Concepts and applications of quantum measurement

George C. Knee
Corpus Christi College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary Term 2014

For my parents.

Concepts and applications of quantum measurement

George C. Knee

Corpus Christi College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Hilary Term 2014

In this thesis I discuss the nature of 'measurement' in quantum theory. 'Measurement' is associated with several different processes: the gradual imprinting of information about one system onto another, which is well understood; the collapse of the wavefunction, which is ill-defined and troublesome; and finally, the means by which inferences about unknown experimental parameters are made.

I present a theoretical extension to an experimental proposal from Leggett and Garg, who suggested that the quantum-or-classical reality of a macroscopic system may be probed with successive measurements arrayed in time. The extension allows for a finite level of imperfection in the protocol, and makes use of Leggett's 'null result' measurement scheme. I present the results of an experiment conducted in Oxford that, up to certain loopholes, defies a non-quantum interpretation of the dynamics of phosphorous nuclei embedded in silicon.

I also present the theory of statistical parameter estimation, and discover that a recent trend to employ time symmetric 'postselected' measurements offers no true advantage over standard methods. The technique, known as weak-value amplification, combines a weak transfer of quantum information from system to meter with conditional data rejection, to surprising effect. The Fisher information is a powerful tool for evaluating the performance of any parameter estimation model, and it reveals the technique to be worse than ordinary, preselected only measurements. That this is true despite the presence of noise (including magnetic field fluctuations causing decoherence, poor resolution detection, and random displacements), casts serious doubt on the utility of the method.

Acknowledgements

I have had the distinct pleasure of many stimulating discussions over the last few years. In alphabetical order, I would like to thank: Konstantin Bliokh, Eliot Bolduc, Tom Close, Bob Coecke, Chris Ferrie, Amir Fruchtman, Joe O'Gorman, Kieran Higgins, John Howell, Andrew Jordan, Yaron Kedem, Ray Lal, Jonathan Leach, Tony Leggett, Daniel Loss, Owen Maroney, Gerard Milburn, Naomi Nickerson, Felix Pollock, Marcus Schaffry, Yutaka Shikano, Nic Teh, Chris Timpson, and Jamie Vicary. They all helped me to better understand, and better enjoy physics. Many of these interactions came about through a series of thirty or so interdisciplinary meetings for the *Quantum Nanoscience: Fundamental Physics, Emerging Structures and Implications for our Ultimate Reality* project.

Most of all I thank my supervisors Andrew Briggs and Simon Benjamin for their advice and guidance. I feel privileged to have worked with all of my co-authors, including Stephanie Simmons and John Morton. I am especially grateful to Erik Gauger, who has been an invaluable source of knowledge and wisdom, and to Georgia Maddocks for her unending support.

Publication list

This thesis is based on the following publications. Please click to access them online.

1. G. C. Knee, S. Simmons, E. M. Gauger, J. J. L. Morton, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, K. M. Itoh, M. L. W. Thewalt, G. A. D. Briggs, and S. C. Benjamin. Violation of a Leggett-Garg inequality with ideal non-invasive measurements. *Nature Communications*, 3:606, January 2012

2. G. C. Knee, E. M. Gauger, G. A. D. Briggs, and S. C. Benjamin. Comment on 'A scattering quantum circuit for measuring Bell's time inequality: a nuclear magnetic resonance demonstration using maximally mixed states'. *New Journal of Physics*, 14(5):058001, May 2012

3. G. C. Knee, G. A. D. Briggs, S. C. Benjamin, and E. M. Gauger. Quantum sensors based on weak-value amplification cannot overcome decoherence. *Phys. Rev. A*, 87:012115, January 2013

4. G. C. Knee and E. M. Gauger. When amplification with weak values fails to suppress technical noise. *Phys. Rev. X*, 4:011032, March 2014

Nomenclature

Symbol	Meaning
$\lvert\psi\rangle$	State vector (a pure quantum state)
ρ	Density operator / density matrix (a pure or mixed quantum state)
δ_{ij}, $\delta(x-y)$	Kronecker delta, Dirac delta
Π_i	Projector
$\mathscr{H}$	A Hilbert space
d	Hilbert space dimension
$\mathbf{H}$	Hamiltonian
t	Time
$\mathcal{H}$	Hadamard operation / gate
U	(Unitary) time evolution operator / gate
$\mathbb{I}$	Identity transformation / operation / gate
$[A, B]$	$AB - BA$ (commutator)
$\bullet$	just a placeholder
$\partial_\bullet$	Derivative with respect to $\bullet$
$\mathcal{P}$	Purity
$\mathcal{K}(\bullet)$	Kraus map
B_i	Kraus operator
E_i	Positive Operator Valued Measure (POVM) element
$\mathcal{S}$	Shannon entropy
D_{KL}, D_H, D_B	Kullback-Leibler divergence, Hellinger distance, Bures distance
$F_\bullet$	Fisher information about $\bullet$
$H_\bullet$	Quantum Fisher information about $\bullet$
$\langle\bullet\rangle$	Expectation value of $\bullet$
$\mathcal{F}^c$, $\mathcal{F}$	Classical fidelity, quantum fidelity
$\Pr(n)$ or P_i	Probability mass function
$p(x)$	Probability density function
$\bar{B}$, μ	Magnetic field, magnetic moment
i	Imaginary unit
N	Number / ensemble size
Tr	Trace
$\sigma_x, \sigma_y, \sigma_z$	Pauli operators / matrices
$\mathbb{R}$	The real numbers

Symbol	Meaning
θ and ϕ	Angles on the Bloch sphere
Q	A macroreal variable
T_1, T_2	Relaxation time, dephasing time
$\oplus$	Sum modulo 2
g	Coupling constant / interaction parameter
G	Measurement strength
K_{ij} or $K(t_i, t_j)$	Two-time correlator
A_w	Weak value
Ξ	Attenuation function
$\star$	Convolution product
$\alpha\,, \beta$	Informational loss due to pixelation, or jitter
$\Delta_\bullet$	Gaussian spread of the wavefunction in $\bullet$ space.
$J_\bullet$	Gaussian spread of the noise in $\bullet$ space.
R	Resolution
x, k	Real space, momentum space
ζ	Venality
$\lvert m_{\mathrm{WVA}}\rangle$	Weak-value 'amplified' meter state
q	Probability of successful postselection
f_{LG}, f_{Bell}, f_{KS}	Leggett-Garg, Bell, Kochen-Specker functions

Acronym	Meaning
NMR	Nuclear magnetic resonance
ESR / EPR	Electron spin resonance / Electron paramagnetic resonance
WVA	Weak-value 'amplification'
LG(I)	Leggett and Garg (inequality)
AAV	Aharonov, Albert, and Vaidman
ABL	Aharonov, Bergmann and Lebowitz
SOS	Souza, Oliveira, and Sarthour
MR, NIM	Macrorealism, non-invasive measurability
SNR	Signal-to-noise ratio
QND	Quantum Non-Demolition
CNOT	Controlled-NOT operation / gate

Contents

Chapter 1

Introduction

The scientific method concerns a continued effort to spot regularities in nature, in order that future observations might be better predicted. If phenomena are to be understood through natural laws, where general principles apply universally, every datum from the laboratory must be accounted for by a theoretical model. By continued observation, the corpus of scientific testimonies is expanded, and the incumbent theories are accordingly re-examined. More often than not, experiments confirm or corroborate; but less frequently the results trigger a more profound re-evaluation of empirical laws. Invariably, results do not match with prediction; but the discrepancy can often be attributed to experimental imperfection, or at worst trigger a modification of 'auxiliary assumptions' rather than the physical theory at hand [203]. Very rarely, the scientific collective choose not to save the theory in this way, but instead opt to discard it in favour of a completely new one.

Quantum theory, along with the theory of relativity, constitutes one of the major advances of 20th century physics. The prevailing view of science is to explain from the 'bottom' upwards: if biology is built upon chemistry, and chemistry in turn upon physics, then the a revolution in physics is a revolution for all of science. The *quantum* revolution was prompted by experimental data of the more disruptive kind; a tumult of anomalous experiments demanded a fresh explanation of the microscopic world. The theory, when it came, was profound in its departure from established concepts of classical physics – which had reigned supreme since the time of Newton. It continues to exhibit remarkable predictive power – not only subjugating the challenges presented at the turn of the previous century, but also surviving extensive tests and trials for the last one hundred years.

Quantum theory was founded to describe atoms and photons: the quintessential quantum systems. In 1913, Niels Bohr suggested that electrons have discrete, or

quantized, energy levels. Since then, 'quantum system' has come to apply to a much larger class of object; including molecules, artificial atoms and nanoscale mechanical systems. In recent decades, applications of quantum physics have been proposed which exploit its most mysterious qualities. Experimentalists around the world have achieved exquisite control of quantum systems, bringing many gedanken-experiments to life. Some of these stellar achievements were recently recognised in the 2012 Nobel Prize for physics [175]. Despite being more than satisfactory in this sense, quantum theory is bedevilled by interpretational dissonance. It is difficult to interpret and difficult to understand in a way that neither classical mechanics nor relativity theory seem to be. More profoundly, there is a singular shortcoming in the formalism of quantum theory that has resisted almost a century of efforts to improve upon it.

A complete history of quantum physics would include accounts of the crises that preceded it (for example: the lack of a consistent model of the atom, and the unexpected behaviour observed in the photoelectric effect), and also accounts of how the theory was placed on firm mathematical foundations (through the work of Schrödinger, Heisenberg, Dirac, von Neumann, and others). Such accounts may be found elsewhere [12]. Instead, in this introductory chapter I shall attempt to introduce only the important concepts necessary for an understanding of this thesis. The main body of the thesis explores the notion of 'measurement' in quantum theory. The objects of my study shall be some rather more modern experiments and theoretical investigations (some of which I have devised, along with my collaborators): studies that raise the same sort of interpretational and foundational questions as the famous works of generations ago, as well as new questions relating to a prophesied age of quantum technology.

Questions of the first kind evoke the possibility of *post-quantum* theories. One can well imagine that given the somewhat paradoxical and surprising nature of quantum theory, a successor was sought not long after its inception. In fact many factors contribute to the pursuit of this (still elusive) goal: in particular the unification of general relativity with quantum mechanics has yet to be achieved, and is strongly suspected to require a radical revision of quantum mechanics. In this thesis we shall nonetheless concern ourselves primarily with that inadequacy which embarrasses quantum theory above all others. The inadequacy is known as 'the measurement problem': various instances of it have been hotly debated for decades, but no clear solution is yet agreed upon [176].

Questions of the second kind revolve around quantum information science: a relatively young field originating in the 1980s. Its most famous proposition is the

universal quantum computer: a machine that generalises the Turing machine (or universal classical computer) by allowing for information to be processed with a new kind of logic. This allows calculations to be performed in new ways which are known to bring huge advantages in efficiency. This technology, and others like it, hold great promise for the application of quantum theory.

These questions ought to be considered together. One should not consider the foundations of a theory without recourse to its application; one cannot hope to realise the most daring of technologies without a clear understanding of the fundamentals. Whether a given experiment or protocol is exhibiting exclusively *quantum* behaviour, rather than the more mundane *classical* behaviour is a quandary that unifies the two research programmes. "This is genuinely quantum" or "that is purely classical" are alacritous decrees; but both in the foundations of quantum theory and in the motivation of quantum technologies one must take great care to fix the exact meanings of such utterances. A simulation of an experiment on a classical computer is an ever-present counter for any claim for 'quantumness'. If a simulation is physically impossible, or hopelessly costly, this is a clue that something special is occurring in the experiment.

This thesis is organised as follows. In the remainder of the introductory chapters, I introduce quantum theory, classical and quantum information theory (Chapter 2), and quantum measurement theory (Chapter 3) – taking care to give full details of the mathematics that will be required for an understanding of this thesis, but no more. I shall, from time to time, try to give pointers toward interpretational issues, many of which will crop up in the succeeding chapters. In Chapter 4 I introduce macrorealism: a class of alternatives to quantum theory defined by Leggett and Garg. I go on to generalise their theory and present experimental data falsifying macrorealism in an ensemble of phosphorus nuclei. I discuss how the results should constrain future theories of physics. In Chapter 5 I introduce the time symmetric re-formulation of quantum theory, and the related idea of weak-value amplification. In Chapters 6 and 7 I show how, in a wide range of experiments, the technological advantages one might expect to arise from the approach are non existent. A final Chapter concludes the thesis.

1.1 Quantum theory

I begin with a presentation of the Copenhagen interpretation, the original, most widespread and popular formulation of quantum theory. Other interpretations are

available. The alternatives can appear, on the face of it, to be radically different to the theory I will set out below. There are significant metaphysical differences. What exists? What is merely a calculation tool? It is my intention in this thesis, as far as is possible, to isolate the special features common to all interpretations of quantum theory, irrespective of their particular ontologies.

Copenhagen quantum theory can be thought of, in essence, as the combination of *the superposition principle*, *quantum dynamics* and *wavefunction collapse*.

1.1.1 The superposition principle

When an object is inspected, it is found in one of a number of mutually exclusive possibilities. I refer to each of these possibilities as a 'classical configuration', or a 'classical state'. The superposition principle states that a system generally inhabits *all* of its classical configurations simultaneously. To avoid making ontological commitments, it is better to say that for each possible configuration that a system may be found in, we must keep track of a certain complex number associated with it in order to correctly predict which configuration the system will be found in at a later time. 'Correctly predict' should be taken in the statistical sense: we require only the probability that the system is found in any particular configuration to be forecast. It is simplest to first think of a system with a finite list of classical configurations.

Imagine, for example, a die. The six classical configurations of the die are labelled by its uppermost face. An elegant notation from Dirac denotes each possibility with a 'ket' $|\bullet\rangle$. For the die we then have $|1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle$. Nothing so far has been exceptional. Application of the quantum superposition principle defines the quantum state of the system, also written as a ket

$$|\psi\rangle = \sum_i c_i |i\rangle , \tag{1.1}$$

with i an index that runs over the classical possibilities. The c_i are complex numbers or amplitudes, satisfying the *normalisation condition* $\sum_i |c_i|^2 = 1$. We shall soon see that this condition is really a statement about the conservation of total probability. Note that, writing the coefficients in polar form

$$c_i = r_i \mathrm{e}^{i\phi_i}, \tag{1.2}$$

the magnitudes of the complex coefficients multiplying each ket can be thought of as 'weights'. The quantum state is smeared out over the classical configurations, but may be more or less concentrated on one or another of them: the concentration being

given by the relative weights on each ket. The reader will appreciate that there are an infinitude of possibilities for $|\psi\rangle$: we might think of them as *quantum* configurations, which are to be sharply distinguished from the classical configurations (at this point being finite in number). Further, identify states which differ only by a global phase

$$\mathrm{e}^{i\phi}\,|\psi\rangle \sim |\psi\rangle \tag{1.3}$$

for any $\phi \in [0, 2\pi]$. We shall soon see that such states are physically indistinguishable.

Returning to our example of the die, we can think of the six possibilities (each labelled with its own ket) as a *basis* for the system: in other words, each is an axis in a six dimensional vector space. One equips the vector space with an inner product: a way of combining kets to give a scalar. We write

$$\langle i|j\rangle = \delta_{ij} \tag{1.4}$$

to express the *orthonormality* of the basis: the Kronecker delta has its usual meaning. The complex vector space is now formally referred to as a Hilbert space $\mathscr{H}$. For a system having d classical states, the dimension of the space is d. I have written $\langle i|$ as the adjoint $|i\rangle^\dagger$: it is the 'bra' to accompany Dirac's ket. One can also write the quantum state (1.1) in a different basis

$$|\psi\rangle = \sum_i c'_i\,|i'\rangle\,. \tag{1.5}$$

'Expanding' the state in this way is equally valid: if the $|i'\rangle$ span $\mathscr{H}$, one can expand any quantum state by a linear combination of them. If the $|i'\rangle$ are orthonormal, they are another basis for the Hilbert space. The freedom of expression here is analogous to giving the state of a classical particle (for example) with reference to different coordinate systems: as long as the transformation between coordinate systems is well defined

$$|i'\rangle = V\,|i\rangle \tag{1.6}$$

we can use either to make predictions. The necessary transformations between the basis elements for our quantum state should be linear and should preserve orthonormality, and hence must be unitary transformations

$$VV^\dagger = \mathbb{I}; \tag{1.7}$$

the product of the transformation with its adjoint must be the identity transformation. Such transformations form the group $U(n)$, and may be represented with complex

$n \times n$ matrices. The group of transformations can be thought of either as relating different bases (or coordinate systems), or as transformations of the quantum state itself in a fixed basis. To be concrete, for our die we can define (for example) the new orthonormal basis $(|1\rangle + |2\rangle)/\sqrt{2}, (|1\rangle - |2\rangle)/\sqrt{2}, |3\rangle, |4\rangle, |5\rangle, |6\rangle$. It is no longer clear how to interpret the first two basis elements. Although no particular basis is privileged mathematically, a particular basis may emerge when motivated with physical arguments – this preference is dictated by the manner in which the object is 'naturally' inspected.

The above example of a change of basis can be thought of as a rotation in a two dimensional subspace of the die. In subsequent chapters of this thesis I shall make heavy use of such subspaces, referred to as quantum bits or 'qubits'. We shall not yet concern ourselves with the physical realisation of the qubit (these are discussed in Chapter 2). It suffices at this point to understand that there are many such realisations, each with certain advantages and disadvantages for technology. Assuming only that there are two mutually exclusive possibilities, for now we imagine that the two possibilities are (arbitrarily, here) labelled 'up' or 'down', so that a general state is written

$$|\psi\rangle = c_\uparrow |\uparrow\rangle + c_\downarrow |\downarrow\rangle . \tag{1.8}$$

It is clear that the following parameterization automatically satisfies the normalisation condition and sacrifices no generality

$$\begin{aligned} c_\uparrow &= \cos\frac{\theta}{2} \\ c_\downarrow &= \mathrm{e}^{i\phi} \sin\frac{\theta}{2}. \end{aligned} \tag{1.9}$$

I have used the freedom of multiplying by an unobservable global phase to set one of the expansion coefficients real. Then by varying θ and ϕ we can reach any $|\psi\rangle$: the two angles cover the entire two dimensional Hilbert space. This oft used trick enables the visualisation of quantum states on the surface of a unit sphere: see Figure 1.1. For any $|\psi\rangle$, there is a $|\psi^\perp\rangle$ at the antipodal point, and $\langle\psi^\perp|\psi\rangle = 0$.

Qubits are the smallest non-trivial quantum system. Now consider the opposite limit, when the number of possibilities is uncountably infinite: for example the position of a particle confined to one dimension. When there are a continuum of possibilities for the quantum state, we require a continuum of basis elements. Clearly the kets cannot be enumerated or even labelled with an integer. Instead we may write $|x\rangle$, and think of x as a continuous index taking any real value. We now replace the

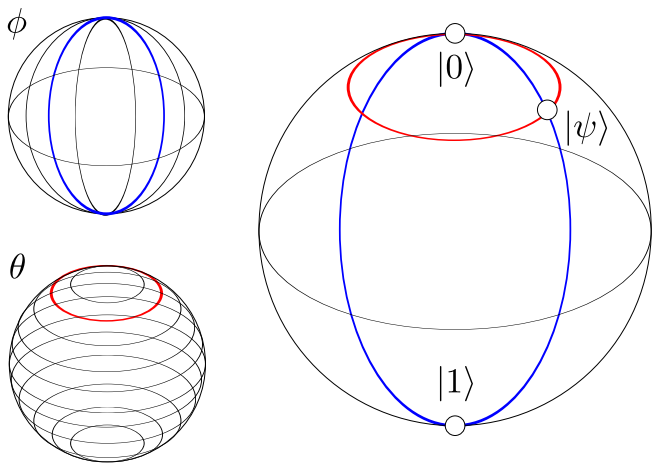


Figure 1.1: The Bloch sphere representation of two dimensional quantum states. Two angles – equivalent to a longitude and latitude coordinate – determine any state $|\psi\rangle$ uniquely. The two classical states, at the North and South pole, are orthogonal in Hilbert space. They are labelled $|0\rangle$ and $|1\rangle$ or $|\uparrow\rangle$ and $|\downarrow\rangle$.

sum in (1.1) with an integral

$$|\psi\rangle = \int c(x)dx\,|x\rangle\,. \tag{1.10}$$

Here $c(x)$ is called the *wavefunction*, and $c(x)dx$ is the continuous version of the expansion coefficients c_i. Orthonormality of the basis elements is now expressed with a Dirac delta function

$$\langle x|y\rangle = \delta(x-y). \tag{1.11}$$

1.1.2 Quantum dynamics

The quantum state of a system is driven from its initial quantum configuration in Hilbert space to some other configuration, according to the Schrödinger equation

$$i\hbar\partial_t\,|\psi\rangle = \mathbf{H}(t)\,|\psi\rangle\,. \tag{1.12}$$

Here ∂_t is the partial time derivative and $\hbar$ is a fundamental constant with units of J s, which sets the scale of quantum effects. For simplicity I shall fix $\hbar = 1$. The Hamiltonian $\mathbf{H}$ is an operator describing the energy of the system. The solution to (1.12) is given by the time-evolution operator:

$$|\psi(0)\rangle \to |\psi(\tau)\rangle = U(\tau)\,|\psi(0)\rangle \tag{1.13}$$

$$U(\tau) = \exp\left(-iT\int_0^\tau \mathbf{H}(t)dt\right). \tag{1.14}$$

The time ordering operator T can be safely ignored in this thesis: for a explanation see, e.g. [26]. In fact the time dependence of $\mathbf{H}$ will always be taken to be impulsive or step-function-like, so that time evolutions can be considered piece-wise with a series of constant Hamiltonians [111]. Thus

$$U(\tau) = \mathrm{e}^{-i\mathbf{H}\tau}. \tag{1.15}$$

By making a suitable unitary transformation, the Hamiltonian may always be diagonalised, and hence written

$$\mathbf{H} = \sum_i \lambda_i \left|e_i\right\rangle \left\langle e_i\right| : \tag{1.16}$$

the $|e_i\rangle$ form a basis (the energy eigenbasis), and the λ_i are eigenvalues (eigenenergies) which are real if $\mathbf{H}^\dagger = \mathbf{H}$ (which we require). Because the Hamiltonian is Hermitian, the time evolution operator $U(\tau)$ is unitary. Finding the unitary similarity transformation which brings the Hamiltonian into diagonal form is generally a challenging task. Given that it is always possible, however, we can plainly see that the time dynamics of the eigenstates is trivial

$$\left|e_i\right\rangle \to \exp(-i\lambda_i\tau)\left|e_i\right\rangle ; \tag{1.17}$$

they merely acquire a phase at a rate given by the relevant eigenenergy. They are therefore called the stationary states. Further, adding to or subtracting from $\mathbf{H}$ an operator proportional to the identity matrix, which is equivalent to a redefinition of the zero energy, will only amount to a change in the global phase of any quantum state. It is therefore unobservable. By convention the Hamiltonian is often taken to be traceless: the unitary operations then have unit determinant and form the subgroup SU(n).

For a two level system, the stationary states must be *antipodal* points on the Bloch sphere, because they are mutually orthogonal under the Hilbert space inner product. For $d = 2$ the Hamiltonian can always be expressed:

$$\mathbf{H} = H_x\sigma_x + H_y\sigma_y + H_z\sigma_z \tag{1.18}$$

for real parameters H_x, H_y, H_z and

$$\begin{aligned}
\sigma_x &= \left|\uparrow\right\rangle\left\langle\downarrow\right| + \left|\downarrow\right\rangle\left\langle\uparrow\right| \\
\sigma_y &= -i\left|\uparrow\right\rangle\left\langle\downarrow\right| + i\left|\uparrow\right\rangle\left\langle\downarrow\right| \\
\sigma_z &= \left|\uparrow\right\rangle\left\langle\uparrow\right| - \left|\downarrow\right\rangle\left\langle\downarrow\right|
\end{aligned} \tag{1.19}$$

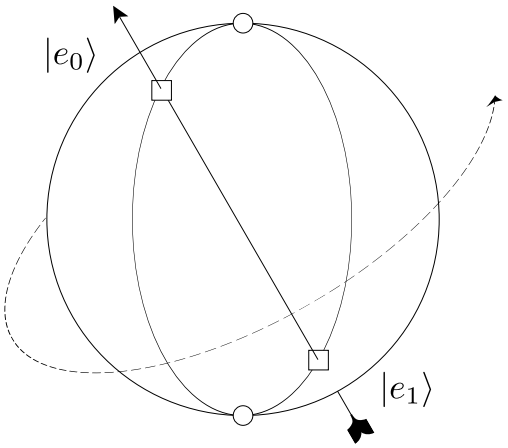


Figure 1.2: Dynamics on the Bloch sphere. The Hamiltonian defines an axis about which all states rotate over time. The energy eigenstates are stationary and shown with squares.

are called Pauli matrices. The ± 1 eigenstates of these operators define six important states in the Bloch sphere (both poles and four equally spaced states on the equator)[1]. It is left as an exercise to the reader to show that a general state, being a superposition of the stationary states, rotates around the axis defined by the Hamiltonian (see Figure 1.2).

It is often very useful to employ a representation of the above algebra. Quantum states in Hilbert space dimension d become complex vectors in $\mathbb{C}^d$, and operators become $d \times d$ complex matrices.

1.1.3 The collapse postulate

We have seen how, according to quantum theory, systems are described by a mathematical object (the quantum state) which seems as if smeared out over the classical configurations of the system. We remain, for now, agnostic about the metaphysical status of the quantum state: one might venture to think of it either as a state of reality (taking quantum theory at face-value) or as a state of knowledge (because of the formal similarities of quantum states to probability distributions). The first position states that the wavefunction is *ontic*, the second that it is *epistemic*.

[1]The Pauli matrices obey commutation relations $[\sigma_i, \sigma_j] = 2i \sum_k \epsilon_{ijk} \sigma_k$, and when multiplied by the imaginary unit (see the Schrödinger equation) compose the generators of a Lie algebra $\mathfrak{su}(2) = \text{span}\{i\sigma_x, i\sigma_y, i\sigma_z\}$. There is a surjective homomorphism from SU(2) to the three dimensional rotations in real space $SO(3)$.

We have seen how the wavefunction can change, evolving over time or with our choice of expansion basis. The truly intriguing parts of quantum theory arise because the quantum state is itself not observable – one cannot simply inspect it. It remains to be shown what observable consequences follow from the superposition principle and quantum dynamics.

The process of inspecting objects in quantum theory is described as a *measurement*. The word already has a meaning in our everyday language, and not all of this meaning survives the transition to quantum mechanics. Arguably, a different word should have been used: but 'measurement' has stuck [19].

A measurement in quantum theory is a probabilistic mapping from the quantum state to a set of 'outcomes'. One associates an operator with each element of the set of outcomes. In the first instance we shall consider operators Π which are projectors, satisfying $\Pi^2 = \Pi$. For our example of a die, an example of a measurement is one which answers 'is the dice in state 1 or in state 2...or in state 6?'. Note the measurement is constructed, at first, as an inspection of the classical configurations of the system: in fact the 'outcomes' are in perfect correspondence with the classical configurations. The projectors form a resolution to the identity

$$\sum_i \Pi_i = \sum_i |i\rangle \langle i| = \mathbb{I}. \tag{1.20}$$

Quantum theory then furnishes the probability of each outcome occurring

$$P_i = \langle\psi|\Pi_i|\psi\rangle = |\langle i|\psi\rangle|^2. \tag{1.21}$$

This is the Born rule. The normalisation of quantum states, and the resolution of measurement operators to the identity guarantee that an outcome always occurs, with all probabilities manifestly positive.

Just as a quantum state can be transformed into another, different measurements can be defined similarly. Relaxing the need to project onto one of the set of the classical possibilities, employing a 'rotated' basis $|i'\rangle$ one has

$$P_{i'} = |\langle i'|\psi\rangle|^2. \tag{1.22}$$

An important kind of dynamics intervenes when a measurement is performed. After the measurement, the quantum state is discontinuously updated, and 'projected' into whichever state corresponds to the measurement outcome

$$|\psi\rangle \to |\psi'\rangle = |k\rangle. \tag{1.23}$$

This process, sometimes called a quantum jump, or referred to as wavefunction collapse, is mysterious – it has a similar flavour to Bayes' rule, which describes how subjective beliefs are updated in light of new data [93]. Discontinuous and instantaneous effects are not considered to be features of a good physical theory, but they do describe our subjective opinions about physical systems. If a measurement simply reveals a pre-existing but unknown value that the system had before the measurement, there is no mystery. But if one is compelled to think of wavefunction collapse as a real dynamical process, one runs into trouble with special relativity (which forbids instantaneous effects). Is the collapse triggered by information flow, or by a physical process?

One can attach a label λ_i to each measurement outcome, and so define an *observable*

$$\mathbf{A} := \sum_i \lambda_i \left|i\right\rangle \left\langle i\right| . \tag{1.24}$$

By construction, λ_i are eigenvalues, and $|i\rangle$ the corresponding eigenvectors (or eigenstates), of $\mathbf{A}$. We can now say that the observable $\mathbf{A}$ takes the value λ_i when the system is in $|i\rangle$. The expectation value is $\langle \mathbf{A} \rangle = \sum_i P_i \lambda_i = \langle \psi | \mathbf{A} | \psi \rangle$.

Returning to our case of the quantum bit, or two-level system, the Bloch sphere picture is once more a useful aid. States $|\psi\rangle$ are points on the surface of the sphere. Measurements, of the kind discussed thus far, are in one-to-one correspondence with states: a measurement with two outcomes must correspond to two states on the sphere. It is simple to see that, given the necessary restrictions on projective measurements (namely that the projectors must be mutually orthogonal) the two states in question must be *antipodal*. A choice of measurement observable, or set of projectors, is really a choice of axis passing through the centre of the Bloch sphere: it is often referred to as a measurement direction. Observables are therefore in one-one correspondence with generators of time evolution, c.f. Figures 1.2 and 1.3. This is also true in higher dimensions.

Given a quantum state and an appropriate measurement direction, the Born rule probabilities are given by the projection of the state (i.e. its shadow) onto the axis defined by the measurement. Hence the angle between the state and the measurement outcome determines the probability; if the system is in one of the eigenstates of the measurement observable, the associated probability will be unity. If the state and measurement outcome are unbiased (i.e. if one is the north/south pole, the other is on the equator) then the probability will be one half. After the measurement the

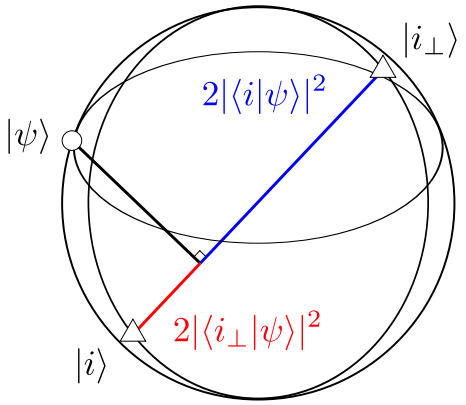


Figure 1.3: For qubits, the Born rule probability $|\langle i|\psi\rangle|^2$ is equal to the projection of the state (shown with a circle) onto the measurement outcome (shown with a triangle). The complementary probability $1 - |\langle i|\psi\rangle|^2 = |\langle i_\perp|\psi\rangle|^2$ is the projection of the state onto the antipodal measurement outcome (also shown as a triangle). One can visualise the apportioning of probability by the perpendicular sectioning of the Bloch sphere diameter (which has length 2) by the state of interest.

state is spontaneously snapped onto one antipode or the other, to remain consistent with the measurement result.

1.1.4 Interference

We are now equipped to understand one of the special features of quantum theory: interference. The phenomenon has long been understood in waves: for example the ripples on a pond can combine with each other constructively (peak with peak or trough with trough) or destructively (peak with trough). It is important the the waves are coherent: if they have the same frequency and a constant phase offset, the phase offset determines what kind of interference to expect. According to quantum theory, however, *all* objects can exhibit interference because of the superposition principle, which associates a smeared out *wave*function to an object. This is deeply mysterious, and is made more so by the fact that interference shows up in experiments involving only a single particle at a time. Objects can interfere with *themselves* in quantum mechanics.

We will consider a Mach-Zehnder interferometer, shown in Figure 1.4. In such an interferometer, one can *coherently* mix quantum states together. A qubit forms a simplified model of the interferometer: the two mutually orthogonal classical pos-

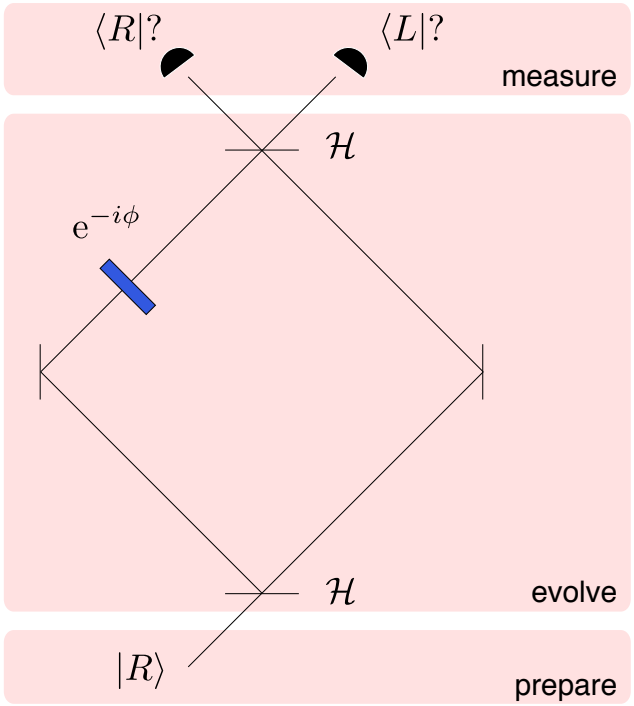


Figure 1.4: An interferometer. Horizontal lines represent beam-splitters, which send the particle left or right with equal probability. Vertical lines are mirrors which reflect the particle with certainty. The block in the left arm represents a unitary transformation on the qubit. The probability at both detectors is affected by the presence of the block, because quantum probability amplitudes can interfere.

sibilities are finding the particle in the left channel $|L\rangle$ or in the right channel $|R\rangle$. The following states then lie on on the equator of the Bloch sphere.

$$|\pm\rangle := \frac{|R\rangle \pm |L\rangle}{\sqrt{2}}. \tag{1.25}$$

In Figure 1.4, a particle is split and sent both ways around a system of mirrors. This operation is called the Hadamard operation: it sets up a superposition $\mathcal{H}|R\rangle = |+\rangle, \mathcal{H}|L\rangle = |-\rangle$. In the left arm, an additional transformation can be applied: the state acquires a phase $\mathrm{e}^{-i\phi}$. The Hadamard operation is applied again, and due to coherent addition of amplitudes

$$\begin{aligned} |R\rangle &\xrightarrow{\mathcal{H}} && |R\rangle + |L\rangle \\ &\xrightarrow{\mathrm{e}^{-i\phi(\mathbb{I}-\sigma_z)/2}} && |R\rangle + \mathrm{e}^{-i\phi}|L\rangle \\ &\xrightarrow{\mathcal{H}} && |R\rangle + |L\rangle + \mathrm{e}^{-i\phi}|R\rangle - \mathrm{e}^{-i\phi}|L\rangle \\ &= && \cos\frac{\phi}{2}|R\rangle + i\sin\frac{\phi}{2}|L\rangle, \end{aligned} \tag{1.26}$$

(I have not normalised the intermediate states), so, letting $U = \mathcal{H}\mathrm{e}^{-i\phi(\sigma_z - \mathbb{I})/2}\mathcal{H}$,

$$P_R := |\langle R|U|R\rangle|^2 = \cos^2\frac{\phi}{2}$$
$$P_L := |\langle L|U|R\rangle|^2 = \sin^2\frac{\phi}{2}. \tag{1.27}$$

The phase introduced by the block in the left arm has been mapped onto a modulation of the outcome probabilities for a measurement distinguishing 'particle through the left channel' and 'particle through the right channel'. This fundamentally different way of combining complex amplitudes, which differs from how probabilities are combined, is thought by some to be the definitive quantum phenomenon. Notice that, if the phase angle ϕ is chosen just right, one can prevent the particle from ever being detected in the right channel. But how can the phase shifter, located in the left arm, have an effect on the $|R\rangle$ state? When the particle is injected into the interferometer, something must traverse both branches. This revelation led Richard Feynman to proclaim interference as "the only mystery" in quantum mechanics. Note that interference is only possible when intermediate states have a nonzero 'overlap'. Restricting to orthogonal states, and transformations between orthogonal states, enforces classicality on the experiment (in the sense that no interference may be exhibited).

This completes my short introduction to the essence of quantum theory. Although there are some augmentations the formalism to be made below, many effects considered especially 'quantum' are included in the above, which has far reaching and profound consequences. The main conceptual problem is to decide when one should apply the second kind of dynamics – the quantum jumps. The choice is not specified by quantum theory, and is the essence of the measurement problem (discussed at further length in Chapter 4). The main technological challenge is to choose preparations, dynamics and measurements to process information in a useful way. A constant challenge to experimental investigations of either kind is to keep the *classical* uncertainties to a tolerable minimum. Only then is there an observable divergence between quantum and classical.

1.1.5 A few other ingredients

The simple linearity of the quantum formalism discussed so far conceals a multitude of subtleties. Consider the innocuous idea that we are to allow a ket for each possible classical state of the object under consideration. A coin has two kets, a die has six and we are must ensure orthonormality: no big deal! But what should be said, for example, about the state of two dice? A naive approach would simply allow six kets

for each die, making twelve in total. But the correct approach is to allow a ket for each possible configuration of the joint system of both dice. We must drop the assumption that each particle can be described independently of other particles. This assumption is so deep-rooted into classical physics that it is not obvious that it is even being made! After surrendering the assumption, one must allow a combinatorial number of kets: in this case $6^2 = 36$. Observe how quickly the number of kets grows: for n dice, we must track 6^n complex numbers. Quantum theory is thus said to have *exponential complexity*. This thesis does not permit a discussion of why this must be so: but note that the issue is related to whether we take the quantum state to be ontic or epistemic: certainly if the quantum state is epistemic then this part of the formalism is less of a surprise. This is because probability distributions are also exponentially complex [189].

1.1.5.1 Entanglement

The overall Hilbert space of a multi-party (or multi-particle) system is generated through the tensor product of the constituent Hilbert spaces: $\mathscr{H}_{AB} = \mathscr{H}_A \otimes \mathscr{H}_B$. The basis elements of the resultant space are given by all possible combinations of the elements at each party: for two qubits one has $\{|\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle\}$. The notation is simplified $|a\rangle \otimes |b\rangle \to |a\rangle |b\rangle \to |ab\rangle$. Once one has the idea of a joint system, and has applied the tensor product to generate the (often very large) multi-party basis, more intriguing ideas emerge. For example, there are non-local superpositions which do not factorise party-by-party. Consider the two-qubit state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\Big(|\uparrow\rangle |\uparrow\rangle + |\downarrow\rangle |\downarrow\rangle \Big). \tag{1.28}$$

Schrödinger, in his discussion of an analogous continuous variable state [177, 200], described the two qubits as being *entangled*. In coining 'entanglement', he provided a catchphrase that is incredibly widely used in the quantum information community, it is commonly thought to be an essential ingredient for quantum technologies. It is prohibitively expensive to simulate entanglement on a classical computer, requiring exponential memory resources. Consider the huge efficiency made by assuming the joint state to factorise – the number of complex coefficients needed shrinks from a combinatorial number to being linearly related to the number of parties. Perhaps quantum computers derive their power from the non factorability of states, since this is precisely the thing which is costly to simulate classically. Entanglement is, however, both a resource and an adversary.

Entanglement is a special part of quantum theory that defies efficient classical simulation. But is it a necessary feature of our understanding of nature? Even if the quantum description of an experiment is valid, there is the possibility of a post-quantum theory restoring an efficient and intuitive description of multi-particle systems. It seems possible that, if one could inspect some hidden aspect of a quantum system, one could reinstate the notion that it has a well defined state before measurement, and reinstate factorability (at the new 'hidden-variable' level). Then one would have a good physical theory with no such special feature as entanglement. Bell's theorem (1964) showed that entanglement has observable consequences, and calls the plausibility of such a 'hidden variable' programme into question [17, 18]. See Chapter 4.

Another observation about entanglement: the special feature can be seen as arising from the demarcation of a larger system into subsystems (typically particle by particle). The demarcation seems legitimate if one has a measurement which can distinguish between the subsystems. Otherwise the demarcation seems subjective, even arbitrary. If one cannot distinguish between the subsystems, one of two things happens.

In the first case, if the subsystems remain locked together in a known way, entanglement simply reduces to superposition: we make the identification $|\uparrow\uparrow\rangle = |\Uparrow\rangle$, for example, and the state of two qubits becomes the state of a single qubit. One can then still observe quantum effects, such as interference, between $|\Uparrow\rangle$ and $|\Downarrow\rangle$. Reversing this argumentation raises a question about 'how far down' superposition goes. One imagines that if we are to believe that a proton is made up of three quarks (two up quarks and one down quark), then any superposition of proton configurations (for example two spatial locations) can be unpacked and thought of as an entangled state of the three quarks:

$$\frac{1}{\sqrt{2}}\Big(|p^+_{\text{left}}\rangle + |p^+_{\text{right}}\rangle \Big) = \frac{1}{\sqrt{2}}\Big(|u_{\text{left}}\rangle\,|u_{\text{left}}\rangle\,|d_{\text{left}}\rangle + |u_{\text{right}}\rangle\,|u_{\text{right}}\rangle\,|d_{\text{right}}\rangle \Big). \qquad (1.29)$$

One can imagine hidden degrees of freedom with the same character which 'come along for the ride' in any quantum control of the collective degree of freedom. It is possible that future experiments will be able to exercise control over the hidden degrees of freedom: perhaps using quarks as quantum information carriers in a quantum information processor, and revealing entanglement in a system previously thought to only exhibit superposition.

In the second case, when one is unaware of how the states of the subsystems relate to one another, one loses track of the quantum state of the total system. As

an example, consider a first qubit entangled with a second, unobserved qubit. Since the state of the first depends on the outcome of a measurement on the second, by ignoring the second qubit we create multiple possibilities $|\psi_k\rangle$ for the quantum state of the first. If we imagine the second qubit is measured, but the outcome is not known, each possibility will in fact carry its own probability P_k. When we come to measure, there will be multiple possibilities for the Born-rule-governed distribution over outcomes. By the law of total probability we can simply mix the distributions together according to their probabilities

$$P_i' = \sum_k P_k |\langle i \,|\, \psi_k \rangle|^2 = \sum_k P_k \langle i \,|\, \psi_k \rangle \langle \psi_k \,|\, i \rangle = \langle i| \, \rho \, |i\rangle \, . \tag{1.30}$$

The last step is a useful device to attach the uncertainty to the state of the first particle: $\rho := \sum_k P_k |\psi_k\rangle \langle \psi_k|$ is known as the density operator. It is a positive, Hermitian, trace-one operator. It is also sometimes referred to as 'the quantum state'. It generalises the pure quantum state in the sense that we can recover the latter by setting the above probabilities P_k to a delta function: then it can be written $\rho = |\psi\rangle \langle \psi|$. Otherwise, the density matrix describes a 'mixed state'. The density operator formalism is a handy way of keeping track of the classical probabilities: they can be absorbed with the quantum uncertainty. We ought to bear in mind the ontological difference between a pure and a mixed state – quite independently of our interpretation of $|\psi\rangle$ it is clear that ρ must be thought of as a description of incomplete knowledge about $|\psi\rangle$.

The disorder, or entropy, of the mixing distribution P_k and the 'closeness' of the pure states which are mixed can be understood through a quantity called the purity

$$\mathcal{P}[\rho] = \mathrm{Tr}(\rho^2). \tag{1.31}$$

If the distribution P_k is very sharply peaked (has low entropy), the state is relatively pure. For $P_k = \delta_{kl}$, $\rho = |\psi_l\rangle \langle \psi_l|$, and the purity is unity. For an equal weight (maximum entropy) mixture of orthogonal (maximally different) states $\{|\psi_l\rangle\}, \langle l|m\rangle = \delta_{lm}$ one has the completely mixed state

$$\rho = \sum_l \frac{1}{d} |l\rangle \langle l| = \frac{\mathbb{I}}{d} \tag{1.32}$$

with the lowest possible purity $\mathcal{P} = 1/d$. A general feature of density operators is the existence of a infinite number of decompositions into mixtures of pure states. Two dimensional density operators can be visualised in the Bloch sphere: see Figure 1.5.

