measurement result only, and then substituting the relevant expression for the probability that it has the outcome i_{N-1} in terms of θ and ϕ_x:

$$\begin{aligned} p_0\mathrm{P}(x|0)\text{–}p_1\mathrm{P}(x|1) &= p_0\mathrm{P}(i_{N-1}\dot{x}|0)\mathrm{P}(\dot{x}|0)\text{–}p_1\mathrm{P}(i_{N-1}\dot{x}|1)\mathrm{P}(\dot{x}|1)\\ &= \frac{1}{2}\left(p_0\mathrm{P}(\dot{x}|0)\text{–}p_1\mathrm{P}(\dot{x}|1)\right)\\ &+ \frac{1}{2}(-1)^{i_{N-1}}\cos(2\theta)\cos(2\phi_{\dot{x}})\left(p_0\mathrm{P}(\dot{x}|0)\text{–}p_1\mathrm{P}(\dot{x}|1)\right)\\ &+ \frac{1}{2}(-1)^{i_{N-1}}\sin(2\theta)\sin(2\phi_{\dot{x}})\left(p_0\mathrm{P}(\dot{x}|0)+p_1\mathrm{P}(\dot{x}|1)\right), \end{aligned} \tag{B.15}$$

At this point I substitute in the sine and cosine of the parameter $\phi_{\dot{x}}$ in place of the expressions for probability, using Eq. B.8. After a small amount of manipulation I arrive at

$$p_0\mathrm{P}(x|0)\text{–}p_1\mathrm{P}(x|1) = R(\dot{x})\left(\frac{\cos(2\phi_{\dot{x}})}{\cos(2\theta)}+(-1)^{i_{N-1}}\right). \tag{B.16}$$

The term $R(\dot{x})$ is introduced to simplify the expression and brings together several pieces of the equation. All that is relelvant here is that it is always positive, so the specific form does not contribute to the sign. To determine the sign of this function I need to consider only the piece inside the brackets. By definition, $-1 < \cos(2\phi_{\dot{x}})/\cos(2\theta) < 1$. This can be seen from the definition above. It follows from the fact that $\cos(2\phi_{\dot{x}})$ is the overlap of the

two most-likely states at a given point in the function. In the limit of many copies this must tend towards zero monotonically, and so it will always be smaller in magnitude than the prior overlap, $\cos(2\theta)$, though may have a different sign. For this reason, the sign of the overall object depends upon only i_{N_1}; if this is zero, then the object is positive, and if it is one, then the object is negative. Thus,

$$\mathrm{sgn}(p_0\mathrm{P}(x|0)–p_1\mathrm{P}(x|1)) = (-1)^{i_{N-1}}. \tag{B.17}$$

Bringing this together with Eq. B.14 gives

$$\cos(2\phi_x) = (-1)^{i_{N-1}}\cos(2\theta)\sqrt{\frac{1–4p_0p_1\cos^{2N-2}(2\theta)}{1–4p_0p_1\cos^{2N}(2\theta)}}, \tag{B.18}$$

which is the form of the adaptive measurement, first derived by Acin *et al*, which I require and which is used in the main body of this thesis.

Bibliography

[1] K. Flatt, S. M. Barnett, and S. Croke. Gleason-Busch theorem for sequential measurements. *Phys. Rev. A*, 96:062125, 2017.

[2] K. Flatt, S. Croke, and S. M. Barnett. Two-time state formalism for quantum eavesdropping. *Phys. Rev. A*, 98:052339, 2018.

[3] K. Flatt, S. Croke, and S. M. Barnett. Multiple-copy state discrimination of noisy qubits. In preparation.

[4] O. Freire Junior. *The Quantum Dissidents: Rebuilding the Foundations of Quantum Mechanics*. Springer, 2015.

[5] A. Peres. *Quantum theory: Concepts and Methods*. Kluwer Academic Publishers, 1998.

[6] C.J. Isham. *Lectures on Quantum Theory*. Imperial College Press, 1995.

[7] A.I.M. Rae. *Quantum Mechanics, Fourth Edition*. Taylor & Francis, 2002.

[8] J.J. Sakurai and J. Napolitano. *Modern Quantum Mechanics, Second Edition*. Cambridge University Press, 2017.

[9] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.

[10] S. M. Barnett. *Quantum Information*. Oxford University Press, Oxford, 2009.

[11] M.G.A. Paris. The modern tools of quantum mechanics. *Eur. Phys. J. ST*, 203, 2012.

