

Indirect Observable Measurement: an Algebraic Approach

by

Justin Dressel

Submitted in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Supervised by

Professor Andrew N. Jordan

Department of Physics and Astronomy
Arts, Sciences and Engineering
School of Arts and Sciences

University of Rochester
Rochester, New York

2013

For Samantha.

Biographical Sketch

The author was born in Albuquerque, New Mexico in March, 1983. He began his studies in mathematics and physics at the New Mexico Institute of Mining and Technology (NMT) in 2001, received the Abraham and Esther Brooke Award in 2004, and completed both Bachelor of Science degrees with highest honors in 2005. He then worked as a software engineer for two years at the National Radio Astronomy Observatory (NRAO) on the Atacama Large Millimeter Array (ALMA) radio telescope project, where he received a Star Award in 2007.

He began doctoral studies in physics at the University of Rochester (UR) in 2007 with a Department of Education GAANN Fellowship, received an American Association of Physics Teachers (AAPT) award in 2008, completed a Master of Arts degree in physics in 2009, received the Agnes M. and George Messersmith Fellowship in 2011, and received a Certificate of Gratitude from the David T. Kearns Center in 2012 for helping to tutor undergraduate students. He has carried out his doctoral research in Quantum Information under the supervision of Professor Andrew N. Jordan, with a brief joint appointment in the laboratory of Professor John C. Howell.

The following publications resulted from the work conducted during the doctoral study:

1. Dressel, J., Howell, J.C., Rajeev, S., and Jordan, A.N., *Gravitational Redshift and Deflection of Slow Light*, Physical Review A **79**, 013834 (2009).

2. Dressel, J., Agarwal, S., and Jordan, A.N., *Contextual Values of Observables in Quantum Measurements*, Physical Review Letters **104**, 240401 (2010).
3. Dressel, J., Broadbent, C.J., Howell, J.C., and Jordan, A.N., *Experimental Violation of Two-Party Leggett-Garg Inequalities with Semi-weak Measurements*, Physical Review Letters **106**, 040402 (2011).
4. Dressel, J., and Jordan, A.N., *Sufficient conditions for uniqueness of the weak value*, Journal of Physics A: Mathematical and Theoretical **45**, 015304 (2012).
5. Dressel, J., and Jordan, A.N., *Significance of the imaginary part of the weak value*, Physical Review A **85**, 012107 (2012).
6. Dressel, J., Choi, Y., and Jordan, A.N., *Measuring which-path information with coupled electronic Mach-Zehnder interferometers*, Physical Review B **85**, 045320 (2012).
7. Dressel, J., and Jordan, A.N., *Contextual-value approach to the generalized measurement of observables*, Physical Review A **85**, 022123 (2012).
8. Dressel, J., and Jordan, A.N., *Weak values are universal in von Neumann measurements*, Physical Review Letters **109**, 230402 (2012).
9. Dressel, J., and Jordan, A.N., *Corrigendum: Sufficient conditions for uniqueness of the weak value*, Journal of Physics A: Mathematical and Theoretical **46**, 029501 (2013).

Acknowledgments

Without the unwavering support of many key individuals, this work would never have seen the light of day. Special thanks must be given to my thesis supervisor, Professor Andrew N. Jordan, who has displayed the patience of a saint while weathering my idiosyncracies and incoherent rambblings over the past five years. He has provided enough rope for me to hang myself, but has never forgotten to provide the shears for me to escape again with some measure of dignity intact. His enthusiastic guidance has been a shining beacon during the long dark hours of work required to bring these ideas to fruition.

I also wish to thank my colleagues for the many hours of discussion and debate that helped to shape these ideas more clearly. Shantanu Agarwal and Dr. Curtis Broadbent deserve special recognition for their patience and insight. I also warmly thank Professor Joseph Eberly, Professor John Howell, Professor Sarada Rajeev, Dr. Nathan Williams, Yunjin Choi, Areeya Chantrasi, Kevin Lyons, Dr. P. Ben Dixon, Levi Greenwood, Mohammed Hashemi, Dr. David Starling, Andrew Vigoren, Xiaofeng Qian, Dr. Xu Wang, Burt Betchart, Greg Howland, Julian Martinez, Joseph Choi, Dr. Mehul Malik, Andreas Liapis, Aaron Mislivec, Dr. Melanie Day, Leah Pearson, Darcey Riley, and many others for thoroughly enjoyable and helpful discussions and support.

The department itself deserves my warm gratitude for being such a supportive and welcoming environment. I thank the many people whose hard work has

made that environment possible, especially Barbara Warren, Janet Fogg, Laura Blumkin, Shirley Brignall, Connie Hendricks, Michie Brown, and Ali DeLeon.

I am also indebted to Mike Culver, Professor Joseph Eberly, and Andrew Vigoren for dragging me out of my office to play raquetball periodically. In a similar vein, I must thank my parents, my sister, and especially my fiancée Samantha Newmark for helping me to maintain my sanity and motivation while bringing this work to completion.

Abstract

This thesis presents a general algebraic approach for indirectly measuring both classical and quantum observables, along with several applications. To handle the case of imperfectly correlated indirect detectors we generalize the observable spectra from eigenvalues to contextual values. Eigenvalues weight spectral idempotents to construct an observable, but contextual values can weight more general probability observables corresponding to indirect detector outcomes in order to construct the same observable. We develop the classical case using the logical approach of Bayesian probability theory to emphasize the generality of the concept. For the quantum case, we outline how to generalize the classical case in a straightforward manner by treating the classical sample space as a spectral idempotent decomposition of the enveloping algebra for a Lie group; such a sample space can then be rotated to other equivalent sample spaces through Lie group automorphisms. We give several classical and quantum examples to illustrate the utility of our approach. In particular, we use the approach to describe the theoretical derivation and experimental violation of generalized Leggett-Garg inequalities using a quantum optical setup. We also describe the measurement of which-path information using an electronic Mach-Zehnder interferometer. Finally, we provide a detailed and exact treatment of the quantum weak value, which appears as a general feature in conditioned observable measurements using a weakly correlated detector.

Contributors and Funding Sources

This work was supervised by a dissertation committee consisting of Professors Andrew N. Jordan (advisor), John C. Howell, and Joseph H. Eberly of the Department of Physics and Astronomy, as well as Professor Robert W. Boyd of the Department of Optics and Professor Carl E. Mueller of the Department of Mathematics. All work was completed by the student with generous help from his advisor Professor Andrew N. Jordan. The Introduction and Chapters 2 and 3 are adapted from the publication *Phys. Rev. Lett.* 104, 240401 (2010) done in collaboration with Shantanu Agarwal, as well as the publications *J. Phys. A: Math. Theor.* 45, 015304 (2012), and *Phys. Rev. A* 85, 022123 (2012). Chapter 4 is adapted from the publication *Phys. Rev. Lett.* 106, 040402 (2011); the data for this experiment was collected, analyzed, and plotted by the student in the laboratory of Professor John C. Howell in close collaboration with Dr. Curtis J. Broadbent and Dr. David J. Starling. Chapter 5 is adapted from the publication *Phys. Rev. B* 85, 045320 (2012) done in collaboration with Yunjin Choi. Chapter 6 is adapted from the publications *Phys. Rev. A* 85, 012107 (2012) and *Phys. Rev. Lett.* 109, 230402 (2012).

Graduate study was supported by a Department of Education GAANN Fellowship and an Agnes M. and George Messersmith Fellowship from the University, as well as by Professor Andrew N. Jordan through the following grants: DARPA DSO Slow Light Grant, NSF Grant No. DMR-0844899, and ARO Grant No. W911NF-09-0-01417.

Table of Contents

Biographical Sketch	iii
Acknowledgments	v
Abstract	vii
Contributors and Funding Sources	viii
List of Tables	xiv
List of Figures	xv
List of Symbols	xvii
1 Introduction	1
2 Classical Observable Measurement	8
2.1 Example: Colorblind Detector	9
2.2 Sample space, Boolean algebra, and Observables	11
2.3 States, Densities, and Collapse	16
2.4 Detectors and Probability Observables	23
2.5 Contextual Values	31

2.5.1	Example: Ambiguous Marble Detector	39
2.5.2	Example: Invasive Ambiguous Detector	41
2.5.3	Example: Redundant Ambiguous Detector	44
2.5.4	Example: Continuous Detector	47
3	Quantum Observable Measurement	52
3.1	Sample Spaces and Observables	52
3.1.1	Example: Polarization	56
3.2	States, Densities, and Collapse	59
3.2.1	Example: Polarization State	64
3.3	Detectors and Probability Observables	65
3.3.1	Example: Coverslip Polarization Detector	72
3.4	Contextual Values	75
3.4.1	Example: Coverslip Detector Revisited	80
3.4.2	Example: Calcite Polarization Detector	83
3.5	The Weak Value as a Conditioned Average	89
4	Generalized Leggett-Garg Inequalities	96
4.1	Macro-realism	97
4.2	Leggett-Garg Inequalities	98
4.3	Conditioned Averages	101
4.4	Quantum Formulation	102
4.5	Experimental Setup	103
4.6	Results	106

5	Electronic Mach-Zehnder Interferometry	109
5.1	Coupled MZIs	112
5.2	Measurement Interpretation	118
5.2.1	POVM	118
5.2.2	Contextual Values	120
5.2.3	Parameters	125
5.2.4	Measurement Disturbance	128
5.2.5	Strong Coupling	130
5.2.6	Weak Coupling Limit	131
5.2.7	Fluctuating Coupling	134
5.2.8	Observation Time	135
5.3	Conditioned Measurements	137
5.3.1	Joint Scattering	137
5.3.2	Quantum Erasure	142
5.3.3	Conditioned Averages	146
5.3.4	Deterministic Measurement	149
5.3.5	Strong Coupling	149
5.3.6	Weak Coupling Limit	150
6	The Complex Weak Value	153
6.1	von Neumann Measurement	159
6.1.1	Traditional Analysis	159
6.1.2	Quantum Operations	161
	Unconditioned Measurement	161
	Conditioned Measurement	166

6.2	Joint Weak Values	169
6.2.1	Reduced State Expressions	171
6.3	Linear Response Regime	173
6.4	Time Symmetry	178
6.5	Examples	180
6.5.1	Bohmian Mechanics	180
6.5.2	Qubit Observable	182
6.5.3	Gaussian Detector	187
6.5.4	Hermite-Gauss Modes	193
7	Concluding remarks	198
	Bibliography	203
A	Algebraic Quantum Mechanics	216
A.1	Lie Groups	217
A.1.1	Lie Algebra	218
A.1.2	Lie Exponential Map	220
A.1.3	Universal Enveloping Algebra	221
A.2	Quantum Observables	224
A.3	Adjoint Maps	225
A.4	Idempotents, Spectra, and Irreducible Factors	227
A.5	Trace	231
A.6	Algebraic Norms and Classification	233
A.7	Representations and Hilbert Space	235
A.7.1	Hilbert Space	236

Adjoint Representation	237
Unitary Representation	238
A.7.2 Dirac Notation	240
A.8 Example: Quantum Spin	242
A.8.1 Spacetime Algebra	242
A.8.2 Spinor Algebra	245
B Sufficient Conditions for the Weak Value	249
C Edge Channel Interaction Phase	254
D Joint Weak Value Moments	260
D.1 Detector Wigner Function	262
D.2 Hermite-Gauss superpositions	263

List of Tables

6.1	Laguerre polynomials	196
A.1	Group properties	218
A.2	Enveloping algebra properties	223
A.3	Algebraic element classification	233
A.4	Algebraic factor classification	234

List of Figures

2.1	Diagram of sample space, Boolean algebra, and observable algebras	12
3.1	Coverslip polarization measurement	72
3.2	Calcite polarization CV	93
3.3	Conditioned detector probability densities	94
3.4	Conditioned detector averages	95
4.1	LGI measurement schematic	98
4.2	LGI Experimental Setup	102
4.3	LGI experimental density matrix	104
4.4	LGI conditioned averages	105
4.5	LGI violations 1	106
4.6	LGI violations 2	108
5.1	Coupled MZI schematic	110
5.2	QPC balance parameters	115
5.3	MZI which-path CV	124
5.4	MZI drain probabilities	125
5.5	Raised cosine distribution	133