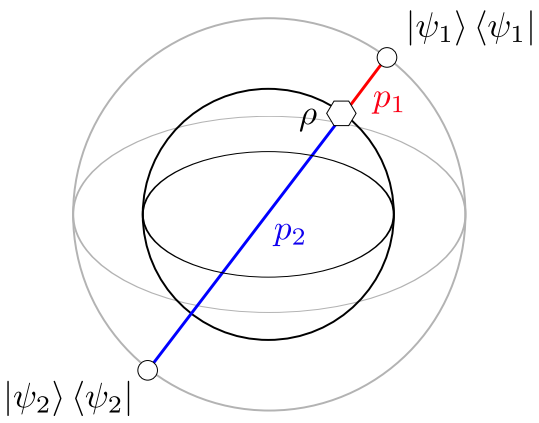


Figure 1.5: Mixed states are represented by density operators $\rho = p_1 |\psi_1\rangle\langle\psi_1| + p_2 |\psi_2\rangle\langle\psi_2|$, which lie in the interior of the Bloch sphere. Shown is the spectral decomposition of a mixed state (shown here as a hexagon), one of many convex sums of pure states giving the same density operator. Mixed states of constant purity will lie on the surface of concentric Bloch spheres with radius $\sqrt{2\mathcal{P} - 1}$, which nest like Russian dolls.

Loss of knowledge about the quantum state is called *decoherence*: it washes out any quantum signatures which would ordinarily be certified by examination of the pure quantum probabilities given by the Born rule. If one sent $\mathbb{I}/2$ through a Mach-Zehnder interferometer (Figure 1.4), the outcome probabilities would be $P_R = P_L = 1/2$ – all noise and no interference! This raises questions about 'how far up' superposition goes – to observe quantum effects on macroscopic scales would require high certainty about all interactions between the system of interest and other degrees of freedom.

In light of this discussion, what then should be made of the often cited dictum that quantum effects are not observed on the scale of our everyday lives – but only near the atomic scale? We can safely conclude that the macroscopic world is more complex than the microscopic world, and hence more likely to involve a large number of uncontrolled interactions. But one can further conjecture that the microscopic world of atoms and molecules does not exhibit a significantly high enough intrinsic complexity to wash out quantum effects. Evidence of quantum behaviour in electrons implies that they do not have a complex hierarchy of non-trivial substructure, because tracing it out (or averaging over it, as is implicit in out current ignorance of any such substructure) would ruin the visibility of quantum effects.

Given the density operator formalism, some upgrades to the rules for dynamics and

measurements are in order. Combining the Schrödinger equation with its Hermitian conjugate

$$\partial_t \langle\psi| = -i \langle\psi| \mathbf{H}; \tag{1.33}$$

(recall $\mathbf{H}^\dagger = \mathbf{H}$), one has

$$\begin{aligned} \partial_t \rho &= |\psi\rangle \partial_t \langle\psi| + (\partial_t |\psi\rangle) \langle\psi| \\ &= i[\mathbf{H}, \rho] \end{aligned} \tag{1.34}$$

where $[\mathbf{H}, \rho] = \mathbf{H}\rho - \rho\mathbf{H}$ is the commutator. Equation (1.34) is the von-Neumann equation for density operator dynamics. The Born rule is given above in (1.30); the measurement update rule is

$$\rho \to \frac{\Pi_k \rho \Pi_k}{P_k}. \tag{1.35}$$

1.1.5.2 Mixed states, dynamics and measurements

Probabilities are the sole observable predictions of quantum theory. When classical uncertainty is 'overlaid' the quantum probabilities mix together. By attaching an index k to a probabilistic prepare/evolve/measure experiment, one has:

$$\begin{aligned} P'_{i,\psi,U} &= \sum_k P_k |\langle i| U |\psi\rangle|_k^2 \\ &= \sum_k P_k \left[\langle i| U |\psi\rangle \langle\psi| U^\dagger |i\rangle\right]_k . \end{aligned} \tag{1.36}$$

The mixing may be absorbed into the state (as we did above with the density operator), or alternatively into the evolution or indeed the measurement itself:

$$\sum_k P_k \left[\langle i| U |\psi\rangle \langle\psi| U^\dagger |i\rangle\right]_k = \langle i| U \underbrace{\left(\sum_k P_k |\psi_k\rangle \langle\psi_k|\right)}_{\rho} U^\dagger |i\rangle ; \tag{1.37}$$

$$= \langle i| \underbrace{\left(\sum_k \sqrt{P_k} U_k |\psi\rangle \langle\psi| U_k^\dagger \sqrt{P_k}\right)}_{\mathcal{K}(\rho)} |i\rangle ; \tag{1.38}$$

$$= \langle\psi| U \underbrace{\left(\sum_k P_k |i_k\rangle \langle i_k|\right)}_{E_i} U^\dagger |\psi\rangle . \tag{1.39}$$

By attaching k to $|\psi\rangle \langle\psi|$, one defines the density operator ρ, discussed above.

Alternatively consider a random Hamiltonian being applied, and hence a random U_k – this way a Kraus map or superoperator $\mathcal{K}(\rho) = \sum_k B_k \rho B_k^\dagger$ is defined, with $B_k = \sqrt{P_k} U_k$. This is a positive map taking density operators to density operators.

Finally, attach k to the measurement itself to model a random choice of measurement basis - so defining an element $E_i = \sum_k P_k \left|i_k\right\rangle \left\langle i_k\right|$ of a POVM (positive operator valued measure). POVM elements are positive and must satisfy $\sum_i E_i = \mathbb{I}$. By virtue of being positive operators the elements may be written $E_i = M_i^\dagger M_i$. The Born rule is then

$$P_{M_k} = \mathrm{Tr}(M_k \rho M_k^\dagger), \tag{1.40}$$

and the measurement update rule, given outcome k was found, is:

$$\rho \to \frac{M_k \rho M_k^\dagger}{P_{M_k}}. \tag{1.41}$$

Each of these is, respectively, a generalised preparation or state; a generalised transformation; and a generalised measurement. Naturally we can have all three problems simultaneously; in the absence of other data besides the probabilities, it would be impossible to tell where to attach k. Each will reduce to the purely quantum case when the probabilities $P_k = \delta_{ik}$. Each inherits properties from its pure counterpart.

Generalised states, evolutions and measurements can also be seen to occur when we ignore some part of a larger system. The extra uncertainty (the distribution P_k) is then due to an ignorance of quantum correlations between the principal system and an 'environment', rather than arising from 'forgetting' the preparation, dynamics or measurement choice. More details can be found in Nielsen and Chuang [146].

1.1.5.3 Example: qubit dephasing

It is instructive at this point to consider an example of a generalised state undergoing generalised evolution. Let us begin with a density operator description of $|+\rangle$, and write it in a matrix representation in the $\{|0\rangle, |1\rangle\}$ basis:

$$\rho = |+\rangle\langle+| = \frac{1}{2}\Big(|0\rangle\langle 0| + |1\rangle\langle 1| + |0\rangle\langle 1| + |1\rangle\langle 0| \Big) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{1.42}$$

Now we will apply a rotation about the σ_z axis with a randomly chosen angle:

$$\begin{aligned} \mathcal{K}(\rho) &= \int p(\theta) d\theta \mathrm{e}^{-i\sigma_z\theta/2} \rho \mathrm{e}^{i\sigma_z\theta/2} \\ &= \int \frac{1}{2} p(\theta) d\theta \Big(|0\rangle\langle 0| + |1\rangle\langle 1| + \mathrm{e}^{-i\theta} |0\rangle\langle 1| + \mathrm{e}^{i\theta} |1\rangle\langle 0| \Big) \\ &= \frac{1}{2}\begin{pmatrix} 1 & \langle \mathrm{e}^{-i\theta} \rangle \\ \langle \mathrm{e}^{i\theta} \rangle & 1 \end{pmatrix}; \end{aligned} \tag{1.43}$$

our Kraus map is written as a integral because θ is a continuous variable; the calculation follows because $\sigma_z |0\rangle = 1, \sigma_z |1\rangle = -1$ by definition. Angle brackets $\langle \bullet \rangle$ denote an average value. It is clear that the off-diagonal elements of the density matrix, beginning at their maximum absolute value, will reduce in magnitude. Calculating the purity yields:

$$\mathcal{P} = \frac{1}{2} + \frac{1}{2} |\langle e^{-i\theta} \rangle|^2. \tag{1.44}$$

Notice that the higher the entropy of $p(\theta)$, the lower the purity of the resultant density operator. If the distribution is uniform over $[0, 2\pi)$, for example, the off-diagonal elements are entirely destroyed, giving $\rho = \mathbb{I}/2$. This evolution could equally well have been derived by entangling the principal spin with other spins which are then ignored or traced over. In either case, ρ moves toward the centre of the Bloch sphere.

Chapter 2

(Quantum) information theory

2.1 Information theory

The field of information theory began with a landmark paper of 1948 by Shannon [179]. The article "A Mathematical Theory of Communication" gave the first rigorous definition of entropy (a word used only casually in the previous chapter). Shannon considered entropy as a measure of the compressibility of a message – the more *order* in the message, the higher the redundancy. With greater redundancy comes the possibility of a more efficient coding of the message. For a probability distribution $\{P_k\}$, Shannon's entropy is defined

$$\mathcal{S} = -\sum_k P_k \log P_k. \tag{2.1}$$

If the logarithm is taken to base 2, the units of entropy are 'bits'. A simple exercise for the reader is to convince themselves that a smooth probability distribution has higher entropy than a 'sharp' function: the uniform distribution has the maximum entropy. Note that Shannon entropy is a global measure: shuffling the probabilities, or relabelling them, does not change $\mathcal{S}$.

Thinking mathematically about information has led to the 'information age' and the proliferation of computers. Digital computers, which are now found throughout modern life, process information stored as bits. Invariably the bits are defined physically by transistors: solid-state electrical switches that have two (classical) states available. Although analogue computers, which process information stored in continuous variables, have been studied, they are not widespread because a method for error correction is not known. Error correction is a fascinating topic which I will not discuss, save to say that it is absolutely crucial in turning ideas about information

processing into real technologies. This is true of classical computers, and even more so for quantum computers, introduced later in this chapter.

2.2 Fisher information as a statistical distance

Continuing to think mathematically, specifically *statistically*, about information leads one to ponder on the 'closeness' of two probability distributions. The uniform distribution has maximum entropy, but if it is slightly 'biased' the entropy decreases. Can a notion of the distance between distributions be quantified? A step in this direction is a quantity known as the Kullback-Leibler divergence [110], or relative entropy [146]:

$$D_{KL}[P_i||Q_i] = \sum_i P_i \log \frac{P_i}{Q_i}. \tag{2.2}$$

for discrete distributions, or

$$D_{KL}[p(s)||q(s)] = \int p(s) \log \frac{p(s)}{q(s)} ds. \tag{2.3}$$

for distributions over a continuous random variable s. It has many of the features one would desire in such a measure: it is non-negative, for example. To qualify as a metric (a 'proper' distance) a measure must i) be positive $D(p,q) \geq 0$ (with equality only when $p = q$), ii) be symmetric $D(p,q) = D(q,p)$ and iii) obey the triangle inequality $D(p,r) \leq D(p,q) + D(q,r)$. The relative entropy is not symmetric, so it is not a metric.

Consider a family of continuous distributions $p(s;a)$ indexed by a parameter a: the Fisher information

$$F = \int (\partial_a \ln p(s;a))^2 p(s;a) ds \tag{2.4}$$

is the curvature of the relative entropy between $p(s;a)$ and $p(s;a+da)$; a proof may be found in [75, p87]. It measures how fast the probabilities tend towards being more distinguishable as a is varied, and it *is* a metric. It has the properties of an *information*: it is positive and increases as $p(s;a)$ (the probability of measuring s given a value of a) becomes less smooth. In this sense it is an inverse measure of entropy; however, it is a local measure. Reshuffling a would generally alter F. If one upgrades to a multiparameter index $p(s;\mathbf{a})$, this defines a manifold of probability distributions. The Fisher information is then a metric on this manifold: it gives the infinitesimal distance between points [108].

I will be chiefly concerned with the Fisher information in this thesis, but there are a number of other metrics that are instructive to explore. The classical fidelity is defined

$$\mathcal{F}^c = \sum_i \sqrt{P_i Q_i}\,, \tag{2.5}$$

or

$$\mathcal{F}^c = \int ds \sqrt{p(s)q(s)}; \tag{2.6}$$

it features in the Hellinger distance $D_H = \sqrt{1-\mathcal{F}^c}$, which is a metric. Such metrics are important ways of quantifying how distinguishable two distributions are.

2.3 Fisher information as a fundamental bound on the performance of statistical estimation

The Fisher information is special among the various metrics or 'statistical distances', because it has a fundamental *operational* meaning, in addition to being a statistical distance metric. To arrive at the former meaning, we need first to consider a *statistical model*. A statistical model is a family of probability distributions over a continuous random variable s

$$p(s;a) \tag{2.7}$$

each indexed by a different value of a – let us assume a to be an unknown parameter, and that our task is find a good estimate for it. In an experiment that samples from an unknown distribution in the family, in general the identity of the distribution amongst the family is not uniquely determined by the sampled data

$$s[0], s[1], s[2], \ldots \tag{2.8}$$

Here $s[n]$ is the value taken by the random variable s in the nth trial. Attempting to discover the parameter which indexes the true distribution is a task known as *parameter estimation*. To arrive at an estimate for a, one postulates an *estimator* $\hat{a}$. An estimator is a function of the data: because the data are random, the estimator itself is a random variable, taking different values (i.e. different estimates) each time the experiment is repeated. A *unbiased* estimator reports the correct value for a on average, although of course one expects fluctuations around the true value in each experiment. Biased estimators give rise to a systematic error in the estimate which does not diminish through averaging.

I follow Kay [97] to derive the ultimate limit on the performance of parameter estimation – similar derivations may be found in Kok and Lovett [108]. Consider an unbiased estimator, which by definition satisfies

$$\int p(s;a)\hat{a}ds = a. \tag{2.9}$$

Differentiate both sides with respect to the unknown parameter

$$\int \hat{a}\partial_a p(s;a)ds = 1 \tag{2.10}$$

$$\int \hat{a}p(s;a)\partial_a \ln p(s;a)ds = 1 \tag{2.11}$$

$$\int (\hat{a}-a)p(s;a)\partial_a \ln p(s;a)ds = 1; \tag{2.12}$$

in the first line I used $\partial_a \hat{a} = 0$ and interchanged the order of differentiation and integration; in the second line I used $\partial_a \ln p(s;a) = \partial_a p(s;a)/p(s;a)$; in the third line I subtracted a quantity equal to zero $\int a\partial_a p(s;a)ds = \partial_a 1 = 0$. Next, the Cauchy-Schwarz inequality implies

$$\int (\hat{a}-a)^2 p(s;a)ds \int (\partial_a \ln p(s;a))^2 p(s;a)ds \geq 1. \tag{2.13}$$

On defining the Fisher information "about a"

$$F_a = \int (\partial_a \ln p(s;a))^2 p(s;a)ds \tag{2.14}$$

one reaches the Cramér-Rao inequality

$$\langle (\hat{a}-a)^2 \rangle \geq \frac{1}{F_a}. \tag{2.15}$$

The Cramér-Rao bound therefore limits, from below, the variance in an estimate $\hat{a}$ of an unknown parameter a given the observed data. The result is a paragon of statistical estimation theory [42]. If N independent observations are made, it is simple to show that

$$F_a^{\text{total}} = NF_a. \tag{2.16}$$

The Fisher information gives insight into the information carried in a probability distribution. It is a functional on such conditional probability distributions, and can also be written:

$$F_a[p(s;a)] = \int_{-\infty}^{\infty} \frac{(\partial_a p(s;a))^2}{p(s;a)} ds\,. \tag{2.17}$$

It is a powerful tool for the appraisal of parameter estimation protocols – it is the definitive measure of metrological performance, when the objective is to 'measure' or estimate a.

2.4 Quantum information theory

Quantum information theory is concerned with a myriad of generalisations of information theory in light of quantum theory. We should not be surprised that there are strong links: already we have seen how similar the quantum state itself is to a probability distribution (and hence to a state of knowledge). The quantum generalisation of probability leads to a quantum generalisation of information.

2.4.1 Quantum statistical distance

Wooters was the first to consider a quantum version of statistical distance. Quantum states are not probability distributions, so we cannot simply import a classical statistical distance. Something new is required [210]. First, consider the quantum fidelity for pure states

$$\mathcal{F} := |\langle\phi|\psi\rangle|, \tag{2.18}$$

which involves $|\psi\rangle$ and $|\phi\rangle$. It is the maximum possible classical fidelity between the probability distributions that arise when the states are combined with a choice of measurement. It is not a metric. The Bures angle, however,

$$D_B := \arccos|\langle\phi|\psi\rangle|, \tag{2.19}$$

introduced by Wooters [210], or the Bures distance (the quantum generalisation of the Hellinger distance)

$$D_B(\phi,\psi) := \sqrt{1-\mathcal{F}^2} \tag{2.20}$$

are metrics on the manifold of quantum states, because they satisfy positivity, symmetry and the triangle inequality.

Indexing quantum states by the unknown parameter a defines a *quantum statistical model*. The quantum Fisher information is another metric on pure quantum states [183]

$$H_a[|\psi\rangle] := 4\langle\partial_a\psi|\partial_a\psi\rangle - 4|\langle\partial_a\psi|\psi\rangle|^2. \tag{2.21}$$

Here $|\partial_a\psi\rangle = \partial_a|\psi\rangle$. If the pure state acquires its dependence on a through a Hamiltonian evolution, $|\psi(a)\rangle = \mathrm{e}^{ia\mathbf{H}}|\psi_0\rangle$, then

$$H_a = 4\langle\psi_0|\mathbf{H}^2|\psi_0\rangle - 4\langle\psi_0|\mathbf{H}|\psi_0\rangle^2, \tag{2.22}$$

proportional to the variance of the Hamiltonian in the initial state. The generalisation to density matrices is

$$H_a := 2\sum_{nm}\frac{|\langle m|\partial_a\rho|n\rangle|^2}{P_n+P_m}, \tag{2.23}$$

where $\rho = \sum P_m |m\rangle \langle m|$. It is the maximum classical Fisher information contained in a state in a variation over all POVMs [24, 152]:

$$H_a = \max_{\{E_k\}} \left(F_a[\mathrm{Tr}[\rho E_k]] \right). \tag{2.24}$$

It therefore inherits the operational meaning from the classical Fisher information: it quantifies the ultimate performance of statistical estimation using a quantum state when the optimal measurement is chosen.

2.4.2 Quantum computation

Digital computers have transformed modern life. They are machines which process digital information: they take an input state and transform it, according to the laws of classical physics, into an output state. In this way they can compute a function of the input. Digital computers represent information with bits, and hence process discrete information – but they can also approximately simulate continuous physical systems [125]. Such machines are constrained by the laws of physics – and given finite resources only a finite set of functions may be efficiently computed. By efficiently, it is meant that the function can be computed with memory (i.e spatial) and time resources that scale polynomially, and not exponentially, with the size of the input to the function.

A quantum computer processes quantum information. It transforms input into output via the laws of quantum physics. Since quantum physics allows for a wider and more general class of states and transformations, quantum computers might be able to efficiently compute a wider class of functions than their classical counterparts. The concept of a quantum computer was first rigorously formalised in 1985 by David Deutsch [46]. Besides explicating a radical extension to the Church-Turing principle ('any physical system can be simulated on a large enough classical computer'), Deutsch illustrated the power of quantum computation with the first example of a quantum algorithm. He showed that a certain (admittedly artificial) decision problem can be solved on quantum computer with fewer 'steps', or operations, than on a classical computer. Although quantum computers are not the main focus of this thesis, a short explanation of this argument will showcase the power of quantum information technology. It will also serve as a useful introduction to the quantum circuit diagram or quantum-computational network, which appear in succeeding chapters.

I follow Nielsen and Chuang [146], and present a slightly improved version of the Deutsch algorithm. Consider an unknown function $f : \{0, 1\} \to \{0, 1\}$. Suppose that

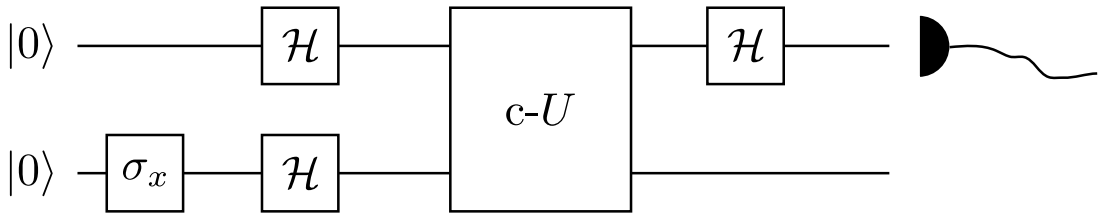


Figure 2.1: A quantum circuit diagram for Deutsch's algorithm. The two 'rails' represent a register of two qubits, and the square boxes represent unitary operations or quantum gates, which are applied sequentially with time progressing left to right. $\mathcal{H}$ is the Hadamard gate. At the end of the algorithm, the first qubit is measured in the computational $|0\rangle, |1\rangle$ basis.

one wishes to learn a global property of the function: for example the sum (modulo 2) of the function with each possible input: $f(0) \oplus f(1)$. This quantity is equal to zero if the function is constant (i.e. if it has the same output for both inputs) and equal to one if it is balanced (i.e. if it has a different output for each input). A classical approach must compute $f(0)$ and $f(1)$, and then also compute $f(0) \oplus f(1)$. A quantum machine can compute the global property using only a single application of f. Observe a two qubit 'register' initialised into $|0\rangle|0\rangle$ shown in Figure 2.1. The register is subjected to a sequence of unitary processes called 'quantum gates'. First, a NOT operation (achieved by a rotation by π around the x-axis) is applied to the second qubit. Then a Hadamard (see Chapter 1) gate is applied to each. A controlled unitary $(c-U)$ gate then applies f to the first qubit and advances the second qubit by the result, modulo 2. The state of the register after each gate follows the chain:

$$\begin{aligned} |00\rangle &\xrightarrow{\mathbb{I}\otimes\sigma_x} |01\rangle \xrightarrow{\mathcal{H}\otimes\mathcal{H}} |+-\rangle \\ &\xrightarrow{\text{c-}U} |0\rangle \underbrace{|0+f(0)\rangle}_{\blacktriangle\triangle} - |0\rangle \underbrace{|1+f(0)\rangle}_{\blacksquare\square} + |1\rangle \underbrace{|0+f(1)\rangle}_{\blacktriangle\square} - |1\rangle \underbrace{|1+f(1)\rangle}_{\blacksquare\triangle}. \end{aligned} \tag{2.25}$$

Now if $f(0) \oplus f(1) = 0$ (the function is constant) then $0 + f(0) = 0 + f(1)$ and $1 + f(0) = 1 + f(1)$: the black shapes label identical states (triangle with triangle, square with square). If $f(0) \oplus f(1) = 1$ (the function is balanced) then $0 + f(0) = 1 + f(1)$ and $1 + f(0) = 0 + f(1)$: the white shapes label identical states. Hence

$$\begin{aligned} &\text{if } f \text{ constant}, \qquad |+\rangle|-\rangle \xrightarrow{\mathcal{H}\otimes\mathbb{I}} |0\rangle|-\rangle \\ &\text{if } f \text{ balanced}, \qquad |-\rangle|-\rangle \xrightarrow{\mathcal{H}\otimes\mathbb{I}} |1\rangle|-\rangle, \end{aligned} \tag{2.26}$$

and inspection of the first qubit reveals the value of $f(0) \oplus f(1)$ with certainty. This is an example of quantum parallelism: somehow, fewer steps were necessary in the quan-

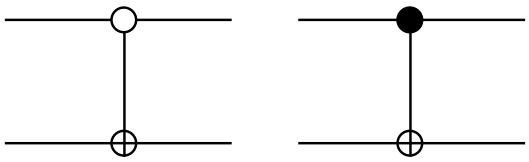


Figure 2.2: Quantum circuit diagram showing two species of controlled-NOT gates. The first is controlled on $|0\rangle$, and the truth table is given in the main text. The second is controlled on $|1\rangle$.

tum algorithm than through the classical algorithm. The Hadamard gate arranged for a pure superposition to exist inside the quantum computer, and is a reversible operation. Attempts to imitate the quantum algorithm on a classical computer, for example by simulating the Hadamard gate with an irreversible randomising process, will fail. The quantum computation combines a kind of parallelism (to compute a function in Hilbert space) with interference (to concentrate the result so it may be read out).

I did not describe how the $c-U$ operation is effected in the lab. It is unitary by virtue of being a concatenation of unitary gates. A universal set of elementary gates is one which can be combined to construct any unitary operation, and hence enable universal computation [47, 124]. An example is the set of arbitrary single qubit unitary gates (achieved by turning on local Hamiltonians), combined with a two-qubit gate such as the CNOT gate. See Figure 2.2. The CNOT gate requires an interaction between the qubits: it is defined by the truth table

$$\begin{aligned} U_{\text{CNOT}}\,|00\rangle &= |00\rangle \\ U_{\text{CNOT}}\,|01\rangle &= |01\rangle \\ U_{\text{CNOT}}\,|10\rangle &= |11\rangle \\ U_{\text{CNOT}}\,|11\rangle &= |10\rangle\,. \end{aligned} \tag{2.27}$$

By combining one and two qubit gates, any unitary operation can be applied to a qubit register: therefore because the $c-U$ gate above is unitary, it can be implemented (although I do not show how, here). More ambitious functions can be computed. The most impressive of these can factor numbers [182], search a database [76] and simulate quantum dynamics [125] more efficiently than is thought possible classically.

2.5 Quantum information media

In this section I present a brief and incomplete survey of candidate quantum information carriers. In classical computing, various media have emerged as the most suitable for different tasks: communication with optical fibres, storage with magnetic disks, and processing in electronic states of silicon nanostructures. Yet more materials and architectures have been proposed for building quantum computers and other quantum technologies. Some of the media are found naturally, and some are engineered. Each enjoys its own benefits.

Important figures of merit vary across quantum technologies. For universal quantum computation, full control over the quantum state is necessary (i.e. a universal set of gates), and scalability is paramount. For large quantum computers, various thresholds on preparation, control, and measurement fidelities must be met to allow for error correction [145]. For quantum sensors, the interaction with the field under study should be strong. Quantum memories are designed to store quantum states over long durations: a high degree of isolation from the environment is necessary to achieve this. For quantum communication, speedy transport of the media is vital. Comprehensive reviews from 2010 [31] and 2011 [112] give more detail.

I will introduce three different units of quantum hardware: the spin qubit, the photon qubit, and the superconducting qubit. For each I describe how information is encoded, and some advantages and disadvantages of architectures that the qubits are to be embedded into. Decoherence times will be an important figure of merit: for a qubit one finds two timescales. T_1 is the timescale on which the energy eigenstate populations of the qubit (ρ_{00} or ρ_{11} elements of the density matrix) relax to their thermal equilibrium values. T_2 is time time taken for the coherences ($\rho_{01} = \rho_{10}^*$ elements) to decay to $1/e$ of their initial values. The timescales describe the severity of, for example, unwanted interactions or thermal fluctuations. T_1 processes, called relaxation or depolarisation, corrupt the θ angle on the Bloch sphere, and T_2 processes, called 'dephasing', corrupt the ϕ angle, see section 1.1.5.3 and Figure 1.1. Both types of process can lead to a loss in purity.

2.5.1 The spin

Spin is an intrinsic property of fundamental particles. The spin of an electron is the simplest example: it was discovered in the famous Stern-Gerlach (SG) experiment. As electrons transit a magnetic field gradient, they are deflected either 'up' or 'down' the gradient by a fixed quantised amount. An SG field gradient can be thought as a

conditional beam-splitter for spins. The two possibilities ('up' and 'down') form the basis of a qubit.

The magnetic field gradient can be used to measure the qubit, by sending a particle to two distinct detectors, depending on its spin state. Hence by the measurement update rule, magnetic fields can also be used to prepare the qubit, as long as the detection does not destroy the particle. Such a scheme is generic to quantum information and is called a Quantum Non-Demolition (QND) measurement[1]. Transverse magnetic fields can be used to control the quantum state of the spin, achieving arbitrary $SU(2)$ rotations (see Chapter 1). One describes the interaction of a spin with an external magnetic field with the Zeeman Hamiltonian:

$$\mathbf{H} = \mu \bar{\sigma}.\bar{B}. \tag{2.28}$$

$\bar{B}$ is a vector describing the strength of the magnetic field in the x, y, z direction, and $\bar{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ is a vector of the Pauli matrices. Because of this Hamiltonian, the Bloch sphere 'spin space' is isomorphic to real space in the sense that a measurement in a certain direction in real space (by orienting the SG apparatus along such an axis) will correspond to the same axis in the Bloch sphere. If we fix one of the directions as our computational basis (an arbitrary choice) then non-parallel magnetic fields induce rotations about the corresponding axes in the Bloch sphere. By control of these magnetic fields one has full control over the quantum state.

This thesis frequently discusses an ensemble of spins. This is a slight abuse of terminology: a true statistical ensemble is a large number of non-interacting identical copies of a physical system: it is a conceptual ideality. The term is also used to describe a large collection of weakly interacting identical spin-bearing molecules. The interactions are often extremely weak, and the number of molecules rather large (10^{23} for example): such a collection is then a good approximation to a statistical ensemble. The elements of the ensemble are understood to be arrayed spatially. Roughly speaking, the larger the ensemble, the better the statistics.

We will be concerned with two types of physical spin ensemble, both addressed through magnetic resonance techniques: nuclear magnetic resonance (NMR) in a liquid state ensemble, and NMR / Electron spin resonance (ESR) in a solid state spin ensemble.

[1]QND is sometimes defined in terms of the back action onto the system, or the disturbance induced into system observables – that is not the intended meaning here.

2.5.1.1 Liquid state NMR

Nuclear Magnetic Resonance is a technique that has been known to chemists for decades. A spatial ensemble of identical molecules is placed in a strong magnetic field. The energy levels of the nuclei in the molecules are split by the Zeeman effect. The nuclei can thus be made to store quantum information, for example in their lowest two Zeeman energy levels, and be driven from one level to the other with resonant radiation. The strong field leads to a very rapid evolution of the Bloch vector around the magnetic field. One then ignores this evolution by working in a co-rotating frame of reference. Slower rotations around other axes are driven by resonantly exciting with a weaker, oscillating transverse magnetic field, inducing a transition (for example) between spin up and spin down. More details can be found in, e.g., Ref [36]; the important point is that full control (rotation about two axes) is possible. Detection is by a current induced in nearby pickup coils by the net magnetisation of the spins.

The local electronic bond configurations in the molecule, and the external field, lead to distinct Zeeman energy splittings at the site of each nucleus. If the neighbouring nuclei of different species interact, either directly or indirectly through the electrons, quantum control of the joint state is possible and entanglement can be generated by clever sequencing of pulsed radiation.

The interaction between nuclei and magnetic fields is weak. The frequencies associated with the Zeeman energy splittings are typically in the MHz range. The resonant radiation must be applied slowly to be selective, necessitating long gate durations. Combined with typical decoherence times of $T_2 \sim 1$s, only of order 100 gates may be achieved before decoherence has 'won'. Also, initialisation is a problem. The energy of a MHz frequency photon is only a few nano electronvolts, and so thermal fluctuations (with energy $k_B T$, k_B is Boltzmann's constant) at room temperature completely swamp the energy splitting between spin up and spin down, leading to a mixed state with low purity. Cooling is not an option, since freezing the molecules causes other problems: for example, the interaction between ensemble elements is not averaged to zero in the solid state as it is in a liquid. Instead of, or along with cooling, a pseudo pure trick is used. There are some conceptual issues with this trick: I will briefly discuss them.

Because of the ratio of Zeeman energy to thermal energy, a nuclear spin at room temperature is in an almost maximally mixed state. When the pseudo-pure approximation is employed, the state ρ of the nuclear spin ensemble decomposes into a pure

part ρ_{pp} and a maximally mixed part $\mathbb{I}/2$:

$$\rho = \epsilon\rho_{pp} + (1-\epsilon)\mathbb{I}/2. \tag{2.29}$$

The maximally mixed subensemble does not contribute to the net magnetisation of the sample, because the moments of the spins cancel out: half are up, and half are down. This subensemble is therefore never observed! Only the ensemble giving a deviation from the identity, ρ_{pp}, contributes a signal. How should this fact impact on our interpretation of NMR experiments? I argue that it depends on the aim of the experiment. Foundational tests may be undermined by the pseudo-pure approximation, but quantum computing may not be.

Souza, Oliveira and Sarthour (SOS), in discussing foundational tests in NMR experiments [188], claim "Since $(1-\epsilon)\mathbb{I}/2$ is not observed, the probe qubit in such a mixed state produces the same result as would be observed if the probe qubit were in a pure state and the detection efficiency of the measurement apparatus were ϵ.". At thermal equilibrium $\epsilon = (1-\alpha)/(1+\alpha)$ with $\alpha = \exp(-\mu_N B/k_b T)$. For typical values of temperature T and magnetic field strength B, $\epsilon < 10^{-7}$ (here μ_N is the magnetic moment of the probe nucleus and k_B is Boltzmann's constant). Assuming the quoted assertion is correct, and ϵ can be interpreted as a detector efficiency, it is rather low. Low detection efficiency is not an intrinsic problem for quantum computing: it just means each computation should be repeated a large-but-fixed number of times to find the result. Foundational tests, on the other hand, are in a sense more demanding of our interpretation of physics. The tests aim to make highly non-trivial claims about nature: for example about post-quantum theories. One cannot necessarily rely, for example, on Equation 2.29, or any other equation presupposing quantum physics. If a subset of the trials of a foundational test are not observed, this opens a loophole: conspiracies arise that exploit the hidden runs to reach alternate conclusions about the nature of reality [155]. Low detector efficiency is also not an intrinsic problem for foundational tests, however, as long as a 'fair sampling hypothesis' is justified.

The fair sampling hypothesis can be stated as follows: If one only measures a fraction of the systems which one has prepared, the gathered statistics faithfully represent the entire ensemble. This may be warranted, for example, in the case of an experiment with photon loss (see below): typically there is no reason to suspect that unobserved photons would have given different results to observed photons, had they indeed been detected. I believe the situation to be different in NMR: the fair sampling assumption is patently false! It is generally accepted that the unobserved component of the nuclear spin ensemble behaves very differently to the measurable

part. It is unobservable precisely because it generates a zero net magnetic field; the field from each spin is cancelled out by other spins in the 'identity' component. If the unobserved spins behaved in the same way as the observed spins, they would become observable - giving a contradiction. This shortcoming in NMR has been discussed at length [102, 25]; the correct conclusion is to worry about NMR for foundational tests, but not for quantum technologies. Souza *et al.* make the first point [186] by claiming only to have *simulated* the violation of a Bell inequality with a room temperature NMR experiment - precisely because of the failure of the fair sampling hypothesis. SOS make the second point by claiming that "every single experiment performed in quantum information processing by liquid-state NMR is based on the power of this technique as a perfect quantum simulator" [188].

NMR has led early progress on quantum computing, with many proof of principle experiments [90, 88]. Techniques to overcome noise have proven useful in other quantum media. The major drawback is that NMR has a scaling problem: the observable signal (i.e. ϵ) diminishes exponentially as the molecule (the quantum computer) becomes larger, limiting NMR quantum computers to registers of around a dozen qubits.

2.5.1.2 Solid state NMR / ESR

There are other places to find spins. For example, they can be found in quantum dots, where electrons are confined in artificial nanostructures by a electrostatic potentials [127]. Many of the same spin resonance techniques can be applied: although control can also be mediated through electric fields, and also optically. Addressing spins in an ensemble has one particular advantage – the strength of the signal is far greater than that of an individual spin, and therefore easier to measure.

One can have the benefits of an ensemble without the drawbacks of liquid state NMR. The first improvement to make is to bring electron spins into the picture. Because the electron spin magnetic moment is typically three orders of magnitude larger than that of a nuclear spin, the Zeemann energy is higher relative to thermal energies. Also, the higher resonant frequency can be addressed with microwaves, which can lead to faster transitions: gate times are around 100 ps. The downside is a shorter coherence time, because the coupling to the environment is also stronger.

A second improvement is found by moving to the solid state. Silicon-28 is a material that has benefited from the huge investments in the computing industry, where global demand has led to the cheap manufacture of semiconductor devices. New ways of using the material have emerged from the Avogadro project, which exploits the

crystalline structure of silicon in an attempt to redefine the kilogram [9]. Isotopically purified host crystals can be engineered for this purpose: they also find application in quantum information science by providing a spin-less 'vacuum in solid' [16] for dopant impurities: for example, phosphorous impurities in silicon [95]. Removing spins, for example those in silicon-29 nuclei, from the environment suppresses decoherence more effectively than the averaging effect seen in liquid state spin ensembles, and also allows for lower temperatures to be reached (freezing is not an issue). At temperatures of a Kelvin or so, the Zeeman energy dominates the thermal energy, allowing for very high spin polarisations via cooling. Entropy is shed to the environment, and the initial state purity is quite high.

There remain drawbacks to ensembles measured with spin resonance techniques. Initialisation by QND readout is not possible, because addressing the ensemble globally leads to only a weak measurement of any given spin. The net magnetisation induces a current in pickup-coil circuits near to the sample. Such a measurement has poor resolution: it cannot distinguish between different micro-states of the ensemble with the same net magnetisation. The measurement, therefore, does not project the ensemble into an eigenstate. See Chapter 3.

Other initialisation techniques can be employed. Hyperpolarisation is a technique to dynamically pump nuclear spins into their ground state via an interaction with electron spins [139]. This has been combined with spin dependent photoionisation to enable world record coherence times; at room temperature $T_2 > 39$ minutes, at cryogenic temperatures $T_2 > 3$ hours [171].

The issue of resource counting is of high importance in motivating quantum technologies. I discussed how the internal structure of a quantum system may or may not be of relevance in Chapter 1: for an ensemble of spin-systems, it seems we should count each spin as a quantum system. But if we consider each spin-system (perhaps it is a large molecule containing a dozen qubits) as a quantum computer, and the fact that it is in a spatial ensemble merely as an incidental fact about how the computer is addressed and read out, the resource count is only one. An ensemble computer merely trades temporal resources for spatial resources– a 'synchronous, single-instruction parallel machine' [41].

Many approaches use spins in the solid state and dispense entirely with a spatial ensemble, and so avoid the pseudo-pure state scaling problem. Defect centres in diamond have been the most successful – there, single shot readout of a spin is possible, and optical transitions allow for the spin to be well initialised into the ground state [173]. The phosphorus in silicon system can also be scaled down to a

device with only one pair of spins – QND initialisation is possible [161] and coherence times remain impressively long ($T_2 > 30$ s) [141].

2.5.2 The photon

Photons are the discrete chunks of energy which compose electromagnetic waves. Quantum information can be stored with light quanta in a number of ways: in the particle number [216], time bin [214] or spatial mode [130]. In this thesis, however, I consider the polarisation of single photons. A photon's polarisation is much like the spin of an electron. There are two classical possibilities: horizontal polarisation $|H\rangle$ or vertical polarisation $|V\rangle$. These can be thought of as the σ_z eigenstates; the eigenstates of σ_y are the right and left circular polarisations, and the eigenstates of σ_x are the diagonal and anti-diagonal polarisations.

A polarising beam splitter (PBS) is the equivalent of a SG gradient, spatially separating $|H\rangle$ and $|V\rangle$. Combined with photodetectors (which typically have only a low efficiency), this enables fully projective (if probabilistic) measurements. Preparations and measurements in rotated bases are achieved with waveplates. A half wave plate will result in a rotation around the σ_y axis, and a quarter wave plate a rotation around the σ_z axis in the Bloch sphere. The angle of rotation is controlled by rotating the waveplate around the optical (direction of propagation) axis. Full control is therefore possible: for more details see [108].

The advantages and disadvantages of photons both derive from the reluctance of a photon to interact. This shyness leads to a extremely long coherence time: because the interaction with the environment is so weak, T_1 and T_2 are essentially infinite. Photons travel at the speed of light, making them excellent for communication technologies. It is very difficult to entangle photons together, but it is possible to couple them through optical nonlinearities: but only weakly and probabilistically [163]. The Knill-Laflamme-Milburn (KLM) architecture removes the need for nonlinearities in an all linear-optical quantum computer [104].

Detection efficiency is low in a traditional single-photon-avalanche-detector: this may prove bothersome for foundational tests, regardless of whether it impedes quantum information processing. This is in strong analogy to the discussion on pseudo-pure states above – unobserved sub-ensembles open loopholes. New superconducting calorimeter detectors can exceed 90% efficiency [143], and by continued improvement look to solve this issue. Photon loss at preparation, evolution and detection remain a constant concern, however: the 'hold time' of an infrared photon in an optical fibre

is approximately 0.1 ms, which limits quantum computations far more severely than any decoherence process.

2.5.3 The superconducting qubit

Superconducting circuits are a synthetic quantum information medium. Cooling a superconducting material below a critical temperature T_c (typically a few Kelvin) allows for current to flow with no resistance: the electrons pair up and condense into a macroscopic wavefunction with a stable phase. Micrometer-size circuits of the inductor-capacitor (LC) type then behave as almost dissipation-less (underdamped) harmonic oscillators. By interrupting the superconductor (for example, niobium or aluminium) with a thin piece of insulating material (for example, aluminium oxide) a *Josephson junction* is formed. These junctions, when introduced to superconducting circuits, introduce non-linear effects and give rise to an anharmonic oscillator: this lifts the energy-level-transition degeneracy. Unequal level spacings allows for targeted control of a 'macroscopic' variable: the wavefunction describes $\sim 10^{10}$ electrons.

The superconducting circuit can be thought of as an artificial atom. Readout can be achieved with nearby magnetometers (typically also superconducting circuits). The circuits are easily synthesised, and so quantum control is achieved at adjustable frequencies, invariably at microwave band. Control is also possible electrically and with magnetic fields. Because of the relative ease of fabrication, a chip based quantum computer is a realistic prospect: coupling nearby qubits with transmission line cavities is possible [50]. Superconducting qubits are also particularly suited to hybridising with other media such as spin ensembles [172, 206].

A flux qubit is based on a SQUID (Superconducting QUantum Interference Device), which is a superconducting loop interrupted by one or two Josephson junctions. The magnetic flux threading the loop can become trapped in a double well potential (a parabola with a cosine 'corrugation'): the left well represents a clockwise persistent current in the loop, and the right well an anticlockwise persistent current.

The charge qubit or 'Cooper-pair box', omits the inductor: the qubit is defined by the two lowest energy levels of a cosine potential. Essentially, the quantum variable is the charge on a small insulating island in the superconductor. A transmon qubit [105] achieves a much higher Josephson energy, compared to the 'kinetic' charging energy, than a standard charge qubit, reducing charge dispersion. Along with other techniques (such as the use of superconducting cavities [149]) this leads to much longer coherence times: at milliKelvin temperatures T_1 and $T_2 \sim 10$ µs. However, the time

taken to perform a quantum gate is around 10 ns, and hence the number of gates that may be performed per coherence time does not compare favourably with spin qubits.

Chapter 3

Quantum measurement theory

What is a measurement? The answer to this question is not easy within the setting of quantum theory. The issue is semantic: the answer often depends on the context [144]. As well as the mysterious irreversible process that gives rise to collapse of the quantum state vector, it is a word also used to describe a reversible, unitary process which does not lead to collapse, but instead brings about correlations between the system and measuring device. For Leggett, the latter use of 'measurement' is a misguided "elevation of a particular form of interaction from a secondary and inessential ingredient of the measurement process to its defining characteristic" [115, 7].

In this section I will explore how the two types of quantum evolution – unitary dynamics and wavefunction collapse – can be combined to enable generalised measurement. Tuning the details of the interaction between a system and a meter – Leggett's inessential ingredient – provides additional insight and opportunities. The analysis provides a model for 'measurement' both in the fundamental sense associated with wavefunction-collapse, and also in the sense associated with parameter estimation – the process by which statistical inferences about the world are made. The latter is a third notion of measurement – the determination of an unknown classical parameter. Imagine a first quantum object (I shall call this the 'system') and a second 'meter' which are coupled through

$$\mathbf{H} = g\delta(t_0)\mathbf{A} \otimes \mathbf{B}; \tag{3.1}$$

here $\mathbf{A}$ is an observable (see Chapter 1) of the system and $\mathbf{B}$ is an observable of the meter. The time profile of the interaction is an unphysical delta function; this choice simplifies the calculations and can be relaxed to a top hat function (or other realistic profile) straightforwardly. The time evolution operator is then

$$U = \exp(-i\mathbf{A} \otimes \mathbf{B}) \neq \exp(-i\mathbf{A}) \otimes \exp(-i\mathbf{B}); \tag{3.2}$$

sometimes the $\otimes$ is dropped for brevity. Notice that if the system begins in an eigenstate of $\mathbf{A} := \sum_i a_i |a_i\rangle \langle a_i|$ and the meter begins in a eigenstate of $\mathbf{B} := \sum_i b_i |b_i\rangle \langle b_i|$, nothing happens save for the accumulation of a global phase $\mathrm{e}^{-ia_j b_k}$ to the joint system, which we can ignore. By arranging for the meter to be in a superposition of eigenstates of $\mathbf{B}$, the coupling induces an evolution of the meter which is conditional on whichever eigenstate the system is in. System and meter therefore become entangled. The unitary process can be thought of as a *quantum coherent premeasurement* of the system by the meter. The premeasurement here is a gradual process (occurring over some time interval) describing the accrual of correlations between two quantum systems through a mutual interaction. The premeasurement model explores the first link in a potentially infinite chain of correlations that might occur through interactions between quantum systems: as is customary, I assume that this chain is terminated, somewhere, by a projective measurement. In this chapter I mainly use 'detection' to refer to the ill-defined process at the end of the chain, which causes the wavefunction to collapse.