[12] T. Maudlin. *Quantum Non-locality & Relativity, Third Edition*. Wiley-Blackwell, 2011.

[13] K. Kraus. *States, Effects and Operations*. Springer, 1983.

[14] S. Croke, S. M. Barnett, and S. Stenholm. Linear transformations of quantum states. *Annals of physics*, 323(4):893–906, 2008.

[15] S.M. Barnett and S. Croke. Quantum state discrimination. *Adv. Opt. Photon.*, 1:238–278, 2009.

[16] J. Bae and L-C. Kwek. Quantum state discrimination and its applications. *J. Phys. A: Math. Theor*, 48:083001, 2015.

[17] C.W. Helstrom. *Quantum detection and estimation theory.* Academic, 1976.

[18] S.M. Barnett and S. Croke. On the conditions for discrimination between quantum states with minimum error. *J. Phys. A*, 42:062001, 2009.

[19] I.D. Ivanovic. How to differentiate between non-orthogonal states. *Phys. Lett. A.*, 123:257–259, 1987.

[20] D. Dieks. Overlap and distinguishability of quantum states. *Phys. Lett. A.*, 126:303–306, 1988.

[21] A. Peres. How to differentiate between non-orthogonal states. *Phys. Lett. A*, 128:19–19, 1988.

[22] C. Jaeger and A. Shimony. Optimal distinction between two non-orthogonal quantum states. *Phys. Lett. A*, 197:83–87, 1995.

[23] A. Chefles. Unambiguous discrimination between linearly independent quantum states. *Phys. Lett. A*, pages 339–347, 1998.

[24] G Van Assche. *Quantum Cryptography and Secret Key Distillation.* Cambridge University Press, 2012.

[25] F. Grosshans, A. Acín, and N.J. Cerf. Continuous-variable quantum key distribution. In *Quantum information with continuous variables of atoms and light*, pages 63–83. Imperial College Press, 2007.

[26] F. Grosshans and P. Grangier. Continuous variable quantum cryptography using coherent states. *Phys. Rev. Lett.*, 88, 2002.

[27] S.L. Braunstein and S. Pirandola. Side-channel-free quantum key distribution. *Phys. Rev. Lett.*, 108:130502, 2012.

[28] H-K. Lo, M. Curty, and B. Qi. Measurement-device-independent quantum key distribution. *Phys. Rev. Lett.*, 108:130503, 2012.

[29] S. Pirandola et al. High-rate measurement-device-independent quantum cryptography. *Nature Photonics*, 9:397–403, 2015.

[30] M. Born. Zur quantenmechanik der stossvorgünge. *Zeitschrift fär Physik*, 37:863–867, 1926. English translation due to F. A. Wheeler and W. H. Zurek.

[31] A. M. Gleason. Measures on the closed subspaces of a Hilbert space. *Indiana Univ. Math. J.*, 6:885–893, 1957.

[32] P. Busch. Quantum states and generalized observables: A simple proof of Gleason's theorem. *Phys. Rev. Lett.*, 91:120403, 2003.

[33] L. Hardy. Quantum theory from five reasonable axioms. v4 used here: available from arXiv:quant-ph/0101012v4, 2001.

[34] S. M. Barnett, J. D. Cresser, J. Jeffers, and D. T. Pegg. Quantum probability rule: a generalization of the theorems of Gleason and Busch. *New J. Phys*, 16:043025, 2014.

[35] C. M. Caves, C. A. Fuchs, K. M. Manne, and J. M. Renes. Gleason-type derivations of the quantum probability rule for generalized measurements. *Foundations of Physics*, 34:193–209, 2004.

[36] R. Cooke, M. Keane, and W. Moran. An elementary proof of gleason's theorem. *Math. Proc. Cam. Phil. Soc.*, 98:117, 1985.

[37] L. Masanes, T.D. Galley, and M.P. Muller. The measurement postulates of quantum mechanics are redundant. *Nat. Commun.*, 10:1361, 2019.

[38] G Lüders. Concerning the state-change due to the measurement process. *Ann. Phys. (Leipzig)*, 15:663–670, 2006.

[39] J. von Neumann. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton, 1955.

[40] H. Granström. Gleason's theorem. Master's thesis, Stockholm University, 2006.

[41] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz. Time symmetry in the quantum process of measurement. *Phys. Rev.*, 134:B1410–1416, 1964.