5.6	MZI joint probabilities $\phi^s = 0$	139
5.7	MZI joint probabilities $\phi^s = \pi/2$	140
5.8	MZI conditional probabilities	141
5.9	MZI quantum erasure	143
5.10	MZI quantum erasure optical analogy	145
5.11	MZI conditioned average	148
5.12	MZI weak and semiweak values	150
6.1	von Neumann measurement schematic	163
6.2	Sequential measurement schematic	167
6.3	Qubit Bloch sphere rotation	186
6.4	Qubit decoherence	191
6.5	Hermite-Gauss measurement schematic	193
6.6	Hermite-Gauss detector profiles	194
6.7	Hermite-Gauss Bloch sphere distortions	195

List of Symbols

AAV	Aharonov-Albert-Vaidman
AB	Aharonov-Bohm
ABL	Aharonov-Bergmann-Lebowitz
CA	Conditioned Average
CV	Contextual Values
HG	Hermite-Gauss
LGI	Leggett-Garg Inequality
MR	Macro-Realism
MSE	Mean-Squared-Error
MZI	Mach-Zehnder Interferometer
POVM	Positive-Operator-Valued Measure
PVM	Projection-Valued Measure
RMS	Root-Mean-Squared
SPAD	Single Photon Avalanche photoDiodes
WV	Weak Value
\mathbb{N}	Natural Numbers
\mathbb{R}	Real Numbers
\mathbb{C}	Complex Numbers
\aleph_0	Cardinality of \mathbb{N}
\aleph_1	Cardinality of \mathbb{R}
\mathfrak{G}	Lie Group

\mathfrak{g}	Lie Algebra
$\mathcal{E}(\mathfrak{g})$	Universal Enveloping Algebra
\mathfrak{A}	General Embedding Algebra
\mathfrak{H}	Hilbert Space
$\mathcal{P}(\mathfrak{H})$	Projective Hilbert Space
$\mathcal{O}(\mathfrak{H})$	Space of Operators on \mathfrak{H}
Φ	Gel'fand Triple Subspace
Φ^*	Gel'fand Triple Dual
\vec{D}, \vec{G}	Lie Algebra Vector
\vec{v}, \vec{w}	Hilbert Space vector
X, Y	Sample Spaces
Σ_X, Σ_Y	Boolean Sigma-algebra
$\Sigma_X^{\mathbb{R}}, \Sigma_Y^{\mathbb{R}}$	Boolean Observable Space
$\text{Ad}_{(\cdot)}$	Lie Group Adjoint Map
$\text{ad}[\cdot]$	Lie Algebra Adjoint Map
ε	Algebraic Idempotent
\hat{A}	Hilbert Space Operator
$\hat{\Pi}$	Projection Operator
$ \cdot\rangle$	Dirac Ket
$\langle\cdot $	Dirac Bra
$[\cdot, \cdot]$	Lie Bracket / Commutator
$(\cdot)^*$	Algebraic Involution
$(\cdot)^\dagger$	Hilbert Space Adjoint
$\langle\cdot, \cdot\rangle$	Hilbert-Schmidt Inner Product
(\cdot, \cdot)	Hilbert Space Inner Product
$\langle\cdot \cdot\rangle$	Dirac Bra-Ket Inner Product
$\text{Tr}(\cdot)$	Trace Functional
$P(\cdot)$	Probability Measure

$\mu(\cdot)$	Reference Measure
$\langle \cdot \rangle$	Expectation Functional
$\langle \cdot \rangle_\mu$	Reference Functional
$\langle \cdot \rangle$	Invasive Expectation Functional
${}_b \langle \cdot \rangle_a$	Expectation Pre-selected in a and Post-selected in b
${}_b \langle \cdot \rangle_a^w$	Generalized Real Weak Value
$(\cdot)^w$	Complex Weak Value Parameter

1 Introduction

It is terrible to see how a single unclear idea, a single formula without meaning, lurking in a young man's head, will sometimes act like an obstruction of inert matter in an artery, hindering the nutrition of the brain, and condemning its victim to pine away in the fullness of his intellectual vigor and in the midst of intellectual plenty.

Charles S. Peirce

Over the past century there has been a remarkable shift in the foundational philosophy of physics. The mechanistic worldview of realistic determinism that drove the development of the classical theory has inexorably succumbed to a stochastic worldview of relational event potentialities. Our most fundamental physical theories now predict only the likelihoods of and correlations between irreversible *measurement events*. The age of certainty has ended [1], yielding to the modern age of probability [2].

This shift in philosophy has not come easily, however. A historical attachment to realism and certainty lingers in our language and our mathematical methods, despite our acknowledgment that our theories must be irreducibly stochastic. This conceptual dualism between certainty and stochasticity manifests most strongly in the quantum theory, where there appear to be two inequivalent modes of evolution. On one hand, the quantum state is often described as a physical entity

that obeys continuous, deterministic, and reversible evolution. It is well described by a wave equation, which specifies a smooth function over a static spacetime. On the other hand, the stochasticity of the measurement process necessarily disrupts this smooth evolution with discontinuous and irreversible ‘quantum jumps’ or ‘state collapse’ events. Debates still rage in the literature over the meaning of this dualistic description, with interpretations ranging from the extremes of a rejection of the physical status of the smoothly evolving quantum state (the quantum Bayesian interpretation [3–5]) to a rejection of the physical status of the discontinuous measurement events (the many-worlds/decoherence interpretation [6–9]).

The traditional quantum theory—unified by Dirac [10] and made more rigorous by von Neumann [11]—copes with the necessary introduction of stochasticity by leveraging existing analogies with classical mechanics. In particular, the formalism emphasizes the role of non-commutative observable operators that are stochastic analogues to the intrinsic measurable properties of deterministic classical particles. Indeed, observables underlie most core concepts in the quantum theory: commutation relations of observables, complete sets of commuting observables, spectral expansions of observables, conjugate pairs of observables, expectation values of observables, uncertainty relations between observables, and time evolution generated by a Hamiltonian observable. Even the quantum state is introduced in terms of the possible values that measurable observables can take. The stochasticity of the theory manifests in a single prescription for averaging the omnipresent observables under a smoothly and deterministically evolving quantum state: the distastefully stochastic quantum jumps corresponding to laboratory measurements are largely hidden by the theoretical formalism in favor of these smoothly evolving averages.

Experimental control of quantum systems has dramatically improved since the early days of quantum mechanics, however, so the discontinuous evolution present in the measurement process can no longer remain hidden and must now

be more carefully investigated. Modern optical and condensed matter systems, for example, can monitor quantum systems using weakly coupled measurement devices (e.g., [12]), resulting in sequences of *nonprojective* quantum jumps that extract partial information from the state and alter it more gently. These gentler jumps exhibit useful features that have no counterpart in the projective jumps assumed by the traditional theory. For example, these gentler jumps can be conditionally reversed by subsequent measurements [13–15] to “uncollapse” the quantum state. More generally, gentle monitoring can be non-destructive, so measurement results can be fed forward into another device that interacts with the system to deliberately control its subsequent evolution [16].

These modern laboratory realities have prompted a pragmatic departure from the observable foundation of the quantum theory. Traditional observables are operationally averaged in terms of projective jumps that strongly affect the state, so it is overtly unclear how to correctly apply such a formalism to measurements that only gently affect the state. The traditional formalism must be refined to address the wider variety of stochastic events that can appear in the laboratory.

In the early 1970’s Davies [17] and Kraus [18] introduced a refinement to the theory of quantum measurement that could address the gentler jumps that can arise from interactions with a detecting device. Their formalism of *quantum operations*, or generalized measurements, has been developed over the past forty years to become a comprehensive and mathematically rigorous complement to the standard quantum theory [19–27]. Quantum operations has seen the most use in the quantum information and quantum computation communities, often with well-supported experimental implementations in quantum optical and condensed matter systems. However, it has not yet seen wide adoption outside of those communities, and so has not yet been widely assimilated into the larger quantum mechanical community.

Unlike the traditional formalism that emphasizes the observables, the quan-

tum operations formalism emphasizes the *states*. Observables are mentioned infrequently in the quantum operations literature, appearing only in the context of projective measurements where they are well-understood. Some references (e.g. [24, 25, 27]) define “generalized observables” in terms of the generalized measurements and detector outcome labels, but give no indication about their relationship to traditional observables, if any. As a result, there has been a growing conceptual gap between the traditional quantum mechanics of observables and the modern treatment of quantum operations that encompasses a much larger class of possible measurements than the traditional observables seemingly allow.

A possible response to this conceptual gap is to declare that measuring traditional observables is a meaningless exercise outside the context of projective measurements. This argument is supported by the fact that any generalized measurement can be understood as a part of a projective measurement being made on a larger joint system (e.g., [24, p. 20]) that can be associated with a joint observable in the usual way. However, this argument neglects that observables still indicate intrinsic symmetries even when measured in such a joint context. Moreover, this argument neglects the parallel research into the “weak measurement” of observables [28–43], which strongly suggests that the notion of linking generalized measurements to traditional observables may not be such an outlandish idea.

The possibility of weak measurements was emphasized in 1988 by Aharonov, Albert, and Vaidman (AAV) [28] as an application of the 1932 von Neumann measurement protocol [11] that uses an interaction Hamiltonian with variable coupling strength to correlate an observable of interest to the generator of translations for a continuous meter observable. The resulting shift in the meter observable is then used to infer information about the observable of interest in a nonprojective manner. The technique has been used to great effect in various laboratories [44–60] to measure physical quantities like pulse delays, beam deflections, phase shifts, polarization, and averaged trajectories. Therefore, we are forced to conclude that

there must be some meaningful way to reconcile generalized measurements with traditional observables more formally.

The primary purpose of the present work is to detail a synthesis between generalized measurements and traditional observables that is powerful enough to encompass standard projective measurements, AAV weak measurements, and any other type of laboratory measurement in between. To accomplish this synthesis, we will develop an algebraic formalism in which the relevant values associated with the measurement of an observable will depend upon *how* that observable is measured. This inability to completely discuss observables without specifying the full measurement strategy is reminiscent of Bell-Kochen-Specker contextual-ity [61–67]—namely, the acknowledgment that quantum observables cannot have pre-determined values without specifying the complete set of compatible measurements that will be performed—which motivates us to name such measurement-context-dependent values the *contextual values* (CV) for an observable. These contextual values will form the necessary bridge between the traditional observable formalism and the modern formalism of quantum operations [68–74].

The secondary purpose of the present work is to outline how both classical probability theory and quantum operations fit together naturally. The developed algebraic approach treats stochastic events and inference in a uniform way, so can describe both classical and quantum probabilistic inference with the same mathematics. Hence, the algebraic approach enables a detailed comparison of the structural similarities between the classical and quantum probabilistic theories. Moreover, the generality ensures that any physical system that can be described by classical Bayesian probability theory will also be able to take full advantage of our general observable measurement technique, and indeed take advantage of the full theory of generalized measurements.

To accomplish these goals, this work will be organized as follows. Chapter 2 will develop the algebraic theory of classical probability starting from an *a priori*

known space of possible events represented as algebraic idempotents. We will pay special attention to generalized classical observable measurements using contextual values. In addition, we will pay special attention to classically invasive measurements, as well as conditioned sequences of measurements.

Chapter 3 will extend the classical probability space to a quantum probability space by identifying the event idempotents as the spectral idempotents for the enveloping algebra of a Lie group. The group structure produces a continuous manifold of incompatible frameworks that are each equivalent to a classical probability space. We will discuss the new quantum features that appear when measuring observables using contextual values due to these incompatible frameworks. Notably, we will show that the real part of a complex quantity known as the *quantum weak value* appears as a natural limit point for a conditioned observable average under fairly general conditions.

Chapter 4 illustrates the use of the contextual value formalism by taking advantage of the strong parallels between the classical and quantum approaches to derive and experimentally violate generalized Leggett-Garg inequalities that test the postulates of macro-realism. We will find that these violations may be understood as a form of classically invasive measurement that can be revealed in sequences of measurements.