Generally neither system nor meter will be in an eigenstate of their respective measurement operators (although I assume that they begin in a factorizable state):

$$U\Big(|\psi_1\rangle |\psi_2\rangle \Big) = U \left(\sum_i c_i |a_i\rangle \sum_j m_j |b_j\rangle \right) \tag{3.3}$$

One can now impose our demarcation of the joint space into system and meter by stipulating the joint detection observable $\mathcal{AB}$. Choosing $\mathcal{A} = \mathbb{I}$ imposes 'system, meter'; making $\mathcal{B} = \mathbb{I}$ imposes 'meter, system'. In each case one *ignores the system.* Here I will stick to the first convention, and further set $\mathcal{B} = \sum_o \lambda_o |o\rangle \langle o|$ which I call the 'meter output observable'[1]. After detection, the probability distribution of the meter will be a mixture of distributions, each corresponding to an ignored classical configuration of the system

$$\begin{aligned} P_o &= \sum_i |\langle a_i| \langle o| U |\psi_1\rangle |\psi_2\rangle|^2 \\ &= \sum_i \langle \psi_1| \langle \psi_2| U^\dagger |a_i\rangle \langle a_i| \otimes |o\rangle \langle o| U |\psi_1\rangle |\psi_2\rangle \\ &= \langle \psi_1| \langle \psi_2| U^\dagger (\mathbb{I} \otimes |o\rangle \langle o|) U |\psi_1\rangle |\psi_2\rangle \,, \end{aligned} \tag{3.4}$$

where I used the fact that $\sum_i |a_i\rangle \langle a_i| = \mathbb{I}$, i.e. the $|a_i\rangle$ are a complete set of states.

[1] It is straightforward to generalise this to an continuous observable.

Now, using the above,

$$\begin{aligned} P_o &= \sum_{ijkl} c_k^* c_i m_j^* m_l \mathrm{e}^{ia_i b_j} \mathrm{e}^{-ia_k b_l} \langle a_i| \langle b_j| (\mathbb{I} \otimes |o\rangle \langle o|) |a_k\rangle |b_l\rangle \\ &= \sum_{ijkl} \delta_{ik} c_k^* c_i m_j^* m_l \mathrm{e}^{ia_i b_j} \mathrm{e}^{-ia_k b_l} \langle b_j|o\rangle \langle o|b_l\rangle \\ &= \sum_{ijl} |c_i|^2 m_j^* \mathrm{e}^{ia_i b_j} \langle b_j|o\rangle \langle o|b_l\rangle \mathrm{e}^{-ia_i b_l} m_l \\ &= \sum_{i} |c_i|^2 |\langle o|U_{\mathbf{B}(a_i)}|\psi_2\rangle|^2 \end{aligned} \tag{3.5}$$

on defining $U_{\mathbf{B}}(a_i)|\psi_2\rangle = \mathrm{e}^{ia_i\mathbf{B}}|\psi_2\rangle = \sum_j m_j \mathrm{e}^{ia_i b_j}|b_j\rangle$. Clearly each eigenstate of $\mathbf{A}$ causes an evolution of the meter initial state generated by $\mathbf{B}$ by a distinct 'distance' a_i (the corresponding eigenvalue). This evolution of the meter can be manifest as a probability density when a suitable meter observable $\mathcal{B}$ is chosen. If the system is in a state other than an eigenstate of $\mathbf{A}$, the quantum state of the meter is probabilistically in each of the possibilities, with the relative probabilities given by $|c_i|^2$. When the meter is then inspected, the distribution of outcomes will thus be mixed together.

Notice that if either $\mathbf{B}$ or $\mathcal{B}$ are poorly chosen the resultant probability distribution will not give information about the meter state. In particular, if $[\mathbf{B}, \mathcal{B}] = 0$, the output variable is a constant of the motion, and is preserved. It cannot then provide any more information after the interaction than it did beforehand. Choosing $\mathbf{B}$ and $\mathcal{B}$ to be mutually unbiased $\langle o|b_j\rangle = \frac{1}{\sqrt{d}}$ for discrete meters, or canonically conjugate $[\mathbf{B}, \mathcal{B}] = i$ for continuous variable meters, will lead to the most pronounced statistical changes. The choice makes the evolution generated by the interaction most visible, and hence is a natural choice for the most effective measurements: the latter was chosen by von-Neumann in his famous model of quantum measurement [144, 133]. For continuous variable quantum systems, the Hamiltonian generates translations in the conjugate variable: for instance $[\mathbf{k}_x, \mathbf{x}] = i$ implies that $\mathrm{e}^{-ia\mathbf{k}_x}|x\rangle = |x - a\rangle$, see Ref. [45].

One may characterise quantum measurements by their statistical properties – including the detection observables in the definition of a measurement. Another approach would be to remove the measurement choice from the analysis, characterising quantum measurements by induced trajectories in Hilbert space. Because, roughly speaking, the error in the measurement (at the meter) is related to a disturbance, or back action (in the system), one could alternatively characterise measurements by their induced back action.

3.1 Strong measurements

If the coupling strength, meter initial state, and output variable all conspire to trigger a maximally informative change in probability in the meter detection, we then have an explicit dynamical model for what goes in in a projective measurement as introduced in Chapter 1. When the meter is inspected, it is found in one of a number of orthogonal possibilities: from Equation (3.5) the relevant probability is $|c_i|^2$. Furthermore, because system and meter are perfectly correlated, observation of outcome λ_o causes the system to collapse into the corresponding eigenvalue of $\mathbf{A}$.

Importantly, this line of reasoning does not solve the measurement problem! One can continue to chain measurements together using the premeasurement model, but eventually one has to assume a projective measurement to collapse the chain of correlations.

In the classical world, measurement is synonymous with simply 'reading-off' values displayed on a suitable meter. One must carefully consider how the concept of measurement changes in the quantum domain. A strong measurement is akin to simply inspecting an object, discovering its state among a number of possibilities. Imagine that one is given an unknown state drawn from an orthonormal basis. What inferences may be made about the identity of the state after only a single measurement, i.e. after only single sample of the resultant probability space? Since the state $|j\rangle$ is unknown, the probability distribution of the meter is also unknown: is it δ_{o1} or δ_{o2}? The distinguishability of the distributions helps. Since there is a one-one relationship between the measurement outcome and the initial state, the latter can be inferred safely from the former by a single sample. This can be formalised by calculating a suitable statistical distance: the Hellinger distance, for example (see Chapter 2):

$$D_H := \sqrt{1 - \sum_i \sqrt{P_i Q_i}} \tag{3.6}$$

between the probability distribution corresponding to the possibility under question, and the complementary probability of the system being in any other possibility. For the delta function this evaluates to 1: the possibilities are perfectly distinguishable, and we have a strong measurement.

3.1.1 Example: CNOT gate as a strong measurement

If one wishes to retain a model for the measurement, including a description of the meter itself, a two qubit example is demonstrative. One may use the CNOT gate,

introduced in Chapter 2. Take the interaction Hamiltonian Equation (3.1), and make the following choices. Let $\mathbf{A} = |a_1\rangle\langle a_1|$, a projector type operator; let the initial meter state be an eigenstate of the meter output variable $|\psi_2\rangle = |o\rangle$; let $\mathbf{B}$ be a mutually orthogonal observable, so that $\langle o|\mathrm{e}^{-i\pi\mathbf{B}}o\rangle = 0$ – in other words, a NOT operation happens if $g = \pi$. By inspection

$$U = \mathrm{e}^{-i\pi|a_1\rangle\langle a_1|\mathbf{B}}\left(\sum_i c_i|a_i\rangle\,|o\rangle\right) \tag{3.7}$$

the meter qubit is flipped only if the system qubit is in $|a_1\rangle$; otherwise the identity transformation $\mathrm{e}^{i\hat{0}} = \mathbb{I}$ is applied. Thus we have found a Hamiltonian which generates a CNOT operation. Clearly the measurement of $\mathcal{B}$ can distinguish which eigenstate of $\mathbf{A}$ the system was in, in a single shot.

3.1.2 Example: null result measurements

Null result measurements are those which probe for a particular measurement outcome, and discard all experimental runs where that outcome was actually found. One only keeps runs where the detector did not click. Consider the interferometer introduced in Chapter 2. By using a beam splitter to spatially separate two classical states, one can place a detector in only one branch. In fact by simply blocking this branch with a barrier one can perform a null result measurement: if a particle passes the beam splitter and is not absorbed into the barrier, then one has determined which branch it is in without a measurement result (i.e. absorption) actually having been reported.

One might imagine a null result measurement being modelled in the following way: if a positive result is recorded, the state becomes updated according to the collapse postulate; but if a negative result is recorded (i.e. no result is recorded) then the identity transformation is applied. For a two level system with basis $\{|0\rangle, |1\rangle\}$, when the result is positive, the POVM element is $|0\rangle\langle 0|$. When it is negative, however, by the completeness relation of POVMs, the complementary POVM element must also be a projector $|1\rangle\langle 1|$, which can cause the largest possible disturbance to the quantum state. The completeness of quantum measurements means that a null result measurement is entirely equivalent to an inspection of *both* possibilities! There is no way, within quantum theory, to make an incomplete measurement: the logical inference *about* the particle is not distinguished from the physical influence *upon* the particle. A consequence is that the quantum state of the particle is significantly

disturbed despite the apparent lack of any physical interaction. A CNOT gate can be used to perform null result measurements, see Chapter 4.

3.2 Weak measurements

A single shot inference about the state of the system is not possible if the distinguishability of the associated meter states is broken. It can be broken, for example, by adding quantum noise to the meter by initialising it into a suboptimal pure state. Alternatively, noise can be added incoherently, by randomising over meter states. In either case, the statistical distance between the meter probability distributions for each eigenstate of the control observable becomes smaller.

One may define a weak measurement as a combination of unitary premeausurement and choice of detection on the meter that leads to a nonzero but less than maximally informative change in probability at the chosen output variable. The Hellinger distance, or some other metric, can be used to quantify this idea.

It is insightful here to consider the evolution of the system state, conditional on the detection outcome at the meter. The POVM description of general quantum measurements (see Chapter 1) will be employed. After the premeasurement, when the meter is found in state $|o\rangle$ the system is updated to

$$\begin{aligned} |\psi_1\rangle \to |\psi_o\rangle &\propto \Big(\mathbb{I}\otimes\langle o|\Big)U\Big(\,|\psi_1\rangle\,|\psi_2\rangle\,\Big) \\ &\propto M_o\,|\psi_1\rangle \end{aligned} \tag{3.8}$$

with

$$M_o = \Big(\mathbb{I}\otimes\langle o|\Big)U\Big(\mathbb{I}\otimes|\psi_2\rangle\,\Big). \tag{3.9}$$

This is an operator on the Hilbert space of the system only. Note that the system is updated to a pure state. The proportionality constant is the norm of the post measurement state:

$$\sqrt{\langle\psi_1|M_o^\dagger M_o|\psi_1\rangle} \tag{3.10}$$

its square

$$P_o = \langle\psi_1|M_o^\dagger M_o|\psi_1\rangle = \langle\psi_1|E_o|\psi_1\rangle \tag{3.11}$$

is the probability of outcome o occurring. We can calculate the Hellinger distance between distributions P_o for different orthogonal states $|\psi_1\rangle\,,|\phi_1\rangle$ of the system:

$$D_H = \sqrt{1-\sum_o \sqrt{\langle\psi_1|E_o|\psi_1\rangle\,\langle\phi_1|E_o|\phi_1\rangle}} \tag{3.12}$$

by my definition, $D_H < 1$ implies a weak measurement. This kind of 'coherently weak' measurement translates into a variable back action on the system – roughly speaking, weaker measurements have less back action that stronger ones. If $E_o = \mathbb{I}/d$, the system is not updated at all after the measurement, and $D_H = 0$.

To deal with the possibility of a poorly chosen meter output variable, one may introduce a metric on the meter states themselves, rather than the meter probability distributions. In this way, use of the optimal meter output variable is implied. This induces a notion of measurement strength that I use in Chapters 5, 6 and 7; it may be quantified, for example, by the Bures distance D_B (see Chapter 2).

3.2.1 Example: Gaussian probe

I will give an example with Gaussian wavefunctions. Let the system be a qubit and let the meter be described by a continuous variable with initial state

$$\psi(x) = \sqrt{\frac{1}{\sqrt{2\pi}\Delta_x}} \exp\left(-\frac{x^2}{4\Delta_x^2}\right) : \tag{3.13}$$

the standard deviation, or 'width' of this state Δ_x will be important. Now couple qubit and meter through

$$\mathbf{H} = g\mathbf{A}\mathbf{k}_x \tag{3.14}$$

so that one has, upon inspecting the meter in the position basis (note $[\mathbf{B}, \mathcal{B}] = [\mathbf{k}_x, \mathbf{x}] = i$)

$$p(x) = \sum_i |c_i|^2 |\psi(x - ga_i)|^2, \tag{3.15}$$

i.e. a mixture of shifts in the meter x variable. Take two neighbouring eigenstates of $\mathbf{A}$:

$$D_H = \sqrt{1 - e^{-\frac{g^2(a_1-a_2)^2}{8\Delta_x^2}}}. \tag{3.16}$$

The statistical distance depends on the ratio $g : \Delta_x$; the measurement strength can be controlled by varying either quantity. A strong measurement, $D_H = 1$, is possible through either: $\Delta_x \to 0$ or $g \to \infty$. Strong measurements here relate to Gaussians separated by a large distance (compared to their width). A weak measurement is anything else, where the Gaussians overlap. One may have course to consider the limit $g \ll \Delta_x$, where the Gaussians overlap almost completely: see Chapter 5. This is a coherently weak measurement: it has the property than information gain from the meter and disturbance to the system are in a tight relationship.

3.2.2 Noisy measurement

Characterisation of quantum measurements by a metric on probability distributions, for example the Hellinger distance, implies that a strong measurement with a corrupted output should also qualify as weak. The back action on the system is not reduced, however: in fact the system can end up completely mixed. Consider the CNOT strong measurement – system and meter are coherently coupled, and the meter is strongly measured. The Hellinger distance can now take any value, including the maximum, strong measurement value $D_H = 1$. Generally speaking this leads to the most disturbance in the system: it may be projected onto an eigenstate of the system observable $\mathbf{A}$. But if the result of the measurement is now passed through a scrambler, the Hellinger distance may decrease with no associated change in the degree of back action. The measurement result is deleted by the scrambler: the meter is in a mixed state, and the system is therefore also in a mixed state. It is thus impossible to observe quantum interference in either system (unless the measurement result is recovered somehow). The underlying quantum states of the meter do not overlap. Only our subjective knowledge of the meter states renders them indistinguishable. The strength of the measurement may not be correlated with the induced back action at all.

Characterisation by a metric on the quantum states themselves, for example with the Bures distance, changes things. Now a noisy measurement can still be classified as strong because the metric is applied before the scrambling has taken place. Of course one may define mixed quantum states of the meter after the scrambling, and the quantum metrics now also classify the measurement as weak.

Chapter 4

Leggett and Garg's theorem

In the introductory chapter, I discussed quantum theory's measurement problem. In this chapter, we will see how the problem can be addressed in the laboratory. Such investigations are motivated by Schrödinger's cat paradox: a thought experiment which highlights the apparent absurdity of extending the predictions of quantum theory beyond the atomic world to the everyday world of human experience [177]. The thought experiment is as follows. Imagine a macroscopic object – a cat, for example – which is coupled to a microscopic object – an atom, say – through a series of intermediate interactions. This is called a von-Neumann chain. If the microscopic object obeys the laws of quantum mechanics, and begins in a superposition of its classical configurations, then the chain will act to correlate the micro- and macroscopic objects. The atom and cat become entangled. If the decay of the atom triggers a chain of events leading to the death of the cat (a scintillation detector is triggered, an electrical circuit is formed, a hammer drops, a vial of poison is released, etc), the smearing of the atom over its configurations (decayed and not decayed) is amplified to the level of the cat, and the whole edifice of atom and cat is in a macroscopic superposition. The cat can non longer be said to be either dead or alive. The *gedanken-experiment* is paradoxical in two key respects, given that quantum mechanical interference effects are not seen in everyday life at the macro scale.

On the one hand, an obvious question arises: 'how does the classical world arise out of the quantum world?', or, 'why don't we observe interference of macroscopic objects?'. This is *the soft measurement problem*. The phenomenon of decoherence, discussed in Chapter 1, describes the loss of knowledge about a quantum state: the resultant reduction in purity is then seen to be responsible for the lowering of the visibility of interference. Decoherence is thus an adequate solution to the soft measurement problem. As in classical mechanics, some effects are hard to see if good

control, or fine enough measurement accuracy cannot be achieved. In the cat paradox, if one ignores the atom, for example, the density operator of the cat takes on a diagonal form (in the $|\text{alive}\rangle, |\text{dead}\rangle$ basis), and no interference is possible.

On the other hand, there is the *hard measurement problem*, which is more difficult to solve. It is possible in principle to arrange for a superposition of dead cat and alive cat, but upon observation only a single outcome will be observed. The harder question, then, is 'what triggers wavefunction collapse?'. Decoherence is not a satisfactory answer. It cannot account for a how a superposition ends up in a classical possibility. An objective micro/macro distinction is needed. As Bell has said, "So long as the wave packet reduction is an essential component [in quantum mechanics], and so long as we do not know exactly when and how it takes over from the Schrödinger equation, we do not have an exact and unambiguous formulation of our most fundamental physical theory" [19].

There are a number of proposals that have been proffered to address the problem: for a comprehensive overview see, e.g. [205, 15]. These include those that claim that wavefunctions never collapse (the many worlds theory [202]) and those that interpret the wavefunction as an expression of knowledge (the ψ-epistemic program [78]). These proposals fight for supremacy on a purely intellectual battleground. Proponents from each school may argue for the virtues of self-consistency, elegance, or intuitiveness: but ultimately it is personal preference that will be the kingmaker for the quantum interpretations.

Yet another class of proposals make empirically testable changes to our physical laws: for example adding non-linear terms to the Schrödinger equation. These theories, known collectively as 'objective collapse theories' are in a sense *more scientific* than other approaches, since, as we shall see below, they make predictions that are amenable to experimental scrutiny and that disagree with the predictions of current models.

Leggett and Garg (LG) defined *macrorealism* as a class of theories containing the objective collapse theories. The unifying essence of macrorealist theories is thus: there is an objective separation between the mundane classical world (where objects always realise *one* of their possible configurations) and the more exotic world of quantum theory (where objects in some sense inhabit many classical configurations simultaneously). This is true regardless of the precise mechanism by which the new physics prohibits macroscopic quantum superpositions from persisting: as we shall see below, objective collapse theories achieve the separation through dynamical laws.

This chapter is arranged as follows. The objective collapse theories will be explained in Section 4.1; LG's argument (including their inequality) will be presented in Section 4.2. I then explain and motivate null-result measurements as the most convincing type of measurement for testing the LGI in Section 4.3. LG's arguments are critically analysed in Section 4.4, which offers comparisons to other no-go theorems. Weak measurement and stationarity will be introduced in Sections 4.5 and 4.6 as somewhat misguided approaches to LG's protocol. I survey some notable experimental tests of the inequality in Section 4.7. The final Section 4.8 and 4.9 will introduce my theoretical extension to LG's work, and an experimental test conducted in Oxford in 2011.

4.1 Motivating macrorealism: objective collapse of the wavefunction

Objective collapse theories solve the measurement problem by unifying the two types of dynamics in quantum theory. Instead of arbitrarily choosing when to apply the Schrödinger equation on the one hand, or the measurement update rule on the other, one can simply apply a generalised dynamical law in all cases. A new Schrödinger equation is defined, which augments the unitary evolution with a random process which which acts to drive the state vector into a classical state. This powerful and elegant idea was first rigorously developed by Gihrardi, Rimini and Weber [70] (GRW), and later by Pearle [155]. GRW's theory is such that 'collapse' becomes more probable the more particles are involved in a superposition. To recover the classical and quantum limits to the new equation, certain free parameters must be fixed. The free parameters must be such that there is no contradiction with experiments showing quantum interference. To solve the hard measurement problem, the choice of parameters must also prohibit macroscopic superpositions. Schrödinger's cat is then only dead and alive for a split second, and the von-Neumann chain has an end. New fundamental constants of nature (the free parameters in the non-linear Schrödinger equation) dictate when each of the previous equations is the appropriate limiting case.

Roger Penrose is another who argues in the same vein: his idea is that gravity will not permit superposition of mass distributions whose gravitational self-energy exceeds that of, e.g. a single graviton [156, 157]. Penrose's theory and that of GRW-Pearle's come under the heading of macrorealism and the spirit in which they approach the cat paradox is clear: as Leggett has put it, they hold that "QM is not the complete

truth about the world; at some level between that of an atom and that of human consciousness, other non-quantum mechanical principles intervene." [118].

For the remainder of the chapter we shall not be so concerned with the physics of any particular theory of macrorealism. The essential feature is that the hard measurement problem is solved by explicit demarcation of the quantum world (where superposition is permitted) from the classical world (where superposition is forbidden) from within the new physical theory. I shall only be concerned with the logic which is common to all such theories, and how they may be constrained through experiment.

4.2 Derivation of a temporal inequality

A quarter of a century ago, LG published a landmark paper entitled "Quantum Mechanics versus Macroscopic Realism: is the Flux There when Nobody Looks?" [120]. In this paper they set out a procedure to pit the predictions of quantum mechanics against macrorealism. The context of their discussion is experiments with Superconducting Quantum Interference Devices (SQUIDs), which are micrometer sized circuits where the current can be in a superposition of flowing in opposite directions simultaneously (see Chapter 2). LG's theorem is general enough to apply to any physical system, however. It addresses the conjunction of the following assumptions:

1. (MR) *Macrorealism per se*: A macroscopic system with two or more macroscopically distinguishable states available to it will at all times *be* in one or the other of these states.

2. (NIM) *Non Invasive Measurability at the macroscopic level*: It is possible in principle, to determine the state of the system with arbitrarily small perturbation on its subsequent dynamics.

In later work a third assumption [117] was included for completeness

3. (IND) *Induction*: The properties of ensembles are determined exclusively by initial conditions (and in particular not by final conditions).

MR can be split into two parts.

4.2.1 The macro- prefix

This prefix denotes that something applies at the macroscopic scale that does not apply on other scales. When it is prepended to 'realism', this is exactly what is intended: that the property of realism (defined below) applies to macroscopic objects but not to microscopic ones. LG have provided two tools: the first is a test of realism; the second a methodology for probing the nature of the quantum/classical boundary. The methodology, sometimes called the Leggett program [195], is simple: since the test of realism, to be explained below, is capable of convincingly ruling it out, one applies it to larger and larger objects. Macrorealist theories, if they are to have any scientific content, will be required to provide parameter regimes for which quantum theory breaks down (see Figure 4.1). If $\{r_i\}$ is a set of parameters which classify the degree of macroscopicity, then progressively falsifying 'realism at the level of $\{s_i\}$' can extend the necessity of a quantum description of nature (or more strictly, the inadequacy of the macrorealist perspective) into that particular region of parameter space. In this way, any given macrorealist theory which holds 'realism is true at the level of $\{q_i\}$' can be ruled out by violating the Leggett-Garg inequality when one of the parameters extends into a forbidden region. Of course if one fails to violate an inequality, then nothing much is said about the whole problem: apart from 'I tried to unequivocally show the existence of macroscopic quantum coherence in my device, having two states differing in mass distribution by so many kilograms and involving so many particles [for example], but failed.'

Since one can consider the logic of the Leggett-Garg inequality independently of the micro/macro distinction, I shall often do so in this chapter. The ultimate aim of LG is undoubtedly for tests in the macroscopic domain – but in the interim before technology enables this goal, experiments may be confined to the microscopic realm, where realism (R) feels a more appropriate term. I shall thus use R or MR depending on the context.

4.2.2 Realism

'Realism' captures most of our intuition about the classical world insofar as it is contrasted by quantum theory. Although it is a term which carries some philosophical baggage, LG gave it an unambiguous meaning: for a system with a multitude of configurations available to it, it occupies exactly one of these configurations at any given time[1]. For example, a system with two classical configurations labelled $\uparrow$ and

[1]We might consider 'value definite' as a better choice of words.

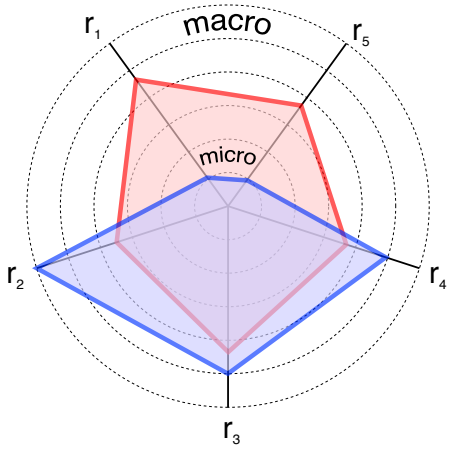


Figure 4.1: Schematic for the parameter space on which macrorealist theories impose limits on the validity of quantum theory (for example the blue polygon). Experimental data can falsify such theories as long as they falsify macrorealism at a level involving a set of parameters (for example the red polygon), some of which are larger than is allowed by the considered theory of macrorealism. The parameters $\{r_i\}$ typically correspond to mass, number of particles etc.

$\downarrow$, one can describe its dynamics fully by specifying one of these possibilities for every point in time. If one considers an ensemble of such systems, the collective state at each time will be a statistical 'mixture' of the two possibilities - in other words a subensemble characterised by probability p will inhabit the $\uparrow$ configuration, and the complementary subensemble characterised by $1-p$ will inhabit $\downarrow$. The system may transform from one configuration to the other and back again: this may happen at some characteristic rate, or not at all. In the following we will consider the configuration of the system on a coarse grid of three instants. The most general evolution will be probabilistic – one can fully describe the dynamics by a set of (in this case eight) probabilities $\mathbb{P}_{abc} = \mathbb{P}(Q(t_1) = a, Q(t_2) = b, Q(t_3) = c)$ where $Q(t_i) = \pm 1$ when the state of the system is $\uparrow$ or $\downarrow$ at time t_i accordingly. Naturally these probabilities are non negative and form a complete set in the sense that $\sum_{abc=\pm 1} \mathbb{P}_{abc} = 1$. Table 4.1 shows how the probabilities attach to the possible time evolutions.

Clearly the system samples the eight possible evolutions with relative frequency determined by the distribution $\mathbb{P}$. In other words a fraction $\mathbb{P}_{abc}$ of the ensemble evolves according to the evolution $Q(t_1) = a, Q(t_2) = b, Q(t_3) = c$. If one calculates the quantity

$$F_{\mathrm{LG}} = Q(t_1)Q(t_2) + Q(t_2)Q(t_3) + Q(t_1)Q(t_3) + 1 \tag{4.1}$$

for each of the eight possibilities, then one finds $F_{\mathrm{LG}} = 0$ or $F_{\mathrm{LG}} = 4$. When this

t_1	t_2	t_3	$\mathbb{P}$	$f_{\rm LG}$
$\uparrow$	$\uparrow$	$\uparrow$	$\mathbb{P}_{111}$	4
$\uparrow$	$\uparrow$	$\downarrow$	$\mathbb{P}_{11\bar{1}}$	0
$\uparrow$	$\downarrow$	$\uparrow$	$\mathbb{P}_{1\bar{1}1}$	0
$\uparrow$	$\downarrow$	$\downarrow$	$\mathbb{P}_{1\bar{1}\bar{1}}$	0
$\downarrow$	$\uparrow$	$\uparrow$	$\mathbb{P}_{\bar{1}11}$	0
$\downarrow$	$\downarrow$	$\downarrow$	$\mathbb{P}_{\bar{1}\bar{1}\bar{1}}$	0
$\downarrow$	$\downarrow$	$\uparrow$	$\mathbb{P}_{\bar{1}\bar{1}1}$	0
$\downarrow$	$\downarrow$	$\downarrow$	$\mathbb{P}_{\bar{1}\bar{1}\bar{1}}$	4

Table 4.1: Macrorealist theories prescribe probabilities for each classical trajectory of a two level system inspected at three distinct times. $Q = 1$ when the state of the system is $\uparrow$, and $Q = -1$, written $\bar{1}$, when the system is $\downarrow$.

quantity is ensemble-averaged, one trivially finds

$$4 \geq \langle F_{\rm LG} \rangle \geq 0. \tag{4.2}$$

The linearity of the ensemble average $\langle \bullet \rangle$ implies

$$f_{\rm LG} = K_{12} + K_{23} + K_{13} + 1 \tag{4.3}$$

$$K_{ij} := \langle Q(t_i) Q(t_j) \rangle. \tag{4.4}$$

By distributing the expectation over the sum, one defines three ensembles. Determining each correlator K_{ij} in a separate ensemble, where a measurement is only made at t_i and t_j is legitimate – the ensembles are equivalent – unless the very act of measuring (or not measuring) at other times impacts upon the statistics of that ensemble. Conjoining NIM and IND to MR then implies the Leggett-Garg inequality (LGI)

$$4 \geq f_{\rm LG} \geq 0. \tag{4.5}$$

To determine each correlator K_{ij} experimentally, one makes measurements at two distinct times, and counts the proportion of runs which give each of the four possible configurations for the two measurement outcomes:

$$K_{ij} := \langle Q(t_i) Q(t_j) \rangle = \frac{N_{++} + N_{--} - N_{+-} - N_{-+}}{N_{++} + N_{--} + N_{+-} + N_{-+}}. \tag{4.6}$$

$N_{\pm\pm}$ is the number of trials which found the system in state ± 1 at an earlier time t_i and in state ± 1 at a later time t_j. The macrorealist predictions for these correlators may be calculated from the macrorealist table, by choosing the two appropriate rows for each two-time correlator (tracing out the column for whichever time is not needed), see Table 4.2.

t_1	t_2	$\mathbb{P}$	$Q(t_1)Q(t_2)$
$\downarrow$	$\downarrow$	$\mathbb{P}_{\bar{1}\bar{1}1}+\mathbb{P}_{\bar{1}\bar{1}\bar{1}}$	1
$\downarrow$	$\uparrow$	$\mathbb{P}_{\bar{1}11}+\mathbb{P}_{\bar{1}1\bar{1}}$	-1
$\uparrow$	$\downarrow$	$\mathbb{P}_{1\bar{1}1}+\mathbb{P}_{1\bar{1}\bar{1}}$	-1
$\uparrow$	$\uparrow$	$\mathbb{P}_{111}+\mathbb{P}_{11\bar{1}}$	1

Table 4.2: Each correlator may be calculated from the macrorealist table by choosing the two appropriate rows for each two-time correlator (tracing out the column for whichever time is not needed). One multiplies each pairwise sum of probabilities by ± 1 according to whether that row was a correlation or anti-correlation. Shown is the K_{12} case.

According to quantum theory, of course, the quantum state is generally 'reset' by a measurement: the unavoidable back-action in a projective measurement, where the state is updated to en eigenstate of the measured observable, is invasive. This leads quantum mechanics to predict a violation of the LG inequality. For a qubit undergoing dynamics governed by

$$U(t_i, t_j) = U((j-i)\tau) = \cos(j-i)\tau\mathbb{I} + \sin(j-i)\tau\sigma_x, \tag{4.7}$$

i.e. coherent oscillations with a fixed period, the correlator is $K_{ij} = \cos((j-i)\tau)$ and hence

$$f_{\mathrm{LG}} = 2\cos\tau + \cos 2\tau + 1, \tag{4.8}$$

which takes the value $f = -0.5$ for $\tau = 2\pi/3$, violating the inequality $f \geq 0$ predicted under MR $\cap$ NIM $\cap$ IND. This is independent of the initial state, see Figure 4.2. As is plain from inspection of the Tables 4.1 and 4.2, the argument constraining the macrorealist to non-negative values for f also does not depend on the primary system's initial state. For a qubit undergoing Rabi oscillations, the variable τ can be replaced by θ, the angle that the state vector is rotated by in the interval between two measurements.

An important, but frequently overlooked, property of the Leggett-Garg inequality is that it is trivially satisfied if the correlators (and hence f) are determined with only three measurements, averaged over a single ensemble. This follows because each of the measurement results will be one of the eight macrorealist possibilities: see Figure 4.3.

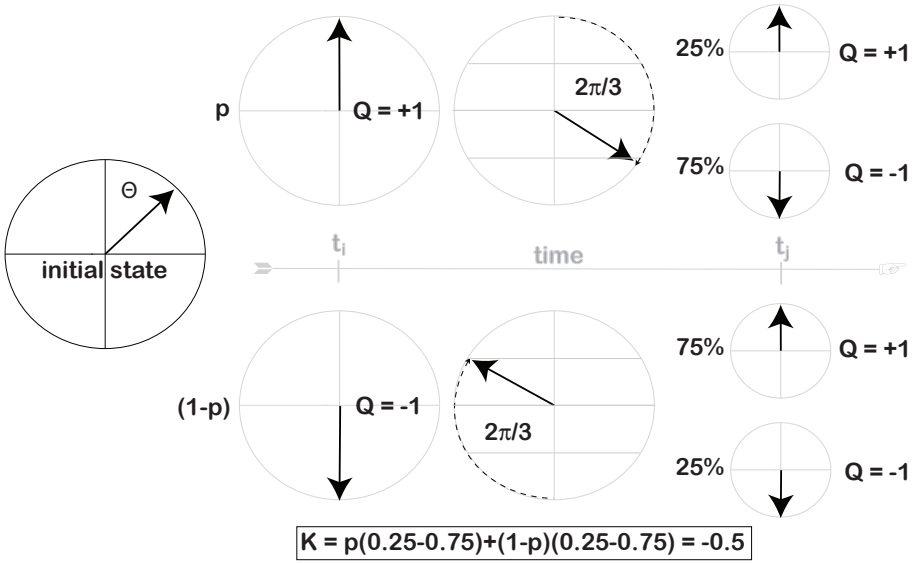


Figure 4.2: According to quantum mechanics, the correlator (Equation 4.6) is independent of the initial state; this is demonstrated on a slice through the Bloch sphere. At the early time t_i, a measurement deletes information about the quantum state – the correlation with a later measurement at t_j is only sensitive to the angle of rotation induced by the time evolution over the interval. Shown is an angle of $2\pi/3$.

4.3 Null-result measurements: ideally non-invasive?

Given NIM, the macrorealist believes that the state of the two level system may be measured in a non-invasive way. The occurrence of a measurement at t_2 in the first and second, but not the third subensemble should not matter unless the measurement is invasive. The physical implementation of the t_2 measurement is therefore of key importance. If the macrorealist is unconvinced by the measurement procedure, he will cry 'clumsy measurement!'. If the macrorealist accepts that the particular physical implementation of the measurement process is non-invasive then either MR or IND must be rejected [207][2].

If the measurements are sufficiently careful that it makes no difference to the behaviour of Q if a measurement is made or not, then the three separate ensembles become equivalent, and the LG inequality may be derived. LG suggested a plausible candidate for such a measurement scheme. Given MR, we can design a measurement apparatus which only couples to one of the possible states of the system (for instance a device which only clicks if the system inhabits $\uparrow$ but does nothing if it inhabits $\downarrow$). As long as the detector does not click, then the system has not been interacted with; but we may still infer the state of the system: it must be in the other state. Repeating

[2]The possibility of IND being false, i.e. backwards causation, is not considered any further.

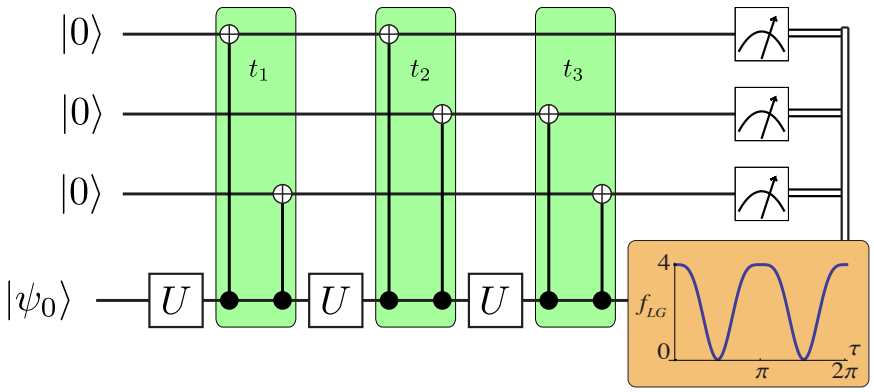


Figure 4.3: A circuit that might be suggested to test the LG inequality in a single ensemble, but which will fail to violate the bound on macrorealism despite the existence of coherent superpositions. The vertical lines are CNOT gates, used to measure the system at each of the instants t_1, t_2, t_3. Two gates map information about the σ_z variable onto three ancillary qubits. In the intervening times the system qubit is evolved according to $U(\tau) = \cos\tau\mathbb{I} + i\sin\tau\sigma_x$. Each of the ancillary qubits is measured to determine (after ensemble averaging) K_{12}, K_{23}, K_{13} respectively. The inset shows that the LG inequality will be obeyed for any value of $\tau := t_3 - t_2 = t_2 - t_1$: therefore the macrorealist sees little to trouble him. Adapted from *New Journal of Physics*, 14(5):058001, May 2012

this procedure with a complementary measuring device which only responds when the system is in the other state, allows one to gather a complete set of measurement results by only keeping results which reported the null outcome (or when the detector failed to click). This procedure is known as an ideal negative result measurement [114]. According to LG, such a procedure is the pinnacle of careful measurement in the eyes of a macrorealist, as it exploits his belief in MR. This is especially the case when the two states can be spatially separated. As an example of such a measurement, consider the Stern Gerlach (SG) apparatus in Figure 4.4. An SG gradient can 'open' a beam of spin-1/2 particles, separating them into spin up and spin down along a certain direction. Two-time correlators of the type necessary for an LG test can be non-invasively determined by a negative result measurement (achieved by blocking one branch of the beam), followed by a unitary evolution and a standard measurement. The first measurement can be thought of as a non invasive preselection of the spin into one state or the other.

To implement such a measurement scheme, one further resolves each of the three ensembles discussed above into two subensembles in order to construct both kinds of negative result measurements.

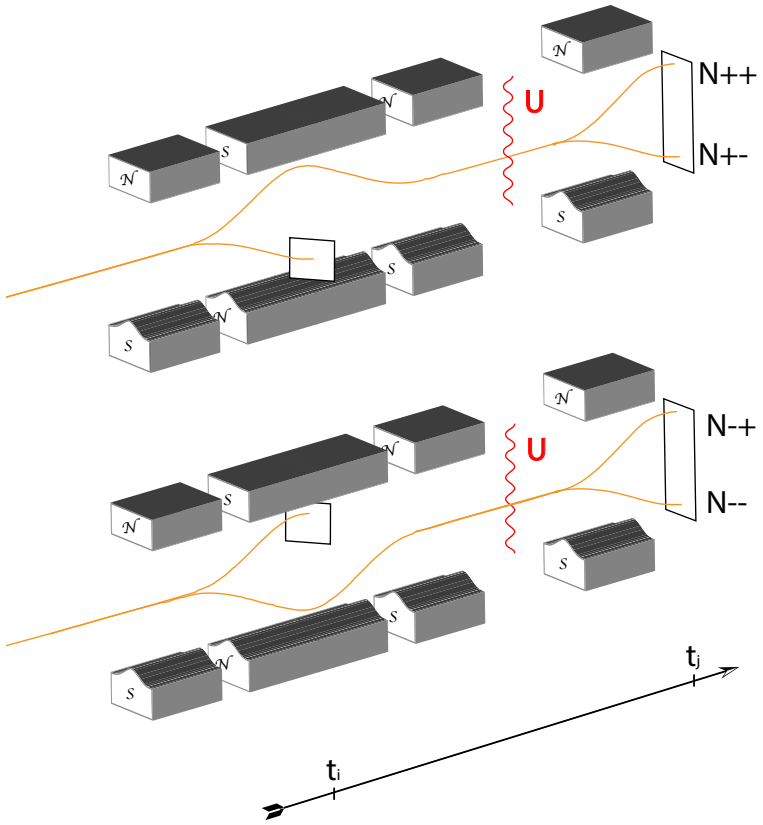


Figure 4.4: Null-result measurements of the autocorrelation of the z-projection of a spin with Stern Gerlach magnets. One must count correlations versus anticorrelations of the same variable over time.

4.4 Loopholes and drawbacks

Comparing the LG inequality to established no-go theorems in quantum mechanics highlights some of its drawbacks. The most striking difference in the formalism is the lack of the phrase 'hidden variables'. Hidden variables, sometimes called 'ontic states' are very prominent both in historical no-go theorems such as Bell's and Kochen and Specker's (introduced below), and in modern works [18, 17, 106, 38, 165]. The ontological-models framework posits such variables, labelled λ, as the real state of affairs in any experiment [78]. The quantum state might feature amongst λ, or it might be considered as merely a probability distribution over λ. The hidden-variable program is motivated by a number of hopes: to restore determinism, for instance, to remove action-at-a-distance, or to solve the measurement problem [57, 123]. Discussion of hidden variables is absent from LG's writings. A treatment of the LG theorem within this framework is given by Timpson and Maroney [199]. Although

a popular conceptual device λ is not a necessary component in the explanation of no-go theorems. Below I describe the Bell and Kochen-Specker theorems, without mentioning hidden variables.

4.4.1 LG vs. Bell

LG's inequality is inspired by the Bell inequality – the former has been called a 'Bell's inequality in time' to reflect the formal similarity. Among the theorems that pit quantum mechanics against rival possibilities, Bell's theorem is thought of as the gold standard. Although historically prior to LG's work (by some twenty years), it can be described within the formalism I have already introduced. Instead of considering the state of a single particle, represented by $Q = \pm 1$, at three distinct times, consider the state of two particles represented by A and B. The two particles are emitted simultaneously from a source, and travel in opposite directions to distant locations. Experimenters at each station, Alice and Bob, then each choose from a set of two possible measurements on their particle: a or a' for Alice, b or b' for Bob. They record the results, which are compared at a later time in coincidence: by repeating this experiment many times, correlators

$$K_{ab} = \langle A(a)B(b)\rangle \tag{4.9}$$

can be determined. Notice that each correlator is defined on a separate, and *a priori* distinct, ensemble. For example, $K_{a'b}$ is determined in an ensemble where Alice chose to measure a' and Bob chose to measure b.

The analogue of NIM is then introduced, to motivate the premise that the distinct ensembles are in fact equivalent. In Bell's theorem, it takes a stronger form, and is called 'locality'. Due to the spacelike separation of Alice's choice from Bob, and his from hers, we are to believe that no influence propagating at subluminal speeds from one location to the other would have time to influence the result. The following linear combination of correlators

$$f_{\text{Bell}} = \langle A(a)B(b)\rangle + \langle A(a')B(b)\rangle + \langle A(a)B(b')\rangle - \langle A(a')B(b')\rangle \tag{4.10}$$

should then be equal to the average taken as if all measurements were performed on a single ensemble:

$$f_{\text{Bell}} \sim \langle A(a)B(b) + A(a')B(b) + A(a)B(b') - A(a')B(b')\rangle\,. \tag{4.11}$$

By a similar argumentation, since A and B can only take values ± 1, one recovers the Bell inequality

$$f_{\text{Bell}} \leq 2. \tag{4.12}$$

This inequality, in its purest logical form, is due to Boole [122, 80, 79, 23]. The original derivation of Equation (4.12) by Bell involved only three correlators – see Bell's original paper [18]. The derivation of a four correlator inequality can be found in Clauser, Horne, Shimony and Holt [38]. Of course in reality, if all measurements were made in each run of the experiment, the inequality would be trivially satisfied[3]. No violation would be found [119]. Experimental tests of the Bell inequality, beginning with Aspect's famous demonstration [11], continue to find violations and are now chiefly aimed at closing various loopholes; for example those associated with low detection efficiency [155, 168].

The LG theorem has certain benefits over the Bell theorem. The former involves only a single system, and can be violated with a maximally mixed (lowest purity) state. The latter requires two systems, and further, that a high purity entangled state is engineered. For many, however, the LG theorem is significantly weaker than the Bell theorem, because the assumption of locality is better motivated than NIM. Special relativity, which denies the possibility of a causal connection between spatially separated locations, is a well confirmed and established theory of nature. NIM is a hangover from classical physics – as quantum theory has become more routine, physics has rather forsaken this idea. Of course, the ideal negative result measurements are intuitively convincing, and if one can spatially separate the states then special relativity can be used to motivate NIM, too.