[42] D. T. Pegg and S. M. Barnett. Retrodiction in quantum optics. *J. Opt. B: Quantum Semiclass. Opt.*, 1:442–445, 1999.

[43] A. Grinbaum. Reconstruction of quantum theory. *Brit. J. Phil. Sci.*, 58:387–408, 2007.

[44] G. Chiribella, G. M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84:012311, 2011.

[45] L. Masanes and M. P. Müller. A derivation of quantum theory from physical requirements. *New J. Phys.*, 13:063001, 2011.

[46] G. Chiribella, G. M. D'Ariano, and P. Perinotti. *Quantum theory from first principles: an Information Approch*. Cambridge University Press, 2017.

[47] G. Birkhoff and J. von Neumann. The logic of quantum mechanics. *Ann. Math.*, 37:823, 1936.

[48] G.W. Mackey. *The mathmetical foundations of quantum mechanics*. W.A. Benjamin, New York, 1963.

[49] G. Cassinelli and N Zanghi. Conditional probabilities in quantum mechanics. *Il Nuovo Cimento*, 73B:237–245, 1983.

[50] S. Shrapnel, F. Costa, and G. Milburn. Updating the Born rule. arxiv: 1702.01845v1 [quant-ph], 2017.

[51] J. Fiutak and J. Van Kranendonk. Impact theory of Raman line broadening. *Canadian Journal of Physics*, 40:1085–1100, 1962.

[52] S. M. Barnett and B. J. Dalton. Liouville space description of thermofields and their generalisations. *J. Phys. A: Math. Gen.*, 20:411–418, 1987.

[53] M-D. Choi. Completely positive linear maps on complex matrices. *Linear Algebra and its Applications*, 10(3):285, 1975.

[54] A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.*, 4:275, 1972.

[55] M. S. Leifer and R. W. Spekkens. Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference. *Phys. Rev. A*, 88:052130, 2013.

[56] C. J. Isham. *Lectures on Groups and Vector Spaces for Physicists.* World Scientific, Singapore, 1989.

[57] N. Gisin. Weinberg's non-linear quantum mechanics and supraluminal communications. *Phys. Lett. A*, 143:1, 1990.

[58] O. Oreshkov, F. Costa, and C. Brukner. Quantum correlations with no causal order. *Nat. Commun.*, 3:1092, 2012.

[59] Y. Aharonov, S. Popescu, J. Tollaksen, and L. Vaidman. Multiple-time states and multiple-time measurements in quantum mechanics. *Phys. Rev. A*, 79:052110, 2009.

[60] Y. Aharonov and L. Vaidman. Complete description of a quantum system at a given time. *J. Phys. A: Math. Gen.*, 24:2315, 1991.

[61] R. Silva, Y. Guryanova, N. Brunner, N. Linden, A. J. Short, and S. Popescu. Pre- and postselected quantum states: Density matrices, tomography, and Kraus operators. *Phys. Rev. A*, 89:012121, 2014.

[62] G. Chiribella, G.M. D'Ariano, and P. Perinotti. Theoretical framework for quantum networks. *Phys. Rev. A*, 80:022339, 2009.

[63] A. Bisio, G. Chiribella, G.M. D'Ariano, P. Perinotti, and S. Facchini. Optimal quantum learning of a unitary transformation. *Phys. Rev. A*, 81:032324, 2010.

[64] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lütkenhaus, and M. Peev. The security of practical quantum key distribution. *Rev. Mod. Phys.*, 81:1301–1350, 2009.

[65] D. Stebila, M. Mosca, and N. Lütkenhaus. The case for quantum key distribution. In A. Sergienko, S. Pascazio, and P. Villoresi, editors, *LNICST 36, Quantum Communication and Quantum Networking*, pages 283–296. Springer, Berlin, 2010.

[66] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, 175, 1984.

[67] C. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, 68:3121–3124, 1992.

[68] S. J. D. Phoenix, S. M. Barnett, and A. Chefles. Three-state quantum cryptography. *Journal of Modern Optics*, 47:507–516, 2000.

[69] Christopher A. Fuchs, Nicolas Gisin, Robert B. Griffiths, Chi-Sheng Niu, and Asher Peres. Optimal eavesdropping in quantum cryptography. i. information bound and optimal strategy. *Phys. Rev. A*, 56:1163–1172, 1997.

[70] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin. Experimental quantum cryptography. *Journal of Cryptology*, 5:3–28, 1992.