Chapter 5 provides another simple but nontrivial example of the contextual values formalism by carefully analyzing the measurement of which-path information using a pair of coupled electronic Mach-Zehnder interferometers. The contextual values formalism makes the discussion particularly transparent, despite the thorny interpretational issues that arise. Notably, the symmetry of the detecting setup leads to a complementarity between particle-like behavior in the system and wave-like behavior in an efficient detector.

Chapter 6 presents a case study of the complex quantum weak value that is enabled by our exact algebraic solution of the von Neumann measurement pro-

ocol. This case study permits the concrete interpretation of the imaginary part of the weak value as a logarithmic directional derivative, which directly involves the observable in its role as a Lie algebraic group generator. This interpretation augments the contextual values interpretation of the real part of the weak value as a conditioned observable average. We also demonstrate that the real part of the generalized complex weak value completely characterizes all von Neumann measurements, making them a universal feature of that measurement protocol.

Finally, Chapter 7 provides a few concluding remarks.

2 Classical Observable Measurement

Common language—or, at least, the English language—has an almost universal tendency to disguise epistemological statements by putting them into a grammatical form which suggests to the unwary an ontological statement. A major source of error in current probability theory arises from an unthinking failure to perceive this. To interpret the first kind of statement in the ontological sense is to assert that one's own private thoughts and sensations are realities existing externally in Nature. We call this the 'mind projection fallacy', and note the trouble it causes many times in what follows. But this trouble is hardly confined to probability theory; as soon as it is pointed out, it becomes evident that much of the discourse of philosophers and Gestalt psychologists, and the attempts of physicists to explain quantum theory, are reduced to nonsense by the author falling repeatedly into the mind projection fallacy.

Edwin T. Jaynes, (2003) [2]

Before delving into the measurement of quantum observables, we first develop the theory of classical observable measurement. Our treatment acknowledges that probability theory, in its most general incarnation, is a system of formal reasoning about Boolean logic propositions [2, 75, 76]. We shall see that the algebraic form of a mutually exclusive set of logical propositions is an idempotent-valued measure, which will be our contact point with quantum mechanics in Chapter 3 and Appendix A.

To make the transition from the classical theory to the quantum theory more transparent, we shall develop the classical theory using the same language as

the modern theory of quantum operations [17–27]. This strategy will highlight exactly where the classical and quantum theories differ and will provide useful analogies for understanding those differences. We will pay special attention to invasive measurements and conditioned sequences of measurements, which will be necessary to understand the quantum generalization.

2.1 Example: Colorblind Detector

A particularly significant contribution we shall make to the operational approach is the *contextual values* formalism that enables the indirect measurement of observables using imperfectly correlated logical propositions, such as those of an ancilla detector. Without this contribution, we feel that the standard operational approach has been incomplete. To make this new idea clear before a more formal and comprehensive exposition, we briefly describe an illustrative example of an *ambiguous detector*.

Suppose we wish to measure a marble that may be colored either red or green. A person with normal vision can distinguish the colors unambiguously and so would represent an ideal detector for the color state of the marble. A partially colorblind person, however, may only guess the color correctly some percentage of the time and so would represent an ambiguous detector of the color state of the marble.

If the person is only mildly colorblind, then the estimations will be strongly correlated to the actual color of the marble. The ambiguity would then be perturbative and could be interpreted as *noise* introduced into the measurement. However, if the person is strongly colorblind, then the estimations may be only mildly correlated to the actual color of the marble. The ambiguity becomes *non-perturbative*, so the noise dominates the signal in the measurement.

We can design an experimental protocol where an experimenter holds up a

marble and the colorblind person gives a thumbs-up if he thinks the marble is green or a thumbs-down if he thinks the marble is red. Suppose, after testing a large number of known marbles, the experimenter determines that a green marble correlates with a thumbs-up 51% of the time, while a red marble correlates with a thumbs-down 53% of the time. The experimental outcomes of thumbs-up and thumbs-down are thus only weakly correlated with the actual color of the marble.

Having characterized the detector in this manner, the experimenter provides the colorblind person with a very large bag of an unknown distribution of colored marbles. The colorblind person examines every marble, and for each one records a thumbs-up or a thumbs-down on a sheet of paper, which he then returns to the experimenter. The experimenter then wishes to reconstruct what the average distribution of marble colors in the bag must be, given only the ambiguous output of his colorblind detector.

For simplicity, the clever experimenter decides to associate the colors with numerical values: 1 for green (g) and -1 for red (r). In order to compare the ambiguous outputs with the colors, he also assigns them *different* numerical values: a for thumbs-up (u), and b for thumbs-down (d). He then writes down the following probability constraint equations for obtaining the average marble color, $\langle \text{color} \rangle$, based on what he has observed,

$$\langle \text{color} \rangle = 1P(g) - 1P(r) = aP(u) + bP(d), \quad (2.1)$$

$$P(u) = (.51)P(g) + (.47)P(r),$$

$$P(d) = (.49)P(g) + (.53)P(r).$$

By inserting $P(u)$ and $P(d)$ into (2.1), he can rewrite the equation as a matrix

equation in the basis of the color probabilities $P(g)$ and $P(r)$,

$$\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} .51 & .49 \\ .47 & .53 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}, \quad (2.2)$$

which must be true for all $P(g)$ and $P(r)$. After solving this equation, he finds that he must assign the amplified values $a = 25.5$ and $b = -24.5$ to the outcomes of thumbs-up and thumbs-down, respectively, in order to compensate for the detector ambiguity. After doing so, he can confidently calculate the average color of the marbles in the large unknown bag using the identity (2.1).

The classical color observable has chosen characteristic values of 1 and -1 that correspond to an ideal measurement. The amplified values of 25.5 and -24.5 that must be assigned to the ambiguous detector outcomes are *contextual values* for the same color observable. The *context* of the measurement is the characterization of the colorblind detector, which accounts for the degree of colorblindness. The expansion (2.1) relates the spectrum of the observable to its generalized spectrum of contextual values. With this identity, both an ideal detector and a colorblind detector can measure the same observable; however, the assigned values must change depending on the context of the detector being used.

2.2 Sample space, Boolean algebra, and Observables

The space of classical observables is a commutative algebra over the reals that we will denote $\Sigma_X^{\mathbb{R}}$. This choice of notation is motivated by the fact that the observable algebra is built from and contains two related spaces, X and Σ_X , that are conceptually distinct and equally important to the theory. The three are illustrated in Fig. 2.1 for reference. We will now briefly construct these three

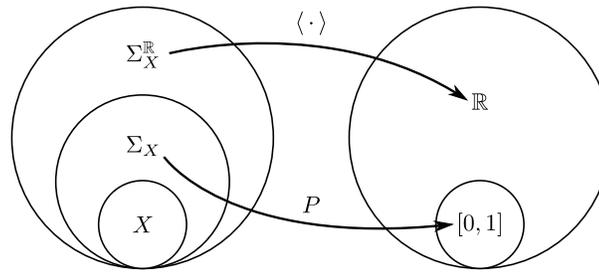


Figure 2.1: Diagram of the relationship between the sample space of atomic propositions X , the Boolean algebra of propositions Σ_X , and the algebra of observables $\Sigma_X^{\mathbb{R}}$. The probability state P is a measure from Σ_X to the interval $[0, 1]$. The expectation functional $\langle \cdot \rangle$ is a linear extension of P that maps $\Sigma_X^{\mathbb{R}}$ to the reals \mathbb{R} ; by construction $\langle \cdot \rangle = P(\cdot)$ whenever both are defined.

spaces.

Sample spaces.—The core of a probability space is a set of mutually exclusive logic propositions, X , known as the *sample space of atomic propositions*. In other words, elements of the sample space, such as $g, r \in X$, represent “yes or no” questions that cannot be answered “yes” simultaneously and cannot be broken into simpler questions. For example, $g =$ “Does the marble look green?” and $r =$ “Does the marble look red?” are valid mutually exclusive atomic propositions. To be a proper sample space, the propositions should form a complete set, meaning that there must always be exactly one true proposition. Physically, such propositions typically correspond to mutually independent outcomes of an experiment that probes some system of interest. Indeed, any accessible physical property must be testable by some experiment, and any experiment can be described by such a collection of yes or no questions.

Boolean algebra.—The atomic propositions in X can be extended to more complex propositions by logical combination in order to form the larger space Σ_X . Specifically, we promote them to an *algebra* under a logical OR and a logical AND. The AND operation takes the algebraic form of a product, $x \text{ AND } y = xy$. Propositions are *idempotent* under the product, so $x^2 = x$. Moreover, *disjoint*

propositions are trivially false under an AND, $xy = 0$, where 0 is the trivially false proposition. The OR operation takes the algebraic form of addition such that the redundant overlap is subsequently removed, $x \text{ OR } y = x + y - xy$. Propositions are idempotent under the logical OR, but not under the raw addition. The OR of all mutually exclusive propositions in X , which has the form of a simple sum, is trivially true. We denote the trivially true proposition as 1_X since its product with any proposition $x \in X$ leaves that proposition invariant, $1_X x = x$. The logical operation of NOT, or complementation (x^c) with respect to X , can then be defined as the subtraction from the trivially true proposition $x^c = 1_X - x$ since $x + x^c = 1_X$ must be true for complementary disjoint propositions. The proposition space Σ_X contains X and is closed under the operations of AND, OR, and NOT; hence, it forms a *Boolean logic algebra*¹. Both the sum and the product of this algebra commute.

The set X of atomic propositions is precisely the set of *primitive idempotents* of Σ_X that partition the trivially true proposition, $1_X = \sum_X x$. As idempotents, they form a partial order under the product as discussed in Appendix A.4. For a continuum of propositions in X then this partitioning process leads to an *idempotent-valued measure*, $1_X = \int_X d\varepsilon(x)$, that takes a measurable set from a measurable space that indexes X , such as the Borel sets on the real line, and uniquely assigns it a corresponding proposition in Σ_X . In what follows, we will usually denote propositions simply as x for simplicity, but we shall also denote idempotent-valued measures as $d\varepsilon(x)$ to disambiguate them from real-valued measures when necessary.

Observables.—We then extend Σ_X linearly over the real numbers to obtain the commutative algebra of *observables* $\Sigma_X^{\mathbb{R}}$. That is, any linear combination of

¹In the measure theoretic version of probability theory [77–80] Σ_X is known as a σ -algebra over the set X , which motivates the notation. There Σ_X is a nonempty set of subsets of X that is closed under countable unions (OR), countable intersections (AND), and set complementation (NOT).

propositions $F = ax + by$ with $a, b \in \mathbb{R}$ and $x, y \in \Sigma_X$ is an observable in $\Sigma_X^{\mathbb{R}}$; similarly any linear combination of observables $H = a'F + b'G$ with $a', b' \in \mathbb{R}$ and $F, G \in \Sigma_X^{\mathbb{R}}$ is also an observable in $\Sigma_X^{\mathbb{R}}$. Countable sums are permitted provided the coefficients converge. The three spaces X , Σ_X , and $\Sigma_X^{\mathbb{R}}$ are illustrated in Fig. 2.1.

The observables combine logical propositions with numbers that describe the relation of each proposition to some meaningful reference. For example, one could define a simple observable $A = (1)g + (-1)r$ that assigns a value of 1 to the proposition asking whether a marble looks green and assigns a value of -1 to the proposition asking whether that same marble looks red in order to distinguish the colors by a sign. Alternatively, one can bestow a physical meaning to the color propositions by defining a wavelength observable instead: $B = (550\text{nm})g + (700\text{nm})r$. One could even define an observable $C = (\$2)g + (-\$3)r$ that indicates a monetary bet made on the color of the marble, with $\$2$ being awarded for a color of green and $\$3$ being lost for a color of red. Such numerical labels are always assigned by convention, but indicate physically relevant information about the type of questions being asked by the experimenter that are answerable by the independent propositions.