4.4.2 LG vs. Kochen-Specker

Another touchstone for the special nature of quantum theory is Kochen and Specker's proof of the failure of non-contextuality [106]. Quantum mechanics is more easily shown to be at odds with non-contextuality than it is with locality: in the latter case one must prepare a special quantum state (which will be entangled, i.e. non-separable), and also achieve the experimentally challenging goal of space-like separation of measurements. The Kochen-Specker theorem can be made state-independent, and requires no spatial separation. It is perhaps inferior to the Bell theorem however, because it relies more heavily on quantum-mechanical concepts in its description.

[3]Because a and a' correspond to non-commuting observables in quantum theory, they are deemed incompatible and may not be co-measured.

	c_1	c_2	c_3	$\prod$
r_1	$\sigma_z\mathbb{I}$	$\mathbb{I}\sigma_z$	$\sigma_z\sigma_z$	$+\mathbb{I}$
r_2	$\mathbb{I}\sigma_x$	$\sigma_x\mathbb{I}$	$\sigma_x\sigma_x$	$+\mathbb{I}$
r_3	$\sigma_z\sigma_x$	$\sigma_x\sigma_z$	$\sigma_y\sigma_y$	$+\mathbb{I}$
$\prod$	$+\mathbb{I}$	$+\mathbb{I}$	$-\mathbb{I}$	

Table 4.3: Reproduced from [140]. $\{\mathbb{I}, \sigma_x, \sigma_z, \sigma_y\}$ are the single qubit Pauli operators, and, e.g. $\sigma_z\sigma_x := \sigma_z \otimes \sigma_x$ indicates a measurement of the Pauli-σ_z on the first qubit and Pauli-σ_x on the second. There is no non-contextual assignment of ± 1 values to these nine observables.

Consider three observables A, B, C satisfying the following commutation relations:

$$[A, B] = [A, C] = 0 \tag{4.13}$$

$$[B, C] \neq 0 \tag{4.14}$$

i.e., B and C are not compatible. Recall that $[X, Y] \equiv XY - YX = 0$ for compatible observables, meaning that both observables may be measured simultaneously, and one does not affect the other. If the commutator is nonzero, the observables are incompatible. Position and momentum are incompatible quantum observables for example, since measuring one disturbs the other. A non-contextual theory holds that the expectation value of A is the same whether it is measured alongside either B or C. Cabello developed an inequality which places restrictions on the correlations that such a theory could exhibit, and one that is violated by any quantum state [34]. In so doing Cabello opened the Kochen-Specker theorem to experimental test, and furthermore enabled tests to be performed on poorly controlled systems with finite precision measurements (where the quantum state is difficult to manipulate arbitrarily). A succinct summary of Cabello's work is given by experimentalists Moussa, Ryan, Cory and Laflamme in their realisation of Cabello's test in an NMR ensemble [140]. Consider the two-qubit observables listed in the Table 4.3. One can check that in each row or column all observables are compatible. Now consider every possible way of assigning either $+1$ or -1 to each observable in the table in a non-contextual way; that is to say, in such a way that whenever an observable is measured on the same system it is assigned the same value, regardless of which other measurements are being made on the other qubit. For all possibilities of this kind,

$$f_{\mathrm{KS}} = \langle \pi_{r_1} \rangle + \langle \pi_{r_2} \rangle + \langle \pi_{r_3} \rangle + \langle \pi_{c_1} \rangle + \langle \pi_{c_2} \rangle - \langle \pi_{c_3} \rangle \leq 4 \,, \tag{4.15}$$

where $\langle \pi_{r_j/c_j} \rangle$ is the expectation value of the product of the outcomes of observables listed in row j / column j. Hence, all non-contextual realist theories are subject to this bound (since they hold that each observable possesses an objective value of $+1$ or -1 regardless of the measurement context). The quantum mechanical prediction is that $f_{\mathrm{KS}} = 6$, violating inequality (4.15).

Moussa et al. successfully violate this inequality experimentally, and besides ruling out non-contextual hidden variable theories provide inspiration to other researchers hoping to perform similar tests of quantum theory and related hidden variable theories. What makes the demonstration particularly interesting is that the experiment was performed in a spin ensemble, where one has no access to individual qubit systems but can only collect information about a large number of them. This is achieved by encoding the correlations which make up f_{KS} into the phase of a single probe qubit, which is well initialised into a predetermined state. This enables an ensemble measurement to obtain a genuine correlation rather than the product of two averages; i.e $\langle \sigma_z \sigma_x \rangle$ rather than the useless $\langle \sigma_z \rangle \langle \sigma_x \rangle$. The fact that the other qubits in the demonstration are in an unknown state is completely mitigated by the generality of Cabello's inequality, which will be violated by any quantum state. The following motto unifies the LG and Kochen-Specker theorems: '$\mathbb{I}$ over 2 will do'.

'One clean qubit' is needed however, to act as a measuring device so that experimental data can be extracted with confidence. There is a very interesting connection to a model of quantum computing (called DQC1) which also operates with only one well initialised qubit in a register of maximally mixed qubits [148, 44]. A possible criticism of the approach of Moussa et al. is that the information about correlation is encoded into the relative phase of the clean qubit: a concept that is difficult to understand without first understanding and accepting quantum mechanics.

Another important issue addressed in this work is the problem of initial state preparation. The probe qubit is required to be initialised in a pure state, but this is in general impossible at finite temperature, where thermal fluctuations ensure that portions of the ensemble are incorrectly prepared. Moussa et al. address this by determining experimentally which portion of the ensemble suffers from this malady, and making a 'fair sampling assumption' correct for it by post-processing the data they extract. It is not clear that this is a legitimate solution to the initialisation problem: the fair sampling assumption states that observed results are representative of unobserved results: this seems to be open to question (see Chapter 2). The authors do calculate what is expected quantum mechanically from the imperfectly initialised

ensemble: but what is less clear is how a sceptic of quantum theory might interpret the shortcoming.

The LG, Bell and Kochen-Specker theorems feel interrelated: the non-commutation of observables seems to underlie them all – even in the LG test, the single observable Q does not commute with itself at different times in the Heisenberg picture (where states are static and operators evolve in time). I will not discuss this point any further: I refer the reader to some papers that attempt to formalise the connections [52, 53, 43, 29, 35, 37, 131].

4.4.3 Degree of macroscopicity

Peres published a paper in 1988 criticising LG's work [158]. He questions whether the states which are superposed in the SQUID (LG's prototypical macrosystem, see Chapter 2) are truly macroscopic. Peres's argument about the macroscopicity of SQUID states highlights an important point: the criteria by which something is deemed to be micro- or macroscopic are not defined (by LG or by others) in a universally accepted way. Indeed even if a SQUID qualifies by e.g. the magnitude of its superconducting current, the superposed states themselves may not also qualify. Peres gives a gravitational wave detector as an example: the antenna of the detector may have a mass of several tons, but if it is well isolated it responds to the minute vibrations of spacetime as a quantum harmonic oscillator. This should presumably be contrasted with a Schrödinger's cat type superposition, where there is a macroscopic difference between the dead cat and the alive cat states: in LG's terminology the states should be *macroscopically distinct*. One might emphasise this point with a more recent example: the proposed superposition of a 'cantilever' or small mirror [99, 132]. Although perhaps of impressive size the two states which are superposed may be the zero- and one-phonon vibrational modes, which are surely far from being macroscopically distinguishable. A superposition involving a higher Fock state – for example the 100 phonon mode, would be (presumably) more macroscopic.

Leggett has suggested a measure of macroscopicity, or degree of 'Schrödinger's catiness', less as definitive metric but rather more of an *ansatz* for how such a measure may operate [117, 113]. The measure consists of two parts: the 'extensive difference' Λ and the 'disconnectivity' D. The first is the maximum quantity picked from a set Λ_i, which are the differences in expectation value of various observables in the two branches of the superposition, each expressed in units of a reference value. The reference value represents the typical value of the expectation value at the atomic scale. For example for the case of magnetic moments this would be the Bohr magneton

μ_B. The second quantity is very much like a measure of the degree of entanglement, and represents the number of particle correlations that must be measured in order to distinguish the superposition from a statistical mixture. The combination of these quantities (i.e. their product or similar) seems to do a decent job of sorting objects into a hierarchy of macroscopicity, although there may be other ways of doing the same job in a more satisfactory way [109]. Nimmrichter and Hornberger appeal directly to GRW-Pearle theory in their measure [147]. The issue of how to count the number of microscopic degrees of freedom echoes the discussion in Chapter 1; how should we deal with our ignorance about the ultimate structure of matter? In my opinion, appeal to 'number of particles' seems to be a contingent idea, which cannot provide a good foundation for dealing with the conceptual difficulties with quantum mechanics.

4.4.4 Isolating realism

Another criticism of LG's work is from Ballentine [13], who expresses scepticism over the extent to which their procedure can test realism (or macrorealism). Ballentine points out that, when confronted with a violation of the LGI, NIM may always be rejected instead of macrorealism. He also stresses that NIM does not in fact hold for QM and it is this feature which leads to its violation of the LGI. The very same criticism has been made by others [58, 64, 20, 77] and so constitutes the single most powerful argument against the work of LG.

There are a number of theories that will violate the LG inequality, but which maintain MR. The Kochen-Specker model for a qubit [106] (which cannot be extended to higher dimensions, and is ψ-epistemic [78]), and the De-Broglie-Bohm pilot wave theory [21, 22] (which reproduces all quantum mechanical predictions, including violation of the Bell inequality) are pre-eminent examples. Arguably the simplest model has been constructed by Montina [138], who posits a single 'memory' bit which is sensitive to measurements and can influence the trajectory of the system. All such models sacrifice NIM to reproduce a violation of the LGI.

LG rebut Ballentine's criticisms in a Comment [121]. LG suggest that NIM is a natural corollary to MR: and provide the following statement regarding Ballentine's preferred conclusion to a LGI violation:

> *Should anyone wish to interpret the results of our proposed experiment... by saying the macroscopic object is indeed in a macroscopic state, but is nevertheless affected by an interaction which we know could have occurred only if it had been in a different macroscopic state, he or she is of course*

> *free to do so; we leave it to the reader to judge whether such an interpretation in any way diminishes the force of the quantum measurement paradox, or the significance of our proposed experiment.*

This quote implies that LG require ideal negative result measurements for their procedure to be most effective.

A supportive paper of 1996 [87] would later state that, if MR implies NIM then the negation of their conjunction implies the failure of MR and not NIM! Although this is enticing, it is not clear whether a strict logical implication holds. Leggett has never claimed that MR strictly *logically implies* NIM; he admits that MR can be never be tested completely in isolation from other assumptions [119].

Timpson and Maroney question whether null-result measurements (where information is extracted from the system without a physical interaction) can be motivated in a theory independent way [199]. Essentially their objection is that, due to the possibility of hidden variables (being important in some as-yet-undiscovered theory of nature), one could never be sure that a null-result measurement actually occurred. Even if NIM can be motivated by control experiments (which show that a null-result measurement introduces no appreciable statistical disturbance to Q), there is no guarantee that the situation does not change in the full experiment. In the control experiments there may have been no detectable change in Q: but Q is not necessarily the only variable at play. Still, they argue, the fact these variables must then become active only when quantum theory would predict a superposition (and inactive when the system in an eigenstate, i.e. when $Q = 1$ or $Q = -1$ with certainty) is strongly suggestive of a quantum effect.

4.5 Weak measurement – a red herring

This section describes work that has been done in a spirit inspired by LG, but with important departures from the original proposal. The first suggestion of applying weak measurements to the LGI was in 1990. Tesche suggested a gedanken-experiment to enable macrorealism to be tested in SQUIDs [198]. He suggests that, alongside the ideal negative result scheme, the coherent oscillations in the SQUID might be preserved through using only a 'weak' coupling to the measuring devices (other nearby SQUIDS, see Chapter 3.). Had Tesche's experiment been performed, no violation would have been recorded! It is precisely the *interruption* of coherent oscillations that gives rise to a violation of LG's inequality – see Chapter 3.

4.5.1 Continuous weak measurement

In 2006 there was a renewed surge of interest in the LG theorem, because of the growing field of quantum computation. Ruskov, Korotkov and Mizel published a paper showing that weak measurements can be used to measure autocorrelations (and hence can be used to test a LGI) [170]; and this is motivated by the premise that strong, projective measurements are often hard to realise in practice.

The authors derive 'weak measurement Bell inequalities in time' by introducing a quantity $-1 < Q(t) < +1$. Note that this quantity is different from the original dichotomic variable of LG since it can take any value in this range (LG have $Q(t) = \pm 1$). The following expression is said to derive from macrorealism:

$$Q(t)Q(t+\tau_1) + Q(t+\tau_1)Q(t+\tau_1+\tau_2) - Q(t)Q(t+\tau_1+\tau_2) \leq 1. \tag{4.16}$$

It is equivalent to Equation (4.1) with the above generalisation from a dichotomic to a continuous variable. It is claimed that in a continuous weak measurement, one collects

$$I(t) = I_0 + (\Delta I/2)Q(t) + \xi(t) \tag{4.17}$$

with I_0 a background signal, ΔI the difference between signals representing $Q = +1$ and $Q = -1$ and $\xi(t)$ is white noise with vanishing time average. One now proceeds to measure the average of the product of the (background corrected) signal at two times separated by τ, i.e.

$$K_I(\tau) := \langle [I(t) - I_0][I(t+\tau) - I_0] \rangle \tag{4.18}$$

$$= (\Delta I/2)^2 \langle Q(t)Q(t+\tau) \rangle + (\Delta I/2)\langle \xi(t)Q(t+\tau) \rangle, \tag{4.19}$$

the second line is obtained by substituting (4.17), and also assuming $\langle Q(t)\xi(t+\tau) \rangle = 0$. This claim is that the system should not anticipate or influence the future noise of the detector. One makes a further assumption $\langle \xi(t)Q(t+\tau) \rangle = 0$, effectively there is no detector back-action. These assumptions does not motivate NIM then, but rather blindly assume it. Note that, according to quantum theory this weak measurement is of the noisy type, and not of the back action evading type [60], see Chapter 3. With this assumption in place one derives

$$K_I(\tau_1) + K_I(\tau_2) - K(\tau_1 + \tau_2) \leq (\Delta I/2)^2 \tag{4.20}$$

by 'averaging inequality (4.16) over time t'. An extra assumption called 'stationarity' is implicit. Stationarity enables the authors to claim a 'compelling advantage' to the

their scheme; the fact that no separation into different ensembles seems necessary. A single system is continuously monitored for many multiples of τ so that a time average can be used. Violations of this inequality have a critically different meaning to violations of the original inequality – which we saw to be a consequence of the inequivalence of separate subensembles (and hence of the failure of NIM). Stationarity is further discussed below in Section 4.6.

4.5.2 Instantaneous weak measurement

In 2006 Jordan, Korotkov and Buttiker extend the continuous weak measurement scheme to the case of instantaneous weak measurements (discussed in terms of a double quantum dot with a capacitively coupled quantum point contact) [91]. The ability to extract data from a single set-up is retained, but the stationarity assumption need not be made. Following from this, a 2008 a theory paper argues for a one to one correspondence between 'strange weak values' and Leggett-Garg inequality violations [208]. The motivation seems to be that, if one performs the first and third measurements strongly, and the intermediate measurement weakly, one has a situation very similar to that described by Aharonov, Albert and Vaidman in 1988 [2] whilst maintaining the quantum violation of the LGI. In this (in)famous work, the average value of a weak measurement is shown to exceed the eigenspectrum of the measurement observable *when the results are appropriately pre- and post-selected.* See Chapter 5. Clearly one can fix an initial and final state in the Leggett-Garg test, and by discarding some of the data can construct the so called 'weak values', which are seen to be 'strange' (i.e. outside of the eigenspectrum) whenever (for whichever time intervals) the Leggett-Garg inequality is violated. There is much controversy surrounding strange weak values [115, 159], and the 'instantaneous weak measurement' strand of Leggett-Garg theoretical work may share in this controversy. One might question whether 'measurement' is the correct interpretation to attach to weak values – see the comment by Leggett [115]. Williams and Jordan claim a 1-1 relationship [208] between weak values being strange and Leggett-Garg inequalities being violated. Although it is an interesting observation, it cannot be the whole story since it is possible to violate the LG inequality with strong measurements, and then no weak values can arise.

4.6 Stationarity – a capitulation

In a paper of 1995 by Huelga et al., the problem of describing a non-invasive measurement process is seemingly circumvented by introducing an extra assumption [85]. Given the assumption of (macro)realism, they argue

> *... it seems plausible* (though nothing more) *that the evolution from* t_1 *to* t_2 *is governed by the same stochastic differential equation as the evolution from* t_2 *to* t_3*, and this implies stationarity; that is*
>
> $$K(t_1, t_2) = K(t_1 - t_2). \tag{4.21}$$

Here $K(t_1, t_2) = \langle Q(t_1)Q(t_2)\rangle$. The stationarity assumption restricts the class of macrorealistic theories that may be tested. Theories that sign up to stationarity make up a minority of the possibilities originally captured by LG's analysis. The assumption is claimed to usurp the NIM assumption [86]. In my opinion the inequality is now a completely different beast, and should be clearly delineated as such. Conflation with (4.2) should be avoided. I shall refer to it as the *stationary Leggett-Garg inequality*. Crucially, one cannot reach this inequality with MR ∩ STAT alone. One must follow the chain

$$\begin{aligned}
&\langle Q(t_1)Q(t_2) + Q(t_2)Q(t_3) + Q(t_1)Q(t_3)\rangle + 1 \geq 0 \\
&\downarrow \text{NIM} \\
&\langle Q(t_1)Q(t_2)\rangle + \langle Q(t_2)Q(t_3)\rangle + \langle Q(t_1)Q(t_3)\rangle + 1 \geq 0 \\
&\downarrow \text{STAT} \\
&2\langle Q(t_1)Q(t_2)\rangle + \langle Q(t_1)Q(t_3)\rangle + 1 \geq 0.
\end{aligned} \tag{4.22}$$

The stationary inequality therefore follows from MR ∩ NIM ∩ STAT: a violation can then be due to the failure of at least one of these three assumptions [59]. It may seem that the inequality can be tested without NIM, since K_{23} is not measured but only inferred: this is the correlator which may be disturbed (relative to K_{13}) by an invasive measurement. Of course the slip is that one is reasoning counterfactually about an experiment that was never performed: to perform the experiment one would need to invoke NIM.

One can construct macrorealist predictions which explicitly break STAT easily. Consider an evolution described by populating Table 4.1 with $\mathbb{P}_{1\bar{1}\bar{1}} = 1$, and all others zero. On the original Leggett-Garg inequality, $f_{\text{LG}} = 0$. On the stationary Leggett-Garg inequality, $f_{sLG} = -2$. This violates even the maximum violation of the original

Leggett-Garg inequality achievable with an ideal quantum mechanical system! No hidden mechanism to track whether the system has been measured is needed [138]. Stationary continuous measurement macrorealist models that also violate the sLGI are also readily constructed. The stationarity assumption, which is linked to the idea of Markovianity [185], therefore seems overly restrictive. The class of theories that remain open to scrutiny with a stationary Leggett-Garg inequality is much narrower than with the original. As Emary et al. [60] put it:

> *. . . NIM requires only that the system has no memory of whether it has been measured or not, Markovianity requires amnesia of its entire history.*

One can check for STAT in separate control experiments [217]. Then, quantum mechanics actually provides STAT by sacrificing NIM. The invasiveness of a projective measurement can enforce stationarity on the correlators, by deleting information about the previous state of the system, making the dynamics Markovian (memoryless).

4.7 Experimental implementations

Since 2010 there has been a number of experimental demonstrations of LGI tests, each finding a violation. Few of the demonstrations have been completely satisfactory, and fewer still have involved systems that might be deemed macroscopic. I will review a small number of experiments here: for a fuller survey, please see the recent review article [60].

4.7.1 The Saclay experiment

The work of Palacios-Laloy et al. [150] describes experiments performed on a superconducting circuit known as a transmon (see Chapter 2). It is an artificial two level system, and is certainly the most macroscopic system considered here: it consists of the two lowest energy states of a Cooper-pair box, and it is claimed to be macroscopic by the authors since 'the dipole moment of the [ground state - excited state] transition is of the order 10^4 atomic units.'. The circuit is capacitively coupled to a stripline resonator (see Figure 4.5), which is used to continuously and weakly measure it in the way prescribed by Ruskov et al. [170], see Figure 4.6. For this reason any theoretical drawback suffered by that work is shared by this experiment: for example the stationarity assumption is made implicitly.

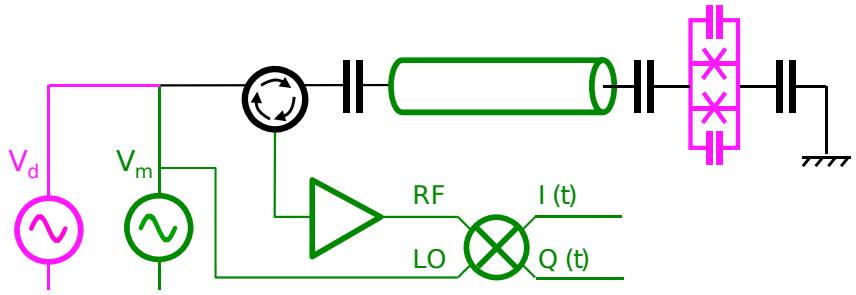


Figure 4.5: In the Saclay experiment a transmon qubit (featuring two Josephson junctions, magenta) is capacitively coupled to a microwave resonator (green) [150]. The microwave source V_m is used to measure and V_d used to drive the transmon. The $I(t)$ and $Q(t)$ signals carry information about the state of the transmon, and are obtained by routing the reflected microwave signal through an amplifier. After demodulation, the phase of the transmon can be inferred from these signals. Reprinted by permission from Macmillan Publishers Ltd: Nature Physics 6:442-447, © 2010.

After describing the experimental setup, the authors perform some control experiments to verify the effect of the weak measurement process on the state of the transmon. This involves driving oscillations in the qubit and measuring (with a strong measurement) the extra noise in its phase after a fixed interval. They then tune the weakness of the measurement (by altering the number of photons stored in the resonator) so that "for sufficiently low measurement strength, the coherent dynamics is only weakly affected by the measurement. This is the regime where the Bell's inequality in time can be tested.". In fact this limit is opposite to the limit where the quantum Zeno effect [135] can be observed (where the measurement(s) are strong enough and frequent enough to prevent coherent dynamics entirely). Noise in the output signal is removed by processing in the frequency domain (i.e. by making a Fourier transform).

The paper asserts that $K(0)$ represents the variance of $z(t)$ (the dichotomic variable). After finding that it is close to one experimentally, the authors claim confirmation that the spin only takes on $z = \pm 1$, unlike a classical magnet which would have a variance of $z = 1/2$. This seems correct: if one considers a genuinely two-level system, its average 'distance' from the mean ($\bar{z} = 0$) will be unity. A simple arrow (or 'macrospin'), on the other hand, can point in any direction in three-dimensional space and hence its average distance from the mean will be less than one. The conclusion of the paper reports a violation by over five standard deviations. The authors take decoherence into account and so have good agreement between experiment and appro-

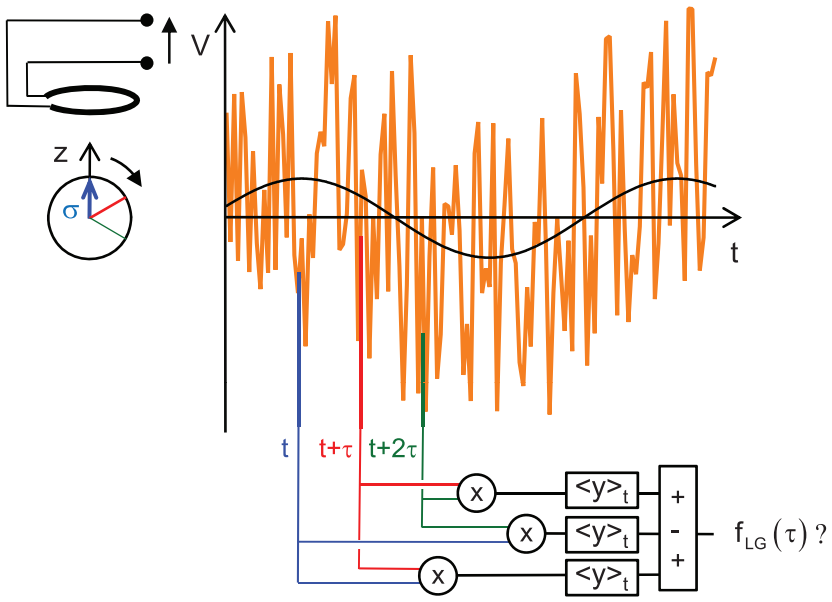


Figure 4.6: Measurement scheme in the Saclay experiment. Via the continuous weak measurement of a transmon qubit, correlations are determined with a time average [150]. The stationary LG function is then evaluated. Reprinted by permission from Macmillan Publishers Ltd: Nature Physics 6:442-447, © 2010.

priately adjusted theoretical predictions. The violation is explained by the quantum mechanical nature of the transmon qubit, and in particular the "invasive nature of the measuring process, which projects partially, but continuously, the [transmon] towards the state corresponding to the detector output. It is the interplay between this continuous projection and the coherent dynamics that yields the violation.". For an intermediate strength measurement, the measurement back action will speed up the evolution from one state to the other as the system approaches an eigenstate, and slow them down as it reaches its maximum distance from an eigenstate [91]. But Palacios-Laloy et al. perform the experiment with a very weak measurement strength, where this effect will be highly suppressed. The violation is thus not likely to be due to the back-action of the measurement, as it is in violations of the original LGI.

4.7.2 The Brisbane experiment

The work of Goggin et al. [73] reports both a violation of a Leggett-Garg inequality, and a confirmation of Williams and Jordan's conjecture [208, 209] that strange

weak values are in one-one correspondence with Leggett-Garg inequality violations. The experimental system is the polarisation state of single photons (see Chapter 2). Measurements are made instantaneously and weakly, by coupling each photon with another 'meter' photon which is strongly measured using the usual polarising beam splitter and photodetector set-up. The coupling between 'signal' and 'meter' photon is achieved with non-linear optical media (realising a probabilistic controlled-sign gate) [164, 163] and is tuned by altering the input state of the meter photon (see Figure 4.7). Note that this is the kind of weak measurement that reduces quantum back action, see Chapter 3.

The inequality that is quoted in this work is $-3 \leq \langle \mathcal{M}_a \mathcal{M}_b \rangle + \langle \mathcal{M}_b \mathcal{M}_c \rangle - \langle \mathcal{M}_a \mathcal{M}_c \rangle \leq 1$, which implies that the demonstration does not use the stationarity assumption (the subscript indices denote three consecutive times). Here $\mathcal{M}_a \sim Q(t_1)$ etc. Only the second measurement $\mathcal{M}_b$ is implemented weakly; it is preceded by a strong preselection and followed by a strong measurement. This is reasonable: if the weakness of the measurement is to be used as an argument for the non-invasiveness of the measurement, then the derivation of the Leggett-Garg inequality will be preserved since the first and third measurements do not need to be non-invasive. Since in this case we have a weak measurement between two strong measurements, by appropriately preselecting and postselecting the weak measurement results one can extract the aforementioned strange weak values from the experiment, quantities which seem to have no classical interpretation. Goggin et al. argue that for this reason the macrorealist description of objects can be brought into question; and so the Leggett-Garg inequality and strange weak values "explore the same concepts (or raise the same problems).". The authors report a violation 14 standard deviations above the limits imposed by macrorealism.

The conclusion of this paper is that as the measurement gets weaker, the LGI is more violated, not less; and that this implies NIM holds and MR should be rejected. One must question in this case how the violation is achieved (since in most cases the quantum mechanical back action is what does it, and in the weak measurement limit this is decreased). The dependence on the measurement strength seems the opposite of what one would expect. In contrast to LG's original proposal (where the inequivalence of subensembles, one of which contains an invasive measurement, is ultimately the cause of the violation), this work only uses a single ensemble to perform all measurements. The real cause of this violation is the rescaling of the correlator which is measured weakly, rather than the ordinary expectation value. For example if $P(D)$ and $P(A)$ represent probabilities of measuring the photon in the

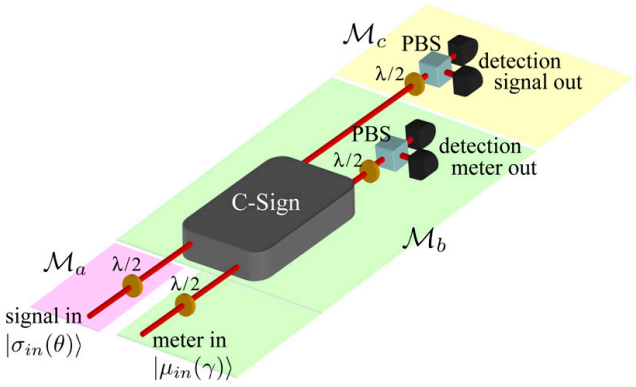


Figure 4.7: A 'signal' photon may be measured by a 'meter' photon through a probabilistic two-photon interference procedure. When successful, this is a controlled-sign gate. The strong measurements $\mathcal{M}_a$ and $\mathcal{M}_c$, and the weak measurement $\mathcal{M}_b$ find application in both the Leggett-Garg inequality and in the weak-values formalism. $\lambda/2$ denotes half waveplates, used to perform unitary transformations. Reproduced by permission from Proceedings of the National Academy of Sciences, January 25, 2011 vol. 108 no. 4 1256-1261.

state corresponding to outcome $+1$ or -1 respectively, then ordinarily the expectation value would be the $P(D) - P(A)$. Goggin et al. use instead the alternative

$$\langle \mathcal{M}_b \rangle = \frac{P(D) - P(A)}{K}, \tag{4.23}$$

where K (the knowledge [164]) is a number quantifying the strength of the measurement varying between 0 and 1. This post-processing is not necessarily legitimate. Consider the measurement result in any given run: it takes on a value outside of the ordinary range. The correction by dividing by the small number only applies to the ensemble as a whole.

4.7.3 The Rio experiment

The work of Souza, Oliveira and Sarthour (SOS) reports a violation of the inequality with a nuclear magnetic resonance (NMR) experiment [187] – an explanation of NMR can be found in Chapter 2. The stationarity assumption is not made, and weak measurements are not implemented. In contrast to other experiments, effort is made to argue that the measuring procedure is indeed non-invasive. SOS use an ancillary qubit to encode correlations in their principle system, and argue that if the principal system remains in a totally mixed state, the measurement must be non-invasive.

In my view, in a test of the LGI it is not obvious how to arrange the measurements so that a sceptical onlooker will not claim that they have disturbed the system, catastrophically corrupting the experimental data. Unless the assumption is convincingly motivated, the derivation of the LGI has little basis.

SOS are quite justified in preparing their system in the maximally mixed state for the purpose of testing the LG inequality. They are not, however, justified in claiming that this implies that all and any measurements made on this state are non-invasive in the sense that LG intended. The interpretation of a mixed state is clear for both quantum physics and classical physics, as it expresses incomplete knowledge about the state of a system. It is true that in quantum physics there is perhaps a richer interpretation of a mixed state: it is a probability distribution on the Hilbert space. There are a multitude of convex decompositions of a mixed state into pure quantum states. For classical physics it is a probability distribution on the classical state space. In either case the maximally mixed state represents zero information about the two-level system being investigated. In SOS's proposed quantum circuit, the state of the system remains a maximally mixed state throughout. This means that at all times there is zero information available about the state, so that the subjective description of the state will remain constant, although one suspects that the objective, physical state of affairs may be changing. In fact this is the case, if one computes the evolution of either classical state $|\uparrow\rangle$ or $|\downarrow\rangle$ individually *one finds that these states are indeed perturbed*, and moreover that they are perturbed in equal but opposite ways. How can ignorance of the identity of the state (according to macrorealism it is either $|\uparrow\rangle$ or $|\downarrow\rangle$) mitigate the invasiveness of measurements?

To make this point more concrete, consider the following scenario. Alice flips a coin but is blindfolded. She ascribes the maximally mixed state to the coin, as there is an equal probability of it showing heads or tails. Now, whilst remaining blindfolded, Alice turns the coin over, mapping heads into tails or tails into heads, depending on the physical state of the coin. This operation is clearly potentially invasive (the coin may now behave differently, compared to the case where no interaction had taken place), but still the state of the coin is the maximally mixed state. There is a very strong analogy between arguing in this scenario that the interaction is non-invasive, and SOS's argument that their circuit contains non-invasive measurements. This is my chief objection to SOS's approach [102].

Can the experiment of SOS be adapted to include, for example, ideal negative result measurements? Possibly, but the one will encounter the issues of fair sampling in ensemble systems, explained in Chapter 1.

4.8 Accommodating neurosis (on the part of the macrorealist)

In this section I present my theoretical extension to the Leggett-Garg inequality. I begin from the following position: strong measurements should be used, in the sense of Chapter 3. One should be able to infer the state of the system in single shot (in particular, averaging should not be necessary). Null-result measurements should be used, since these are the most convincing realisations of NIM. The stationarity assumption should not be made, since it weakens the conclusions considerably.

Imagine that the null-result measurements sometimes report a null result when they should not have: a false negative. If the null-result measurements are implemented by triggering (or not triggering) a response in an ancillary measuring device (i.e with a CNOT gate, see Chapter 3), this problem occurs if the measuring device begins in the wrong state. Then, when the measurement result is not null, it is transformed into a state associated with a null result. An erroneous false negative is recorded.

According to quantum theory, an incorrectly prepared ancilla will give rise to an incorrect correlator sign. To the macrorealist it will give a false indication that the measurement had been noninvasive, allowing a potentially corrupt element through the postselection. I define the venality ζ as the fraction of the ensemble for which the ancilla is incorrectly prepared. Expressed otherwise, it is the probability that a possibly invasive measurement occurred but was not discarded. With the time evolution now codified by a rotation angle θ, quantum physics predicts that each correlator K_{ij} generalises to $(1-\zeta)K_{ij} - \zeta K_{ij}$, leading to

$$f_{\mathrm{LG}} \to (1-2\zeta)(2\cos\theta + \cos 2\theta) + 1. \tag{4.24}$$

Various macrorealist 'attitudes' are defined with regard to a venal ancilla, which leads to generalisations of the Leggett-Garg inequality. A little algebra reveals thresholds on the venality, which must be exceeded to rule out the various flavours of macrorealism (see Figure 4.8). A 'moderate' view is that any invasively perturbed systems act in a random way, and so the measurement results average to produce zero net correlation. Then $K_{ij} \to (1-\zeta)K_{ij}$ and so with $g = K_{12}+K_{23}+K_{13}$ and $g \geq -1$ for a macrorealist,

$$f_{\mathrm{LG}}^{\mathrm{moderate}} = (1-\zeta)g + 1 \geq \zeta. \tag{4.25}$$

Note f_{LG} is still constrained to be non-negative. A neurotic, or 'adversarial' view is that invasively perturbed elements will, by some unidentified process, act in such a

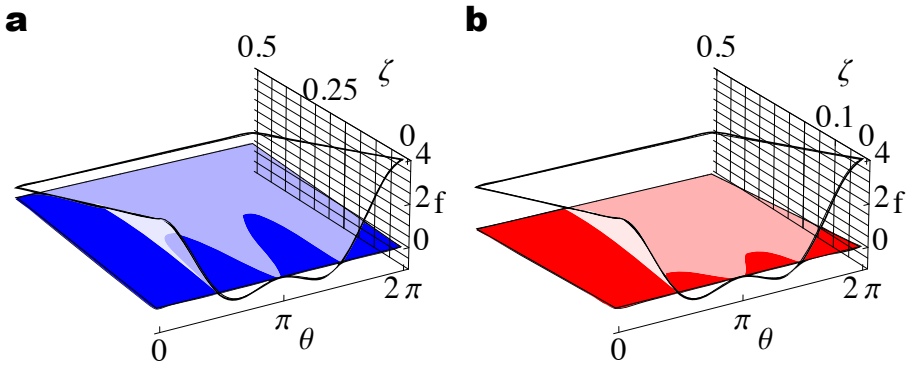


Figure 4.8: The macrorealist predictions corresponding to **a.** moderate and **b.** adversarial attitudes are plotted along with the predictions of quantum theory (white), as a function of time and of venality. Notice that even under a finite level of imperfection in the null-result measurement scheme, convincing demonstrations remain possible (where the white curve dips below the macrorealist bounds). Reproduced from *Nature Communications*, 3:606, January 2012.

manner as to minimise f_{LG}. Consequently $K_{ij} \to (1-\zeta)K_{ij} - \zeta$ so that

$$f_{\mathrm{LG}}^{\mathrm{adversarial}} = (1-\zeta)g - 3\zeta + 1 \geq -2\zeta. \tag{4.26}$$

This is the most aggressive stance available to a macrorealist. Keeping ζ low is important for a convincing test of the LGI. The venality is typically minimised by removing entropy from the measuring device: by cooling to low temperatures, for example.

4.9 The Oxford experiment

In this section I present the results of an experiment conducted in Oxford in 2011 that I helped to design. In the design of the experiment, I followed Paz and Mahler [154] who advocate the use of 'memories', or in the common parlance 'measurement ancillas', to realise ideal negative result measurements in an atomic setting. The experimental system is a solid state spin ensemble – see Chapter 2. Neither stationarity nor weak measurement is relied on: instead null-result measurements are employed. The full six subexperiments (two for each of three separate subensembles) envisaged by LG are for the first time used to violate their inequality. The experiment was performed by my colleague Stephanie Simmons [103].

In this experiment, we employed a method which equips a two level system with a local measuring device: another two-level system [154]. I refer to the system being

tested as the 'primary system' and the associated measuring device as the 'ancilla'. We used the venality concept explained above to take care of any imperfections.

Instead of repeating many times to collect statistics for the correlators, we employed an array of many identical systems – a spin ensemble, as in the experimental test of non-contextuality discussed in Section 4.4.2 above. In a spatial ensemble one has no access to individual elements: expectation values are gathered, rather than measurement outcomes, because averaging is automatic and unavoidable. Crucially, use of a separate ancillary measuring qubit coupled to each element of the ensemble means that the test may still be performed. The correlation information is encoded into the expectation value of the measurement ancilla, making it determinable by spin resonance readout techniques.

Consider a family of three experiments, each one beginning with a primary system in an identical initial state ρ_s and evolving under identical conditions governing the dynamics of the state (see Figure 4.9b). Recall that the ideal negative result measurement scheme requires that the three principal experiments introduced previously are each further resolved into a pair of experiments, one for non-invasive measurement of $\uparrow$, and one for $\downarrow$, see Figure 4.9c. We utilised either a CNOT gate (which will flip the state of the ancilla if the control, i.e. the primary system, is in $\downarrow$) or use an anti-CNOT gate (which will flip the state of the ancilla qubit if the primary is in $\uparrow$). In each case we only postselected experimental runs where the gate was not triggered. The second measurement in each experiment need not be implemented non-invasively, since the subsequent dynamics are irrelevant. Note that it is important that the physical implementation of the CNOT (and anti-CNOT) operation is such that the primary system receives no perturbation when it is in the state associated with a null result.

In the Oxford experiment, the primary system and the ancilla were identified in a coupled spin system. A great number ($\sim 10^{10}$) of phosphorous donor atoms embedded in a millimetre sized silicon crystal compose a solid-state ensemble of spin pairs, called Si:P for short (see Chapter 2). The phosphorous nucleus, which has spin-1/2, is identified as the primary system. When immersed in a magnetic field two classical states are revealed in the NMR spectrum. These are the two states of interest to the macrorealist, and those germane to a measure of Schödinger-catiness. To find a violation of the LGI, the requisite nuclear superposition is generated with a resonant radio-frequency pulse. An electron, another spin-1/2 particle bound to the nucleus, is identified as the measurement device or ancilla. The hyperfine interaction between the nucleus and electron enables the quantum coherent premeasurement of the former by the latter; entanglement, which arises in such a premeasurement,

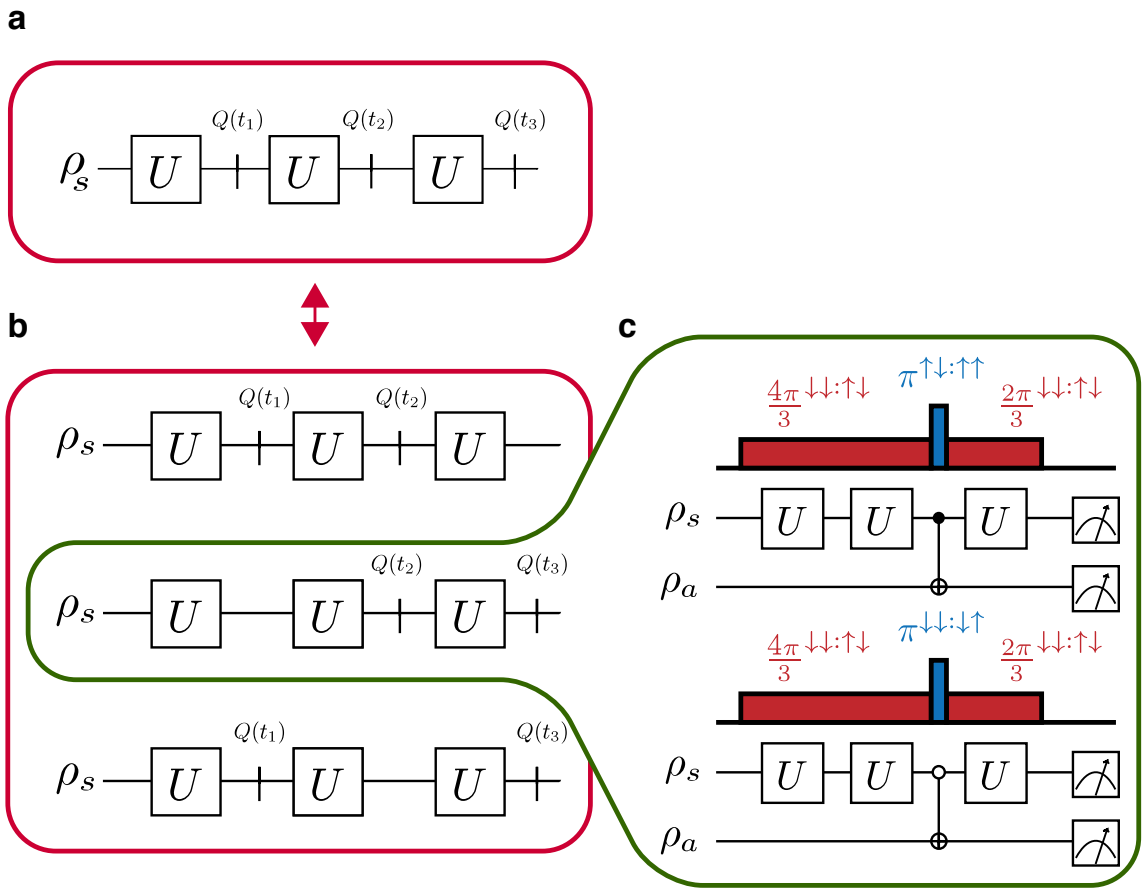


Figure 4.9: Our implementation of the LG test required six sub-experiments. Left panel: the three principal experiments from which the LG function is determined. Right panel: the actual lab implementation for one of the three principal experiments; the other principal experiments are similarly resolved into a pair of complementary sub-experiments. Shown in colour are the corresponding pulses applied to our coupled-spin$\frac{1}{2}$ system. The CNOT / anti-CNOT operations are applied with a single selective pulse. Reproduced from *Nature Communications*, 3:606, January 2012.

was showcased in this system in 2010 [184]. The requisite CNOT gates for the null-result measurement scheme can be achieved through resonant driving with microwave radiation, as described in Chapter 2. Ultimately information is extracted into the classical domain via population tomography of the joint nucleus-electron system at the end of the experiment. See Figure 4.10.