[71] T. Kim et al. Complete physical simulation of the entangling-probe attack on the Bennett-Brassard 1984 protocol. *Phys. Rev. A*, 75:042327, 2007.

[72] N. Lütkenhaus. Security against eavesdropping in quantum cryptography. *Phys. Rev. A*, 54, 1996.

[73] P.W. Shor and J. Preskill. Simple proof of the security of the bb84 quantum key distribution protocol. *Phys. Rev. Lett.*, 85, 2000.

[74] D. Gottesman and H-K. Lo. Proof of security of quantum key distribution with two-way classical communication. *IEEE Transactions on Information Theory*, 49, 2003.

[75] C.A. Fuchs and A. Peres. Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Phys. Rev. A*, 53:2038–2045, Apr 1996.

[76] H.E. Brandt. Quantum-cryptographic entangling probe. *Phys. Rev. A*, 71:042312, Apr 2005.

[77] A.K. Ekert, B. Huttner, G.M. Palma, and A. Peres. Eavesdropping on quantum-cryptographical systems. *Phys. Rev. A*, 50:1047–1052, 1994.

[78] K. Tamaki, M. Koashi, and N. Imoto. Unconditionally secure key distribution based on two nonorthogonal states. *Phys. Rev. Lett.*, 90:167904, 2003.

[79] K. Tamaki and N. Lütkenhaus. Unconditional security of the bennett 1992 quantum key-distribution protocol over a lossy and noisy channel. *Phys. Rev. A*, 69:032316, 2004.

[80] J.-C. Boileau, K. Tamaki, et al. Unconditional security of a three state quantum key distribution protocol. *Phys. Rev. Lett.*, 94:040503, 2005.

[81] J.M. Renes. Equiangular spherical codes in quantum cryptography. *Phys. Rev. A*, 70:052315, 2004.

[82] M. Schiavon, G. Vallone, and P. Villoresi. Experimental realization of equiangular three-state quantum key distribution. *Scientific Reports*, 6, 2016.

[83] A.S. Holevo. Statistical decision theory for quantum systems. *J. Multivariate Anal.*, 3:337–394, 1973.

[84] A. Peres and W.K. Wootters. Optimal detection of quantum information. *Phys. Rev. Lett.*, 66:1119, 1991.

[85] S. Massar and S. Popescu. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.*, 74:1259, 1995.

[86] B.L. Higgins, A.C. Doherty, et al. Multiple-copy state discrimination: Thinking globally, acting locally. *Phys. Rev. A*, 83:052314, 2011.

[87] S. Slussarenko, M.M. Weston, et al. Quantum state discrimination using the minimum average number of copies. *Phys. Rev. Lett.*, 118:030502, 2017.

[88] A. Acin, E. Bagan, et al. Mutliple-copy two-state discrimination with individual measurements. *Phys. Rev. A*, 71:032338, 2005.

[89] D. Brody and B. Meister. Minimum cost decision for quantum ensembles. *Phys. Rev. Lett.*, 76, 1996.

[90] M. Ban, K. Yamazaki, and O. Hirota. Accessible information in combined and sequential quantum measurements on a binary-state signal. *Phys. Rev. A*, 55:22–26, 1997.

[91] R. Blume-Kohout, S. Croke, and M. Zwolak. Quantum data gathering. *Scientific Reports*, 3, 2013.

[92] C.H. Bennett et al. Quantum nonlocality without entanglement. *Phys. Rev. A*, 59.

[93] E. Chitambar and M-H. Hsieh. Revisiting the optimal detection of quantum information. *Phys. Rev. A*, 88:020302, 2013.

[94] R. Blume-Kohout et al. Demonstration of qubit operations below a rigorous fault tolerance threshold with gate set tomography. *Nat. Commun.*, 8:14485, 2017.

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

[96] D. Aharanov and M. Ben-Or. Fault-tolerant computing with biased-noise superconducting qubits: a case study. *SIAM J. Comput.*, 38:1207–1282, 2008.

[97] D. Aharanov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the Thirteenth Annual ACM Symposium on theory of Computation*, pages 20–30. STOC, 1997.

[98] Y.R. Sanders, J.J. Wallman, and B.C. Sanders. Bounding quantum gate error rate based on reported average fidelity. *New J. Phys.*, 18:012002, 2016.

[99] G. Weir et al. Optimal measurement strategies for the trine states with arbitrary prior probabilities. *Quantum Science and Technology*, 3:035003, 2018.