Representation.—Since it consists of idempotents that partition unity, a Boolean algebra Σ_X can be represented as the lattice of projection operators acting on a (rigged) Hilbert space as discussed in Appendix A.7. The elements $\{x\}$ of X correspond to rank-1 projection operators $\{|x\rangle\langle x|\}$ onto orthogonal subspaces spanned by orthonormal vectors $\{|x\rangle\}$ in the Hilbert space, or δ -normalized orthogonal vectors in a rigged Hilbert space. Any sum of n elements of X , $x_1 + \dots + x_n$, corresponds to a rank- n projection operator $|x_1, \dots, x_n\rangle\langle x_1, \dots, x_n|$ onto a subspace spanned by n orthonormal vectors $\{|x_1\rangle, \dots, |x_n\rangle\}$ in the Hilbert space. The observables $\Sigma_X^{\mathbb{R}}$ can similarly be represented as the algebra of commuting Hermitian operators acting on the same Hilbert space. However, the algebras Σ_X and $\Sigma_X^{\mathbb{R}}$

need not be represented in this fashion to be well defined.

Independent Probability Observables.—We will call disjoint elements of the Boolean algebra Σ_X that partition unity *independent probability observables*. Unity can be partitioned into many disjoint sets as $\sum_i x_i = 1_X$, such that $x_i x_j = \delta_{ij}$. Each partitioning corresponds to a particular detector arrangement that only probes those propositions. A maximally refined partitioning is known as a *closure relation*. Note that any physical experiment will involve a *finite* partitioning, since one can only distinguish a finite number of outcomes in a laboratory.

Simple observables can be constructed from independent probability observables by associating a real value $f(x_i)$ to each proposition in the sum, $F = \sum_i f(x_i)x_i$. The product of the observable with any of its constituent probability observables simplifies, $F x_i = f(x_i)x_i$; hence, the associated values form the set of *eigenvalues* for the observable. For a finite observable space $\Sigma_X^{\mathbb{R}}$, the set of atomic propositions X itself is a maximally refined partition that can construct any observable in the space,

$$F = \sum_{x \in X} f(x)x. \quad (2.3)$$

In the continuous case, one can formally construct limiting sequences of simple observables that approximate more general observables defined as in Appendix A.4 using an idempotent-valued measure,

$$F = \int_X f(x)d\varepsilon(x), \quad (2.4)$$

where $f(x)$ is a measurable function that specifies the *spectrum* of the observable. Importantly, however, no experimental protocol can measure such a continuous observable directly—only collections of approximating simple observables constructed from finite partitions may be probed by an experiment. Hence, in what

follows we will largely restrict ourselves to such finite partitions as producing the set of *measurable* observables.

2.3 States, Densities, and Collapse

Probability measures.—A state P is a *probability measure* over the Boolean algebra Σ_X , meaning that it is a linear map from Σ_X to the interval $[0, 1]$ such that $P(1_X) = 1$. Such a state P assigns a numerical value $P(x)$ to each proposition $x \in \Sigma_X$ that quantifies its degree of *plausibility*; that is, $P(x)$ formally indicates how likely it is that the question x would be answered “yes” were it to be answered, with 1 indicating a certain “yes” and 0 indicating a certain “no.” The value $P(x)$ is called the *probability* for the proposition x to be true. Normalizing $P(1_X) = 1$ ensures that exactly one proposition in the sample space must be true.

For continuous spaces, the state becomes an integral. For an experimentally accessible finite partitioning of a continuous space, each interval of the partition can be assigned probability from a theoretically continuous state via the integral,

$$P(x_0) = \int_{x_0 \in \Sigma_X} P(d\varepsilon(x)) = \int_{x_0 \in \Sigma_X} dP(x), \quad (2.5)$$

where the integration measure $dP(x) = P(d\varepsilon(x))$ is defined in terms of its limiting values on the primitive propositions.

Frequencies.—Empirically, one can check a *finite set* of probabilities by repeatedly asking a proposition in Σ_X to identically prepared systems and collecting statistics regarding the answers. For a particular proposition $x \in \Sigma_X$, the ratio of yes-answers to the number of trials will converge to the probability $P(x)$ as the number of trials becomes infinite. However, the probability has a well-defined meaning as a plausibility prediction even without actually performing such a repeatable experiment. Indeed, designing good quality repeatable experiments to

check the probabilities assigned by a predictive state is the primary goal of experimental science, and is generally quite difficult to achieve.

Expectation functionals.—The linear extension of a state P to the whole observable algebra $\Sigma_X^{\mathbb{R}}$ is an *expectation functional* that averages the observables, and is traditionally notated with angled brackets $\langle \cdot \rangle$. For a simple observable $F = \sum_{x \in X} f(x)x$, then,

$$\langle F \rangle = \sum_{x \in X} f(x)P(x), \quad (2.6)$$

is the *expectation value*, or average value, of F under the functional $\langle \cdot \rangle$ that extends the probability state P . Since $\langle \cdot \rangle$ is linear, it passes through the sum and the constant factors of $f(x)$ to apply directly to the propositions x . The restriction of $\langle \cdot \rangle$ to Σ_X is P , so $\langle x \rangle = P(x)$ as written in (2.6). The probability state P and its linear extension $\langle \cdot \rangle$ are illustrated in Fig. 2.1. For continuous spaces the sum (2.6) becomes an integral of the measurable function $f(x)$, $\langle F \rangle = \langle \int_X f(x)d\varepsilon(x) \rangle = \int_X f(x)dP(x)$; however, this can only be tested experimentally using a *simple* observable that forms a discrete approximation of the continuous one. Only simple observables are strictly measurable as averages of empirical frequencies.

Moments.—The n^{th} statistical moment of a simple observable F is $\langle F^n \rangle = \sum_{x \in X} f^n(x)P(x)$ and quantifies the fluctuations of the observable measurements that stem from uncertainty in the state. Specifically, the observable power has the form,

$$F^n = \sum_{x_1 \cdots x_n} f(x_1) \cdots f(x_n)x_1 \cdots x_n = \sum_x f^n(x)x, \quad (2.7)$$

which empirically corresponds to measuring the observable F n times in a row per trial on identical systems, giving a sense of the natural fluctuation of the

observable. The final equality only holds because the propositions x_i are mutually exclusive so subsequent measurements will agree with the first, $x_1 \cdots x_n = x_1 \delta_{1 \cdots n}$, and can thus be omitted as implicit in experiments. However, we will see in Section 2.5 that this simplification of the sum need not happen for more general measurements.

Densities.—States can often be represented as *densities* with respect to some *reference measure* μ from Σ_X to \mathbb{R}^+ , which can be convenient for calculational purposes. Just as the state P can be linearly extended to an expectation functional $\langle \cdot \rangle$, any reference measure μ can be linearly extended to a functional $\langle \cdot \rangle_\mu$.

As a familiar example from continuous spaces, such a reference functional takes the form of an integral $\langle F \rangle_\mu = \int_X f(x) d\mu(x)$. The representation of a state as a density follows from changing the integration measure for the state to the reference measure,

$$\langle F \rangle = \int_X f(x) dP(x) = \int_X f(x) \frac{dP}{d\mu}(x) d\mu(x). \quad (2.8)$$

The Jacobian conversion factor $dP/d\mu$ from the integral over the probability measure $dP(x)$ to the integral over the reference measure $d\mu(x)$ is the *probability density* for P with respect to μ , also known as a *Radon-Nikodym derivative* [77, 81–83]. From this probability density, we can then define a *state density observable*,

$$P_\mu = \int_X \frac{dP}{d\mu}(x) d\varepsilon(x), \quad (2.9)$$

using the appropriate idempotent-valued measure $d\varepsilon$ as discussed in Section A.4. This state density directly relates the expectation functional $\langle \cdot \rangle$ to the reference functional $\langle \cdot \rangle_\mu$ according to the relation (2.8) rewritten as an observable product,

$$\langle F \rangle = \langle P_\mu F \rangle_\mu. \quad (2.10)$$

For discrete partitions we can more intuitively define a state density observable in terms of a ratio of measures,

$$P_\mu = \sum_{x \in X} \frac{P(x)}{\mu(x)} x. \quad (2.11)$$

Then by definition and linearity, $\langle P_\mu F \rangle_\mu = \langle F \rangle$, as required. Evidently, the measure μ must be nonzero for all propositions x for which P is nonzero in order for such a state density to be well defined. This definition as a ratio of measures will correctly reproduce the Radon-Nikodym derivative (2.9) in the continuous case using a limiting prescription [77].

Trace.—The reference measure that is constant for all propositions of the same size according to their natural partial order is the *trace* Tr . For finite spaces the trace is usually defined as the counting measure which evaluates to the rank of any proposition in Σ_X ; for example, given $x, y, z \in X$ then $(x + y + z) \in \Sigma_X$ is a rank-3 proposition and $\text{Tr}(x + y + z) = \text{Tr}(x) + \text{Tr}(y) + \text{Tr}(z) = 1 + 1 + 1 = 3$. Since the trace evaluates to unity on any atomic proposition in a finite space, *any* state has a trace-density defined by equation (2.11) that is traditionally notated as ρ .

$$\rho = \sum_{x \in X} P(x) x. \quad (2.12)$$

This trace-density is the only state density that is always defined and exactly determined by the probabilities of the atomic propositions $P(x)$. Because of this, the trace-representation of a state can be naturally interpreted as an inner product,

$$\langle \rho, F \rangle = \text{Tr}(\rho F) = \langle F \rangle, \quad (2.13)$$

between the trace-density and the observable. The trace-density ρ will be completely equivalent to the quantum mechanical density operator when extended to

the quantum case in Chapter 3.

For continuous spaces the counting measure diverges, so the trace is redefined to be a multiple of the translationally-invariant Lebesgue-Stieltjes measure dx as a reference, as discussed in Appendix A.5. The probability density with respect to the Lebesgue-Stieltjes measure is given the simple notation $p(x)$, so (2.8) may be rewritten as $\langle F \rangle = \int_X f(x) p(x) dx$. It is important to remember that only finite approximations to this state density can be empirically verified, however.

State collapse.—If a question on the probability space is answered by some experiment, then the state indicating the plausibilities for future answers must be updated to reflect the acquired answer. The update process is known as *Bayesian state conditioning*, or *state collapse*. Specifically, if a proposition $y \in \Sigma_X$ is verified to be true, then the experimenter updates the expectation functional to the conditioned functional,

$$\langle F \rangle_y = \frac{\langle y F \rangle}{P(y)}, \quad (2.14)$$

that reflects the new information. For a proposition $x \in \Sigma_X$, the conditional probability $\langle x \rangle_y = P(yx)/P(y)$ has the traditional notation $P(x|y)$ and is read as “the probability of x given y .”

From (2.14), any state density corresponding to P will be similarly updated to a new density via a product,

$$P_{\mu|y} = \frac{P_{\mu} y}{P(y)}. \quad (2.15)$$

Notably, conditioning the trace-density ρ on an atomic proposition $y \in X$ will collapse the density to become the proposition itself, $\rho_y = \rho y / P(y) = y$.

Note that the proposition y serves a dual role in the conditioning procedure. First, it is used to compute the normalization probability $P(y)$. Second, it di-

rectly updates the state via a product action. The product indicates that future questions will be logically linked to the answered question with the *and* operation; that is, the knowledge about the system has been refined by the answered question. The process of answering a question about the system and then conditioning the state on the new information is called a *measurement*; moreover, since the proposition y is a projection acting on the density, this kind of measurement is called a *projective measurement*.

Bayes' rule.—If we pick another proposition $z \in \Sigma_X$ as the observable in (2.14) we can derive *Bayes' rule* as a necessary consequence by interchanging y and z and then equating the joint probabilities $P(yz)$,

$$P(z|y) = P(y|z) \frac{P(z)}{P(y)}. \quad (2.16)$$

Bayes' rule relates conditioned expectation functionals to one another and so is a powerful logical inference tool that drives much of the modern emphasis on the logical approach to probability theory.