Null-result measurements can be difficult to realise in NMR systems [187]. CNOT gates can be built from composing several other gates, some of which may be unconditional on the control system. Such an approach is equivalent to a single conditional operation (i.e. according to quantum theory it achieves the same result), but will not typically convince a macrorealist that an ideal negative result measurement is

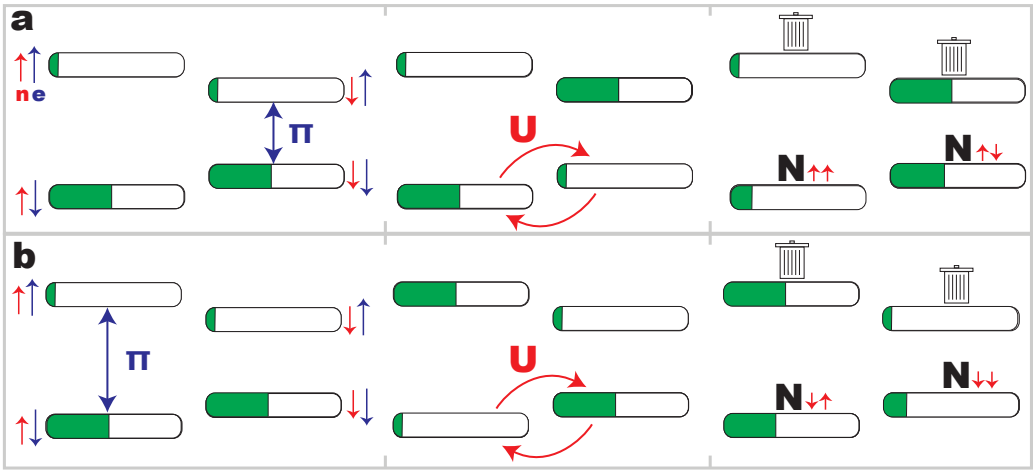


Figure 4.10: Energy levels diagram for one of the six experiments implemented in the Oxford experiment. The four classical configurations are the energy eigenstates. They are indexed both by the spin up or down state of the nucleus (red arrow) and the electron (blue arrow) in a strong magnetic field. Transitions between the states are possible with NMR / ESR spectroscopy techniques: resonant pulses at radio- or microwave frequency. The population of the ensemble in each of the states is shown schematically by the length of a green bar. The protocol begins by flushing out either **a** spin down or **b** spin up on the nucleus into a trash state (by a CNOT π-pulse). In this setting, the initial population of the trash state is given by the venality ζ. The dynamics are driven by a radio-frequency pulse U and then the correlator determined through population tomography. This scheme is analagous to that described in Figure 4.4 with Stern-Gerlach apparatuses.

being conducted. The Si:P system used in the Oxford experiment benefits from the uncomplicated realisation of such measurements: a single, truly selective operation is all that is needed.

4.9.1 Reducing the venality through hyperpolarisation

The unitary nuclear rotation U may be performed in a manner which is conditional on the system being in the 'correct' ancilla state $\downarrow$ because the postselected data will always correspond to the unitary operation U having been applied. If the rotation is conditional in this way, one of the two 'bad' populations becomes inactive and will not experience any evolution whatsoever in the course of the protocol (specifically state $|\downarrow\uparrow\rangle$ for the CNOT circuits and $|\uparrow\uparrow\rangle$ for the anti-CNOT circuits). The inactive state does not participate in the experiment and may be ignored. By minimising the population of the single *active* venal population we can reach a reduced effective venality.

If the population distribution of all four energy levels is the same for the initial state of both circuits in each pair we have e.g. in the $\{|\downarrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\uparrow\uparrow\rangle\}$ basis

$$\rho_C = \rho_A = \frac{1}{Z}\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & c & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & d \end{pmatrix} \tag{4.27}$$

where ρ_C and ρ_A are initial states prepared for CNOT and anti-CNOT circuits respectively, and $Z = a + b + c + d$ in both cases. The following expressions describe the lower bounds on quantum mechanical (QM), moderate and adversarial macrorealist predictions:

$$g_{\rm LG}^{\rm QM} \geq \frac{1}{Z}(a + b - c - d)(\cos 2\theta + 2\cos\theta) \tag{4.28}$$

$$g_{\rm LG}^{\rm moderate} \geq -\frac{1}{Z}(a + b) \tag{4.29}$$

$$g_{\rm LG}^{\rm adversarial} \geq -\frac{1}{Z}(a + b + 3c + 3d) \tag{4.30}$$

where $g = K_{12} + K_{13} + K_{23}$ and $f = g + 1$. The venality $\zeta = (c + d)/Z$ allows one to write

$$g_{\rm LG}^{\rm QM} \geq (1 - 2\zeta)(\cos 2\theta + 2\cos\theta) \tag{4.31}$$

$$g_{\rm LG}^{\rm moderate} \geq -(1 - \zeta) \tag{4.32}$$

$$g_{\rm LG}^{\rm adversarial} \geq -(1 - \zeta) - 3\zeta. \tag{4.33}$$

In thermal equilibrium $(a, b, c, d) = (1, \alpha, 1, \alpha)$ and so in general $\zeta = 2\alpha/(2 + 2\alpha)$. When oscillations are only driven on those primary systems which were paired with a correctly initialised ancilla, one (system,ancilla) state always remains unused throughout the experiment. We exploited this fact by hyperpolarising the system so that the remaining active state has a lower population than is possible in thermal equilibrium at a given temperature. If the population distribution is identical across only the three active levels of the experiment we have

$$\rho_C = \frac{1}{Z}\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & [c] & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & d \end{pmatrix} \tag{4.34}$$

for the CNOT circuit and

$$\rho_A = \frac{1}{Z}\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & [c] \end{pmatrix} \tag{4.35}$$

for the anti-CNOT circuit with $Z = a + b + c + d$ as usual. The inactive state is denoted with [ ]. These different initial states, although physically distinct, are logically congruent because the relevant active energy levels have the same population distribution. The predictions are now

$$g_{\mathrm{LG}}^{\mathrm{QM}} \geq \frac{1}{Z}(a+b-2d)(\cos 2\theta + 2\cos\theta) \tag{4.36}$$

$$g_{\mathrm{LG}}^{\mathrm{moderate}} \geq -\frac{1}{Z}(a+b) \tag{4.37}$$

$$g_{\mathrm{LG}}^{\mathrm{adversarial}} \geq -\frac{1}{Z}(a+b+6d). \tag{4.38}$$

Note that all predictions are independent of the inactive state with population c, except for in the normalisation Z. The normalisation can be arbitrarily scaled without affecting the comparison of the three predictions for g (or for f) since they will all be affected linearly in the same fashion. We choose to multiply g by $Z/(a+b+2d)$ so that there is a normalisation of $a + b + 2d = Z_r$ and no longer any dependence on c. This allows one to define the venality as $\zeta = 2d/Z_r$ and to recover equations (4.31),(4.32),(4.33). This technique is equivalent to supplying the single four level population distribution

$$\rho' = \frac{1}{Z_r}\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & d & 0 & 0 \\ 0 & 0 & b & 0 \\ 0 & 0 & 0 & d \end{pmatrix} \tag{4.39}$$

to both types of circuit. Using hyperpolarisation we achieve $(a, b, c, d) = (1, \alpha, \alpha, \alpha^2)$ so that $\zeta = 2\alpha^2/(1 + \alpha + 2\alpha^2)$.

4.9.2 Results

We performed two experimental tests. The results shown in Figure 4.9. The first used a simple state in thermal equilibrium at 2.6 K with $\zeta = 2\alpha/(2 + 2\alpha) = 0.150$, yielding $f = -0.031$. The second used the hyperpolarisation sequence from an initial state at 2.7 K. Due to the conditional nature of U this technique reduces the venality to $\zeta = 2\alpha^2/(1+\alpha+2\alpha^2) = 0.056$, yielding $f = -0.296$. The adversarial bound at this venality is -0.0112, and so this position is violated by over two standard deviations. In the course of our experiments, the fidelity of the final state populations with respect to the ideal target was never less than 98.9%.

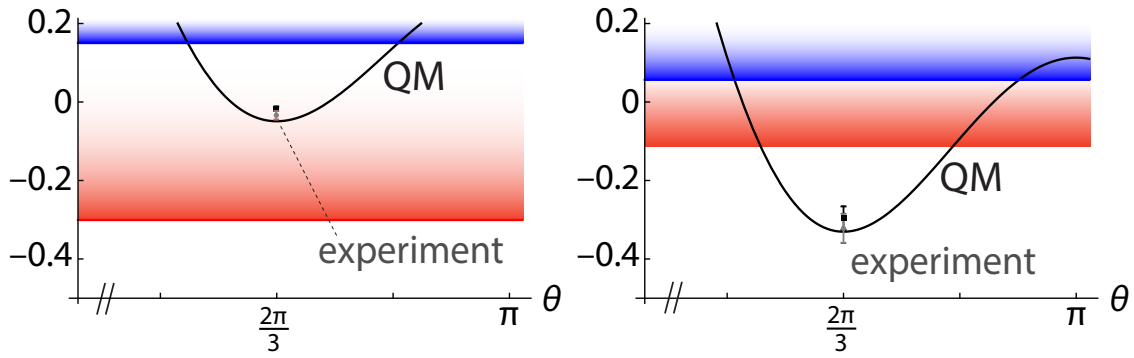


Figure 4.11: The Oxford experiment found a violation of a generalised Leggett Garg inequality [103]. The red and blue areas, and the black line correspond to a slice through Figure 4.8 at the given venality. Reproduced from *Nature Communications*, 3:606, January 2012.

4.10 Discussion

In this chapter I have discussed the theorem of Leggett and Garg – its strengths and weaknesses, its relation to other no-go theorems, and several laboratory implementations. The Oxford experiment, perhaps the most convincing realisation to date, calls into question the possibility of thinking of nuclear spin states in a macrorealist way. One must reject MR or NIM or IND. It paves the way for experiments on larger systems, which could provide genuine insight into the quantum measurement problem. Objective collapse theories, which motivate Leggett and Garg's inquiry, may be constrained or ruled out in the near future. The quantum control of much larger objects, which is necessary to nontrivially constrain theories such as GRW-Pearle's, is a realistic prospect in the coming decade.

To end this chapter, below I discuss points for improvement, not with the LG theorem per se, but in how the Oxford experiment fell short of the ideal LG experiment.

4.10.1 Drawbacks specific to the Oxford experiment

In the ideal scenario, the experimenter applies either the CNOT or anti-CNOT to the primary system-ancilla pair to perform the non-invasive measurement. In real spin resonance experiments each of the pulses will excite finite amplitude in the unwanted transition (i.e. it is not infinitely far off-resonance). The post-selection procedure will remove any pairs from the ensemble which are affected by a microwave pulse, detuned or not; but of course this post-selection is ill-informed for those pairs in which the ancilla is incorrectly initialised. To allow for this one can simply expand the venality

to include a fraction Δ of the inactive state population. Note that this Δ can be arbitrarily minimised in spin-resonance experiments by for example increasing the duration of the pulses which are applied, or using a sample with a larger splitting between the two microwave frequencies. In our experiment the Δ is less than 0.04 and we have confirmed that the corresponding correction to venality makes little difference to the degree of violation of our Leggett-Garg inequality.

On a deeper level, our implementation of the null-result scheme was not ideal. In the Stern-Gerlach example discussed in Section 4.3, the two states became spatially separated before the null-result measurement. In our experiment, by contrast, the two states are at the same location, and spatial separation is difficult. Regardless of how quantum theory, or even classical magnetic resonance theory, predicts the interaction of the spins with an off resonant ac magnetic field, it is at least feasible that some inner structure of the system is nevertheless sensitive to it. The spins bathe in the radiation: their immersion in the field makes the invasiveness of the measurement more plausible.

To address this issue, operational notions of disturbance may be gathered in control experiments. The statistical disturbance of the ensemble, due to a measurement, may be quantified [116]. Wilde and Mizel generalise the Leggett-Garg inequality to account for the finite invasiveness discovered in the control experiment – a violation of the LG inequality is still possible [207]. A similar approach was taken in the experiment of George et al. [67], who made use of a three-level system to violate the LGI (see the supplementary material of that paper). For further discussion of the different role played by NIM in that setting, see Ref [27].

The most obvious objection is of course that nuclei are not deemed macroscopic. Future experiments must strive to push the macroscopicity higher: superconducting qubits [215], nano-mechanical resonators [167] and macroscopic crystals [217] are promising candidates. Matter-wave interferometry with large molecules is another promising direction – the interference of large molecules, for example the C_{60} 'bucky ball' fullerenes [10], and more recently larger molecules such as viruses [68] have arguably been the most macroscopic interference experiments to date. The gradual build up of an interference pattern, through individual scintillations on a screen, is perhaps the most beautiful pageant of quantum theory's central mystery. But does the interference pattern alone force one to conclude that the molecule went through both slits? The LG inequality could lend additional rigour to such demonstrations: for example, by introducing null-result measurements to the experiments.

Chapter 5

Weak-value 'amplification'

This chapter concerns an application of quantum information theory in the laboratory. It marks a transition in this thesis, away from foundational issues and towards technological considerations. I shall no longer pit quantum theory against rival theories of nature; but rather take it 'warts and all' as given, and examine its utility in information processing tasks. The remaining chapters concentrate on quantum sensors: devices that apply the principles of quantum theory to improve metrology.

I shall commit to an evaluation of the advantages of quantum technologies in spite of the issues, raised in preceding chapters, concerning the consistency of the quantum formalism. I shall also endeavour to pinpoint from whence the reputed quantum advantages germinate: which special features of the theory lead to improvements in the laboratory? I begin with a discussion of entanglement enhanced metrology – a well-established method of enhancing precision. The true focus of the next chapters is an approach related to post-selected, weak measurements – a far more controversial proposal. In both cases my chosen figure-of-merit is the Fisher information, introduced in Chapter 2.

5.1 Entanglement-enhanced metrology

Entanglement is a strikingly non-classical feature of quantum theory – see Chapter 1. It gives rise to an increased precision in quantum metrology, which arises through the following recipe:

1. Relax the factorability of states
2. Entanglement emerges as a non-classical feature
3. Evolution under classical fields is *faster* for a subclass of entangled states

4. Quantum speedup!

I will set a benchmark for entanglement enhanced sensors by considering N spin 1/2 particles: a large collection of prototypical quantum bits – see Chapter 1. A collection of N non-interacting copies all experience an identical magnetic field of strength B along the $z \sim \sigma_z$ direction. A strong magnetic field leads to a Zeeman splitting of the spin's energy levels: commonly one is interested in a systematic deviation δB from the ideal field. If the Bloch vector describing the quantum state of the qubit is initialised in the equatorial plane (i.e. into a state unbiased with respect to the 'up' $|0\rangle$ and 'down' $|1\rangle$ eigenstates) it will precess around the z axis at a rate proportional to the strength of B. Moving into a frame that rotates with the same angular speed as this evolution, small deviations δB in the magnetic field strength are seen as slower evolutions of the Bloch vector around the z-axis. We can write

$$|\psi\rangle \propto |0\rangle + \mathrm{e}^{i\mu\delta B t}\,|1\rangle : \tag{5.1}$$

the strength of the interaction between the spin and the field is set by the magnetic moment μ, which is particular to the species of spin involved. The sensing protocol is not complete until the probe itself is measured: one must therefore make a choice of measurement basis. Choosing the basis of eigenstates of σ_z, $|0\rangle\langle 0|$ and $|1\rangle\langle 1|$, would not be smart, since the phase (and hence information about δB) would be completely lost when the probability amplitudes are mod-squared. The probabilities would not depend on δB, and thus the Fisher information 'about δB'

$$F_{\delta B} := \sum_i \frac{(\partial_{\delta B} P_i)^2}{P_i} \tag{5.2}$$

is zero. Choosing the basis of eigenstates of σ_x, $|+\rangle\langle +|$ and $|-\rangle\langle -|$ is better[1], and yields:

$$P_{\langle +|} = \cos^2 \frac{\mu\delta B t}{2} \qquad P_{\langle -|} = \sin^2 \frac{\mu\delta B t}{2} \tag{5.3}$$

which has a Fisher information of

$$\begin{aligned} F_{\delta B} :&= \left(\frac{\partial P_{\langle +|}}{\partial \delta B}\right)^2 \frac{1}{P_{\langle +|}} + \left(\frac{\partial P_{\langle -|}}{\partial \delta B}\right)^2 \frac{1}{P_{\langle -|}} \\ &= \mu^2 t^2. \end{aligned} \tag{5.4}$$

The state vector, through exposure to the magnetic field, acquires an increasing amount of information about B as time progresses. Clearly any measurement axis

[1] $|\pm\rangle = (|0\rangle \pm |1\rangle)/\sqrt{2}$

lying in the equatorial plane of the Bloch sphere is optimal (the choice within the plane only modifies the phase of the oscillatory probabilities, and F is unchanged). Generally the probabilities depend on the measurement choice, and the *quantum* Fisher information is a powerful tool to find the maximal Fisher information over all possible measurement choices (see Chapter 2).

The total Fisher information is additive over the N elements of the ensemble, giving

$$F^{\text{total}} = N\mu^2 t^2. \tag{5.5}$$

According to the Cramér-Rao bound (CRB), the minimum variance of the parameter δB is given by $1/F$ (see Chapter 2). Because the standard error is the square root of the variance, for an ensemble of N spins behaving independently, the CRB on the standard error of the estimate of δB is $1/(\sqrt{N}\mu t)$. The scaling with N is known as the standard quantum limit [72].

Consider now, instead of an ensemble of non-interacting spins, a system of N coupled spins that have been arranged (through quantum control) into the state

$$\begin{aligned} |\psi_N\rangle &\propto |00\ldots0\rangle + |11\ldots1\rangle \\ &\propto |0\rangle^{\otimes N} + |1\rangle^{\otimes N}. \end{aligned} \tag{5.6}$$

This is called a GHZ (Greenberger, Horne, Zeilinger) state, or a $N00N$ state (for N 'up', zero 'down' plus zero 'up', N 'down') [89]. It is also referred to as a 'cat' state, because it is a collective superposition of a potentially large number of degrees of freedom as N increases. Note that this type of state is entangled. Under the same magnetic field it will evolve into

$$|\psi_N\rangle \to |\psi'_N\rangle \propto |0\rangle^{\otimes N} + \mathrm{e}^{iN\mu\delta Bt}\,|1\rangle^{\otimes N} \tag{5.7}$$

and with access to the right measurement one has

$$P_{\langle+|^{\otimes n}} = \cos^2\frac{N\mu\delta Bt}{2} \qquad P_{\langle-|^{\otimes n}} = \sin^2\frac{N\mu\delta Bt}{2} \tag{5.8}$$

and the Fisher information is

$$F = N^2\mu^2 t^2. \tag{5.9}$$

When taken independently, it was as if the N spins had been replaced with a single spin with a higher magnetic moment $\mu \to \sqrt{N}\mu$. If cooperating through entanglement, the N spins become a single 'super-spin' with $\mu \to N\mu$.

Entanglement gives rise to a square root improvement in the standard error, over the standard approach given the same resources t and N [71, 160, 89, 136, 174]. The

superior scaling with N, which is 'beyond the standard quantum limit', is known as the Heisenberg limit. Recall that spins are archetypal qubits – the Heisenberg scaling achievable with entanglement is a general feature of quantum theory, and may be exploited in, for example, optical systems, superconducting qubits, and other quantum information carrying media.

The entanglement enables faster evolution through Hilbert space, which translates, with the right measurement, into the probabilities having a larger gradient in the δB direction. This increased speed of evolution thereby increases the Fisher information measure (equation 5.2). The phenomenon of weak values, or the AAV effect (to be introduced shortly) also exhibits such a feature: but there is a cost. One can achieve only a *probabilistic* quantum speedup. Before showing how this works, I will discuss some of the history behind the technique.

5.2 Time symmetric quantum mechanics

In one sense, quantum theory is time symmetric. Probabilities, its only empirical predictions, are preserved when the preparation (the initial state) and the measurement (the final state) are interchanged:

$$| \langle \psi_f | U | \psi_i \rangle |^2 = | \langle \psi_i | U^\dagger | \psi_f \rangle |^2 \tag{5.10}$$

if the time evolution operator is also reversed ($U^\dagger$ is the solution to the time reversed $t \to -t$ Schrödinger equation). Experiments are nevertheless usually described in a certain temporal order: one *begins* with a state preparation, evolves *forwards* according to the Schrödinger equation, and *ends* with a measurement: of course this preference is not the only option.

Aharonov, Bergmann and Lebowitz (ABL), study this issue in their landmark paper of 1964 "Time symmetry in the process of quantum measurement" [3]. They point out that to characterise an ensemble by final conditions is just as natural as to do so by initial conditions. Given the final state a system was found in, one can retrodict the probability of it having been in any particular state immediately prior to measurement. The same is true of any classical experiment. An important difference, however, is that quantum theory is a probabilistic theory at the microscopic level. Evolution under unitary dynamics is deterministic, and hence time reversible: but the collapse postulate introduces asymmetry. Specification of the initial state does not determine the final state after a quantum jump, and *vice versa*.

The concept of pre and postselected ensembles is a natural product of this inquiry. Such ensembles are defined doubly: by both a shared initial state and a shared final state. Far from being strange concepts, ABL argue that these ensembles are quite ubiquitous in experiments – their example is the detection of charged particles in a cloud chamber [3].

ABL considered measurements performed in such ensembles: i.e. measurements in the interval between preselection (into $|\psi_i\rangle$) and postselection (into $|\psi_f\rangle$). The probability of finding the state (at the intermediate time) in $|a_k\rangle$ is:

$$P_{a_k|f,i} = \frac{|\langle\psi_f|a_k\rangle\langle a_k|\psi_i\rangle|^2}{\sum_j |\langle\psi_f|a_j\rangle\langle a_j|\psi_i\rangle|^2}. \tag{5.11}$$

Equation (5.11) is the infamous ABL rule, a case of Bayes' rule [6, p141]. Aharonov and his collaborators are keen to use their time symmetric viewpoint to reinstate the objective reality of a quantum system's properties, including those described by noncommuting operators [8]. Applying the ABL rule counterfactually, however, – as in, 'a measurement would have found such-and-such *had* an experiment actually been performed', when no actual experiment was performed – is controversial [137, 96]. The time symmetric formulation of quantum mechanics has emerged from Aharanov and his collaborators' investigations. Although there are no new physical predictions, the reformulation makes a metaphysical gambit: namely, that along with the usual state vector which evolves forwards in time from the past to the future, there is another state vector evolving backwards in time from the future towards the past [5].

The possibility that relaxing the preference for preselected-only descriptions could improve statistical inferences can be found in the 1964 paper:

> *Hence, if both the initial and final state of a system are known, use of the* [preselection only] *prediction formulas . . . will lead to a loss in precision of the probabilistic statement concerning the intermediate observation.*

This idea clearly has application in precision technologies: it highlights the increase in entropy brought about by ignoring, or averaging over, many different postselections when in fact the actual postselection is known. Forgetting the result of the final measurement is disadvantageous given a pre and postselected ensemble.

Notice that projective measurements destroy any influence that the preselection can have on the postselection. What if the intermediate measurement is not projective? The next section combines weak measurements, which were introduced in Chapter 1, with the pre and postselected ensemble.

5.3 The weak value and 'amplification'

There are some intriguing consequences of combining preselection, weak measurement, and postselection. The unexpected results have lead research away from ABL's suggestion: instead of avoiding loss of precision by ignoring the final state in a pre and post selected experiment, workers have attempted to improve precision by enforcing postselection on otherwise preselected-only experiments.

In a seminal paper of 1988, Aharonov, Albert and Vaidman (AAV) presented a curious quantum mechanical thought experiment giving rise to a quantity the authors called 'the *weak value* of a quantum variable' [2]. AAV defined the quantity

$$A_w := \frac{\langle\psi_f|\mathbf{A}|\psi_i\rangle}{\langle\psi_f|\psi_i\rangle}, \tag{5.12}$$

generalizing the usual expectation value $\langle\mathbf{A}\rangle = \langle\psi_i|\mathbf{A}|\psi_i\rangle$. The expectation value is recovered when $|\psi_f\rangle = |\psi_i\rangle$. Notice that the weak value is complex, $A_w \in \mathbb{C}$. 'Obtaining' A_w involves i) initializing a quantum system of interest (the 'system') into the state $|\psi_i\rangle$; ii) coupling the system weakly to an ancillary measuring device (the 'meter') through the system operator $\mathbf{A}$; and iii) post-selecting the system into a definite final state $|\psi_f\rangle$. The meter can then be interrogated at full strength to reveal something about the system. When the meter wavefunction is of a certain form, the projection of the weak value in the complex plane gives an effective average shift of the meter in the appropriate phase space variable. For example the position variable is shifted by the real part of the weak value and the momentum variable by the imaginary part [94].

AAV's paper was entitled "How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100". Because the deflection induced in the meter is much less than its inherent uncertainty, when the pre- and postselected states are close to orthogonal ($\langle\psi_f|\psi_i\rangle \to 0$) destructive interference occurs in the meter, and the magnitude of the weak value A_w becomes much larger than $\max(\langle\mathbf{A}\rangle)$. They arrive at this conclusion as follows:

$$\begin{aligned} |m_{\mathrm{WVA}}\rangle :&= \langle\psi_f|\mathrm{e}^{-ig\mathbf{A}\mathbf{k}_x}|\psi_i\rangle|m\rangle \\ &= \left(\langle\psi_f|\psi_i\rangle - ig\langle\psi_f|\mathbf{A}|\psi_i\rangle\mathbf{k}_x + \ldots\right)|m\rangle \\ &\approx \langle\psi_f|\psi_i\rangle\left(1 - igA_w\mathbf{k}_x\right)|m\rangle\,, \end{aligned} \tag{5.13}$$

so that to linear order in g

$$|m_{\mathrm{WVA}}\rangle = \mathrm{e}^{-igA_w\mathbf{k}_x}\,|m\rangle\,; \tag{5.14}$$

it is as if the system has undergone a straightforward shift, but due to a coupling to a classical field of strength A_w. In this thesis, I refer to (5.14) as the AAV approximation, which is accurate when certain conditions on g, A_w and $|m\rangle$ hold [56]. The motivation behind using postselection is that, where the AAV approximation is valid and when $\langle\psi_f|\psi_i\rangle \to 0$, the average displacement is much larger than is otherwise possible[2]. The large displacement is often claimed to be an advantage, especially for overcoming sources of technical noise. AAV's expression for A_w has a limited range of validity [56]: however, exact treatments reveal that the qualitative effect persists outside this range [211, 107, 142]. Weak values have been obtained experimentally, see e.g. Refs. [166, 164].

The weak values formalism revived interest in pre and postselected ensembles in the late 1980s. Since then it has grown popular and controversial in equal measure. It has been suggested as the basis for all of quantum theory [81]; as a resolution to Hardy's non-locality paradox [128]; as leading to violations of Heisenberg [169] and Leggett-Garg inequalities [208, also Chapter 4]; as capable of separating an object from its properties [4], and above all as a technological tool [83, 55]. I like to refer to the bevy of mysteries as 'Quantum Weirdness 2.0'. The various experiments and thought-experiments have an uncanny knack of confounding even seasoned quantum physicists, just as they had become accustomed to non-locality, contextuality, and the rest of it. Aharonov, the chief architect of weak values, has called them 'Quantum Paradoxes' in a book co-authored with Daniel Rohrlich [6]. Many of the thought experiments laid out in the book have been implemented in the laboratory.

Shortly after the 1988 AAV paper, Duck, Stevenson and Sudarshan published "The sense in which a "weak measurement" of a spin-1/2 particle's spin component yields a value 100" [56]. The paper sets about clearing up some of the conceptual difficulties with the AAV effect, as well as confirming that the predictions would be borne out experimentally. They placed 'weak measurement' in inverted commas so as to "remain neutral on the semantical and philosophical issue" of its meaning. Aligning myself with the spirit of this precedent, I will likewise suspend 'amplification' in commas, because I doubt that the effects to which it has now become inseparably attached truly deserve the name. However, *weak measurement* shall be spared this treatment: when it is properly distinguished from the weak values formalism, it is a perfectly uncontroversial generalisation of projective measurement[3], see Chapter 3.

[2]The shift may not be made arbitrarily larger, as is revealed by exact calculations [56].

[3]Given the multiple meanings of 'measurement' itself in quantum physics, its use ought to be tacitly cautioned, too: but I do not wish to drown this thesis in a sea of punctuation marks. See the article by Bell [19].

The controversial approach taken by Aharonov and his collaborators has many names: time symmetric quantum mechanics, the ABL rule, the two-state-vector formalism, weak values, and so on. It is important to realise that the approach does not make any new predictions over standard quantum theory. The approach only highlights some surprising effects that were already present in quantum theory. It is unlikely to solve any significant conceptual problem, although it has the potential to spawn new insights [134]. Busch referred to the AAV effect (which was first discovered in 1987 [1]) as "a surprise but no miracle" [33].

In beginning my discussion of the possible technological insights, I will demonstrate that weak-value 'amplification' (henceforth WVA) does not bring an overall benefit to metrology in the absence of classical noise. In subsequent chapters I will allow imperfections to be modelled.

5.4 Noise-free arguments against weak-values

Before presenting my main results, I firstly set the scene of parameter estimation in the time symmetric setting. Firstly, I define two distinct estimation tasks for study. The first is the estimation of the weak value itself (which is related to the problem of estimating the initial state $|\psi_i\rangle$). The second is the estimation of the strength of the coupling between system and meter.

5.4.1 Estimation heuristics

It is important to distinguish two estimation tasks that may be attempted in an experiment using the AAV effect. Given the approximate evolution in (5.14), one can either estimate the interaction parameter g, or the weak value A_w itself. The quantum Fisher information (QFI) provides an excellent figure of merit for the performance of the estimation of either. For pure states:

$$H_\bullet[|\psi\rangle] \equiv 4\left\langle \partial_\bullet\psi|\partial_\bullet\psi\right\rangle - 4|\left\langle \partial_\bullet\psi|\psi\right\rangle|^2. \tag{5.15}$$

In the estimation of the weak value itself, one has

$$|\partial_{A_w} m_{\mathrm{WVA}}\rangle = -ig\mathbf{k}_x \mathrm{e}^{-igA_w\mathbf{k}_x}\,|m\rangle\,, \tag{5.16}$$

and the QFI about the weak value, using Equation (5.15) is

$$\begin{aligned} H_{A_w}[|m_{\mathrm{WVA}}\rangle] &= 4g^2\left\langle m|\mathbf{k}_x^2|m\right\rangle - 4g^2\left\langle m|\mathbf{k}_x|m\right\rangle^2 \\ &= 4g^2\mathrm{Var}(\mathbf{k}_x). \end{aligned} \tag{5.17}$$

Notice that the size of the weak value is irrelevant! Note also that the QFI scales with the interaction strength (equivalently the measurement strength). This means that at a given value of g, the use of eigenvalues (arising from a standard approach using preselection only) or the larger-in-magnitude weak values, is immaterial. The QFI per run is the same – it is given by the variance of $\mathbf{k}_x$ in the meter. If the measuring device is finite dimensional the observable could just as well be, for example, a qubit observable. Further, the reduced postselection probability will be costly: there are fewer runs that succeed, and the total information is lower.

Imagine that one has perfect knowledge of $g, |\psi_f\rangle$,and $\mathbf{A}$: then estimation of the weak value is effectively estimation of the initial state of the system $|\psi_i\rangle$. This idea has been developed by Lundeen and coworkers [129]. If the initial state encodes useful information about a classical field, then with all other parameters known, the strength of this field may be estimated. This is the situation studied in Chapter 6, which describes one of the first analyses of WVA using Fisher information [100].

For the remainder of this chapter I concentrate on the more intriguing case of estimating g, and analyse that situation once more with noise in Chapter 7. Here, in analogy to entanglement above, the quantum speedup due to weak values can be derived through the following recipe:

1. Relax preference for preselected-only ensembles
2. Interference arises as a non-classical effect
3. Evolution is faster when postselection succeeds
4. *Probabilistic* quantum speedup!

5.4.2 Benchmarking

In the evaluation of a newly proffered experimental technique, the issue of a good benchmark is highly important. Tanaka and Yamamoto calculate the QFI about g in the postselected, WVA quantum state $|m_{\mathrm{WVA}}\rangle$ [196]. Crucially, their benchmark applies to *the joint system-meter state immediately after their interaction.* A rather trivial argument reveals WVA to be inferior. Since the QFI maximises a quantity (namely the classical Fisher information) over all possible measurements, Tanaka and Yamamoto's benchmark relates to the 'best' measurement that can be performed on the joint system meter state. The best measurement might be, for example, a projection into an entangled basis. As long as it possible to think of the WVA strategy as a generalised POVM on the joint system-meter state, the associated classical Fisher

information cannot exceed their benchmark by construction! It is indeed possible to think of the WVA strategy as a POVM with elements

$$\begin{aligned} E_s^{\checkmark} &:= |\psi_f\rangle \langle\psi_f| \otimes |s\rangle \langle s| , \\ E_s^{\times} &:= \mathbb{I} - \sum_s E_s^{\checkmark} \end{aligned} \tag{5.18}$$

where $\{|s\rangle\}$ is the chosen measurement basis for the meter. The information associated with this choice of measurement is not increased by discarding runs where the postselection fails – this is a consequence of the data-processing inequality [40]. I suggest a different benchmark is the natural one, which makes the question of whether there could be an advantage with WVA less straightforward to answer. The benchmark should be *preselected only*. That is, the system is entirely ignored: the POVM elements are now

$$E_s := \mathbb{I} \otimes |s\rangle \langle s| . \tag{5.19}$$

It is not obvious that the WVA strategy is worse: whilst the benchmark only makes contact with the system once, through the interaction with the meter, the WVA strategy is permitted to interrogate the system twice – the postselection is a second bite of the cherry, which might increase the information.

5.4.2.1 First order argument

Tanaka and Yamamoto's results [196] can easily be adapted to my preferred benchmark. Let us first consider a 'first order' application of Equation (5.15): i.e. when the AAV approximation holds. If one expands the unnormalised post-selected state of the meter, one finds

$$|m_{\mathrm{WVA}}\rangle \propto \langle\psi_i|\psi_i\rangle |m\rangle - ig \langle\psi_i|\mathbf{A}|\psi_i\rangle \mathbf{k}_x |m\rangle \tag{5.20}$$

with the proportionality constant given by

$$\frac{1}{\langle\psi_i|\psi_i\rangle \sqrt{1 - g^2 A_w}}. \tag{5.21}$$

One can see that the norm of $|m_{\mathrm{WVA}}\rangle$ is independent of g to first order *if the weak value is not too large*. Upon dismissing the term in g^2, the derivative takes on a simple form

$$\begin{aligned} |m_{\mathrm{WVA}}\rangle &= |m\rangle - igA_w \mathbf{k}_x |m\rangle \\ \Rightarrow |\partial_g m_{\mathrm{WVA}}\rangle &= -iA_w \mathbf{k}_x |m\rangle \end{aligned} \tag{5.22}$$

and hence

$$\begin{aligned} H_g[|m_{\rm WVA}\rangle] &= 4|A_w|^2 \langle m|\mathbf{k}_x|m\rangle - |iA_w^* \langle m|\mathbf{k}_x|m\rangle + g|A_w|^2 \langle m|\mathbf{k}_x^2|m\rangle|^2 \\ &= 4|A_w|^2 \left(\langle m|\mathbf{k}_x^2|m\rangle - \langle m|\mathbf{k}_x|m\rangle^2\right) \\ &= 4|A_w|^2 {\rm Var}(\mathbf{k}_x) \\ &= 4|A_w|^2 \langle \mathbf{k}_x^2\rangle \end{aligned} \tag{5.23}$$

The variance is taken in the initial meter state, and I discarded a term in g^2. With no loss of generality I set $\langle \mathbf{k}_x\rangle = \langle m|\mathbf{k}_x|m\rangle = 0$.

Now for the preselected-only benchmark. If we put the system in an eigenstate $|\psi_{i_*}\rangle$ of the measurement observable $\mathbf{A}$ with largest magnitude eigenvalue λ_*,

$$|\psi_{i_*}\rangle\,|m\rangle \to {\rm e}^{-ig\mathbf{A}\mathbf{k}_x}\,|\psi_{i_*}\rangle\,|m\rangle = |\psi_{i_*}\rangle\,{\rm e}^{-ig\lambda_*\mathbf{k}_x}\,|m\rangle\,. \tag{5.24}$$

It is as if the meter has been exposed to a classical field of strength $g\lambda_*$. We may use

$$\begin{aligned} H_g^{\rm benchmark} &= 4\langle \mathbf{H}^2\rangle - 4\langle \mathbf{H}\rangle^2 \\ &= 4\lambda_*^2 \langle \mathbf{k}_x^2\rangle\,. \end{aligned} \tag{5.25}$$

The equations are simple: to go from the WVA strategy to the preselected-only benchmark, one replaces the the weak value of $\mathbf{A}$ with the largest eigenvalue of $\mathbf{A}$. Since the former can be much larger, this would seem to point to an advantage: compare (5.25) with (5.23). The problem is that preselection works every time, but postselection does not always occur. One must therefore multiply by the postselection probability q for a fair comparison. As will be proved in the next chapter, $q|A_w|^2 \leq \lambda_*^2$. There is no overall advantage.

5.4.2.2 The Tanaka-Yamamoto argument

Tanaka and Yamamoto give a full order argument, valid for any g. The exact state is

$$|m_{\rm WVA}\rangle = \frac{B\,|m\rangle}{\sqrt{q}} \tag{5.26}$$

with the Kraus operator $B = \langle\psi_i|{\rm e}^{-ig\mathbf{A}\mathbf{k}_x}|\psi_i\rangle$ and $q = \langle m|B^\dagger B|m\rangle$ which is not independent of g. Hence one must employ the quotient rule to differentiate:

$$|\partial_g m_{\rm WVA}\rangle = \frac{1}{q}\left(\sqrt{q}\hat{B}\,|m\rangle - \frac{\hat{q}}{2q}B\,|m\rangle\right); \tag{5.27}$$

I denote the derivative with respect to g with a caret '$\hat{}$', and I used $\partial_g \sqrt{q} = \hat{q}/2q$. Now

$$\langle \partial_g m_{\mathrm{WVA}} | \partial_g m_{\mathrm{WVA}} \rangle = \frac{1}{q^2} \left(q \langle m | \hat{B}^\dagger \hat{B} | m \rangle + \frac{\hat{q}^2}{4q} - \frac{\hat{q}^2}{2\sqrt{q}} \right); \tag{5.28}$$

note that $\hat{q} = \langle m | \hat{B}^\dagger B | m \rangle + \langle m | B^\dagger \hat{B} | m \rangle$. Also

$$| \langle \partial_g m_{\mathrm{WVA}} | m_{\mathrm{WVA}} \rangle |^2 = \frac{1}{q^2} \left| \langle m | \hat{B}^\dagger B | m \rangle - \frac{\hat{q}}{2\sqrt{q}} \right|^2; \tag{5.29}$$

now when applying Equation (5.15), the cross terms of (5.29) will cancel with the final two terms from (5.28) so that

$$H_g = \frac{4 \langle m | \hat{B}^\dagger \hat{B} | m \rangle}{q} - \frac{4 \langle m | \hat{B}^\dagger B | m \rangle^2}{q^2}. \tag{5.30}$$

Tanaka and Yamamoto argue that, since the second term is non-negative, the following inequality holds

$$q H_g \leq 4 \langle m | \hat{B}^\dagger \hat{B} | m \rangle \tag{5.31}$$

then, expanding

$$\begin{aligned} \langle m | \hat{B}^\dagger \hat{B} | m \rangle &= \langle m | \langle \psi_i | \partial_g \mathrm{e}^{ig\mathbf{H}} | \psi_f \rangle \langle \psi_i | \partial_g \mathrm{e}^{-ig\mathbf{H}} | \psi_i \rangle | m \rangle \\ &\leq \langle m | \langle \psi_i | \mathbf{H}^2 | \psi_i \rangle | m \rangle \\ &= \langle \mathbf{A}^2 \rangle \langle \mathbf{k}_x^2 \rangle \end{aligned} \tag{5.32}$$

the second inequality follows by replacing the projector $|\psi_f\rangle \langle \psi_f|$ with the identity projector which is a sum over a complete set of postselection states, with all terms positive. The information is corrected by the lowered postselection probability, in accordance with the methodology laid out in Ref. [100]: this leads to the same conclusion as the first order argument above, namely that there is no overall advantage to be found.

5.5 Discussion

Imagine two betting games at a casino. Both games cost £1 to play, and involve flipping a pair of coins. In Game 1, the casino pays a £50 prize for each 'heads', and £0 for each 'tails'. In Game 2, the prize is £200 for two heads, and £0 for the other three possibilities. Which game should be preferred? Clearly the games are

equivalent if an infinite number of trials is possible. If only a small number of trials is possible, one might consider Game 1 to be a surer bet, and Game 2 to be a riskier option. Furthermore, if the prize in the second game was only £199.99, it would seem even less attractive when compared with the first.

This scenario is analogous to noise-free weak-value metrology. It is philosophically interesting that the Fisher information – analogous to the cash prize – can be concentrated into a small number of events: but this is not (by itself) a compelling technological reason for employing WVA. They are even less attractive when one realises that a performance cost will inevitably be paid by the increased experimental complexity of postselection, and the fact that a bias or systematic error is introduced by any finite g, because of the inexactness of the AAV approximation.

The situation might change when noise is ON. For spin-based sensing, when noise was OFF, entanglement was of great benefit. With noise ON, for example when there are unknown random magnetic fields present, the advantage may disappear altogether [178, 84, 174]. For weak-value metrology, noise may make the protocol hard to implement: preselection, postselection, and so on may be imperfect. But there is another intriguing possibility: what if weak-value metrology is more robust to noise than standard techniques? It may be worth paying the price for the nonlinearity bias, and for the low success of probability, if the reward is much higher because noise is mitigated. This question will be the central inquiry of the remaining chapters of this thesis.

Chapter 6

Phase-parameter estimation with dephasing

This chapter is a first foray into applying the Fisher information metric to weak-value 'amplification'. Being inspired by the spin-based scenario in the preceding chapter, it proceeds with an analysis of the performance of a magnetic field sensor. A second spin, nearby and with a switchable interaction, is a natural quantum coherent meter which can enable the AAV effect [212]. Weak values have been extracted with qubit meters experimentally [164].

The chapter makes the following departures from AAV's original discussion: the coherent measuring device is a qubit, rather than a continuous pointer, and I perform only full order calculations. The chapter is organised as follows. First, I introduce a *noisy* spin-based magnetic field sensor, and calculate the quantum Fisher information (QFI) using strong and direct measurements to provide a benchmark for performance. The object of the parameter estimation (to which the QFI relates) is a property of the first spin, and not of the interaction between the spins. Next I allow the second spin to measure the first at arbitrary strength, and again calculate the QFI. Finally I introduce postselection on to the first spin, enabling the AAV effect. I calculate the QFI and compare it to that found for indirect and for direct measurements. I draw conclusions about the relative performance at the end of the chapter.

6.1 A strong and direct benchmark

I extend the spin-based field sensor paradigm, previously introduced in Chapter 5, to include unknown random fields. The random fields coexist with the fixed field.

The magnitude of the latter is to be estimated; the former will be averaged over in the course of repeated measurement of the spin and hence lead to decoherence. The evolution of the spin can then described by a density operator (see Section 1.1.5.3):

$$\begin{aligned} 2\rho_{00}(t) &= 1 \\ 2\rho_{01}(t) &= -i\mathrm{e}^{-i\mu\delta Bt}\Xi(t); \end{aligned} \tag{6.1}$$

ρ is the 2-dimensional density matrix of the spin, and μ is its magnetic dipole moment which I set to unity without loss of generality. The density matrix is written in the spin up $|0\rangle$, spin down $|1\rangle$ basis. As usual, completeness demands $\rho_{11} = 1 - \rho_{00}$ and $\rho_{10} = \rho_{01}^*$. I take $\Xi(t) \in [0, 1]$ to be a real, non-negative function (assumed to be independent of δB) which describes the attenuation of the off-diagonal terms. By choosing this function appropriately one can model, for example, pure dephasing ($\Xi(t) = \mathrm{e}^{-\Gamma t}$) or 1/f noise type decay ($\Xi(t) = \mathrm{e}^{-\Gamma t^2}$) with a characteristic rate Γ [32]. The dependence on t is important but sometimes suppressed in my notation. δB is an unknown quantity and is the subject of the parameter estimation problem. Whilst for clarity I describe the tangible example of field sensing with a spin, the results apply to many other phase estimation scenarios.

The density matrix, through the spin's exposure to the magnetic field, acquires an increasing amount of information about δB as time progresses. However, the build-up of useful information is in competition with the decoherence induced by the random fields. The latter wash out all information after a sufficiently long time. Statistical inferences about δB may still be made after finite time by performing measurement on many identical preparations of the density matrix. The precision of the inferences depends on certain properties of the density matrix and the measurement choice: it is related to the Fisher information (see Chapter 2). Fixing the measurement POVM to a sharp measurement in the $\sigma_x := |0\rangle\langle 1| + |1\rangle\langle 0|$ basis, one finds for a spin in a field

$$F_{\mathrm{d}} = \frac{t^2 \cos^2(t\delta B)\Xi^2}{1 - \Xi^2 \sin^2(t\delta B)}. \tag{6.2}$$

Throughout this chapter, the Fisher information is always 'about' δB. Here, the subscript 'd' is used to denote quantities pertaining to a direct sensing of δB, i.e. through the density matrix described by Equation (6.1). This expression exhibits oscillations over time as the angle between the quantum state and the measurement basis varies between pessimal (when the measurement basis is parallel to the final state, $t\delta B = [n + 1/2]\pi$) to optimal (when the measurement basis is perpendicular to the final state, $t\delta B = n\pi$).