Disturbance.—Conditioning, however, is not the only way that one can alter a state. One can also *disturb* a state without learning any information about it, which creates a transition to an updated expectation functional that we denote with a tilde $\langle \tilde{\cdot} \rangle$ according to,

$$\langle \tilde{F} \rangle = \langle \mathcal{D}(F) \rangle, \quad (2.17a)$$

$$\mathcal{D}(F) = \sum_{x \in X} \langle F \rangle_{D_x} x, \quad (2.17b)$$

$$\langle F \rangle_{D_x} = \sum_{x' \in X} f(x') D_x(x'). \quad (2.17c)$$

Here the disturbance \mathcal{D} is a map from $\Sigma_X^{\mathbb{R}}$ to $\Sigma_X^{\mathbb{R}}$ that is governed by a collection of states $\{D_x\}$ that specify *transition probabilities* $D_x(x')$ from old propositions x to new propositions x' . To be normalized, the transition states must

satisfy $D_x(1_X) = 1$, so that $\langle 1_X \rangle_{D_x} = 1_X$ and therefore $\mathcal{D}(1_X) = 1_X$. Updating the state according to (2.17) is also known as *Bayesian belief propagation* [5] and is more commonly written in the fully expanded form $\langle \tilde{F} \rangle = \sum_{x \in X} P(x) \sum_{x' \in X} D_x(x') f(x')$.

Time evolution.—As an important special case, the time evolution of a Markovian stochastic process is a form of disturbance \mathcal{D}_t , parametrized by a time interval t . No information is learned as the system evolves, so the knowledge about the system as represented by the expectation functional can only propagate according to the laws governing the time evolution. If the process is reversible, then this disturbance is the adjoint map for a one-parameter *Lie group* that leaves the space $\Sigma_X^{\mathbb{R}}$ invariant. For example, if we define a time-evolving observable as $F(t) = \mathcal{D}_t(F)$ on a continuous phase space then we have $dF(t)/dt = \{F(t), H\}_p$, where $\{\cdot, \cdot\}_p$ is defined point wise as the Poisson bracket (Lie bracket of the group) and H is the Hamiltonian (generator in the Lie algebra). This differential equation is a phase space representation of the group equation (A.12) discussed in Appendix A.3 using a truncation of the Moyal product as the algebraic product between functions, and has the exponential solution $\mathcal{D}_t(F) = \exp(t\{\cdot, H\}_p)(F)$, where the Liouvillian $\{\cdot, H\}_p$ is the induced adjoint map corresponding to H .

Correlation functions.—Correlations between observables at different times can be obtained by inserting a time-evolution disturbance between the observable measurements,

$$\langle F(0)G(t) \rangle = \langle F \mathcal{D}_t(G) \rangle = \sum_{x \in X} P(x) f(x) \sum_{x' \in X} D_{x,t}(x') g(x'). \quad (2.18)$$

Operationally this corresponds to measuring the observable F , waiting an interval of time t , then measuring the observable G . Similarly, n -time correlations can be defined with $n - 1$ time-evolution disturbances between the observable measurements $\langle F_1 \mathcal{D}_{t_1}(F_2 \cdots \mathcal{D}_{t_{n-1}}(F_n) \cdots) \rangle$. Computing the correlation of an observable

with itself without intervening disturbance will produce a higher moment $\langle F^n \rangle$.

Invasive measurement.—A system may also be disturbed during the physical process that implements conditioning, which will alter the state above and beyond the pure conditioning expression (2.14). With such an *invasive measurement*, one conditions a state after a disturbance induced by the measurement process has occurred; hence, one obtains a new state,

$$\langle \tilde{F} \rangle_y = \frac{\langle \mathcal{D}(y F) \rangle}{\langle \mathcal{D}(y) \rangle} = \frac{\sum_{x \in X} P(x) \sum_{x' \in X} D_x(y x') f(x')}{\sum_{x \in X} P(x) D_x(y)}, \quad (2.19)$$

which is a composition of the measurement disturbance (2.17) followed by the pure conditioning (2.14). We shall see that this sort of invasive measurement is inescapable in the quantum case.

2.4 Detectors and Probability Observables

For a single ideal experiment that answers questions of interest with perfectly correlated independent outcomes, knowing the spectrum of an observable for that experiment is completely sufficient. However, in many (if not most) cases the independent propositions corresponding to the experimental outcomes are only *imperfectly correlated* with the questions of interest about the system. Since in such a case one may not have direct access to the questions of interest, one also may not have direct access to the observables of interest. One must instead *infer* information about the observables of interest *indirectly* from the correlated outcomes of the detector to which one does have access.

Joint sample space.—To handle this case formally, we first enlarge the sample space to include both the sample space of interest, which we call the *system*, X and the accessible sample space, which we call the *detector*, Y . Questions about the system and the detector can be asked independently, so every question for

the system can be paired with any question from the detector; therefore, the resulting joint sample space must be a product space, $XY = \{xy \mid x \in X, y \in Y\}$, where the products of propositions from different sample spaces commute. The Boolean algebra Σ_{XY} and observable algebra $\Sigma_{XY}^{\mathbb{R}}$ are constructed in the usual way from the joint sample space, and contain the algebras Σ_X , Σ_Y , $\Sigma_X^{\mathbb{R}}$, and $\Sigma_Y^{\mathbb{R}}$ as subalgebras. When represented as operators on a Hilbert space, the corresponding joint representation exists within the tensor product of the system and detector space representations. Each sample space thus produces an irreducible *factor* for the observable algebra, exactly as discussed in Appendix A.4, where the total product of factors is still commutative.

Product states.—If the probabilities of the system propositions are *uncorrelated* with the probabilities of the detector propositions under a joint state P on the joint sample space, then the joint state can be written as a *composition* of independent states that are restricted to the sample spaces of the system and detector, $P = P_X \circ P_Y$. Just as the state P has a linear extension to $\langle \cdot \rangle$, its restrictions P_X and P_Y have linear extensions $\langle \cdot \rangle_X$ and $\langle \cdot \rangle_Y$ to their corresponding factors. Thus, for any joint observable F an uncorrelated expectation has the form $\langle F \rangle = \langle \langle F \rangle_Y \rangle_X = \langle \langle F \rangle_X \rangle_Y$. Such an uncorrelated joint state is known as a *product state*. The name stems from the fact that for a simple product $F_X F_Y$ of system and detector observables the corresponding joint expectation decouples into a product of system and detector expectations separately, $\langle F_X F_Y \rangle = \langle F_X \rangle_X \langle F_Y \rangle_Y$.

Similarly, general measures on the joint sample space can be product measures. Notably, the trace $\text{Tr} = \text{Tr}_X \circ \text{Tr}_Y$ on XY factors into *partial traces*, Tr_X and Tr_Y . On continuous spaces the Lebesgue-Stieltjes measure also factors, $\langle F \rangle = \int_X [\int_Y f(x, y) p(x, y) dy] dx = \int_Y [\int_X f(x, y) p(x, y) dx] dy$.

Correlated states.—In addition to product states, the joint space admits a much larger class of *correlated* states where the detector and system questions are dependent on one another. With such a correlated state a measurement on

the detector cannot be decoupled in general from a measurement on the system. Information gathered from a measurement on a detector under a correlated state will also indirectly provide information about the system, thus motivating the term “detector.”

Reduced states.—For a pure system observable F_X or a pure detector observable F_Y , the average under a joint state will be equivalent to the average under a state restricted to either the system or the detector space, known as a *reduced state*, or a *marginalized state*. We can define such a reduced state by using the joint state density under any reference *product* measure $\mu = \mu_X \circ \mu_Y$, such as the trace Tr . It then follows that,

$$\langle F_X \rangle = \left\langle \langle P_\mu \rangle_{\mu_Y} F_X \right\rangle_{\mu_X} = \langle P_{\mu_X} F_X \rangle_{\mu_X}, \quad (2.20a)$$

$$\langle F_Y \rangle = \left\langle \langle P_\mu \rangle_{\mu_X} F_Y \right\rangle_{\mu_Y} = \langle P_{\mu_Y} F_Y \rangle_{\mu_Y}. \quad (2.20b)$$

The quantities $P_{\mu_X} = \langle P_\mu \rangle_{\mu_Y}$ and $P_{\mu_Y} = \langle P_\mu \rangle_{\mu_X}$ are the *reduced state densities* that define the reduced states P_X and P_Y with expectation functionals,

$$\langle F_X \rangle_X = \langle P_{\mu_X} F_X \rangle_{\mu_X}, \quad (2.21a)$$

$$\langle F_Y \rangle_Y = \langle P_{\mu_Y} F_Y \rangle_{\mu_Y}. \quad (2.21b)$$

By definition, $\langle F_X \rangle = \langle F_X \rangle_X$ and $\langle F_Y \rangle = \langle F_Y \rangle_Y$. However, in general $\langle F \rangle \neq \langle \langle F \rangle_Y \rangle_X$ and $\langle F \rangle \neq \langle \langle F \rangle_X \rangle_Y$ unless P is a product state. The resulting reduced expectations $\langle \cdot \rangle_X$ and $\langle \cdot \rangle_Y$ are independent of the choice of reference product functional μ .

Probability observables.—Any correlation between the system and detector in the joint state allows us to directly relate propositions on the detector to *observables* on the system. We can compute the relationship directly by using a closure

relation and rearranging the conditioning procedure (2.14) to find,

$$P(y) = \sum_{x \in X} P(x)P(y|x) = \left\langle \sum_{x \in X} P(y|x) x \right\rangle = \langle E_y \rangle_X, \quad (2.22)$$

$$E_y = \sum_{x \in X} P(y|x) x. \quad (2.23)$$

The resulting set of system observables $\{E_y\}$ exactly correspond to the detector outcomes $\{y\}$. Analogously to a set of independent probability observables, they form a partition of the system identity, but are indexed by detector propositions rather than by system propositions, $\sum_{y \in Y} E_y = 1_X$. Such a set $\{E_y\}$ has the common name *positive operator-valued measure* (POVM) [22], since it forms a measure over the detector sample space Y consisting of positive operators when represented on a Hilbert space. However, we shall make an effort to refer to them as general *probability observables* to emphasize their physical significance. As long as the detector outcomes are not mutually exclusive with the system, the probability observables (2.23) will be a faithful representation of the reduced state of the detector in the observable space of the system.

Process tomography.—The probability observables are completely specified by the *conditional likelihoods* $P(y|x)$ for a detector proposition y to be true given that a system proposition x is true. Such conditional likelihoods are more commonly known as *response functions* for the detector and can be determined via independent *detector characterization* using known reduced system states; such characterization is also known as *detector tomography*, or *process tomography*.

State tomography.—Any good detector will maintain its characterization when it is coupled to any *unknown* reduced system state. That is, a noninvasive coupling of such a good detector to an unknown system produces a correlated joint state according to $P(xy) = P_X(x)P(y|x)$, where P_X is the unknown reduced system state prior to the interaction with the detector. This is useful precisely because

the unknown state can then be determined from the measured joint probabilities $P(xy)$ and the known response functions $P(y|x)$ as $P_X(x) = P(xy)/P(y|x)$, which is a procedure known as *state tomography*.

Generalized state collapse.—In addition to allowing the computation of detector probabilities, $P(y) = \langle E_y \rangle_X$, probability observables also have the dual role of updating the reduced system state following a measurement on the detector. To see this, we apply the general rule for state collapse (2.14) for a detector proposition y on the joint state to find,

$$\langle F_X \rangle_y = \frac{\langle y F_X \rangle}{P(y)} = \sum_{x \in X} f_X(x) P(y|x) \frac{P_X(x)}{P(y)} = \frac{\langle E_y F_X \rangle_X}{\langle E_y \rangle_X}, \quad (2.24)$$

which can be seen as a generalization of the Bayesian conditioning rule (2.14) to account for the effect of an imperfectly correlated detector, and can also be understood as a form of *Jeffrey's conditioning* [84]. For this reason, probability observables are commonly called *effects* of the *generalized measurement*. A reduced state density P_{μ_X} for the system updates as $P_{\mu_X|y} = P_{\mu_X} E_y / \langle E_y \rangle_X$. Such a *generalized measurement* is nonprojective, so is not constrained to the disjoint questions on the sample space of the system. As a result, it answers questions on the system space *ambiguously* or *noisily*.

Weak measurement.—The extreme case of such an ambiguous measurement is a *weak measurement*, which is a measurement that does not (appreciably) collapse the system state. Such a measurement is inherently ambiguous to the extent that only a minuscule amount of information is learned about the system with each detection. Formally, the probability observables for a weak measurement are all nearly proportional to the identity on the system space. Typically, an experimenter has access to some control parameter ϵ (such as the correlation

strength) that can alter the weakness of the measurement such that,

$$\forall y, \lim_{\epsilon \rightarrow 0} E_y(\epsilon) = P_Y(y)1_X, \quad (2.25)$$

where $P_Y(y) \in (0, 1)$ is the nonzero probability of obtaining the detector outcome y in the absence of any interaction with the system. Then for small values of ϵ the measurement leaves the system state nearly unperturbed, $P_{\mu_X|y} = P_{\mu_X} E_y(\epsilon) / \langle E_y(\epsilon) \rangle_X \approx P_{\mu_X}$. The limit as such a control parameter $\epsilon \rightarrow 0$ is known as the *weak measurement limit* and is a formal idealization not strictly achievable in an experiment.

Strong measurement.—The opposite extreme case is a *strong measurement* or projective measurement, which is a measurement for which all outcomes are independent, as in (2.3). The projective collapse rule (2.14) can therefore be seen as a special case of the general collapse rule (2.24) from this point of view.

Measurement sequences.—A further benefit of the probability observable representation of a detector is that it becomes straightforward to discuss sequences of generalized measurements performed on the same system. For example, consider two detectors that successively couple to a system and have the outcomes y and z measured, respectively. To describe the full joint state of the system and both detectors requires a considerably enlarged sample space. However, if the detectors are characterized by two sets of probability observables $\{E_y\}$ and $\{E'_z\}$ we can immediately write down the probability of both outcomes to occur as well as the resulting final collapsed system state without using the enlarged sample space,

$$P(yz) = \langle E'_z E_y \rangle_X, \quad (2.26a)$$

$$\langle F_X \rangle_{yz} = \frac{\langle E'_z E_y F_X \rangle_X}{\langle E'_z E_y \rangle_X}. \quad (2.26b)$$

Similarly, a conditioned density takes the form $P_{\mu_X|yz} = P_{\mu_X} E'_z E_y / \langle E'_z E_y \rangle_X$. The

detectors have been *abstracted* away to leave only their effect upon the system of interest.

Generalized invasive measurement.—The preceding discussion holds provided that the detector can be noninvasively coupled to a reduced system state P_X to produce a joint state $P(xy) = P_X(x)P(y|x)$. However, more generally the process of coupling a reduced detector state P_Y to the reduced system state P_X will *disturb* both states as discussed for (2.17). The disturbance produces a joint state from the original product state of the system and detector according to,

$$\langle \widetilde{xy} \rangle = \langle \langle \mathcal{D}(xy) \rangle_Y \rangle_X, \quad (2.27)$$

$$\mathcal{D}(xy) = \sum_{x' \in X} \sum_{y' \in Y} D_{x',y'}(xy) x'y', \quad (2.28)$$

where $D_{x',y'}$ are states specifying the joint transition probabilities for the disturbance. The noninvasive coupling $P(xy) = P_X(x)P(y|x)$ is a special case of this where the reduced system state is unchanged by the coupling.

As a result, we must slightly modify the derivation of the probability observables (2.22) to properly include the disturbance,

$$\langle \widetilde{y} \rangle = \langle \langle \mathcal{D}(y) \rangle_Y \rangle_X = \langle \widetilde{E}_y \rangle_X, \quad (2.29a)$$

$$\widetilde{E}_y = \langle \mathcal{D}(y) \rangle_Y = \sum_{x \in X} \sum_{y' \in Y} P_Y(y') D_{x,y'}(y) x. \quad (2.29b)$$

The modified probability observable \widetilde{E}_y includes both the initial detector state P_Y and the disturbance from the measurement. Detector tomography will therefore find the effective characterization probabilities $\widetilde{P}(y|x) = \sum_{y' \in Y} D_{x,y'}(y) P_Y(y')$.

The generalized collapse rule similarly must be modified to include the distur-

bance,

$$\left\langle \widetilde{F}_X \right\rangle_y = \frac{\langle \langle \mathcal{D}(y F_X) \rangle_Y \rangle_X}{\langle \langle \mathcal{D}(y) \rangle_Y \rangle_X} = \frac{\langle \mathcal{E}_y(F_X) \rangle_X}{\left\langle \widetilde{E}_y \right\rangle_X}, \quad (2.30)$$

$$\mathcal{E}_y(F_X) = \langle \mathcal{D}(y F_X) \rangle_Y = \sum_{x' \in X} x' \sum_{y' \in Y} P_Y(y') \sum_{x \in X} D_{x', y'}(y x) f(x). \quad (2.31)$$

Surprisingly, we can no longer write the conditioning in terms of just the probability observables \widetilde{E}_y ; instead we must use an *operation* \mathcal{E}_y that takes into account both the coupling of the detector and the disturbance of the measurement in an active way. The measurement operation is related to the effective probability observable according to, $\mathcal{E}_y(1_X) = \widetilde{E}_y$.

The change from observables to operations when the disturbance is included becomes particularly important for a sequence of invasive measurements. Consider an initial system state P_X that is first coupled to a detector state P_Y via a disturbance \mathcal{D}_1 , then conditioned on the detector proposition y , then coupled to a second detector state P_Z via a disturbance \mathcal{D}_2 , and finally conditioned on the detector proposition z . The joint probability for obtaining the ordered sequence (y, z) can be written as

$$\langle \langle \mathcal{D}_1(y \langle \mathcal{D}_2(z) \rangle_Z) \rangle_Y \rangle_X = \left\langle \mathcal{E}_y(\widetilde{E}'_z) \right\rangle_X. \quad (2.32)$$

The effective probability observable $\mathcal{E}_y(\mathcal{E}'_z(1_X)) = \mathcal{E}_y(\widetilde{E}'_z)$ for the ordered measurement sequence (y, z) is no longer a simple product of the probability observables \widetilde{E}_y and \widetilde{E}'_z as in (2.26a), but is instead an ordered *composition of operations*.

The ordering of operations also leads to a new form of *postselected* conditioning. Specifically, if we condition only on the second measurement of z in an invasive

sequence (y, z) , we obtain,

$${}_z\langle\tilde{y}\rangle = \frac{\langle\mathcal{E}_y(\tilde{E}'_z)\rangle_X}{\sum_{y'\in Y}\langle\mathcal{E}_{y'}(\tilde{E}'_z)\rangle_X} = \frac{\langle\mathcal{E}_y(\tilde{E}'_z)\rangle_X}{\langle\mathcal{E}(\tilde{E}'_z)\rangle_X}, \quad (2.33)$$

$$\mathcal{E}(\tilde{E}'_z) = \sum_{y'\in Y}\mathcal{E}_{y'}(\tilde{E}'_z) = \langle\mathcal{D}(\tilde{E}'_z)\rangle_Y. \quad (2.34)$$

The different position of the subscript serves to distinguish the postselected probability ${}_z\langle\tilde{y}\rangle$ from the preselected probability $\langle\tilde{y}\rangle_z = \langle\mathcal{E}'_z(\tilde{E}_y)\rangle_X / \langle\tilde{E}'_z\rangle_X$ corresponding to the reverse measurement ordering of (z, y) . The operation \mathcal{E} appearing in the denominator is called a *nonselective measurement* since it includes the disturbance induced by the measurement coupling, but does not condition on any particular detector outcome. When the disturbance to the reduced system state vanishes, the conditioning becomes order-independent and both types of conditional probability reduce to $P(y|z) = \langle E_y E'_z \rangle_X / \langle E'_z \rangle_X$.

The two forms of conditioning for invasive measurements in turn lead to a modified form of Bayes' rule that relates the preselected conditioning of a sequence to the postselected conditioning of the same sequence,

$${}_z\langle\tilde{y}\rangle = \langle\tilde{z}\rangle_y \frac{\langle E_y \rangle_X}{\langle\mathcal{E}(\tilde{E}'_z)\rangle_X}. \quad (2.35)$$

When the disturbance to the reduced system state vanishes, the nonselective measurement \mathcal{E} reduces to the identity operation, ${}_z\langle\tilde{y}\rangle$ reduces to $P(y|z)$, $\langle\tilde{z}\rangle_y$ reduces to $P(z|y)$, and we correctly recover the noninvasive Bayes' rule (2.16).

2.5 Contextual Values

Observable correspondence.—With the preliminaries about generalized state conditioning out of the way, we are now in a position to discuss the measurement

of observables in more detail. First we observe an important corollary of the observable representation of the detector probabilities $P(y) = \langle E_y \rangle_X$ from (2.22): *detector* observables can be mapped into equivalent *system* observables,

$$\langle F_Y \rangle = \sum_{y \in Y} f_Y(y) P(y) = \langle F_X \rangle_X, \quad (2.36)$$

$$F_X = \sum_{y \in Y} f_Y(y) E_y. \quad (2.37)$$

Note that the eigenvalues $f_X(x) = \sum_{y \in Y} f_Y(y) P(y|x)$ of the equivalent system observable F_X are not the same as the eigenvalues $f_Y(y)$ of the original detector observable F_Y , but are instead their average under the detector response. If the system propositions were accessible then the system observable F_X would allow nontrivial inference about the detector observable F_Y , provided that the probability observables were nonzero for all y in the support of F_Y .

Contextual values.—A more useful corollary of the expansion (2.37) is that any *system* observable that can be expressed as a combination of probability observables may be equivalently expressed as a *detector* observable,

$$F_X = \sum_{y \in Y} f_Y(y) E_y \implies F_Y = \sum_{y \in Y} f_Y(y) y, \quad (2.38)$$

which is the classical form of our main result. Using this equivalence, *we can indirectly measure such system observables using only the detector.* We dub the eigenvalues of the detector observable $f_Y(y)$ the **contextual values** (CVs) of the system observable F_X under the *context of the specific detector* characterized by a specific set of probability observables $\{E_y\}$. The CVs form a *generalized spectrum* for the observable since they are associated with general probability observables for a generalized measurement and not independent probability observables for a projective measurement; the eigenvalues are a special case when the probability observables are the spectral projections of the observable being measured.

With this point of view, we can understand an observable as an *equivalence class* of possible measurement strategies for the same average information. That is, using appropriate pairings of probability observables and CVs, one can measure the same observable average in many different ways, $\langle F_X \rangle = \sum_{x \in X} f_X(x)P(x) = \sum_{y \in Y} f_Y(y) \langle E_y \rangle_X$. Each such expansion corresponds to a different experimental setup.

Moments.—Similarly, the n^{th} statistical moment of an observable can be measured in many different, yet equivalent, ways. For instance, the n^{th} moment of an observable F_X can be found from the expansion (2.38) as,

$$\langle (F_X)^n \rangle = \left\langle \left(\sum_{y \in Y} f_Y(y) E_y \right)^n \right\rangle_X = \sum_{y_1, \dots, y_n \in Y} f_Y(y_1) \cdots f_Y(y_n) \langle E_{y_1} \cdots E_{y_n} \rangle_X. \quad (2.39)$$

By examining the general collapse rule for measurement sequences (2.26a) we observe that the quantity $\langle E_{y_1} \cdots E_{y_n} \rangle_X$ must be the joint probability for a sequence (y_1, \dots, y_n) of n *noninvasive* measurements that couple the same detector to the system n times in succession. As we anticipated in Section 2.3, the multiple sums do not collapse in the general case, so the average in (2.39) is explicitly different from $\sum_{y \in Y} (f_Y(y))^n P(y) = \langle (F_Y)^n \rangle$, which is the n^{th} statistical moment of the *raw* detector results.

We conclude that, *for imperfectly correlated noninvasive detectors, we can perform measurement sequences to obtain the correct statistical moments of an observable using a particular set of CVs.* Only for unambiguous measurements with independent probability observables do such measurement sequences reduce to simple powers of the eigenvalues being averaged with single measurement probabilities. If a single measurement by the detector is done per trial, then only the statistical moments of the *detector* observable F_Y can be inferred from that set of CVs, as opposed to the true statistical moments of the inferred system observable

F_X .