To eliminate the dependence on the measurement choice one can deploy the optimal POVM: since the estimation procedure involves many samples of the probability distribution, some of them may be used to adaptively update the POVM after an initial guess, causing it to rapidly converge on an optimum [28]. A quantity that captures the maximum F in a variation over all POVMs is the quantum Fisher information (QFI), defined as

$$H := 2\sum_{n,m} \frac{|\langle m|\partial_{\delta B}\rho|n\rangle|^2}{P_n + P_m}. \tag{6.3}$$

The above sum only includes terms for which $P_n + P_m \neq 0$ and where n, m index the basis states in the spectral decomposition of the density matrix, $\rho = \sum_i P_i |i\rangle \langle i|$ [153]. In our case

$$H_{\mathrm{d}} = \Xi^2 t^2, \tag{6.4}$$

which is, notably, independent of the parameter δB. The oscillations in time have disappeared, and this leaves only an envelope with turning points which can be found by solving $\dot{\Xi}t^2 = -\Xi$. For example when $\Xi = \mathrm{e}^{-\Gamma t}$ the maximum of H occurs at $t^* = 1/\Gamma$.

According to the quantum Cramér-Rao bound [42, 97], the minimum variance of the parameter δB is given by $1/(NH)$ where N is the number of trials. A larger value of H thus entails a smaller minimum variance, which is obtained efficiently through maximum likelihood estimation [63].

6.2 Arbitrary strength ancilla measurement

An ancillary spin can be used as part of the measurement process, enabling a broader class of measurements (see Chapter 3). Consider an ancillary qubit (the meter) initialised in the equatorial (x-y) plane of the Bloch sphere, coupled to the system qubit so that one has control over the joint system. I consider a measurement operation which effects the following unitary transformation on the system and meter

$$\mathscr{M}(G) = \Pi_+ \otimes \mathbb{I} + \Pi_- \otimes \exp(iG\pi\sigma_z/2). \tag{6.5}$$

$\Pi_\pm$ are projectors onto the ± 1 eigenstates of a traceless 'control observable' (of the system) that lies in the plane, $\Pi_+ - \Pi_- = \cos\Theta\sigma_y + \sin\Theta\sigma_x$. The choice of $\sigma_z := |1\rangle\langle 1| - |0\rangle\langle 0|$ as a meter observable in the second term is chosen because the ancilla is initialised in plane. Best results are obtained when the ancilla measurement basis is unbiased with respect to the axis about which the ancilla rotates (see Chapter 3). The measurement strength can be varied independently of the initial state of the

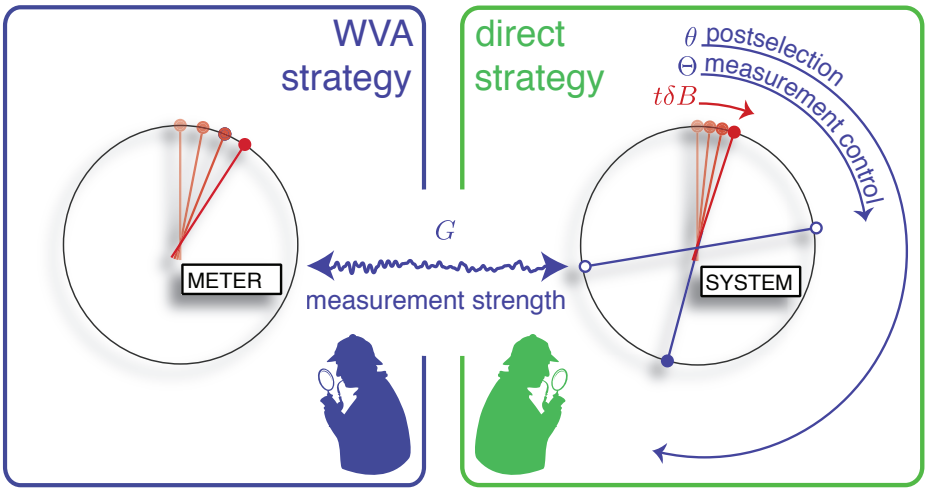


Figure 6.1: A schematic for two different approaches to parameter estimation using spins; the view is into the equatorial plane of the Bloch sphere. A direct strategy relies only on interrogations of the system spin, which picks up a field dependent phase $\exp(it\delta B)$ over time. A WVA strategy makes use of an ancillary spin in the hope of gaining an advantage. The latter technique involves coupling the first spin with a second 'meter' spin with variable strength, and then interrogating the meter spin only if the system spin is successfully postselected into a certain final state. Such a postselection can be achieved by the use of projective measurements in an appropriate basis given by θ, and is known to lead to a larger than expected deflection of the meter spin if the coupling G is weak. Whilst δB is an unknown quantity of interest, θ, Θ, G are tunable by the experimenter in the WVA strategy. Reproduced from *Phys. Rev. A*, 87:012115, January 2013.

meter: the measurement is strong when $G = 1$ and gets weaker as $G \to 0$. After tracing out the system spin, one finds that the QFI of the meter spin is

$$H_{\text{anc}} = \frac{\Xi^2 t^2 \sin^2(G\pi/2)\sin^2(\Theta - t\delta B)}{1 - \Xi^2 \cos^2(\Theta - t\delta B)}. \tag{6.6}$$

The subscript 'anc' denotes quantities pertaining to an indirect sensing protocol, involving the use of an ancilla spin coupled to the system spin with arbitrary measurement strength. Assuming the experimenter has control over the measurement operation, it is clear that she arranges $\Theta = t\delta B + \pi/2$ for best results (which corresponds to the control observable being unbiased w.r.t. the system state), whence

$$H_{\text{anc}} \to \Xi^2 t^2 \sin^2(G\pi/2). \tag{6.7}$$

This matches the performance of the direct strategy when the measurement corresponds to a full controlled $\pi-$rotation, $G = 1$. This is as expected: when $G = 1$ we have simply an explicit model for strong and direct measurements. The WVA scheme

would seem to sacrifice performance by operating near to $G = 0$ (where the dependence of $H_{\rm anc}$ on G is roughly quadratic), however, we shall see that the postselection step of the protocol leads to an 'amplification' which mitigates this apparent loss of information.

6.3 Postselected strategy

Preselection and weak measurement are in place. Next I introduce postselection into the protocol, to enable the AAV effect. The system spin is allowed to pick up phase in the weak field as usual. After a given time the measurement $\mathscr{M}$ is triggered: this entangles the system spin with the meter, which is initialised into the -1 eigenstate of σ_y, $\eta(0) = |i^-\rangle\langle i^-|$. The system spin is then measured, and only if it is found in a certain postselection state $|\psi_f\rangle = (|0\rangle + \mathrm{e}^{i\theta}|1\rangle)/\sqrt{2}$, the ancillary spin is interrogated in the usual manner (using an adaptive maximum likelihood estimation procedure). In the most general case the angle between pre- and postselection can be varied independently of the measurement control angle (as sketched in Figure 6.1). The density matrix of the meter after the measurement interaction and postselection will be given by

$$\eta(t) \propto \langle\psi_f|_s \mathscr{M}\left[\rho(t)\otimes\eta(0)\right]\mathscr{M}^\dagger|\psi_f\rangle_s, \tag{6.8}$$

with the proportionality constant fixed by normalization. Identity matrices are implied in the above expression, i.e. $\langle\psi_f|_s = \langle\psi_f|\otimes\mathbb{I}$ etc. Under these operations, the evolution of the meter is confined to the x-y plane, which is optimal for phase estimation [197]. The full evolution of η is (from Mathematica, cos and sin represented with c and s, $\mathscr{A}_1 = \theta - 2\Theta + t\delta B$, $\mathscr{A}_2 = \theta - t\delta B$, $\mathscr{A}_3 = G\pi/2$, $\mathscr{A}_4 = \theta - \Theta$)

$$\begin{aligned}\eta_{00}(t) &= \frac{1}{2} \\ \eta_{01}(t) &= \frac{ie^{-\frac{1}{2}iG\pi}\left(\mathrm{c}\mathscr{A}_3 + i\mathrm{c}\mathscr{A}_4\mathrm{s}\mathscr{A}_3 + (\mathrm{c}^2[\mathscr{A}_3/2]\mathrm{c}\mathscr{A}_2 - \mathrm{c}\mathscr{A}_1\mathrm{s}^2[\mathscr{A}_3/2] + i\mathrm{c}\mathscr{A}_2\mathrm{s}\mathscr{A}_3)\,\Xi\right)}{2 + 2\left(\mathrm{c}^2[\mathscr{A}_3/2]\mathrm{c}\mathscr{A}_2 + \mathrm{c}\mathscr{A}_1\mathrm{s}^2[\mathscr{A}_3/2]\right)\Xi}.\end{aligned} \tag{6.9}$$

Here I specialise to $\theta = \Theta + \pi/2$ (corresponding to a postselection state that is unbiased w.r.t. the control observable). Due to the symmetry between pre and post selection, one can choose instead to have the preselection unbiased w.r.t. the control observable, as for the $H_{\rm anc}$ above. One does no better in that case. I obtain, after normalizing, the simplified expression:

$$\begin{aligned}\eta_{00}(t) &= \frac{1}{2} \\ \eta_{01}(t) &= \frac{ie^{-\frac{1}{2}iG\pi}(\cos(G\pi/2) + (\cos(\theta - t\delta B) - i\sin(G\pi/2)\sin(\theta - t\delta B)\Xi)}{2 + 2\Xi\cos(G\pi/2)\cos(\theta - t\delta B)}.\end{aligned} \tag{6.10}$$

Of course, $\eta_{11} = 1 - \eta_{00}$ and $\eta_{10} = \eta_{01}^*$. Calculating the QFI can be difficult for arbitrary density matrices, but I found that applying Equation (6.3) to a general qubit state in the equatorial plane leads to

$$H = \frac{(y^2-1)\hat{x}^2 - 2xy\hat{x}\hat{y} + (x^2-1)\hat{y}^2}{x^2+y^2-1}, \tag{6.11}$$

where $x + iy = 2\eta_{01}(t)$ and '$\hat{}$' denotes partial derivative w.r.t. δB. One straightforwardly obtains, taking real and imaginary parts of (6.10),

$$H_{\mathrm{WVA}} = H_{\mathrm{d}} \sin^2(G\pi/2)A. \tag{6.12}$$

The subscript 'WVA' denotes the use of an indirect sensing strategy through the density matrix (6.10) (i.e. an exact AAV type measurement). I have treated the measurement strength entirely generally. The expression admits the following interpretation: the first factor is the bare information available through direct techniques; the second factor represents the cost of having a finite strength measurement, and is present with and without postselection; the final factor

$$A = (1 + \Xi(t)\cos(G\pi/2)\cos(\theta - t\delta B))^{-2} \tag{6.13}$$

is due to the weak value 'amplification' effect. A becomes large when both $\theta - t\delta B$ is close to an odd integer multiple of π and $\cos(G\pi/2) \approx 1$, i.e. for a weak measurement strength and almost orthogonal pre- and postselection. The upper two panels of Figure 6.2 demonstrate that in this regime $H_{\mathrm{WVA}} > H_{\mathrm{d}}$, and the WVA strategy would seem to outperform the direct one.

However, the much reduced postselection probability

$$q = Tr[(|\psi_f\rangle\langle\psi_f| \otimes \mathbb{I})\mathscr{M}\rho(t)\mathscr{M}^\dagger] \tag{6.14}$$

must be taken into account, evaluating to $q = 1/(2\sqrt{A})$. Note that q is nonzero when pre- and postselection are orthogonal due to the back action of the weak measurement [211]. Once the probability is properly accounted for,

$$qH_{\mathrm{WVA}} = \frac{\Xi^2 t^2 \sin^2(G\pi/2)}{2(1 + \Xi\cos(G\pi/2)\cos(\theta - t\delta B))}. \tag{6.15}$$

A fair comparison between the efficiency of the weak value approach and the direct approach can be made by considering the ratio of the CRBs relating to the two strategies. The ratio is the inverse of the Fisher information ratio, which is shown in the lower right panel of Figure 6.2 and given by

$$\frac{qH_{\mathrm{WVA}}}{H_{\mathrm{d}}} = \frac{\sin^2(G\pi/2)}{2(1 + \Xi\cos(G\pi/2)\cos(\theta - t\delta B))}. \tag{6.16}$$

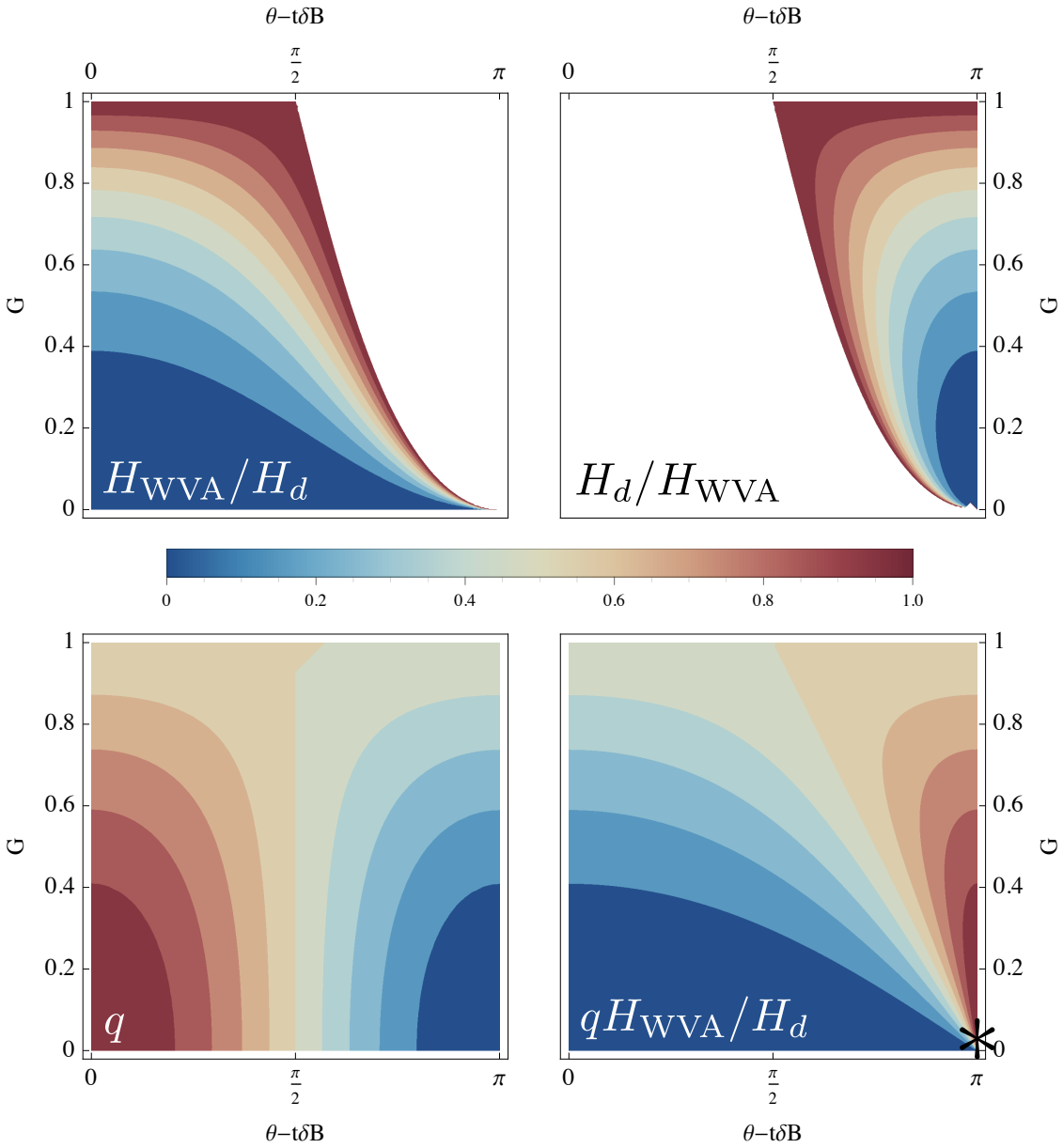


Figure 6.2: Contour lines of four important quantities are plotted against measurement strength G and the angle $(\theta - t\delta B)$ between pre- and postselection. *UL pane:* $H_{\mathrm{WVA}} : H_{\mathrm{d}}$ uncorrected for the low postselection probability. White regions show a ratio higher than one and hence the apparent superiority of the WVA scheme over a direct scheme; *UR pane:* the inverse ratio, white regions now show where the WVA scheme is inferior; *LL pane:* the postselection success probability q; *LR pane:* shows the corrected ratio $qH_{\mathrm{WVA}} : H_{\mathrm{d}}$, note there are no longer white regions and the weak strategy is always worse. The $*$ denotes a choice of (G, θ) that is generalised to include dephasing noise in Figure 6.4. Adapted from *Phys. Rev. A*, 87:012115, January 2013.

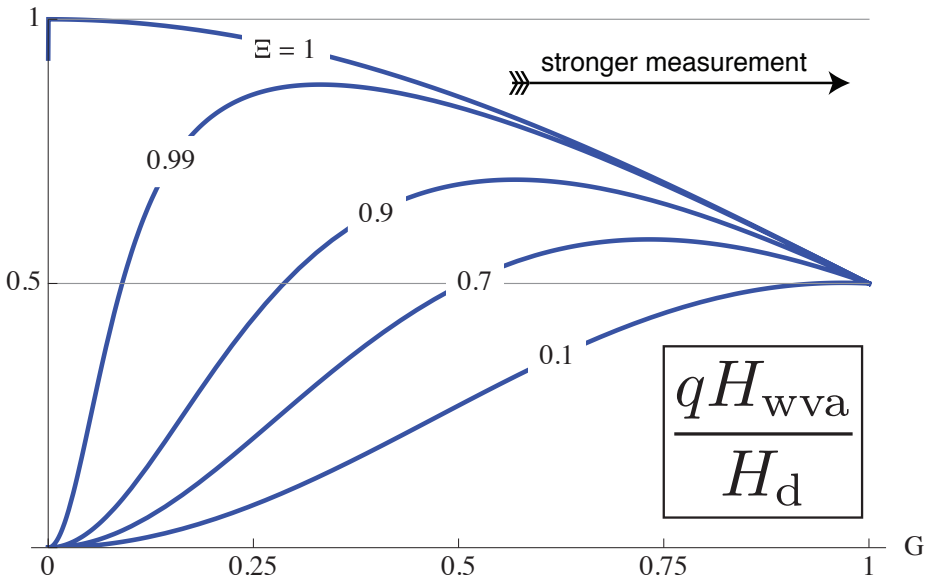


Figure 6.3: Ratio of information available in the WVA strategy and in its direct counterpart, as a function of the measurement strength G. The different curves correspond to various values of the attenuation function $\Xi(t)$. The postselection is fixed to $\theta = t\delta B + \pi$. When $G = 1$, one should match the direct strategy, only the postselection probability $q = 1/2$ implies by symmetry that half of the information is thrown away. Reproduced from *Phys. Rev. A*, 87:012115, January 2013.

By inspection, this expression never exceeds unity for *any* function $\Xi(t) \in [0, 1]$. This argument is sufficient to establish that the WVA technique can never produce an estimate of δB with a lower variance than through a strong, direct technique [100].

When dephasing noise is completely absent then one can reach a ratio of unity for the correct choice of G and θ (see Figure 6.2). This is quite surprising given the heuristic noise-free argument in Section 5.4.1. I attribute the surprising advantage to the effects not captured by the AAV approximation. When pre and postselection are too close to orthogonal, the approximation (5.14) breaks down (specifically gA_w is no longer small compared with unity) and the evolution of the meter becomes non-linear. When preselection and postselection are chosen exactly orthogonal, the weak value is undefined. Because of the breakdown of the approximation, the results of the previous chapter (Equation (5.17)) may not apply. My calculations reveal that the non-linear behaviour, which appears far outside AAV's regime of validity, makes up for the performance penalty associated with a weak measurement. However, here I found that even a small attenuation is catastrophic to the weak value technique because for small G there is a rapid decay of the ratio $qH_{\mathrm{WVA}} : H_{\mathrm{d}}$ as Ξ decreases (see Figure 6.3), and a strong measurement quickly becomes favourable.

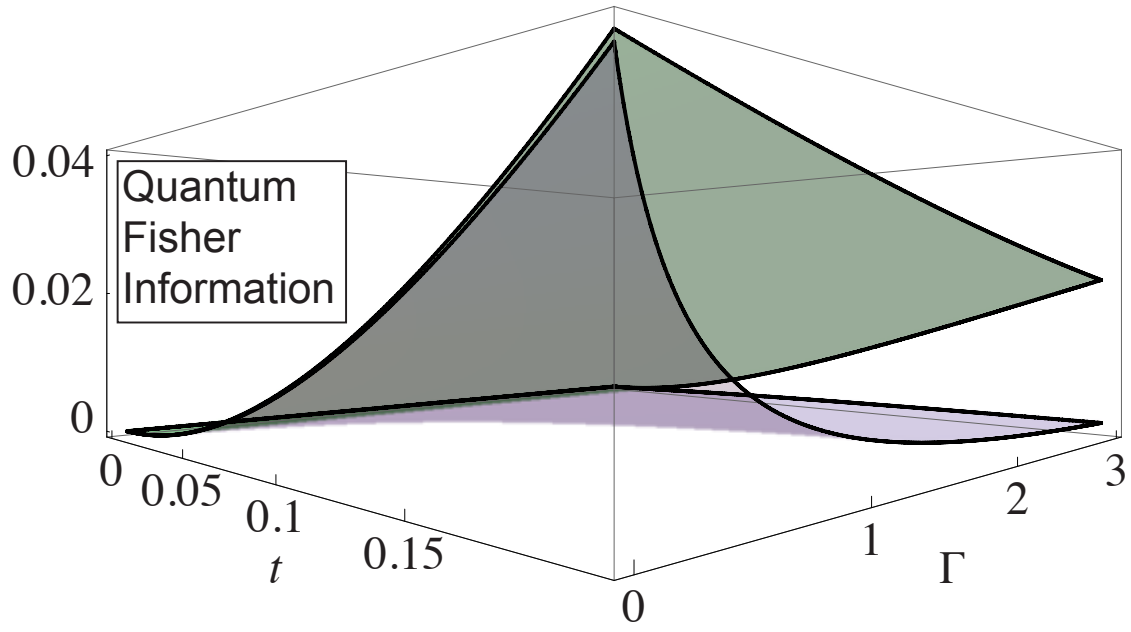


Figure 6.4: Quantum Fisher information (in units of Tesla^{-2}) for the two competing strategies is plotted as a function of time t and the dephasing rate Γ (for $\Xi = \mathrm{e}^{-\Gamma t}$). The upper green surface corresponds to the direct strategy H_{d} and the lower purple surface corresponds to qH_{WVA}, the corrected weak-value amplified strategy. The postselection and measurement strength are fixed to $\theta = t\delta B + \pi, G = 0.02$, respectively, corresponding to the $*$ in Figure 6.2. Even a moderate amount of dephasing has a catastrophic effect on the weak value scheme. Reproduced from *Phys. Rev. A*, 87:012115, January 2013.

To illustrate this behaviour with a concrete example, the time dependence of H_{d} and H_{WVA} is shown in Figure 6.4 for phenomenological dephasing noise, $\Xi(t) = e^{-\Gamma t}$.

6.4 Using all the data

I now address the possibility of keeping *all* of the data after the postselection measurement. Since the WVA technique can get close to the performance of a direct strategy and involves discarding the majority of experimental runs, one might imagine that some of the discarded data may be used to increase the information, perhaps even allowing the technique to outperform the direct strategy. In addition to the QFI arising from successful postselection Equation (6.15), one now has

$$(1-q)H^{\perp}_{\mathrm{WVA}} = \frac{t^2\Xi^2 \sin^2(G\pi/2)}{2 - 2\Xi\cos(G\pi/2)\cos(\theta - t\delta B)} \tag{6.17}$$

resulting from runs that would ordinarily be discarded.

One can see that the two quantities are complementary in the following sense. In the regime where the WVA effect is strongest, the discarded runs carry less and less information. Since the total information is additive, one can achieve

$$\begin{aligned} H_{\text{total}} :=& qH_{\text{WVA}} + (1-q)H^{\perp}_{\text{WVA}} \\ =& \frac{t^2 \sin^2(G\pi/2)}{1 - \Xi^2 \cos^2(G\pi/2)\cos^2(\theta - t\delta B)}. \end{aligned} \tag{6.18}$$

This quantity cannot exceed H_{d}, but has some interesting features. It is greater than qH_{WVA} in regions where the WVA effect is small, notably reaching $qH_{\text{WVA}}/H_{\text{d}} = 1$ (rather than 1/2) when the measurement is fully strong. It converges on qH_{WVA} when the WVA effect is pronounced, see Figure 6.5.

6.5 Discussion

I have analysed the utility of weak value 'amplification' for the purpose of estimating an unknown phase parameter appearing in the Hamiltonian of a two level quantum system, finding no advantage over strong and direct techniques for the broad class of noise models captured by Equation (6.1). This includes any kind of dephasing noise. When decoherence is completely absent the WVA strategy can match the performance of the direct approach, encouraging the motto: 'one postselected run acting as though many unpostselected runs' [61]. In contrast to entanglement or discord enhanced sensing protocols, however, which are robust against a degree of mixing of the quantum state [136], any level of dephasing noise ruins the performance of the WVA approach.

While I have described the system-meter qubit pair as spin-1/2 particles, they are isomorphic to many other physical systems: for example one can use the polarization states of photons to measure a phase shift introduced by a crystal, and couple photons together to enable weak measurement [164].

These results do not seem to contradict studies that have put WVA to use experimentally [83, 51, 190] or the many theoretical proposals for improving signal to noise [219, 30, 98, 61, 82, 181] since in those cases the quantity of interest is an interaction parameter, and only technical noise is claimed to be overcome. When the limiting disturbance is to the quantum state however, rather than to the classical information following the measurement, and when the subject of the parameter estimation is related to A_w, then there is no advantage to be gained by using a weak-value amplified approach. Technical noise, which is the dominant noise in optical systems, is the subject of the next chapter.

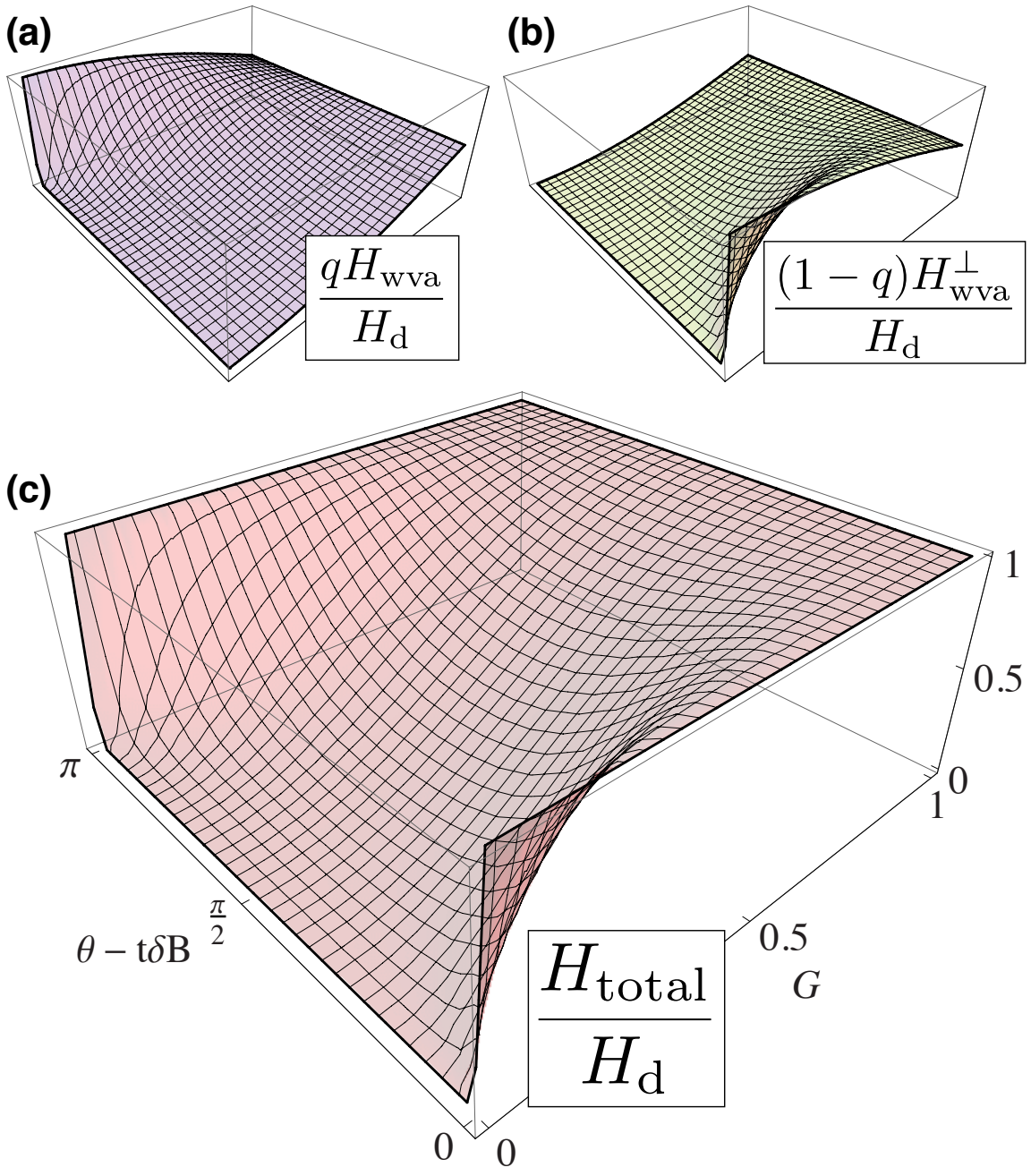


Figure 6.5: Quantum Fisher information for a weak-value amplified sensing strategy **(a)** when the postselection is 'successful' **(b)** when the postselection is 'unsuccessful' **(c)** when both are considered together; the plots shown correspond to the decoherence-free case. Reproduced from *Phys. Rev. A*, 87:012115, January 2013.

Chapter 7

Interaction-parameter estimation with technical noise

In Chapter 3 we saw that after a quantum-coherent premeasurement, the meter wave function encodes information about the 'system' through their mutual interaction. By employing postselection, the weak-value 'amplification' technique can probabilistically generate a larger-than-usual displacement of the wave function. Is this of of use when technological constraints are present? Most of the experimental interest in weak-value 'amplification' (WVA) has been in optical scenarios: see for example [192]. In this chapter I focus on the foremost technological application of the Aharonov, Albert and Vaidman (AAV) effect: in the estimation of the small constant (known as an interaction parameter) found to couple a suitably defined system and meter. Typically the system of interest has a finite spectrum, and the measuring device is a continuous degree of freedom. In 2008 Hosten and Kwiat detected the spin Hall effect of light (which is coupling between the polarisation and transverse momentum of photons at an air-glass boundary) using WVA [83]. Chapter 6 modelled decoherence, which is the dominant and limiting noise in solid state systems. This chapter models technical noise, or detector imperfection, which limits performance in optical systems.

This chapter is arranged as follows. First, I present a full order, noise free argument contra weak-value 'amplification' for a qubit coupled to a continuous system initialised into a Gaussian state. I plot the relative performance (against a preselected-only benchmark) across the entire parameter space of the problem, and also allow for the WVA technique to operate even when postselection fails. Next, I adopt an approximate, first-order approach to the AAV effect, and repeat the calculations. This lends greater transparency to the results, and allows for the qubit to be upgraded to a *qudit* (a d-level system). The main results of this chapter concern noise. Using the

approximate approach, I model detector pixelation and random displacements (jitter) as the most prevalent types of technical noise. The results of the second half of this chapter are general enough to apply to a wavefunction of arbitrary shape in most cases. I specialise to a Gaussian wavefunction where convenient, or when instructive.

Experiments have been performed that suggest advantages over noise: the most prominent claim a suppression of technical issues such as pointing stability, detector saturation, finite resolution, unwanted displacements and noise with long correlation times [166, 83, 51, 30, 74]. I will show that, for the most ostensible technical issues, namely random transverse displacements and pixelation, such an advantage only prevails whilst sub-optimal estimation procedures are employed. If the fundamental limit on the uncertainty in the estimate of the interaction parameter is attainable (by optimal estimation), the advantage disappears.

Begin with a particle described by a product of its internal state and its transverse spatial wavefunction. Expanding the initial state $|\psi_i\rangle$ in the eigenbasis of an internal *control observable* $\mathbf{A}|a_j\rangle = \lambda_j|a_j\rangle$ one has

$$|\psi_i\rangle|m\rangle = \sum_j c_j|a_j\rangle|m\rangle\,, \tag{7.1}$$

where c_j are an appropriate set of normalised amplitudes and $|m\rangle$ is the normalised transverse component of the particle's spatial degree of freedom, which will serve as a quantum coherent measuring device. One can expand the spatial degree of freedom, for example, $|m\rangle = \int m(x)dx|x\rangle$ in the basis of $\mathbf{x}$ eigenstates. Allow the particle to undergo a displacement, transverse to its direction of propagation and dependent on the internal state. The time evolution operator is

$$U = \mathrm{e}^{-ig\mathbf{A}\mathbf{k}_x}\,, \tag{7.2}$$

where $\mathbf{k}_x$ is the operator corresponding to the x-component of the particle's momentum $[\mathbf{k}_x, \mathbf{x}] = i$, and g is the coupling constant or 'interaction parameter' [82]. The dynamics is unitary since $\mathbf{A}$ is Hermitian and has real eigenvalues. Thus it generates translations in the meter variable conjugate to $\mathbf{k}_x$. In this case we have

$$|a_j\rangle \int m(x)dx|x\rangle \to |a_j\rangle \int m(x-\lambda_j g)dx|x\rangle\,, \tag{7.3}$$

i.e. displacements in x. The analyses in this chapter apply equally if $\mathbf{k}_x$ and $\mathbf{x}$ are interchanged, so that the interaction induces shifts in momentum space.

Under this dynamics, each of the eigenstates of the control observable becomes correlated to a separate wavefunction, which, up to the proportionality constant g, is

centred on the corresponding eigenvalue. When $\lambda_j g$ is much larger than the 'width' of $m(x)$, this constitutes a strong premeasurement of the internal state of the particle by its spatial wavefunction. Formally, this occurs when the overlap $\mathcal{O}_{ij} := \int \psi^*(x - \lambda_i g) m(x - \lambda_j g) dx$ between each pair of shifted wavefunctions is vanishingly small. When $\lambda_j g$ is relatively small and the wavefunctions are no longer well resolved, the premeasurement is said to be *weak*, see Chapter 3.

7.1 Benchmarking

The key figure of merit that we shall be concerned with is the Cramér-Rao bound on an estimate of g (see Chapter 2). I shall compare that of the WVA strategy (which gives rise to p_{WVA} with probability q) to a standard strategy (which gives rise to p_{std} with certainty). In the limit of $N \to \infty$, their ratio is equal to $qF_g[p_{\mathrm{WVA}}]$ to $F_g[p_{\mathrm{std}}]$. Here

$$F_g[p] := \int \frac{(\partial_g p(x))^2}{p(x)} dx \tag{7.4}$$

is the Fisher information 'about g', and $p(x) = |m(x)|^2$ (the distribution of meter readings). The Fisher information should be corrected (multiplied) by the probability of successful postselection. This is in accordance with the additivity of the Fisher information over N independent events (of which only qN are available to the WVA strategy).

My *standard* strategy benchmark performs preselection only. This means one should average over a complete set of postselection states. Beginning from Equation (7.3), one traces over the internal degree of freedom and measures the particle in the x basis to give

$$p_{\mathrm{std}}(x) = \sum_j |c_j|^2 |m(x - \lambda_j g)|^2 \,. \tag{7.5}$$

The weighted sum is a convex combination: by controlling the c_j one can mix the probability densities corresponding to the different eigenvalues of $\mathbf{A}$. An immediate optimization for the standard strategy presents itself: The Fisher information metric is convex [39], meaning that any such mixing of probability distributions will be sub-optimal. One should therefore choose the c_j to filter the probability distribution with eigenvalue λ_* that gives the highest Fisher information. Now $F_g[p_{\mathrm{std}}] = F_g[|m(x - \lambda_* g)|^2]$.

If the system is a qubit, then $\lambda = \pm 1$, and if the wavefunction is a Gaussian,

$$m(x) = \sqrt{\frac{1}{\sqrt{2\pi}\Delta_x}} \exp\left(-\frac{x^2}{4\Delta_x^2}\right) , \tag{7.6}$$

then

$$F_g[p_{\text{std}}] = \frac{1}{\Delta_x^2}. \tag{7.7}$$

This quantity is my benchmark: it is the standard by which WVA will be judged.

7.1.1 Fisher information or signal-to-noise?

The Fisher information (see Chapter 2) is a well defined metric on statistical models, with a powerful operational meaning. The signal-to-noise ratio (SNR) has previously been the most commonly used metric in the weak value literature [218, 98, 190]. The signal is defined as the average result on the detector $\langle x \rangle$, and the noise is defined as the standard deviation at the detector $\sqrt{\langle x^2 \rangle - \langle x \rangle^2}$. For a transversally shifted Gaussian distribution, the square of the signal to noise ratio is equal to the Fisher information. A Gaussian distribution is completely characterised by its first and second moments: for more complicated distributions, one might expect the SNR and Fisher information to diverge, since the latter is generally dependent on all moments of the distribution.

Note also that if, through some strange mechanism, the shifted Gaussian distribution undergoes a fixed shift, of 1000 say, the SNR increases but the Fisher information does not (since this extra shift does not depend on the parameter of interest g). This fact undermines the SNR as a good measure of performance – although it does agree with the Fisher information in many cases. When there is useful information in the higher moments of the meter distribution, the Fisher information should be favoured over the SNR.

7.2 Full order WVA model

Rather than ignore the system after the system-meter interaction, the WVA strategy instead allows the particle to undergo another internal-state-dependent shift, this time in the y-direction and at full strength, i.e. with an interaction Hamiltonian $\propto \mathbf{B}\mathbf{k}_y$. The product of coupling constant and interaction time should be large enough such that the eigenstates $|\psi_f\rangle$ of observable $\mathbf{B}$ are well separated into distinct 'branches'. By selecting only one branch for investigation, one *postselects* the particle into a

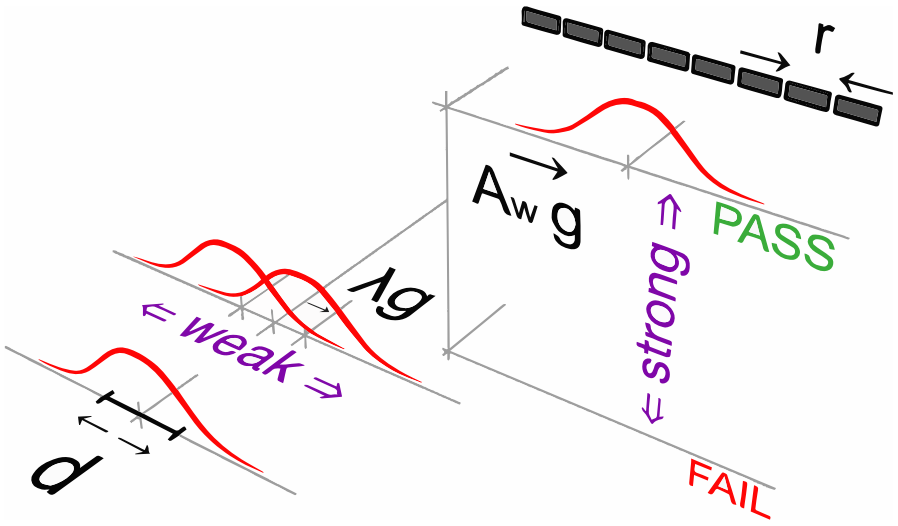


Figure 7.1: The AAV setup. A particle's spatial wavefunction is weakly coupled to an internal observable **A**, causing it to undergo a small lateral shift conditional on the internal state. It is then strongly post-selected into the eigenstate of another observable **B** before impinging on a array of pixels with width r. By tuning certain parameters one can induce larger-than-usual average displacements at the detection stage. This figure illustrates the $\dim(\mathscr{H}) = 2$ case. Reproduced from *Phys. Rev. X*, 4:011032, March 2014.

single final internal eigenstate of **B**, see Figure 7.1. Alternatively, postselection may be achieved with the use of a polarising filter, or suitable device which destroys the particle unless it is in the correct eigenstate. Any such state can be expanded in the basis of **A**, $|\psi_f\rangle = \sum_j c'_j |a_j\rangle$. The final state in the 'success' branch will then be given by[1]

$$|m_{\mathrm{WVA}}\rangle = \frac{1}{\sqrt{q}} \sum_j c_j \bar{c}'_j \int m(x - \lambda_j g) dx |x\rangle \,, \tag{7.8}$$

where the normalization q represents the probability of successful postselection, and depends on g through $\mathcal{O}_{ij}$.

The AAV approximation will be made later in this chapter; it ignores certain non-linear effects of the system-meter interaction and can overestimate the centroid of the probability distribution [48, 107, 126]. When the approximation breaks down, the position of the centroid can deviate from A_w and the wavefunction shape changes, sometimes featuring a secondary peak [56, 211, 69]. Such subtleties motivate the use of exact, full order calculations. Comparing against the preselection-only benchmark, I first show that there can be no advantage from WVA without using the AAV approx-

[1] $\bar{c}$ is the complex conjugate of c.

imation. For simplicity, and because it is almost exclusively the case in experiments, I restrict the dimension of the system to 2.

7.2.1 How to make A_w real or imaginary for a qubit

As discussed in Chapter 5, AAV prove a simplified expression for the final meter state which is valid when g is sufficiently small. Although an exact treatment makes discussion of their simplifying weak value

$$A_w = \frac{\langle\psi_f|\mathbf{A}|\psi_i\rangle}{\langle\psi_f|\psi_i\rangle} \tag{7.9}$$

rather redundant, the distinction between real and imaginary weak values translates into a natural condition on the relationship between preselection, postselection and control observable. Here I parameteris the initial and final states of the qubit by

$$\begin{aligned}|\psi_i\rangle &= \cos\frac{\theta_i}{2}|+\rangle + \mathrm{e}^{i\phi_i}\sin\frac{\theta_i}{2}|-\rangle \\ |\psi_f\rangle &= \cos\frac{\theta_f}{2}|+\rangle + \mathrm{e}^{i\phi_f}\sin\frac{\theta_f}{2}|-\rangle\end{aligned} \tag{7.10}$$

and define $\delta\phi = \phi_i - \phi_f$. Here $\mathbf{A}|\pm\rangle = \pm|\pm\rangle$. Begin by rationalising

$$\begin{aligned}A_w =& \frac{\langle\psi_f|\mathbf{A}|\psi_i\rangle\langle\psi_i|\psi_f\rangle}{|\langle\psi_f|\psi_i\rangle|^2} \\ \propto& \left(\cos\frac{\theta_i}{2}\cos\frac{\theta_f}{2} - \sin\frac{\theta_i}{2}\sin\frac{\theta_f}{2}\mathrm{e}^{i\delta\phi}\right)\left(\cos\frac{\theta_i}{2}\cos\frac{\theta_f}{2} + \sin\frac{\theta_i}{2}\sin\frac{\theta_f}{2}\mathrm{e}^{-i\delta\phi}\right)\end{aligned} \tag{7.11}$$

where the constant of proportionality is real. Now

$$A_w \propto \cos^2\frac{\theta_i}{2}\cos^2\frac{\theta_f}{2} - \sin^2\frac{\theta_i}{2}\sin^2\frac{\theta_f}{2} + 2i\cos\frac{\theta_i}{2}\cos\frac{\theta_f}{2}\sin\frac{\theta_i}{2}\sin\frac{\theta_f}{2}\sin\delta\phi. \tag{7.12}$$

By inspection

$$A_w \begin{cases} \propto i & \text{if } \cos^2\frac{\theta_i}{2}\cos^2\frac{\theta_f}{2} = \sin^2\frac{\theta_i}{2}\sin^2\frac{\theta_f}{2} \\ \in \mathbb{R} & \text{if } \delta\phi = 0. \end{cases} \tag{7.13}$$

The arguments in the remainder of this chapter are often made twice: both for a choice of pre- and postselection that guarantees a pure imaginary weak value, and for a choice that guarantees a real weak value.

7.2.2 Imaginary-weak-value 'amplification'

Here I choose $\theta_i = \theta_f = \pi$ which guarantees a purely imaginary weak value. The position wavefunction is then

$$m_{\mathrm{WVA}}(x) = \frac{1}{\sqrt{2q}}\left(m(x+g) + \mathrm{e}^{i(\phi_i-\phi_f)}m(x-g)\right). \tag{7.14}$$

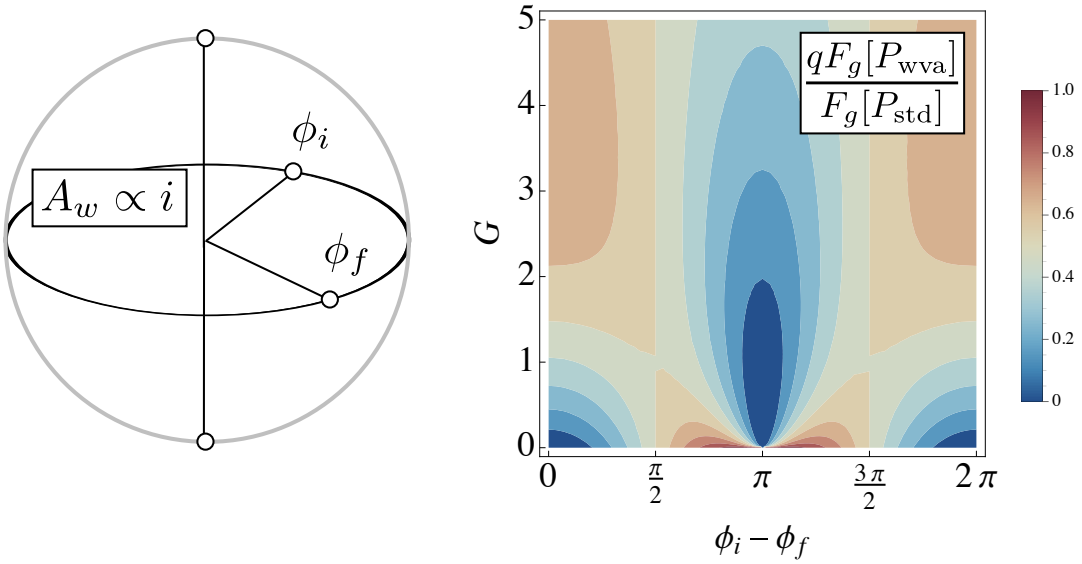


Figure 7.2: Contour plot of the corrected ratio of Fisher information for the weak-value 'amplification' strategy to Fisher information for the standard strategy, expressed as a function of the measurement strength G and angle between pre- and postselection $\phi_i - \phi_f$. The plot relates to imaginary weak values. Adapted from *Phys. Rev. X*, 4:011032, March 2014.

But detection is not usually performed in the position basis when the weak value is imaginary. Instead the Fourier transform yields the momentum wavefunction

$$\tilde{m}_{\text{WVA}}(k_x) = \frac{1}{\sqrt{2q}}\tilde{m}(k_x)\left(\mathrm{e}^{-igk_x} + \mathrm{e}^{i(\phi_i-\phi_f+gk_x)}\right) . \tag{7.15}$$

Because the spatial wavefunction is a Gaussian with waist Δ_x, the momentum wavefunction is also Gaussian with waist $\Delta_{k_x} = 1/(2\Delta_x)$. Thus the measurement strength $G = g^2/\Delta_x^2 = 4g^2\Delta_{k_x}^2$. Computing the corrected ratio of informations leads to

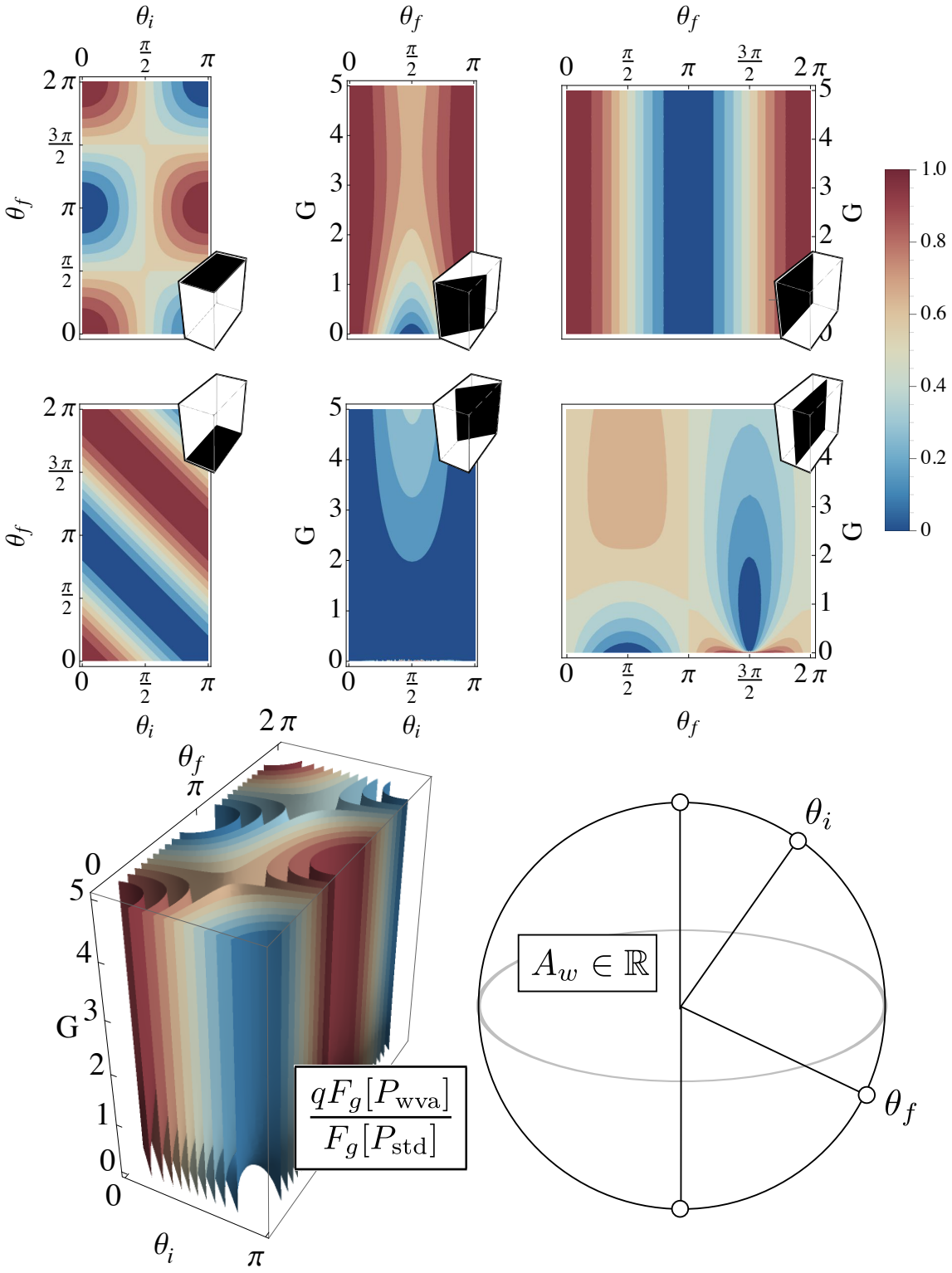
$$\frac{qF_g[p_{\text{WVA}}(k_x)]}{F_g[p_{\text{std}}(x)]} = \frac{2e^{G/2}G\cos(\delta\phi) + 2e^G - \cos(2\delta\phi - 1)}{4e^{G/2}\cos(\delta\phi) + 4e^G}. \tag{7.16}$$

I once more compare against the standard strategy which measures the particle in the x-basis. By inspection the ratio can never exceed unity, see Figure 7.2.

7.2.3 Real-weak-value 'amplification'

Setting $\delta\phi = 0$ guarantees a purely real weak value: an optimal choice for detection in the position basis. With such a choice one finds, without approximation and by Equation (7.8), that

$$m_{\text{WVA}}(x) := \langle x|m_{\text{WVA}}\rangle = \frac{1}{\sqrt{q}}\left(\cos\frac{\theta_i}{2}\cos\frac{\theta_f}{2}m_+ + \sin\frac{\theta_i}{2}\sin\frac{\theta_f}{2}m_-\right) , \tag{7.17}$$


θ_i
θ_f
G
$\frac{qF_g[P_{\text{wva}}]}{F_g[P_{\text{std}}]}$
$A_w \in \mathbb{R}$

where $m_\pm = m(x \pm g)$. In the absence of noise, one can find an analytic expression for the Fisher information ratio, again taking $m(x)$ as a Gaussian (7.6),

$$\frac{qF_g[p_{\mathrm{WVA}}]}{F_g[p_{\mathrm{std}}]} = \frac{1}{2}\Bigg(1 + \cos\theta_f \cos\theta_i - e^{-\frac{G}{2}} \sin\theta_f \sin\theta_i + \frac{G(1+\cos\theta_f\cos\theta_i)}{1 + e^{G/2}\mathrm{cosec}\,\theta_f \mathrm{cosec}\,\theta_i (1+\cos\theta_f\cos\theta_i)}\Bigg). \tag{7.18}$$

I have defined $G = g^2/\Delta_x^2$ as the *measurement strength*, see Chapter 3. It is instructive to examine (7.18) in certain limiting cases, shown in Figure 7.3. Treating the situation at this level of generality allows for the study of intermediate parameter regimes – for example when the weak measurement back-action is non negligible [61]. The ratio is entirely symmetric with respect to interchanging $\theta_i \leftrightarrow \theta_f$, and is also invariant under $\theta_i + \pi, \theta_f + \pi$. Whilst the ratio (7.18) can reach unity, it never exceeds it.

7.2.4 Failed postselection

One might wonder whether the (potentially large) number of discarded events occurring at the postselection stage may be of use. One recent proposal suggested 'recycling' the rejected particles for another run [54]. By contrast, in Chapter 6 I considered looking at the statistics in both 'pass' and 'fail' branches of the postselection. Calculating the sum of Fisher informations in both branches, I found that

---

Figure 7.3 *(preceding page)*: The corrected Fisher information for the real-WVA strategy is shown in units of the standard strategy information as a contour plot in the measurement strength G and the pre- and postselection angles θ_i, θ_f. Two-dimensional slices through this three-dimensional plot are taken for interesting parameter combinations. **The three uppermost contour plots** indicate parameter combinations for which WVA can mimic the standard technique (red regions). From left to right: In the limit of a strong measurement $G \to \infty$; when preselection and postselection are parallel $\theta_i = \theta_f$; and when preselection is into an eigenstate of the control observable $\theta_i = 0$. **The three lowermost contour plots** indicate combinations for which the WVA technique can exhibit characteristically amplified centroids. From left to right: in the limit of a weak measurement $G \to 0$; when pre- and postselection are chosen orthogonal $\theta_f = \theta_i + \pi/2$; and when preselection is into a state unbiased w.r.t. the control observable $\theta_i = \pi/2$. The three-dimensional plot isosurfaces have colors corresponding to fractional values of $qF_g[p_{\mathrm{WVA}}]/F_g[p_{\mathrm{std}}]$: in the two-dimensional plots the same colors shade regions between contour lines. The symmetries of the parameter space allow one to restrict $\theta_i \in [0, \pi]$ with no loss in generality. Adapted from *Phys. Rev. X*, 4:011032, March 2014

(although the Fisher information is by definition greater) the ratio of informations remains at unity or below [100]. Here I consider the use of data in both branches as well as their correlations with a 'which branch variable' (call this b). In other words, I upgrade the WVA strategy; replacing postselection with a final measurement and allowing any postprocessing of the result. The ultimate performance of this class of estimation strategies will be given by the Fisher information calculated in the *joint* probability distribution over b, s:

$$F_g[\mathcal{P}(b,x)] := \sum_b \int ds \frac{(\partial_g \mathcal{P}(b,s))^2}{\mathcal{P}(b,s)}. \tag{7.19}$$

Here s may be the position x or momentum k_x variable. A joint distribution generally contains more structure (and more information) than is revealed by its marginal distributions. The postselection branch index b carries information about g. When the system is a two-dimensional Hilbert space, b is a binary pass/fail variable.

In two subsections below, I present an argument that the information in the joint distribution is still never greater than the standard benchmark information. First I examine real weak values, and then imaginary weak values. In both cases the wavefunction can take an arbitrary shape. In both cases I shall make use of the following fact: if one has a joint distribution

$$\mathcal{P}(b,x) = \delta_{b1}\bar{p}^{\text{pass}} + \delta_{b0}\bar{p}^{\text{fail}}, \tag{7.20}$$

where the bar denotes an unnormalised probability distribution, the Fisher information will be

$$\begin{aligned} F_g\left[\mathcal{P}(b,x)\right] :=& \sum_b \int dx \frac{\left(\partial_g \mathcal{P}(b,x)\right)^2}{\mathcal{P}(b,x)} \\ =& \int dx \frac{\left(\partial_g \bar{p}^{\text{pass}}\right)^2}{\bar{p}^{\text{pass}}} \\ &+ \int dx \frac{\left(\partial_g \bar{p}^{\text{fail}}\right)^2}{\bar{p}^{\text{fail}}}\,. \end{aligned} \tag{7.21}$$

7.2.4.1 Real weak values

Let $b = 1$ when postselection passes and $b = 0$ when postselection fails: by Equation (7.17), the joint probability distribution for system and meter is then

$$\begin{aligned}\mathcal{P}(b,x) &= \delta_{b1}\left(\cos\frac{\theta_i}{2}\cos\frac{\theta_f}{2}m_+ + \sin\frac{\theta_i}{2}\sin\frac{\theta_f}{2}m_-\right)^2 \\ &+ \delta_{b0}\left(-\cos\frac{\theta_i}{2}\sin\frac{\theta_f}{2}m_+ + \sin\frac{\theta_i}{2}\cos\frac{\theta_f}{2}m_-\right)^2 \\ &= \delta_{b1}\bar{p}^{\text{pass}} + \delta_{b0}\bar{p}^{\text{fail}}\,, \end{aligned} \tag{7.22}$$

where $m_\pm$ is a real wavefunction with arbitrary shape that has undergone a positive (negative) shift by some function of g. The probability densities $\bar{p}(x) = m^2$, with $m = \sum_i e_i m_i$, may not be normalised. The Fisher information integrand

$$\frac{(\partial_g \bar{p})^2}{\bar{p}} = \frac{(2m\partial_g m)^2}{m^2} = 4\left(\partial_g m\right)^2\,, \tag{7.23}$$

which follows by the chain rule. Now write

$$\left(\partial_g m\right)^2 = \left(\sum_i e_i \partial_g m_i\right)\left(\sum_j e_j \partial_g m_j\right) \tag{7.24}$$

by taking the derivative inside the sum. Expanding the sum into diagonal and off-diagonal terms, multiplying by four and integrating over x gives

$$\begin{aligned}\int dx\left[\left(\partial_g\bar{p}\right)^2/\bar{p}\right] &= \sum_i e_i^2 F_g[m_i^2] \\ &+ 4\sum_{i\neq j} e_i e_j \int dx \partial_g m_i \partial_g m_j\,, \end{aligned} \tag{7.25}$$

where $F_g[\bullet]$ is the Fisher information functional. The expression is not an information unless $\bar{p}$ is normalised: otherwise it may be negative. Combining Equations (7.21) and (7.25), we have

$$\begin{aligned}F_g\left[\mathcal{P}(b,x)\right] &= \left(\cos^2\frac{\theta_i}{2}\cos^2\frac{\theta_f}{2} + \sin^2\frac{\theta_i}{2}\sin^2\frac{\theta_f}{2}\right)F_g[m_+^2] \\ &+ \left(\cos^2\frac{\theta_i}{2}\sin^2\frac{\theta_f}{2} + \sin^2\frac{\theta_i}{2}\cos^2\frac{\theta_f}{2}\right)F_g[m_-^2] \\ &+4\left(2\cos\frac{\theta_i}{2}\cos\frac{\theta_f}{2}\sin\frac{\theta_i}{2}\sin\frac{\theta_f}{2} - \right. \\ &\quad \left. 2\cos\frac{\theta_i}{2}\sin\frac{\theta_f}{2}\sin\frac{\theta_i}{2}\cos\frac{\theta_f}{2}\right) \\ &\quad \times \int dx \partial_g m_+ \partial_g m_-\,. \end{aligned} \tag{7.26}$$

The final term is zero. By symmetry we have $F_g[m_+^2] = F_g[m_-^2]$ giving:

$$F_g[\mathcal{P}(b,x)] = F_g[m_\pm^2]\,, \tag{7.27}$$

which can be obtained through a purely preselected strategy.

7.2.4.2 Imaginary weak values

By Equation (7.15), we see that the joint probability distribution is

$$\begin{aligned}\mathcal{P}(b,k_x) &= \frac{1}{2}|\tilde{m}(k_x)|^2\left[\delta_{b1}\left(1+\cos(\phi_i-\phi_f+2gk_x)\right)\right.\\ &\quad\left.+\,\delta_{b0}\left(1-\cos(\phi_i-\phi_f+2gk_x)\right)\right].\\ &= \delta_{b1}\bar{p}^{\text{pass}} + \delta_{b0}\bar{p}^{\text{fail}}.\end{aligned} \tag{7.28}$$

We can once more employ Equation (7.21) to arrive at

$$\begin{aligned}F_g[\mathcal{P}(b,k_x)] =& \frac{1}{2}\int dk_x \frac{(\tilde{p}(k_x))^2}{\tilde{p}(k_x)}\left[\frac{(-2k_x\sin(2gk_x+\phi_i-\phi_f))^2}{1+\cos(2gk_x+\phi_i-\phi_f)}\right.\\ &\left.+\,\frac{(-2k_x\sin(2gk_x+\phi_i-\phi_f))^2}{1-\cos(2gk_x+\phi_i-\phi_f)}\right]\\ =& \int dk_x|\tilde{m}(k_x)|^2(-2k_x)^2\\ =& 4\langle\mathbf{k}_x^2\rangle\\ =& F_x[p(x)].\end{aligned} \tag{7.29}$$

The last step follows from the Heisenberg uncertainty principle [65]. This information can be obtained through preselection only, see Equation (7.17).

7.3 First order WVA model

In order to study noise (in the form of detector imperfections), it is convenient to reintroduce some approximations at this stage. AAV showed that, when g is small, the effect (of weak measurement and postselection) on the wavefunction is captured by a single quantity $A_w = \langle\psi_f|\mathbf{A}\,|\psi_i\rangle\,/\langle\psi_f|\psi_i\rangle$ [2], their well known 'weak value'. For details, see Chapter 5. To linear order in g, the distribution over x becomes

$$\begin{aligned}p_{\text{WVA}}(x) &:= |\langle x|m_{\text{WVA}}\rangle|^2\\ &\approx |m(x-g\text{Re}(A_w))|^2\,,\end{aligned} \tag{7.30}$$

where I assume only that the initial spatial wavefunction is real-valued but do not make assumptions about its shape [193, 49, 194]. If we take the initial momentum

space wavefunction to be a Gaussian $\tilde{m}(k_x) \propto \mathrm{e}^{-k_x^2/4\Delta_{k_x}^2}$, the distribution over k_x transforms to [107]

$$\begin{aligned}\tilde{p}_{\mathrm{WVA}}(k_x) &:= |\langle k_x|m_{\mathrm{WVA}}\rangle|^2 \\ &\approx |\tilde{m}(k_x - 2g\Delta_{k_x}^2 \mathrm{Im}(A_w))|^2. \end{aligned} \tag{7.31}$$

Under these assumptions, the shape of the meter wavefunction is not changed, and the perturbation effect can be modelled through the simple shifts above: for a detailed discussion of departures from this behaviour, see Ref [107]. The AAV approximation is heavily relied on in experiments [166, 83, 30, 204, 191, 82, 190, 213]: but many theory papers investigate where it can break down [56, 69, 211, 151].

The motivation behind using postselection is that, where the AAV approximation is valid and when $\langle\psi_f|\psi_i\rangle \to 0$, the average displacement is much larger than is otherwise possible. The large displacement is often claimed to be an advantage, especially for overcoming sources of technical noise. A simple thought experiment is seductive: imagine that ordinarily the shift induced by the weak measurement $\lambda_j g$ is smaller than the width of a single pixel in a digital sensor (such as a CCD)[2]. An amplified shift, if large enough, could then be distinguished from no shift at all, whereas previously no such distinction was possible [30]. This thought experiment presupposes an approach to parameter estimation where only the mean of the probability distribution is important. However, the mean is not always a *sufficient* statistic. As we shall see, if one makes use of the entire statistical distribution to inform one's estimate of g, then a larger displacement is not *per se* advantageous. Further, no single-shot distinction between possibilities will be likely in an experiment employing the AAV effect – the final meter state has the same quantum uncertainty as it did initially, which must be much greater than g. Any inferences must therefore proceed statistically, and make use of a large ensemble. Surprisingly, small shifts do not make g harder to estimate than large shifts.

7.3.1 Ideal Detector

Although the aim of this section is to study certain detector imperfections, to begin I will consider a stable detector having infinite resolution. It is instructive to investigate the Fisher information of a general distribution $p(s') = |\psi(s')|^2$ whose argument is

[2]The shift might instead be smaller than the width of a photodiode stepped through finite displacements [166]

shifted $s' = s - \nu g$. Again s may be either a position or momentum variable. Consider the derivative w.r.t. g, which by the chain rule

$$\begin{aligned} \partial_g p(s - \nu g) &= \frac{\partial s'}{\partial g}\frac{\partial p(s')}{\partial s'} \\ &= -\nu \partial_{s'} p(s') \,, \end{aligned} \tag{7.32}$$

and so

$$\begin{aligned} F_g[p(s - \nu g)] &= \int \nu^2 \frac{(\partial_{s'} p(s'))^2}{p(s')} ds \\ &= \nu^2 F_s[p(s)] \,, \end{aligned} \tag{7.33}$$

as $ds = ds'$ under the change of integration variable. $F_s[p(s)]$ is the information about s, and is related to the variance for Gaussian distributions. Notably, the information is *independent of the size of the parameter* g [66]. The multiplier ν, however, acts as a sort of velocity for the statistical model: it represents the rate at which the probability density changes as g is swept through parameter space. The independence of F on g is an important difference to other measures of metrological performance: for example the signal-to-noise ratio given in [14] and employed by Refs. [190, 51, 98] shows a dependence on the size of the shift itself rather than its rate of change.

Equation (7.33) implies that in the standard strategy ν should be given by the largest eigenvalue λ_*. The standard strategy, then, prepares the system in the eigenstate of $\mathbf{A}$ having the largest eigenvalue, and performs *no* final measurement on the system.

In the WVA strategy, a final measurement gives rise to the AAV effect, and by Equation (7.30) or (7.31) one of the branches experiences a statistical velocity given by an anomalously large effective eigenvalue $\mathrm{Re}(A_w)$ or $2\Delta_{k_x}\mathrm{Im}(A_w)$, respectively. By assuming that the experimenter chooses either to measure in position space or momentum space, and therefore will choose either $A_w = \mathrm{Re}(A_w)$ or $A_w = \mathrm{Im}(A_w)$, respectively, one can replace the real or imaginary part of the weak value with its modulus. Then taking a ratio of the WVA strategy to the standard strategy gives

$$\frac{F_g[p_{\mathrm{WVA}}]}{F_g[p_{\mathrm{std}}]} = \frac{|A_w|^2}{\lambda_*^2} \tag{7.34}$$

for real weak values or

$$\frac{F_g[\tilde{p}_{\mathrm{WVA}}]}{F_g[p_{\mathrm{std}}]} = \frac{|A_w|^2}{\lambda_*^2}\frac{4\Delta_{k_x}^4 F_{k_x}[\tilde{p}(k_x)]}{F_x[p(x)]} \tag{7.35}$$

for imaginary weak values. Recall that A_w can be much larger than λ_*. It appears then, that in the postselected runs, not only is the average displacement of the meter

amplified but so too the measure of sensitivity as captured by the Fisher information. When pre- and postselection are tuned close to orthogonal this triggers destructive interference in the meter, causing cancellation where the superposed meter states overlap most. This leaves only a wavefunction where the meter states overlap least, i.e. where the meter states are most distinguishable. It is therefore not surprising that the postselected wavefunction can give better performance, because the Fisher information is a measure of distinguishability of neighbouring meter states as g is varied.

The $4\Delta_{k_x}^4 F_{k_x}[\tilde{p}(k_x)]/F_x[p(x)]$ term in Equation (7.35) must be carefully interpreted. Notice that I continue to use the standard strategy information as a benchmark, which necessarily involves detection in the x-basis. If we take $p(x)$ and $\tilde{p}(k_x)$ to be generated from Gaussian wavefunctions, we obtain by Heisenberg's relation [65]

$$\frac{4\Delta_{k_x}^4 F_{k_x}[\tilde{p}(k_x)]}{F_x[p(x)]} = \frac{4\Delta_{k_x}^2}{F_x[p(x)]} = 1. \tag{7.36}$$

Employing an imaginary weak value then offers no advantage over a real one.

Of course the benefit due to 'amplification' is not without cost: we must correct the 'amplification' factor for the reduced probability of data being selected $|\langle\psi_f\,|\psi_i\rangle\,|^2$ – a necessary drawback of WVA. Then one finds

$$\frac{|A_w|^2}{\lambda_*^2}|\langle\psi_f\,|\psi_i\rangle\,|^2 = \frac{|\langle\psi_f|\mathbf{A}\,|\psi_i\rangle\,|^2}{\lambda_*^2} \leq 1. \tag{7.37}$$

The inequality follows by expanding $|\psi_f\rangle$ and $|\psi_i\rangle$ in the eigenbasis of $\mathbf{A}$ and applying the Cauchy-Schwarz inequality:

$$\begin{aligned} |\langle\psi_f|\mathbf{A}|i\rangle|^2 = \left|\sum_k \bar{c}'_k\lambda_k c_k\right|^2 &\leq \sum_n |c'_n|^2 \sum_m |\lambda_m c_m|^2 \\ &= \sum_m \lambda_m^2 |c_m|^2 \\ &= \langle\mathbf{A}^2\rangle \leq \lambda_*^2\,. \end{aligned} \tag{7.38}$$

Thus, combining Equations. (7.34), or (7.35) and (7.36), with (7.37) we have

$$\frac{qF_g[p_{\mathrm{WVA}}]}{F_g[p_{\mathrm{std}}]} \leq 1; \tag{7.39}$$

at least under the AAV approximation and in the absence of noise, there can be no advantage through WVA for the purpose of estimating g. The above analysis holds for a system of any finite dimension.

7.3.2 Failed postselection

As above, one can consider making use of information in all branches of the postselection. To compute the total information, begin with a joint probability distribution

$$\mathcal{P}(b,x) = \sum_f \delta_{fb} q_b |m(x - \mathrm{Re}(A_w^b)g)|^2, \tag{7.40}$$

or

$$\tilde{\mathcal{P}}(b,k_x) = \sum_f \delta_{fb} q_b |\tilde{m}(k_x - 2\Delta_{k_x}\mathrm{Im}(A_w^b)g)|^2, \tag{7.41}$$

once more employing the AAV approximation: b indexes the outcomes of the final, strong measurement. In this regime, q does not depend on g, although in an real experiment the postselection probability itself carries information about the interaction parameter [216]. Because of the Kronecker δ, each output branch can be considered independently, leading to

$$\begin{aligned} F_g[\mathcal{P}(b,x)] &= \sum_f q_f |A_w^f|^2 F_x[p(x)] \\ &= \langle \mathbf{A}^2 \rangle F_x[p(x)] \end{aligned} \tag{7.42}$$

for real weak values, or indeed

$$\begin{aligned} F_g[\tilde{\mathcal{P}}(b,k_x)] &= \sum_f q_f |A_w^f|^2 (2\Delta_{k_x})^2 \\ &= \langle \mathbf{A}^2 \rangle 4\Delta_{k_x}^2 \end{aligned} \tag{7.43}$$

for imaginary weak values. The last step is in strong analogy with the usual resolution of probability-weighted weak-values to the expectation value $\sum_f q_f A_w^f = \langle \mathbf{A} \rangle$ [180]. Notice that once the initial state of system and meter are chosen, the expected value of the square of the control observable along with the shape of initial meter wavefunction set a fixed amount of information in the joint system-meter state after their interaction. The final strong measurement may distribute the information (which we should think of as a conserved quantity) among the output branches in various ways. The choice of basis for the final measurement may lead to a balanced distribution if, for example, the final measurement basis is unbiased with respect to the initial state. If on the other hand the basis contains an element which is almost orthogonal to the initial state, an arbitrarily large portion of the information may be concentrated into the corresponding branch of the final measurement, leaving a negligible amount in the failed postselection mode. This agrees qualitatively with the exact results of

the previous section, see Figures 7.3 and 7.2, and also of the previous Chapter, see Figure 6.5.

When the AAV effect gives rise to a significant 'amplification' in one branch, there is little to be gained from monitoring other branches: henceforth we shall thus only be concerned with the probability distribution p_{WVA} in a suitably chosen 'success' branch. In such a situation an important question remains: is the postselected meter wavefunction (carrying almost all of the information about g) more robust to technical noise than an un-postselected wavefunction?

7.3.3 Pixelation

We are now in a position to introduce realistic imperfections to the scenario. I consider first the effects of pixelation on an arbitrary wavefunction which has been subject to a simple shift. This is of relevance to real weak values, and also to imaginary weak values for a restricted class of momentum space wavefunctions. I then specialise to Gaussian wavefunctions, which serves both as an instructive example and also as the canonical wavefunction used in weak-value 'amplification' experiments for both real and imaginary weak values.

7.3.3.1 For any wavefunction

Imagine that the particle (after weak measurement and postselection) impinges on a detector comprising pixels of finite width. To model such a device, I build up a discrete probability mass $\Pr(n)$ by dividing the s-axis into pixels of size r_s. Each pixel carries an integer label $n = \lfloor s/r_s \rceil$ (so that values of s are rounded to the nearest integer multiple of r_s). The total Fisher information then becomes

$$F_g[\Pr(n)] = \sum_n \frac{1}{\Pr(n)} (\partial_g \Pr(n))^2 \,. \tag{7.44}$$

The width of the detector is modelled by varying the range of this sum, but here I take it to be infinite. For this reason any relabelling of the pixels will not change the result. For example, adding a fixed integer to each pixel's label is a bijection $f : \mathbb{Z} \to \mathbb{Z}$, which preserves the value of the sum in analogy with the above change of integration variable, equation (7.33). The probability of a click in pixel n is

$$\Pr(n) = \int_{r_s(n-1/2)}^{r_s(n+1/2)} p(s) ds \,. \tag{7.45}$$

Because of the invariance under a relabelling of the pixels, any shifts in s can now be taken modulo r_s.

To understand the effect of pixelation on $p(s-\nu g)$ we may use the chain rule (7.32) again, taking the derivative under the integral sign:

$$F_g[\Pr(\lfloor s'/r_s \rceil)] = \sum_n \frac{\left(\int_{r_s(n-1/2)}^{r_s(n+1/2)} \partial_g p(s')ds\right)^2}{\int_{r_s(n-1/2)}^{r_s(n+1/2)} p(s')ds} = \nu^2 \sum_n \frac{\left(\partial_{s'} \int_{r_s(n-1/2)}^{r_s(n+1/2)} p(s')ds\right)^2}{\int_{r_s(n-1/2)}^{r_s(n+1/2)} p(s')ds}\,. \tag{7.46}$$

The pixelation effect is almost decoupled from the dependence on ν. However, when making the change of integration variable,

$$\Pr(\lfloor s'/r_s \rceil) = \int_{r_s(n-1/2)}^{r_s(n+1/2)} p(s-\nu g)ds = \int_{r_s(n-1/2-\nu g)}^{r_s(n+1/2-\nu g)} p(s')ds'\,, \tag{7.47}$$

the finite limits of integration prevent removing the dependence on ν altogether. Allowing the pixelated detector to be displaced through a controllable quantity μ:

$$\Pr(\lfloor s'/r_s \rceil) = \int_{r_s(n-1/2)}^{r_s(n+1/2)} p(s+\mu-\nu g)ds = \int_{r_s(n-1/2)}^{r_s(n+1/2)} p(s-h)ds\,, \tag{7.48}$$

it becomes clear that the importance of ν is screened by the alignment $h := (\nu g - \mu) \bmod r_s$. The fraction h indicates the relative alignment of the centroid with the pixel boundaries. For example, if $h = 0$ then the mean is aligned exactly in the middle of a pixel; if $h = 0.5$ then the mean is on a pixel boundary – this is shown schematically in the inset of Figure 7.4. One then obtains

$$F_g[\Pr(\lfloor (s-\nu g)/r_s \rceil)] = \nu^2 F_s[\Pr(\lfloor (s-h)/r_s \rceil)]. \tag{7.49}$$

The relationship of ν to h is purely incidental: there is no reason to suppose that larger values of ν will make alignment easier than smaller values. One should therefore treat h identically for the WVA and standard strategies. One can consider adaptive techniques to asymptotically achieve $h = 0.5$, or – if alignment control is impossible – one can average over $h \in [0, 0.5]$. Any degradation due to pixelation commutes with the 'amplification' effect.

Correcting by the success probability and dividing by the standard strategy information, for real weak values one has, taking h identical in both strategies,

$$\frac{qF_g[\Pr(\lfloor (x - \mathrm{Re}(A_w)g)/r_x \rceil)]}{F_g[\Pr(\lfloor (x - \lambda_* g)/r_x \rceil)]} = \frac{\mathrm{Re}(\langle f|\mathbf{A}\,|\psi_i\rangle)^2}{\lambda_*^2} \leq 1\,. \tag{7.50}$$

The pixelation effect has completely cancelled, reducing the problem to that of the ideal detector above – see Equation (7.37).

For imaginary weak-values, we have

$$\frac{qF_g[\Pr(\lfloor (k_x - 4\Delta_{k_x}\mathrm{Im}(A_w)g)/r_{k_x} \rceil)]}{F_g[\Pr(\lfloor (x - \lambda_* g)/r_x \rceil)]} \leq \frac{\alpha(r_{k_x}, \Delta_{k_x})}{\alpha(r_x, \Delta_x)}\,, \tag{7.51}$$

where I define $\alpha(r_s, \Delta_s) = F_s[\Pr(\lfloor s \rceil)]/F_s(p(s))$ as the fraction of information remaining after pixelation, and have applied Equations. (7.36) and (7.37). Clearly there will not be a perfect cancellation if the pixelation in the k_x direction is more or less severe than in the x direction: but as I have shown this is independent of the magnitude of the mean. Moreover I give evidence in the next section that α is monotonically decreasing in r_s/Δ_s, and in fact equal to unity to first order. Then $\alpha(r_{k_x}, \Delta_{k_x}) : \alpha(r_x, \Delta_x) \approx 1$.

7.3.3.2 For Gaussian wavefunctions

It is instructive to fix the input wavefunction to get an idea of exactly how the pixelation affects its informational content. By letting the initial state be described by

$$m(x) = \sqrt{\frac{1}{\sqrt{2\pi}\Delta_s}}\exp\left(-\frac{s^2}{4\Delta_s^2}\right)\,, \tag{7.52}$$

i.e. a Gaussian centred on the origin with waist Δ_s, one finds

$$F_s[p(s)] = \frac{1}{\Delta_s^2}\,. \tag{7.53}$$

so that $F_g[p(s')] = \nu^2/\Delta_s^2$. Under pixelation $p(s')$ becomes a discrete probability mass. By the above arguments it is sufficient to study

$$\Pr(n) = \frac{1}{2}\left[\mathrm{erf}\left(\frac{r_s}{\Delta_s}\gamma_+\right) - \mathrm{erf}\left(\frac{r_s}{\Delta_s}\gamma_-\right)\right]\,, \tag{7.54}$$

where I have introduced convenient quantities $\gamma_\pm = (h - n \pm 0.5)/\sqrt{2}$ and applied Equation (7.45). Define α by dividing the diminished Fisher information of the pixelated probability mass by the un-pixelated information

$$\alpha(r_s, \Delta_{r_s}) = \frac{F_s[\Pr(n)]}{F_s[p(s)]} = \frac{1}{F_s[p(s)]}\sum_n f_n\,, \tag{7.55}$$

where the summand is to be understood as the pixel-wise information. This function shows how much information remains after pixelation has occurred. We may take the division inside the sum, attaching it to each term. In pixel n

$$\frac{f_n}{F_s[p(s)]} = \frac{1}{\pi}\frac{\left(e^{-R^2\gamma_+^2} - e^{-R^2\gamma_-^2}\right)^2}{\operatorname{erf}(R\gamma_+) - \operatorname{erf}(R\gamma_-)}\,, \tag{7.56}$$

where I have chosen to define $R := r_s/\Delta_s$ as an inverse measure of resolution. It represents the granularity of the pixels against the characteristic width of the Gaussian. One expects the ratio of $F_s[\Pr(n)]$ to $F_s[p(s)]$ to converge to unity as $R \to 0$.

All that remains is to sum these relative informations (7.56) over all pixels. I numerically plot the dependence of the sum on R in Figure 7.4. The limit when $h = 0.5$ and $r_s \to \infty$ is analytically known: the Fisher information is a fraction $2/\pi$ of the non-pixelated case [162]. This limit corresponds to a perfectly aligned 'split-detector' with only two pixels [190]. In the opposite limit, $\alpha \approx 1$ to first order in R, meaning that pixelation is a non-issue as long as the pixel size is at least as small as the width of the wavefunction. My results agree qualitatively with Ref. [191], which found that pixelation does not set a lower bound on the magnitude of a shift that can be estimated.

When pixelation is severe and alignment is poor one does better with a broader statistical distribution than with a narrower one. Taking the wavefunction once more to be Gaussian (7.52) in Figure 7.5 I fix the pixel width and show how the Fisher information depends on Δ_s in such a way as to offer a trade-off between uncertainty and misalignment. If perfect alignment cannot be achieved, there is a limit to how low Δ_s may be taken before the worsened effective resolution combines with the bad alignment to kill off the Fisher information. In circumstances of extreme pixelation, it can thus be advantageous to increase Δ_s – introducing extra uncertainty leads to a superior performance. One could increase Δ_s through a unitary process, to present a broader meter wavefunction for the system to interact with. Another option would be to intentionally introduce random displacements to the meter (or detection apparatus) immediately prior to detection. The latter option is an instance of a well-known image- and audio-processing technique known as dithering. The distinction is reminiscent of the discussions in Chapter 3.

7.3.4 Transverse Jitter

Here I model another prevalent source of imperfection: random lateral displacements of the particle. Again I begin with an arbitrary wavefunction subject to a simple

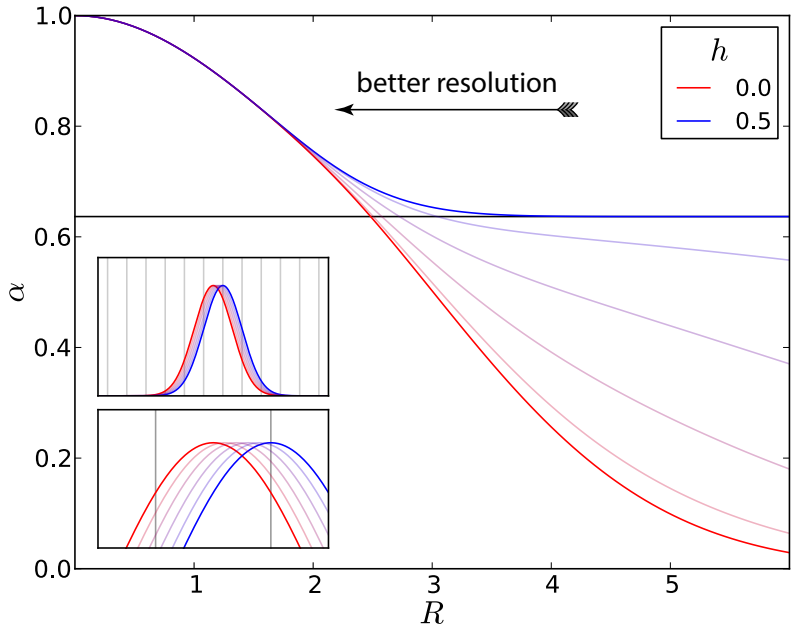


Figure 7.4: Numerically obtained relationships between relative Fisher information α and inverse resolution R. The different curves (blue to red, top to bottom and right to left in the inset) correspond to different alignments which are schematised in the insets. The worst and best cases at $h = 0$ and $h = 0.5$, respectively, are shown in bold. Aside from misalignment effects (which become important when $R \gtrsim 3$) finite resolution does not dramatically limit parameter estimation. In fact with only two pixels the penalty paid is only a loss of about a third of the information, as long as good alignment is possible. The curves are equally relevant for the standard strategy or the WVA strategy. Reproduced from *Phys. Rev. X*, 4:011032, March 2014

shift (composed of a g dependent part and also a random part), but then specialise to Gaussian wavefunctions in order to make a concrete connection with the majority of WVA experiments.

7.3.4.1 For any wavefunction

I define 'jitter' – random displacements of the measuring apparatus, or equivalently the incident beam – by convolving the probability distribution with a suitable noise kernel $\mathcal{N}_s$:

$$\mathcal{N}_s \star p(s) := \int p(s+\mu)\mathcal{N}_s(\mu)d\mu\,. \tag{7.57}$$

This is perhaps the most prominent source of technical noise that has been studied in the WVA context, until now only using the signal-to-noise ratio metric [14, 190, 98, 107]. In those cases WVA gave a relative advantage, due to identifying the 'signal'

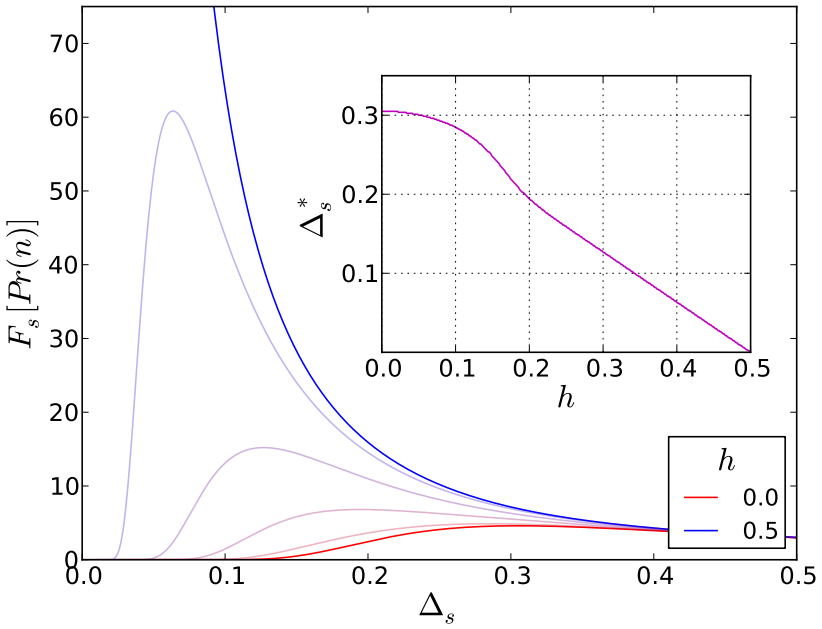


Figure 7.5: Numerically obtained relationships between Fisher information and Gaussian spread Δ_s for fixed pixel width $r_s = 1$. The bold curves are for $h = 0$ (lowermost) and $h = 0.5$ (uppermost), with fainter curves interpolating linearly. Without pixelation, narrower Gaussian distributions are superior – the delta function offers the ultimate precision. However, the misalignment effect introduced with pixelation combines with the low effective resolution as Δ_s is reduced, preventing the continued improvement. For $\Delta_s \gtrsim r_s/3$ alignment ceases to be important. At any given degree of misalignment (which is described by h), there will be an optimum choice $\Delta_s = \Delta_s^*$. The inset shows how this optimum changes with h. Reproduced from *Phys. Rev. X*, 4:011032, March 2014.

with the mean of the probability distribution, which becomes invisible when it falls through the noise floor [190]. The Fisher information metric informs us about the performance of a superior estimation strategy, which allows for (in principle) *any* data to contribute to the estimate of the unknown parameter. In my application of the metric, I assume the noise is independent of g so that, by another application of the chain rule (7.32), we have

$$\begin{aligned} F_g[\mathcal{N}_s \star p(s - \nu g)] &= \int \frac{(\partial_g [\mathcal{N}_s \star p(s')])^2}{\mathcal{N}_s \star p(s')} ds \\ &= \nu^2 \int \frac{(\partial_{s'} [\mathcal{N}_s \star p(s')])^2}{\mathcal{N}_s \star p(s')} ds \\ &= \nu^2 F_x[\mathcal{N}_s \star p(s)]\,. \end{aligned} \tag{7.58}$$

Once more the application of noise commutes with the 'amplification' factor ν^2, the second main result of this chapter. For real weak values, the noise cancels out in the ratio of Fisher informations:

$$\frac{qF_g[\mathcal{N}_x \star p(s - \mathrm{Re}(A_w)g)]}{F_g[\mathcal{N}_x \star p(s - \lambda_* g)]} = \frac{\mathrm{Re}(\langle\psi_i|\mathbf{A}\,|\psi_i\rangle)^2}{\lambda_*^2} \leq 1\,. \tag{7.59}$$

For imaginary weak values, once more there may not be a complete cancellation due to the possibility of different width wavefunctions and different noise kernels in position and momentum space. We then have

$$\frac{qF_g[\mathcal{N}_{k_x} \star p(k_x - 4\Delta_{k_x}\mathrm{Im}(A_w)g)]}{F_g[\mathcal{N}_x \star p(x - \lambda_* g)]} \leq \frac{\beta(\Delta_{k_x}, \mathcal{N}_{k_x})}{\beta(\Delta_x, \mathcal{N}_x)} \tag{7.60}$$

on defining $\beta(\Delta_s, \mathcal{N}_s)$ as the attenuation factor of Fisher information under jitter: a function which depends on the width of the wavefunction and on the noise kernel in s-space (see below). In reality $\beta(\Delta_{k_x}, \mathcal{N}_{k_x})$ may or may not be independent of $\beta(\Delta_x, \mathcal{N}_x)$, depending, e.g., on how detection is achieved in the laboratory. Exploiting the relative severity of noise in position and momentum space with imaginary weak values has been proposed by Kedem [98]; but this possibility does not derive from the 'amplification' of the mean.