We can, however, change the CVs to define new observables that correspond to powers of the original observable, such as $G_X = (F_X)^n = \sum_{y \in Y} g_Y(y) E_y$. These new observables can then be measured indirectly using the same experimental setup without the need for measurement sequences. The CVs $g_Y(y)$ for the n^{th} power of F_X will not be simple powers of the CVs $f_Y(y)$ for F_X unless the measurement is unambiguous.

Invasive measurements.—If the measurement is invasive, then the disturbance forces us to associate the CVs with the measurement *operations* $\{\mathcal{E}_y\}$ and not solely with their associated probability operators $\{\tilde{E}_y\}$ in order to properly handle measurement sequences as in (2.31). Specifically, we must define the *observable operation*,

$$\mathcal{F}_X = \sum_{y \in Y} f_Y(y) \mathcal{E}_y, \quad (2.40)$$

which produces the identity $\mathcal{F}_X(1_X) = \sum_{y \in Y} f_Y(y) \tilde{E}_y = F_X$ similar to (2.38).

Correlated sequences of invasive observable measurements can be obtained by composing the observable operations,

$$\langle (\mathcal{F}_X)^n(1_X) \rangle_X = \sum_{y_1, \dots, y_n} f_Y(y_1) \cdots f_Y(y_n) \left\langle \mathcal{E}_{y_1}(\mathcal{E}_{y_2}(\cdots(\tilde{E}_{y_n})\cdots)) \right\rangle_X. \quad (2.41)$$

Such an n -measurement sequence reduces to the n^{th} moment (2.39) when the disturbance vanishes.

If time evolution disturbance \mathcal{D}_t is inserted between different invasive observable measurements, then we obtain an invasive *correlation function* instead,

$$\left\langle F_X \widetilde{G_X}(t) \right\rangle = \langle \mathcal{F}_X(\mathcal{D}_t(\mathcal{G}_X(1_X))) \rangle_X. \quad (2.42)$$

When the observable measurements become noninvasive, then this correctly reduces to the noninvasive correlation function (2.18). Similarly, n -time invasive correlations can be defined with $n - 1$ time-evolution disturbances between the invasive observable measurements $\langle \mathcal{F}_1(\mathcal{D}_{t_1}(\mathcal{F}_2(\cdots \mathcal{D}_{t_{n-1}}(\mathcal{F}_n(1_X)) \cdots))) \rangle$.

Conditioned averages.—In addition to statistical moments of the observable, we can also use the CVs to construct principled *conditioned averages* of the observable. Recall that in the general case of an invasive measurement sequence we can condition the observable measurement in two distinct ways. If we condition on an outcome z before the measurement of F_X we obtain the *preselected conditioned average* $\langle \widetilde{F}_X \rangle_z$ defined in (2.30). On the other hand, if the invasive conditioning measurement of z happens after the invasive observable measurement then we must use the postselected conditional probabilities (2.33) to construct a *postselected conditioned average*,

$${}_z \langle \widetilde{F}_X \rangle = \sum_{y \in Y} f_Y(y) {}_z \langle \widetilde{y} \rangle = \frac{\sum_{y \in Y} f_Y(y) \langle \mathcal{E}_y(\tilde{E}'_z) \rangle_X}{\sum_{y \in Y} \langle \mathcal{E}_y(\tilde{E}'_z) \rangle_X} = \frac{\langle \mathcal{F}_X(\tilde{E}'_z) \rangle_X}{\langle \mathcal{E}(\tilde{E}'_z) \rangle_X}. \quad (2.43)$$

The observable operation \mathcal{F}_X and the nonselective measurement \mathcal{E} encode the relevant details from the first measurement. When the disturbance to the reduced system state vanishes, both the preselected and the postselected conditioned averages simplify to the pure conditioned average $\langle F_X \rangle_z$ defined in (2.24) that depends only on the system observable F_X .

While the pure conditioned average $\langle F_X \rangle_z$ is independent of the order of conditioning and is always constrained to the eigenvalue range of the observable, the postselected invasive conditioned average ${}_z \langle \widetilde{F}_X \rangle$ can, perhaps surprisingly, stray outside the eigenvalue range with ambiguous measurements. The combination of the amplified CVs and the disturbance can lead to a postselected average that lies anywhere in the full CV range, rather than just the eigenvalue range. We will see an example of this in Sec. 2.5.2.

Inversion.—So far we have treated the CVs in the expansion (2.38) as known quantities. However, for a realistic detector situation, the CVs will need to be experimentally determined from the characterization of the detector and the observable that one wishes to measure. The reduced system state P_X will generally not be known *a priori*, since the point of a detector is to learn information about the system in the *absence* of such prior knowledge. We can still solve for the CVs without knowledge of the system state, however, since the probability observables are only specified by the conditional likelihoods $P(y|x)$ that can be obtained independently from detector tomography.

To solve for the CVs when the system state is presumed unknown, we rewrite (2.38) in the form,

$$F_X = \sum_{x \in X} x \sum_{y \in Y} P(y|x) f_Y(y) = \sum_{x \in X} x \langle F_Y \rangle_x = \mathcal{S}(F_Y), \quad (2.44)$$

where $\mathcal{S} = \sum_x x \langle \cdot \rangle_x$ is the map that converts observables in the detector space to observables in the system space $\mathcal{S} : \Sigma_Y^{\mathbb{R}} \rightarrow \Sigma_X^{\mathbb{R}}$. Our goal is to invert this map and solve for the required spectrum of F_Y given a desired system observable F_X . However, the inverse of such a map is not generally unique; for it to be uniquely invertible it must be one-to-one between system and detector spaces of equal size. If the detector space is smaller than the system, then no exact inverse solutions are possible; it may be possible, however, to find course-grained solutions that lose some information. Perhaps more alarmingly, if the detector space is larger than the system, then it is possible to have an infinite set of exact solutions.

When disturbance is taken into account as in (2.29), the equality (2.44) becomes,

$$F_X = \langle \mathcal{D}(F_Y) \rangle_Y = \mathcal{S}(F_Y), \quad (2.45)$$

so the composition of the disturbance \mathcal{D} and the detector expectation $\langle \cdot \rangle_Y$ produces the map \mathcal{S} that must be inverted. Equation (2.44) is a special case when the reduced system state is unchanged by the coupling disturbance.

Pseudoinversion.—The entire set of possible solutions to (2.45) may be completely specified using the *Moore-Penrose pseudoinverse* of the map \mathcal{S} , which we denote as \mathcal{S}^+ . The pseudoinverse is the inverse of the restriction of \mathcal{S} to the space $\Sigma_Y^{\mathbb{R}} \setminus \{F \in \Sigma_Y^{\mathbb{R}} \mid \mathcal{S}(F) = 0\}$; that is, the null space of \mathcal{S} is removed from the detector space before constructing the inverse. We will show a practical method for computing the pseudoinverse using the singular value decomposition in the examples to follow.

Using the pseudoinverse, all possible solutions of (2.45) can be written compactly as,

$$F_Y = \mathcal{S}^+(F_X) + (\mathcal{J} - \mathcal{S}^+\mathcal{S})(G), \quad (2.46)$$

where \mathcal{J} is the identity map and $G \in \Sigma_Y^{\mathbb{R}}$ is an arbitrary detector observable. The solutions specified by the pseudoinverse in this manner contain exact inverses and course-grainings as special cases.

Detector variance.—Since $(\mathcal{J} - \mathcal{S}^+\mathcal{S})$ is a projection operation to the null space of \mathcal{S} , the second term of (2.46) lives in the null space of \mathcal{S} and is orthogonal to the first term. Therefore, the norm squared of F_Y has the form,

$$\|F_Y\|^2 = \sum_y (f_Y(y))^2 = \|\mathcal{S}^+(F_X)\|^2 + \|(\mathcal{J} - \mathcal{S}^+\mathcal{S})(G)\|^2, \quad (2.47)$$

making the $G = 0$ solution have the smallest norm.

The norm $\|F_Y\|$ of the CV solution is relevant because the second moment of the detector observable F_Y is simply bounded by the norm squared $\langle (F_Y)^2 \rangle = \sum_y P(y)(f_Y(y))^2 \leq \|F_Y\|^2$. The second moment is similarly an upper bound for

the variance of the detector observable $\text{Var}(F_Y) = \langle (F_Y)^2 \rangle - (\langle F_Y \rangle)^2 \leq \langle (F_Y)^2 \rangle$. Therefore, the norm squared is a reasonable upper bound for the detector variance that one can make without prior knowledge of the state.

Mean-squared error.—The variance of F_Y governs the mean-squared error of any estimation of its average with a finite sample, such as an empirically measured sample in a laboratory. Specifically, one measures a sequence of detector outcomes of length n , (y_1, y_2, \dots, y_n) , and uses this finite sequence to estimate the average of F_Y via the *unbiased estimator*,

$$\overline{F_Y} = \frac{1}{n} \sum_i^n f_Y(y_i), \quad (2.48)$$

that converges to the true mean value $\langle F_Y \rangle_Y = \langle F_X \rangle$ as $n \rightarrow \infty$. The mean squared error of this estimator $\text{MSE}(\overline{F_Y})$ from the true mean is the variance over the number of trials in the sequence $\text{Var}(F_Y)/n$. Hence, the maximum mean squared error for a finite sequence of length n must be bounded by the norm squared of the CVs divided by length of the sequence,

$$\text{MSE}(\overline{F_Y}) = \frac{\text{Var}(F_Y)}{n} \leq \frac{\|F_Y\|^2}{n}. \quad (2.49)$$

That is, the norm bounds the number of trials necessary to obtain an experimental estimation of observable averages to a desired precision using the imperfect detector.

Pseudoinverse prescription.—Choosing the arbitrary observable to be $G = 0$ therefore not only picks the solution $F_Y = \mathcal{S}^+(F_X)$ that is uniquely related to F_X by discarding the irrelevant null space of \mathcal{S} , but also picks the solution with the smallest norm, which places a reasonable upper bound on the statistical error. Without prior knowledge of the system state, the pseudoinverse solution does a reasonable job at obtaining an optimal fit to the relation (2.45). Moreover,

when (2.45) is not satisfied by the direct pseudoinverse then an exact solution is impossible, but the pseudoinverse still gives the “best fit” coursegraining of an exact solution in the least-squares sense. As such, we consider the direct pseudoinverse of F_X to be the preferred solution in the absence of other motivating factors stemming from prior knowledge of the state being measured.

2.5.1 Example: Ambiguous Marble Detector

As an illustrative example similar to the one given in the introduction, suppose that one wishes to know whether the color of a marble is green or red, but one is unable to examine the marble directly. Instead, one only has a machine that can display a blue light or a yellow light after it examines the marble color. In such a case, the marble colors are the propositions of interest, but the machine lights are the only accessible propositions. The lights may be correlated imperfectly with the marble color; for instance, if a blue light is displayed one may learn something about the possible marble color, but it may still be partially *ambiguous* whether the marble is actually green or actually red.

The relevant Boolean algebra for the system is $\Sigma_X = \{0, g, r, 1_X\}$, where g is the proposition for the color green, r is the proposition for the color red, and $1_X = g + r$ is the logical *or* of the two possible color propositions. We consider the task of measuring a simple color observable $F_X = (+1)g + (-1)r$ that distinguishes the colors with a sign using an imperfectly correlated detector.

The relevant Boolean algebra for the detector is $\Sigma_Y = \{0, b, y, 1_Y\}$, where b is the proposition for the blue light, y is the proposition for the yellow light, and $1_Y = b + y$. In order to measure the marble observable F_X using only the detector, the experimenter must determine the proper form of the corresponding detector observable F_Y .

First, the experimenter characterizes the detector by sending in known samples

and observing the outputs of the detector. After many characterization trials, the experimenter determines to some acceptable precision the four conditional probabilities,

$$P(b|g) = 0.6, \quad P(y|g) = 0.4, \quad (2.50a)$$

$$P(b|r) = 0.2, \quad P(y|r) = 0.8, \quad (2.50b)$$

for the detector outcomes b and y given specific marble preparations g and r . These characterization probabilities completely determine the detector response in the form of its *probability observables* (2.23),

$$E_b = P(b|g)g + P(b|r)r, \quad (2.51a)$$

$$E_y = P(y|g)g + P(y|r)r. \quad (2.51b)$$

By construction, $E_b + E_y = g + r = 1_X$.