While we require $\partial_g \mathcal{N}_s = \partial_s \mathcal{N}_s = 0$, otherwise the noise kernel $\mathcal{N}_s(\mu)$ can have arbitrary properties (up to normalization and positivity); in particular it may have non-zero mean.

7.3.4.2 For Gaussian wavefunctions

As an example, choose the noise kernel to be normally distributed (with zero mean and variance J_s), giving

$$\mathcal{N} \star p(s) = \int_{-\infty}^{\infty} \frac{p(s-\mu)\exp\left(-\frac{1}{2}\frac{\mu^2}{J_s^2}\right)}{\sqrt{2\pi}J_s} d\mu\,. \tag{7.61}$$

If $p(s;\Delta_s)$ is Gaussian with width Δ_s, the application of jitter is equivalent to a redefinition of the spread

$$\mathcal{N} \star p(s - \nu g; \Delta_s) = p(s - \nu g; \sqrt{\Delta_s^2 + J_s^2}) \tag{7.62}$$

so that the Fisher information takes on a Lorentzian dependence on Δ_s, with scale parameter proportional to J_s

$$F_g[\mathcal{N}_s \star p(s - \nu g; \Delta_s)] = \frac{\nu^2}{\Delta_s^2 + J_s^2}\,. \tag{7.63}$$

I then find, for this example,

$$\beta(\Delta_s, J_s) = \frac{1}{1 + \frac{J_s^2}{\Delta_s^2}}. \tag{7.64}$$

7.4 Discussion

In this chapter, I have studied the metrological advantages of weak-value 'amplification' from a parameter estimation perspective, and derived the ultimate limits on uncertainty that apply in the canonical setting involving a continuous meter. My discussion began with noise-free results holding outside the regime of validity of the AAV approximation.

I went on to analyse the situation in the presence of jitter and of finite detector resolution. Under my choice of metric, and under these prevalent examples of technical imperfection, I have found there to be no advantage whatsoever for real weak-values.

For imaginary weak-values I have shown that any advantage – if one exists at all – will be menial: due, for example, to an (often incidental) mismatch of noise in position and momentum space rather than an anomalously large mean. I therefore conclude that the 'amplification' aspect of the weak-value formalism offers no fundamental benefits over suitable standard strategies.

In Chapter 6, I have studied WVA in the context of phase estimation, and found that limitations imposed by decoherence are not mitigated by the technique. The combination of those results (concerning quantum noise affecting the wavefunction prior to detection), with the findings of this chapter (concerning noise and imperfections relating to the detection apparatus itself) restrict the class of noise where one expects superiority of WVA in metrology scenarios. The very recent work of Ferrie and Combes shows that even errors associated with a finite sample size do not favour a weak-value approach [62].

Nonetheless, when circumstances make increased use of resources preferable to complicated post-processing, WVA may remain an attractive option. My results also do not rule out that aspects other than the 'amplification' itself may still be useful in certain experimental circumstances [92]. Alongside the opportunity to induce changes in the direct meter observable (rather than its conjugate), other possibly useful applications include the reduction in signal intensity; either because this reduces the absolute systematic error (e.g. due to imperfect optics) or because this avoids detector saturation [201]. However, these effects are quite distinct from the notion

of 'amplification'. More theoretical work is required in order that their utility be assessed and quantified.

Chapter 8

Conclusion

Most of this thesis has been devoted to quantum measurement in its various guises. I began with a description of quantum theory itself, which is a radical departure from classical physics. Although there are many perspectives on the special nature of quantum theory, I concentrated on the consummate quantum phenomenon, interference, and on the insights provided by information theory and estimation theory. The contrast between classical and quantum physics is revealed by the three facets of quantum measurement. First, quantum systems are coupled together with unitary dynamics; second, the wavefunction collapses stochastically and data are produced; thirdly, the data are processed and parameters are estimated.

These tools and concepts have enabled two streams of analysis. The first concerns experimental investigations of the the foundations of quantum theory – and on the problem of measurement. The second concerns the application of generalised quantum measurements – particularly those that are weak and postselected – in the field of quantum sensors.

On the first topic, I have contributed to the wider field in the following ways. I have helped to design an experiment which tests the Leggett-Garg inequality in a solid state spin ensemble. It was the first experiment to implement the original null-result measurements prescribed by Leggett and Garg. This advantage was unlocked by the use of two complementary controlled-NOT gates; each an instance of a strong measurement. Use of such a measurement scheme allowed the experiment to steer away from both weak measurements and the stationarity assumption, which weaken the conclusions found in previous experiments. Further, I generalised the Leggett-Garg inequality to allow for corruption in the null-result measurements. This was necessary for an ensemble experiment performed at finite temperature, but the introduction of the venality ζ allows for a universal treatment of imperfection in foundational tests

of quantum theory. The theoretical extension therefore extends far beyond the Oxford Leggett-Garg experiment, and will strengthen future foundational experiments on more macroscopic objects. Such experiments, if carefully performed, have the potential to revolutionise physics once more. Should the Schrödinger equation, which underpins our understanding of matter, energy, and life itself, be replaced by a non-linear usurper?

On the second topic, my main contributions are as follows. I have been the first to apply the Fisher information metric as a measure of performance of weak-value amplification. This has brought rigour to the claims that the technique, which uses a very specialised form of weak measurement and postselection, offers no advantage in metrological settings. My work is distinguished by the inclusion of noise in the analyses. Studying both dephasing, which is a limitation typical to solid state physics, and random displacements and pixelation, which are imperfections mostly found in optical systems, I have uncovered a surprising truth. The notion of weak-value 'amplification' – the characteristically large average displacement of the quantum-coherent measuring device brought about via postselection – is not a method for suppressing noise. At best, the weak-value techniques are a new way of achieving the standard precision. At worst they are a biased and approximate technique which reduces robustness (particularly to decoherence) and increases experimental complexity. I leave open the question of whether other effects arising through the weak-values formalism, such as a lowered photon flux, are useful.

The topics are united by the word 'measurement'. It has several meanings in the context of quantum theory. Discovering how best to interpret, and how best to use quantum mechanics requires the most fastidious treatment of the word. Casual and inexact use can lead to confusion – that is the problem with 'measurement'. The *measurement problem*, on the other hand, is deeply entrenched in the axioms of quantum theory. It is doubtful that greater care with words will solve that mystery.

In the pursuit of the twin goals of a greater understanding, and a more potent application of quantum theory, I have found 'crosstalk' to be an interesting and fruitful happening. Progress in both fields is, I feel, ever greater for the interplay between the foundational and technological considerations.

Bibliography

[1] Y. Aharonov, D. Z. Albert, A. Casher, and L. Vaidman. Surprising quantum effects. *Physics Letters A*, 124(4):199 – 203, 1987.

[2] Y. Aharonov, D. Z. Albert, and L. Vaidman. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Phys. Rev. Lett.*, 60(14):1351–1354, Apr 1988.

[3] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz. Time symmetry in the quantum process of measurement. *Phys. Rev.*, 134(6B):B1410–B1416, Jun 1964.

[4] Y. Aharonov, S. Popescu, D. Rohrlich, and P. Skrzypczyk. Quantum cheshire cats. *New Journal of Physics*, 15(11):113015, 2013.

[5] Y. Aharonov, S. Popescu, and J. Tollaksen. A time symmetric formulation of quantum mechanics. 63(11):27–32, 2010.

[6] Y. Aharonov and D. Rohrlich. *Quantum paradoxes: quantum theory for the perplexed*. Wiley, Weinheim, 2005.

[7] Y. Aharonov and L. Vaidman. Aharonov and Vaidman reply. *Phys. Rev. Lett.*, 62(19):2327, May 1989.

[8] Y. Aharonov and L. Vaidman. Properties of a quantum system during the time interval between two measurements. *Phys. Rev. A*, 41:11–20, Jan 1990.

[9] B. Andreas, Y. Azuma, G. Bartl, P. Becker, H. Bettin, M. Borys, I. Busch, M. Gray, P. Fuchs, K. Fujii, H. Fujimoto, E. Kessler, M. Krumrey, U. Kuetgens, N. Kuramoto, G. Mana, P. Manson, E. Massa, S. Mizushima, A. Nicolaus,

A. Picard, A. Pramann, O. Rienitz, D. Schiel, S. Valkiers, and A. Waseda. Determination of the Avogadro constant by counting the atoms in a Si-28 crystal. *Phys. Rev. Lett.*, 106:030801, Jan 2011.

[10] M. Arndt, O. Nairz, J. Vos-Andreae, C. Keller, G. van der Zouw, and A. Zeilinger. Wave-particle duality of C_{60} molecules. *Nature*, 401(6754):680–682, 10 1999.

[11] A. Aspect, P. Grangier, and G. Roger. Experimental tests of realistic local theories via Bell's theorem. *Phys. Rev. Lett.*, 47(7):460–463, August 1981.

[12] G. Bacciagaluppi and A. Valentini. *Quantum Theory at the Crossroads: Reconsidering the 1927 Solvay Conference*. Cambridge University Press, 2009.

[13] L. E. Ballentine. Realism and quantum flux tunneling. *Phys. Rev. Lett.*, 59(14):1493–1495, Oct 1987.

[14] S. Barnett, C. Fabre, and A. Maitre. Ultimate quantum limits for resolution of beam displacements. *The European Physical Journal D - Atomic, Molecular, Optical and Plasma Physics*, 22(3):513–519, 2003.

[15] A. Bassi and G. Ghirardi. Dynamical reduction models. *Physics Reports*, 379(5–6):257 – 426, 2003.

[16] P. Becker, H.-J. Pohl, H. Riemann, and N. Abrosimov. Enrichment of silicon for a better kilogram. *physica status solidi (a)*, 207(1):49–66, 2010.

[17] J. Bell. On the problem of hidden variables in quantum mechanics. *Reviews of Modern Physics*, 38(3):447–452, 1966.

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

[19] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics*, chapter "Against 'measurement' ". Cambridge University Press, 1987.

[20] F. Benatti, G. Ghirardi, and R. Grassi. On some recent proposals for testing macrorealism versus quantum mechanics. *Foundations of Physics Letters*, 7(2):105–126, 1994.

[21] D. Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. I. *Phys. Rev.*, 85:166–179, Jan 1952.

[22] D. Bohm. A suggested interpretation of the quantum theory in terms of "hidden" variables. II. *Phys. Rev.*, 85:180–193, Jan 1952.

[23] G. Boole. On the theory of probabilities. *Philosophical Transactions of the Royal Society of London*, 152:225–252, 1862.

[24] S. L. Braunstein and C. M. Caves. Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.*, 72:3439–3443, May 1994.

[25] S. L. Braunstein, C. M. Caves, R. Jozsa, N. Linden, S. Popescu, and R. Schack. Separability of very noisy mixed states and implications for NMR quantum computing. *Phys. Rev. Lett.*, 83:1054–1057, Aug 1999.

[26] H. Breuer and F. Petruccione. *The Theory of Open Quantum Systems*. OUP Oxford, 2007.

[27] G. A. D. Briggs. Experimental implementations of quantum paradoxes. In D. C. Struppa and J. M. Tollaksen, editors, *Quantum Theory: A Two-Time Success Story*, pages 367–376. Springer Milan, 2014.

[28] D. Brivio, S. Cialdi, S. Vezzoli, B. T. Gebrehiwot, M. G. Genoni, S. Olivares, and M. G. A. Paris. Experimental estimation of one-parameter qubit gates in the presence of phase diffusion. *Phys. Rev. A*, 81:012305, Jan 2010.

[29] C. Brukner, S. Taylor, S. Cheung, and V. Vedral. Quantum entanglement in time. 2004, quant-ph/0402127v1.

[30] N. Brunner and C. Simon. Measuring small longitudinal phase shifts: Weak measurements or standard interferometry? *Phys. Rev. Lett.*, 105:010405, Jul 2010.

[31] I. Buluta, S. Ashhab, and F. Nori. Natural and artificial atoms for quantum computation. *Reports on Progress in Physics*, 74(10):104401, 2011.

[32] G. Burkard. Non-markovian qubit dynamics in the presence of $1/f$ noise. *Phys. Rev. B*, 79:125317, Mar 2009.

[33] P. Busch. Surprising features of unsharp quantum measurements. *Physics Letters A*, 130(6Äì7):323 – 329, 1988.

[34] A. Cabello. Experimentally testable state-independent quantum contextuality. *Phys. Rev. Lett.*, 101:210401, Nov 2008.

[35] T. Calarco, M. Cini, and R. Onofrio. Are violations to temporal Bell inequalities there when somebody looks? *EPL (Europhysics Letters)*, 47(4):407, 1999.

[36] P. T. Callaghan. *Principles of Nuclear Magnetic Resonance Microscopy*. Oxford science publications. Clarendon Press, 1993.

[37] Z. Chen and A. Montina. Measurement contextuality is implied by macroscopic realism. *Phys. Rev. A*, 83(4):042110, Apr 2011.

[38] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

[39] M. Cohen. The Fisher information and convexity. *IEEE Transactions on Information Theory*, 14(4):591–592, 1968.

[40] J. Combes, C. Ferrie, Z. Jiang, and C. M. Caves. Probabilistic quantum metrology? Probably not. Sep 2013, 1309.6620v1.

[41] D. G. Cory, A. F. Fahmy, and T. F. Havel. Ensemble quantum computing by NMR spectroscopy. *Proceedings of the National Academy of Sciences*, 94(5):1634–1639, 1997.

[42] H. Cramer. *Mathematical methods of statistics*. Princeton University Press, Princeton, 1946.

[43] S. Das, S. Aravinda, R. Srikanth, and D. Home. Unification of Bell, Leggett-Garg and Kochen-Specker inequalities: Hybrid spatio-temporal inequalities. *EPL (Europhysics Letters)*, 104(6):60006, 2013.

[44] A. Datta, A. Shaji, and C. M. Caves. Quantum discord and the power of one qubit. *Phys. Rev. Lett.*, 100:050502, Feb 2008.

[45] A. C. de la Torre. The ubiquitous XP commutator. *European Journal of Physics*, 27(2):225, 2006.

[46] D. Deutsch. Quantum theory, the Church-Turing principle and the universal quantum computer. *Royal Society of London Proceedings Series A*, 400:97–117, July 1985.

[47] D. Deutsch, A. Barenco, and A. Ekert. Universality in quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 449(1937):669–677, 1995.

[48] A. Di Lorenzo. Full counting statistics of weak-value measurement. *Phys. Rev. A*, 85:032106, Mar 2012.

[49] A. Di Lorenzo. Comment on "Optimal probe wave function of weak-value amplification". *Phys. Rev. A*, 87:046101, Apr 2013.

[50] L. DiCarlo, M. D. Reed, L. Sun, B. R. Johnson, J. M. Chow, J. M. Gambetta, L. Frunzio, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf. Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature*, 467(7315):574–578, 09 2010.

[51] P. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell. Ultrasensitive beam deflection measurement via interferometric weak value amplification. *Phys. Rev. Lett.*, 102:173601, Apr 2009.

[52] J. Dressel, C. J. Broadbent, J. C. Howell, and A. N. Jordan. Experimental violation of two-party Leggett-Garg inequalities with semiweak measurements. *Phys. Rev. Lett.*, 106(4):040402, Jan 2011.

[53] J. Dressel and A. N. Korotkov. Avoiding loopholes with hybrid Bell-Leggett-Garg inequalities. Nov 2013, 1310.6947v1.

[54] J. Dressel, K. Lyons, A. N. Jordan, T. M. Graham, and P. G. Kwiat. Strengthening weak value amplification with recycled photons. May 2013, 1305.4520v1.

[55] J. Dressel, M. Malik, F. M. Miatto, A. N. Jordan, and R. W. Boyd. Understanding quantum weak values: Basics and applications. May 2013, 1305.7154v1.

[56] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan. The sense in which a "weak measurement" of a spin-1/2 particle's spin component yields a value 100. *Phys. Rev. D*, 40(6):2112–2117, Sep 1989.

[57] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 47(10):777–780, May 1935.

[58] A. Elby and S. Foster. Why SQUID experiments can rule out non-invasive measurability. *Physics Letters A*, 166(1):17 – 23, 1992.

[59] C. Emary. Private communication, 2013.

[60] C. Emary, N. Lambert, and F. Nori. Leggett-Garg inequalities. *Reports on Progress in Physics*, 77(1):016001, 2014.

[61] A. Feizpour, X. Xing, and A. M. Steinberg. Amplifying single-photon nonlinearity using weak measurements. *Phys. Rev. Lett.*, 107:133603, Sep 2011.

[62] C. Ferrie and J. Combes. Weak value amplification is suboptimal for estimation and detection. *Phys. Rev. Lett.*, 112:040406, Jan 2014.

[63] R. A. Fisher. Theory of statistical estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22(05):700–725, 1925.

[64] S. Foster and A. Elby. A SQUID no-go theorem without macrorealism: What SQUIDs really tell us about nature. *Foundations of Physics*, 21:773–785, 1991.

[65] B. Frieden. Fisher information and uncertainty complementarity. *Physics Letters A*, 169(3):123 – 130, 1992.

[66] B. Frieden. *Science from Fisher Information: A Unification.* Cambridge University Press, 2004.

[67] R. E. George, L. M. Robledo, O. J. E. Maroney, M. S. Blok, H. Bernien, M. L. Markham, D. J. Twitchen, J. J. L. Morton, G. A. D. Briggs, and R. Hanson. Opening up three quantum boxes causes classically undetectable wavefunction collapse. *Proceedings of the National Academy of Sciences*, 2013.

[68] S. Gerlich, S. Eibenberger, M. Tomandl, S. Nimmrichter, K. Hornberger, P. J. Fagan, J. TÃxen, M. Mayor, and M. Arndt. Quantum interference of large organic molecules. *Nat Commun*, 2:263, 04 2011.

[69] T. Geszti. Postselected weak measurement beyond the weak value. *Phys. Rev. A*, 81:044102, Apr 2010.

[70] G. C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Phys. Rev. D*, 34(2):470–491, Jul 1986.

[71] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum-enhanced measurements: Beating the standard quantum limit. *Science*, 306(5700):1330–1336, 2004.

[72] V. Giovannetti, S. Lloyd, and L. Maccone. Quantum metrology. *Phys. Rev. Lett.*, 96:010401, Jan 2006.

[73] M. E. Goggin, M. P. Almeida, M. Barbieri, B. P. Lanyon, J. L. O'Brien, A. G. White, and G. J. Pryde. Violation of the Leggett-Garg inequality with weak measurements of photons. *Proceedings of the National Academy of Sciences*, 108(4):1256–1261, 2011.

[74] Y. Gorodetski, K. Y. Bliokh, B. Stein, C. Genet, N. Shitrit, V. Kleiner, E. Hasman, and T. W. Ebbesen. Weak measurements of light chirality with a plasmonic slit. *Phys. Rev. Lett.*, 109:013901, Jul 2012.

[75] C. Gourieroux and A. Monfort. *Statistics and Econometric Models*, volume 1. Cambridge University Press, 1995.

[76] L. K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Phys. Rev. Lett.*, 79(2):325–328, Jul 1997.

[77] L. Hardy, D. Home, E. J. Squires, and M. A. B. Whitaker. Realism and the quantum-mechanical two-state oscillator. *Phys. Rev. A*, 45:4267–4270, Apr 1992.

[78] N. Harrigan and R. Spekkens. Einstein, incompleteness, and the epistemic view of quantum states. *Foundations of Physics*, 40:125–157, 2010.

[79] K. Hess, K. Michielsen, and H. D. Raedt. Possible experience: From Boole to Bell. *Europhysics Letters*, 87(6):60007, 2009.

[80] K. Hess, K. Michielsen, and H. D. Raedt. Reply to the comment by A. J. Leggett and Anupam Garg. *Europhysics Letters*, 91(4):40002, 2010.

[81] H. F. Hofmann. Weak valued statistics as fundamental explanation of quantum physics. Apr 2013, 1305.0053v1.

[82] H. F. Hofmann, M. E. Goggin, M. P. Almeida, and M. Barbieri. Estimation of a quantum interaction parameter using weak measurements: theory and experiment. *Phys. Rev. A*, 86:040102, Oct 2012.

[83] O. Hosten and P. Kwiat. Observation of the spin Hall effect of light via weak measurements. *Science*, 319(5864):787–790, 2008.

[84] S. F. Huelga, C. Macchiavello, T. Pellizzari, A. K. Ekert, M. B. Plenio, and J. I. Cirac. Improvement of frequency standards with quantum entanglement. *Phys. Rev. Lett.*, 79:3865–3868, Nov 1997.

[85] S. F. Huelga, T. W. Marshall, and E. Santos. Proposed test for realist theories using Rydberg atoms coupled to a high-Q resonator. *Phys. Rev. A*, 52(4):R2497–R2500, Oct 1995.

[86] S. F. Huelga, T. W. Marshall, and E. Santos. Temporal Bell-type inequalities for two-level Rydberg atoms coupled to a high-Q resonator. *Phys. Rev. A*, 54(3):1798–1807, Sep 1996.

[87] G. Jaeger, C. Viger, and S. Sarkar. Bell-type equalities for SQUIDs on the assumptions of macroscopic realism and non-invasive measurability. *Physics Letters A*, 210(1-2):5 – 10, 1996.

[88] J. A. Jones. Quantum computing with NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 59(2):91 – 120, 2011.

[89] J. A. Jones, S. D. Karlen, J. Fitzsimons, A. Ardavan, S. C. Benjamin, G. A. D. Briggs, and J. J. L. Morton. Magnetic field sensing beyond the standard quantum limit using 10-spin N00N states. *Science*, 324(5931):1166–1168, 2009.

[90] J. A. Jones and M. Mosca. Implementation of a quantum algorithm on a nuclear magnetic resonance quantum computer. *The Journal of Chemical Physics*, 109(5):1648–1653, 1998.

[91] A. N. Jordan, A. N. Korotkov, and M. Buttiker. Leggett-Garg inequality with a kicked quantum pump. *Phys. Rev. Lett.*, 97:026805, 2006, cond-mat/0510782.

[92] A. N. Jordan, J. Martinez-Rincon, and J. C. Howell. Technical advantages for weak-value amplification: When less is more. *Phys. Rev. X*, 4:011031, Mar 2014.

[93] J. Joyce. Bayes' theorem. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2008 edition, 2008.

[94] R. Jozsa. Complex weak values in quantum measurement. *Phys. Rev. A*, 76(4):044103, Oct 2007.

[95] B. E. Kane. A silicon-based nuclear spin quantum computer. *Nature*, 393(6681):133–137, 05 1998.

[96] R. E. Kastner. The nature of the controversy over timesymmetric quantum counterfactuals. *Philosophy of Science*, 70(1):pp. 145–163, 2003.

[97] S. Kay. *Fundamentals of Statistical Signal Processing: estimation theory. Volume 1.* PTR Prentice-Hall, 1993.

[98] Y. Kedem. Using technical noise to increase the signal-to-noise ratio of measurements via imaginary weak values. *Phys. Rev. A*, 85:060102, Jun 2012.

[99] D. Kleckner, I. Pikovski, E. Jeffrey, L. Ament, E. Eliel, J. van den Brink, and D. Bouwmeester. Creating and verifying a quantum superposition in a micro-optomechanical system. *New Journal of Physics*, 10(9):095020, 2008.

[100] G. C. Knee, G. A. D. Briggs, S. C. Benjamin, and E. M. Gauger. Quantum sensors based on weak-value amplification cannot overcome decoherence. *Phys. Rev. A*, 87:012115, January 2013.

[101] G. C. Knee and E. M. Gauger. When amplification with weak values fails to suppress technical noise. *Phys. Rev. X*, 4:011032, March 2014.

[102] G. C. Knee, E. M. Gauger, G. A. D. Briggs, and S. C. Benjamin. Comment on 'A scattering quantum circuit for measuring Bell's time inequality: a nuclear magnetic resonance demonstration using maximally mixed states'. *New Journal of Physics*, 14(5):058001, May 2012.

[103] G. C. Knee, S. Simmons, E. M. Gauger, J. J. L. Morton, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, K. M. Itoh, M. L. W. Thewalt, G. A. D. Briggs, and S. C. Benjamin. Violation of a Leggett-Garg inequality with ideal non-invasive measurements. *Nature Communications*, 3:606, January 2012.

[104] E. Knill, R. Laflamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409(6816):46–52, 01 2001.

[105] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf. Charge-insensitive qubit design derived from the cooper pair box. *Phys. Rev. A*, 76:042319, Oct 2007.

[106] S. Kochen and E. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17(1):59–87, 1967.

[107] A. G. Kofman, S. Ashhab, and F. Nori. Nonperturbative theory of weak pre- and post-selected measurements. *Physics Reports*, 520(2):43 – 133, 2012.

[108] P. Kok and B. Lovett. *Introduction to Optical Quantum Information Processing*. Cambridge University Press, 2010.

[109] J. I. Korsbakken, F. K. Wilhelm, and K. B. Whaley. The size of macroscopic superposition states in flux qubits. *EPL (Europhysics Letters)*, 89(3):30003, 2010.

[110] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 03 1951.

[111] I. Kuprov. Lecture notes on spin dynamics. 2011.

[112] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O'Brien. Quantum computers. *Nature*, 464(7285):45–53, 03 2010.

[113] A. J. Leggett. Macroscopic quantum systems and the quantum theory of measurement. *Progress of Theoretical Physics Supplement*, 69:80–100, 1980.

[114] A. J. Leggett. Experimental approaches to the quantum measurement paradox. *Foundations of Physics*, 18(9):939–952, 1988.

[115] A. J. Leggett. Comment on "How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100". *Phys. Rev. Lett.*, 62(19):2325, May 1989.

[116] A. J. Leggett. Some thought-experiments involving macrosystems as illustrations of various interpretations of quantum mechanics. *Foundations of Physics*, 29:445–456, 1999.

[117] A. J. Leggett. Testing the limits of quantum mechanics: motivation, state of play, prospects. *Journal of Physics: Condensed Matter*, 14:R415–R451, 2002.

[118] A. J. Leggett. The quantum measurement problem. *Science*, 307(5711):871, 2005.

[119] A. J. Leggett. Realism and the physical world. *Rep. Prog. Phys.*, 71:022001, 2008.

[120] A. J. Leggett and A. Garg. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Phys. Rev. Lett.*, 54(9):857–860, March 1985.

[121] A. J. Leggett and A. Garg. Comment on "Realism and quantum flux tunneling". *Phys. Rev. Lett.*, 59(14):1621, Oct 1987.

[122] A. J. Leggett and A. Garg. Comment on "Possible experience: from Boole to Bell" by Hess K. et al. *Europhysics Letters*, 91(4):40001, 2010.

[123] P. G. Lewis, D. Jennings, J. Barrett, and T. Rudolph. The quantum state can be interpreted statistically. 01 2012, 1201.6554v1.

[124] S. Lloyd. Almost any quantum logic gate is universal. *Phys. Rev. Lett.*, 75:346–349, Jul 1995.

[125] S. Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

[126] A. D. Lorenzo. Weak values maximizing the output of weak measurements. Jul 2013, 1307.4524v1.

[127] D. Loss and D. P. DiVincenzo. Quantum computation with quantum dots. *Phys. Rev. A*, 57:120–126, Jan 1998.

[128] J. S. Lundeen and A. M. Steinberg. Experimental joint weak measurement on a photon pair as a probe of Hardy's paradox. *Phys. Rev. Lett.*, 102:020404, Jan 2009.

[129] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber. Direct measurement of the quantum wavefunction. *Nature*, 474(7350):188–191, 06 2011.

[130] M. Malik, M. Mirhosseini, M. P. J. Lavery, J. Leach, M. J. Padgett, and R. W. Boyd. Direct measurement of a 27-dimensional orbital-angular-momentum state vector. *Nat Commun*, 5, 01 2014.

[131] S. Marcovitch and B. Reznik. Structural unification of space and time correlations in quantum theory. Mar 2011, 1103.2557v1.

[132] W. Marshall, C. Simon, R. Penrose, and D. Bouwmeester. Towards quantum superpositions of a mirror. *Phys. Rev. Lett.*, 91:130401, Sep 2003.

[133] P. A. Mello and L. M. Johansen. Measurements in quantum mechanics and von Neumann's model. *AIP Conference Proceedings*, 1319(1):3–13, 2010.

[134] D. Mermin. Consistent treatments of quantum mechanics. *Physics Today*, 64(8), 2011.

[135] B. Misra and E. C. G. Sudarshan. The Zeno's paradox in quantum theory. 18(4):756–763, 1977.

[136] K. Modi, H. Cable, M. Williamson, and V. Vedral. Quantum correlations in mixed-state metrology. *Phys. Rev. X*, 1:021022, Dec 2011.

[137] K. Molmer. Counterfactual statements and weak measurements: an experimental proposal. *Physics Letters A*, 292(3):151 – 155, 2001.

[138] A. Montina. Dynamics of a qubit as a classical stochastic process with time-correlated noise: Minimal measurement invasiveness. *Phys. Rev. Lett.*, 108:160501, Apr 2012.

[139] J. J. L. Morton. *Electron Spins in Fullerenes as Prospective Qubits*. PhD thesis, University of Oxford, 2005.

[140] O. Moussa, C. Ryan, D. Cory, and R. Laflamme. Testing contextuality on quantum ensembles with one clean qubit. *Phys. Rev. Lett.*, 104(16):160501, 2010.

[141] J. T. Muhonen, J. P. Dehollain, A. Laucht, F. E. Hudson, T. Sekiguchi, K. M. Itoh, D. N. Jamieson, J. C. McCallum, A. S. Dzurak, and A. Morello. Storing quantum information for 30 seconds in a nanoelectronic device. Feb 2014, 1402.7140v1.

[142] K. Nakamura, A. Nishizawa, and M.-K. Fujimoto. Evaluation of weak measurements to all orders. *Phys. Rev. A*, 85:012113, Jan 2012.

[143] S. Nam, B. Calkins, T. Gerritts, S. Harrington, A. Lita, F. Marsili, V. Verma, I. Vayshenker, R. Mirin, M. Shaw, W. Farr, and J. Stern. Superconducting nanowire avalanche photodetectors. In *Photonics Conference (IPC), 2013 IEEE*, pages 366–367, Sept 2013.

[144] J. V. Neumann. *Mathematical Foundations of Quantum Mechanics*. Investigations in physics. Princeton University Press, 1996.

[145] N. H. Nickerson, Y. Li, and S. C. Benjamin. Topological quantum computing with a very noisy network and local error rates approaching one percent. *Nat Commun*, 4:1756, 04 2013.

[146] M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information (Cambridge Series on Information and the Natural Sciences)*. Cambridge University Press, 1st edition, 2004.

[147] S. Nimmrichter and K. Hornberger. Macroscopicity of mechanical quantum superposition states. *Phys. Rev. Lett.*, 110:160403, Apr 2013.

[148] H. Ollivier and W. H. Zurek. Quantum discord: A measure of the quantumness of correlations. *Phys. Rev. Lett.*, 88:017901, Dec 2001.

[149] H. Paik, D. I. Schuster, L. S. Bishop, G. Kirchmair, G. Catelani, A. P. Sears, B. R. Johnson, M. J. Reagor, L. Frunzio, L. I. Glazman, S. M. Girvin, M. H. Devoret, and R. J. Schoelkopf. Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.*, 107:240501, Dec 2011.

[150] A. Palacios-Laloy, F. Mallet, F. Nguyen, P. Bertet, D. Vion, D. Esteve, and A. N. Korotov. Experimental violation of a Bell's inequality in time with weak measurement. *Nature Physics*, 6:442–447, June 2010.

[151] A. K. Pan and A. Matzkin. Weak values in nonideal spin measurements: An exact treatment beyond the asymptotic regime. *Phys. Rev. A*, 85:022122, Feb 2012.

[152] M. Paris and J. Rehacek. *Quantum State Estimation*. Lecture Notes in Physics. Springer, 2004.

[153] M. G. A. Paris. Quantum estimation for quantum technology. *Int. J. Quant. Inf.*, 07:125, 2009.

[154] J. P. Paz and G. Mahler. Proposed test for temporal Bell inequalities. *Phys. Rev. Lett.*, 71(20):3235–3239, Nov 1993.

[155] P. M. Pearle. Hidden-variable example based upon data rejection. *Phys. Rev. D*, 2:1418–1425, Oct 1970.

[156] R. Penrose. On gravity's role in quantum state reduction. *General Relativity and Gravitation*, 28:581–600, 1996.

[157] R. Penrose. Quantum computation, entanglement and state reduction. *Philosophical Transactions of the Royal Society of London A*, 356, 1998.

[158] A. Peres. Quantum limitations on measurement of magnetic flux. *Phys. Rev. Lett.*, 61(18):2019–2021, October 1988.

[159] A. Peres. Quantum measurements with postselection. *Phys. Rev. Lett.*, 62(19):2326, May 1989.

[160] L. Pezzé and A. Smerzi. Entanglement, nonlinear dynamics, and the Heisenberg limit. *Phys. Rev. Lett.*, 102:100401, Mar 2009.

[161] J. J. Pla, K. Y. Tan, J. P. Dehollain, W. H. Lim, J. J. L. Morton, F. A. Zwanenburg, D. N. Jamieson, A. S. Dzurak, and A. Morello. High-fidelity readout and control of a nuclear spin qubit in silicon. *Nature*, 496(7445):334–338, 04 2013.

[162] K. Pötzelberger and K. Felsenstein. On the Fisher information of discretized data. *Journal of Statistical Computation and Simulation*, 46(3-4):125–144, 1993.

[163] G. J. Pryde, J. L. O'Brien, A. G. White, S. D. Bartlett, and T. C. Ralph. Measuring a photonic qubit without destroying it. *Phys. Rev. Lett.*, 92(19):190402, May 2004.

[164] G. J. Pryde, J. L. O'Brien, A. G. White, T. C. Ralph, and H. M. Wiseman. Measurement of quantum weak values of photon polarization. *Phys. Rev. Lett.*, 94(22):220405, Jun 2005.

[165] M. F. Pusey, J. Barrett, and T. Rudolph. On the reality of the quantum state. *Nat Phys*, 8(6):475–478, 06 2012.

[166] N. W. M. Ritchie, J. G. Story, and R. G. Hulet. Realization of a measurement of a "weak value". *Phys. Rev. Lett.*, 66:1107–1110, Mar 1991.

[167] O. Romero-Isart. Quantum superposition of massive objects and collapse models. *Phys. Rev. A*, 84:052121, Nov 2011.

[168] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland. Experimental violation of a Bell's inequality with efficient detection. *Nature*, 409(6822):791–794, 02 2001.

[169] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg. Violation of Heisenberg's measurement-disturbance relationship by weak measurements. *Phys. Rev. Lett.*, 109:100404, Sep 2012.

[170] R. Ruskov, A. N. Korotkov, and A. Mizel. Signatures of quantum behaviour in single-qubit weak measurements. *Phys. Rev. Lett.*, 96:200404, 2006.

[171] K. Saeedi, S. Simmons, J. Z. Salvail, P. Dluhy, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Pohl, J. J. L. Morton, and M. L. W. Thewalt. Room-temperature quantum bit storage exceeding 39 minutes using ionized donors in silicon-28. *Science*, 342(6160):830–833, 2013.

[172] S. Saito, X. Zhu, R. Amsüss, Y. Matsuzaki, K. Kakuyanagi, T. Shimo-Oka, N. Mizuochi, K. Nemoto, W. J. Munro, and K. Semba. Towards realizing a quantum memory for a superconducting qubit: Storage and retrieval of quantum states. *Phys. Rev. Lett.*, 111:107008, Sep 2013.

[173] M. Schaffry, E. M. Gauger, J. J. L. Morton, and S. C. Benjamin. Proposed spin amplification for magnetic sensors employing crystal defects. *Phys. Rev. Lett.*, 107:207210, Nov 2011.

[174] M. Schaffry, E. M. Gauger, J. J. L. Morton, J. Fitzsimons, S. C. Benjamin, and B. W. Lovett. Quantum metrology with molecular ensembles. *Phys. Rev. A*, 82:042114, Oct 2010.

[175] M. Schirber. Nobel prize—tools for quantum tinkering. *Physics*, 5:114, Oct 2012.

[176] M. Schlosshauer, J. Kofler, and A. Zeilinger. A snapshot of foundational attitudes toward quantum mechanics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3):222 – 230, 2013.

[177] E. Schrödinger. Die gegenwärtige situation in der quantenmechanik. *Naturwissenschaften*, 23:823–828, 1935.

[178] A. Shaji and C. M. Caves. Qubit metrology and decoherence. *Phys. Rev. A*, 76:032111, Sep 2007.

[179] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(4):623–656, 1948.

[180] Y. Shikano and A. Hosoya. Weak values with decoherence. *Journal of Physics A: Mathematical and Theoretical*, 43(2):025304, 2010.

[181] Y. Shikano and S. Tanaka. Estimation of spin-spin interaction by weak measurement scheme. *EPL (Europhysics Letters)*, 96(4):40002, 2011.

[182] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[183] L. Shun-Long. Fisher information of wavefunctions: Classical and quantum. *Chinese Physics Letters*, 23(12):3127, 2006.

[184] S. Simmons, R. M. Brown, H. Riemann, N. V. Abrosimov, P. Becker, H.-J. Phol, M. L. W. Thewalt, K. M. Itoh, and J. J. L. Morton. Entanglement in a solid state spin ensemble. *Nature*, 470(7332):69–72, 2010.

[185] A. M. Souza, J. Li, D. O. Soares-Pinto, R. S. Sarthour, S. Oliveira, S. F. Huelga, M. Paternostro, and F. L. Semião. Experimental demonstration of non-markovian dynamics via a temporal bell-like inequality. Aug 2013, 1308.5761v1.

[186] A. M. Souza, A. Magalhães, J. Teles, E. R. deAzevedo, T. J. Bonagamba, I. S. Oliveira, and R. S. Sarthour. NMR analog of Bell's inequalities violation test. *New Journal of Physics*, 10(3):033020, 2008.

[187] A. M. Souza, I. S. Oliveira, and R. S. Sarthour. A scattering quantum circuit for measuring Bell's time inequality: a nuclear magnetic resonance demonstration using maximally mixed states. *New Journal of Physics*, 13(5):053023, 2011.

[188] A. M. Souza, I. S. Oliveira, and R. S. Sarthour. Reply to Comment on 'A scattering quantum circuit for measuring Bell's time inequality: a nuclear magnetic resonance demonstration using maximally mixed states'. *New Journal of Physics*, 14(5):058002, 2012.

[189] R. W. Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Phys. Rev. A*, 75:032110, Mar 2007.

[190] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell. Optimizing the signal-to-noise ratio of a beam-deflection measurement with interferometric weak values. *Phys. Rev. A*, 80:041803, Oct 2009.

[191] G. Strübi and C. Bruder. Measuring ultrasmall time delays of light by joint weak measurements. *Phys. Rev. Lett.*, 110:083605, Feb 2013.

[192] D. Struppa and J. Tollaksen. *Quantum Theory: A Two-Time Success Story: Yakir Aharonov Festschrift*. Springer, 2013.

[193] Y. Susa, Y. Shikano, and A. Hosoya. Optimal probe wave function of weak-value amplification. *Phys. Rev. A*, 85:052110, May 2012.

[194] Y. Susa, Y. Shikano, and A. Hosoya. Reply to "Comment on 'Optimal probe wave function of weak-value amplification' ". *Phys. Rev. A*, 87:046102, Apr 2013.

[195] S. Takagi. *Macroscopic Quantum Tunneling*. Cambridge University Press, 2002.

[196] S. Tanaka and N. Yamamoto. Information amplification via postselection: A parameter-estimation perspective. *Phys. Rev. A*, 88:042116, Oct 2013.

[197] B. Teklu, M. G. Genoni, S. Olivares, and M. G. A. Paris. Phase estimation in the presence of phase diffusion: the qubit case. *Physi. Scr.*, 2010(T140):014062, 2010.

[198] C. Tesche. Can a noninvasive measurement of magnetic flux be performed with superconducting circuits? *Phys. Rev. Lett.*, 64(20):2358–2361, 1990.

[199] C. Timpson and O. Maroney. Quantum- vs. macro- realism: what does the Leggett-Garg inequality actually test? *The British Journal for the Philosophy of Science*, 2013.

[200] J. D. Trimmer. The present situation in quantum mechanics: A translation of Schrödinger's "cat paradox" paper. *Proceedings of the American Philosophical Society*, 124(5):323–338, 1980. English translation of [177].

[201] L. Vaidman. Comment on "weak value amplification is suboptimal for estimation and detection". Feb 2014, 1402.0199v1.

[202] L. Vaidman. Many-worlds interpretation of quantum mechanics. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Spring 2014 edition, 2014.

[203] W. van Orman Quine. *Two dogmas of empiricism*. Springer, 1976.

[204] G. I. Viza, J. Martínez-Rincón, G. A. Howland, H. Frostig, I. Shomroni, B. Dayan, and J. C. Howell. Weak-values technique for velocity measurements. *Opt. Lett.*, 38(16):2949–2952, Aug 2013.

[205] D. Wallace. The quantum measurement problem: State of play. December 2007, 0712.0149v1.

[206] J. H. Wesenberg, A. Ardavan, G. A. D. Briggs, J. J. L. Morton, R. J. Schoelkopf, D. I. Schuster, and K. Mølmer. Quantum computing with an electron spin ensemble. *Phys. Rev. Lett.*, 103:070502, Aug 2009.

[207] M. Wilde and A. Mizel. Addressing the Clumsiness Loophole in a Leggett-Garg Test of Macrorealism. *Foundations of Physics*, 42:256–265, 2011.

[208] N. S. Williams and A. N. Jordan. Weak values and the Leggett-Garg inequality in solid-state qubits. *Phys. Rev. Lett.*, 100(2):026804, Jan 2008.

[209] N. S. Williams and A. N. Jordan. Erratum: Weak values and the Leggett-Garg inequality in solid-state qubits. *Phys. Rev. Lett.*, 103:089902, Aug 2009.

[210] W. K. Wootters. Statistical distance and Hilbert space. *Phys. Rev. D*, 23:357–362, Jan 1981.

[211] S. Wu and Y. Li. Weak measurements beyond the Aharonov-Albert-Vaidman formalism. *Phys. Rev. A*, 83(5):052106, May 2011.

[212] S. Wu and K. Molmer. Weak measurements with a qubit meter. *Physics Letters A*, 374(1):34 – 39, 2009.

[213] X.-Y. Xu, Y. Kedem, K. Sun, L. Vaidman, C.-F. Li, and G.-C. Guo. Phase estimation with weak measurement using a white light source. *Phys. Rev. Lett.*, 111:033604, Jul 2013.

[214] S. Yokoyama, R. Ukai, S. C. Armstrong, C. Sornphiphatphong, T. Kaji, S. Suzuki, J.-i. Yoshikawa, H. Yonezawa, N. C. Menicucci, and A. Furusawa. Ultra-large-scale continuous-variable cluster states multiplexed in the time domain. *Nat Photon*, 7(12):982–986, 12 2013.

[215] J. Q. You and F. Nori. Atomic physics and quantum optics using superconducting circuits. *Nature*, 474(7353):589–597, 06 2011.

[216] L. Zhang, A. Datta, and I. A. Walmsley. Precision metrology using weak measurements. Nov 2013, 1310.5302v1.

[217] Z.-Q. Zhou, S. F. Huelga, C.-F. Li, and G.-C. Guo. Experimental violation of a Leggett-Garg inequality with macroscopic crystals. 09 2012, 1209.2176.

[218] X. Zhu, Y. Zhang, S. Pang, C. Qiao, Q. Liu, and S. Wu. Quantum measurements with preselection and postselection. *Phys. Rev. A*, 84:052111, Nov 2011.

[219] O. Zilberberg, A. Romito, and Y. Gefen. Charge sensing amplification via weak values measurement. *Phys. Rev. Lett.*, 106:080405, Feb 2011.