Second, the experimenter expands the system observable F_X using the detector probability observables (2.51) and unknown *contextual values* (CVs) $f_Y(b)$ and $f_Y(y)$ (2.38),

$$F_X = (+1)g + (-1)r = f_Y(b)E_b + f_Y(y)E_y. \quad (2.52)$$

After expressing this relation as the equivalent matrix equation,

$$\begin{pmatrix} +1 \\ -1 \end{pmatrix} = \begin{pmatrix} P(b|g) & P(y|g) \\ P(b|r) & P(y|r) \end{pmatrix} \begin{pmatrix} f_Y(b) \\ f_Y(y) \end{pmatrix}, \quad (2.53)$$

it can be directly inverted to find the CVs (2.46),

$$f_Y(b) = 3, \quad f_Y(y) = -2. \quad (2.54)$$

Therefore,

$$F_X = (+1)g + (-1)r = (3)E_b + (-2)E_y, \quad (2.55)$$

so F_X can be inferred from a measurement of the equivalent detector observable $F_Y = (3)b + (-2)y$.

Notably, the CVs (2.54) are amplified from the eigenvalues of ± 1 due to the *ambiguity* of the detector. The amplification compensates for the ambiguity so that the correct *average* can be obtained after measuring an ensemble of many unknown marbles described by the initial marble state P_X . The amplification also leads to a larger upper bound for the variance (2.47) of the detector,

$$\|F_Y\|^2 = 13. \quad (2.56)$$

Hence, we can expect the imperfect detector to display a root-mean-squared (RMS) error (2.49) in the reported average color that is no larger than $\sqrt{13/n} \approx 3.6/\sqrt{n}$ after n repeated measurements. For contrast, a perfect detector would display an RMS error no larger than $\sqrt{2/n} \approx 1.4/\sqrt{n}$ after n repeated measurements.

2.5.2 Example: Invasive Ambiguous Detector

In addition to being ambiguous, the marble color detection apparatus in the last example Section 2.5.1 could be generally *invasive*. That is, the act of making a measurement of the marble color could probabilistically change the color of the marble. In such a case, the characterization probabilities (2.50) composing the probability observables (2.51) would be a combination of the initial state of the detector lights P_Y and a *disturbance* \mathcal{D} from the measurement coupling according

to (2.29),

$$\tilde{P}(b|g) = P_Y(b)(D_{g,b}(gb) + D_{g,b}(rb)) + P_Y(y)(D_{g,y}(gb) + D_{g,y}(ry)), \quad (2.57a)$$

$$\tilde{P}(y|g) = P_Y(b)(D_{g,b}(gy) + D_{g,b}(ry)) + P_Y(y)(D_{g,y}(gy) + D_{g,y}(ry)), \quad (2.57b)$$

$$\tilde{P}(b|r) = P_Y(b)(D_{r,b}(gb) + D_{r,b}(rb)) + P_Y(y)(D_{r,y}(gb) + D_{r,y}(ry)), \quad (2.57c)$$

$$\tilde{P}(y|r) = P_Y(b)(D_{r,b}(gy) + D_{r,b}(ry)) + P_Y(y)(D_{r,y}(gy) + D_{r,y}(ry)), \quad (2.57d)$$

where we have used the marginalization identity $D_{c,d}(b) = D_{c,d}(gb) + D_{c,d}(rb)$ for $c \in \{g, r\}$ and $d \in \{b, y\}$. For a noninvasive detector, the transition probabilities that involve marbles changing color must be zero $D_{g,b}(rb) = D_{g,b}(ry) = D_{g,y}(ry) = D_{g,y}(rb) = D_{r,b}(gb) = D_{r,b}(gy) = D_{r,y}(gb) = D_{r,y}(gy) = 0$. However, they need not be zero for a general invasive detector.

As an example, suppose that the initial detector state is unbiased, $P_Y(b) = P_Y(y) = 1/2$, and that the detector has a 10% chance of flipping the color of a given marble. The following possible values for the *sixteen* transition probabilities would then lead to the same effective characterization probabilities (2.50) as before,

$$D_{g,b}(gb) = 0.5 \qquad D_{g,y}(gb) = 0.5, \quad (2.58a)$$

$$D_{g,b}(gy) = 0.3 \qquad D_{g,y}(gy) = 0.3, \quad (2.58b)$$

$$D_{r,b}(rb) = 0.1 \qquad D_{r,y}(rb) = 0.1, \quad (2.58c)$$

$$D_{r,b}(ry) = 0.7 \qquad D_{r,y}(ry) = 0.7, \quad (2.58d)$$

$$D_{g,b}(rb) = 0.1 \qquad D_{g,y}(rb) = 0.1, \quad (2.58e)$$

$$D_{g,b}(ry) = 0.1 \qquad D_{g,y}(ry) = 0.1, \quad (2.58f)$$

$$D_{r,b}(gb) = 0.1 \qquad D_{r,y}(gb) = 0.1, \quad (2.58g)$$

$$D_{r,b}(gy) = 0.1 \qquad D_{r,y}(gy) = 0.1. \quad (2.58h)$$

Since the effective characterization probabilities are the same, the probability

observables are the same as (2.51), leading to the same CVs as (2.54) to measure the observable $F_X = (+1)g + (-1)r$. However, the number of probabilities needed to completely characterize the measurement being made to account for disturbance has quadrupled from four to sixteen.

The disturbance of the reduced marble state will become apparent only when making a second measurement after the first one. Suppose we make a second measurement of the marble colors g and r directly. The probability of obtaining a detector outcome $d \in \{b, y\}$ and then observing a specific marble color $c \in \{g, r\}$ will then be $P_X(g)(P_Y(b)D_{g,b}(cd) + P_Y(y)D_{g,y}(cd)) + P_X(r)(P_Y(b)D_{r,b}(cd) + P_Y(y)D_{r,y}(cd))$. If we define an *operation* as in (2.31) to be,

$$\begin{aligned} \mathcal{E}_d(c) &= \langle \mathcal{D}(cd) \rangle_Y, \\ &= g(P_Y(b)D_{g,b}(cd) + P_Y(y)D_{g,y}(cd)) + r(P_Y(b)D_{r,b}(cd) + P_Y(y)D_{r,y}(cd)), \end{aligned} \quad (2.59)$$

then we can express the probability for the sequence compactly as $\langle \mathcal{E}_d(c) \rangle_X$.

Averaging the outcomes for the detector lights using the CVs (2.54) and then conditioning on a particular marble color c in the second measurement produces a *postselected conditioned average* of the marble colors (2.43) as reported by the invasive ambiguous detector,

$$\left\langle \widetilde{F}_X \right\rangle_c = \frac{f_Y(b) \langle \mathcal{E}_b(c) \rangle_X + f_Y(y) \langle \mathcal{E}_y(c) \rangle_X}{\langle \mathcal{E}_b(c) \rangle_X + \langle \mathcal{E}_y(c) \rangle_X}. \quad (2.60)$$

If we also preselect the marbles to be a particular color, we can compute the pre- and postselected conditioned averages of the marble colors as reported by the invasive ambiguous detector from (2.54), (2.58), and (2.60),

$${}_g \left\langle \widetilde{F}_X \right\rangle_g = 1.125, \quad (2.61a)$$

$${}_r \left\langle \widetilde{F}_X \right\rangle_g = 0.5, \quad (2.61b)$$

$$\left\langle \widetilde{F}_X \right\rangle_g = 0.5, \quad (2.61c)$$

$$\left\langle \widetilde{F}_X \right\rangle_r = -1.375. \quad (2.61d)$$

Due to a combination of the invasiveness and the ambiguity of the measurement, the postselected conditioned averages can stray outside the eigenvalue range $[-1, 1]$ for the observable F_X . However, they remain within the CV range $[-2, 3]$. When the measurement is noninvasive, then the pre- and postselected conditioned averages in (2.61) that remain well-defined reduce to the pure conditioned averages $\langle F_X \rangle_g = 1$ and $\langle F_X \rangle_r = -1$.

2.5.3 Example: Redundant Ambiguous Detector

As another variation, consider a similar marble detection setup similar to the example in Section 2.5.1, but where the detector apparatus has not two, but *three* independent outcome lights: blue, yellow, and purple. Hence, the detector Boolean algebra is $\Sigma_Y = \{0, b, y, p, b + y, b + p, y + p, 1_Y\}$, where p is the new proposition for the purple light, and $1_Y = b + y + p$. After characterizing the detector the experimenter finds the conditional probabilities,

$$P(b|g) = 0.5, \quad P(y|g) = 0.3, \quad P(p|g) = 0.2, \quad (2.62a)$$

$$P(b|r) = 0.1, \quad P(y|r) = 0.7, \quad P(p|r) = 0.2, \quad (2.62b)$$

that define the probability observables,

$$E_b = P(b|g)g + P(b|r)r, \quad (2.63a)$$

$$E_y = P(y|g)g + P(y|r)r, \quad (2.63b)$$

$$E_p = P(p|g)g + P(p|r)r. \quad (2.63c)$$

By construction, $E_b + E_y + E_p = 1_X$. Furthermore, $E_p = (0.2)1_X$, so the purple outcome cannot distinguish whether the marble is green or red and can be imagined as a generic detector malfunction outcome.

The experimenter now has a choice for how to assign CVs to a detector observable F_Y in order to infer the marble observable $F_X = (+1)g + (-1)r$. A simple choice is to ignore the redundant (and nondistinguishing) purple outcome by zeroing out its CV $f_Y(p) = 0$, and then invert the remaining relationship analogously to (2.53) to find $f_Y(b) = 3.125$ and $f_Y(y) = -1.875$. The variance bound for this simple choice is $\|F_Y\|^2 = 13.2813$, leading to a root-mean-squared error no larger than $\sqrt{13.2813/n} \approx 3.6/\sqrt{n}$ after n repeated measurements.

However, a better choice is to find the preferred values for all three outcomes using the pseudoinverse (2.46) of the map between F_Y and F_X . To do this, we write a matrix equation similar to (2.53) that uses all three outcomes,

$$\begin{pmatrix} +1 \\ -1 \end{pmatrix} = \mathcal{S} \begin{pmatrix} f_Y(b) \\ f_Y(y) \end{pmatrix}, \quad (2.64a)$$

$$\mathcal{S} = \begin{pmatrix} P(b|g) & P(y|g) & P(p|g) \\ P(b|r) & P(y|r) & P(p|r) \end{pmatrix}. \quad (2.64b)$$

The pseudoinverse \mathcal{S}^+ can be constructed by using the singular value decomposition, $\mathcal{S} = \mathcal{U}\Sigma\mathcal{V}^T$, where \mathcal{U} is an orthogonal matrix composed of the normalized eigenvectors of $\mathcal{S}\mathcal{S}^T$, \mathcal{V} is an orthogonal matrix composed of the normalized eigenvectors of $\mathcal{S}^T\mathcal{S}$, and Σ is a diagonal matrix composed of the singular values of \mathcal{S} (which are the square roots of the eigenvalues of $\mathcal{S}\mathcal{S}^T$ and $\mathcal{S}^T\mathcal{S}$). After computing the singular value decomposition, the pseudoinverse can be constructed as $\mathcal{S}^+ = \mathcal{V}\Sigma^+\mathcal{U}^T$, where Σ^+ is the diagonal matrix constructed by inverting all nonzero elements of Σ^T . Performing this inversion we find the following preferred

CV,

$$S^+ = \frac{5}{36} \begin{pmatrix} 15 & -7 \\ -3 & 11 \\ 3 & 1 \end{pmatrix}, \quad (2.65a)$$

$$\begin{pmatrix} f_Y(b) \\ f_Y(y) \\ f_Y(p) \end{pmatrix} = S^+ \begin{pmatrix} +1 \\ -1 \end{pmatrix} = \frac{5}{18} \begin{pmatrix} 11 \\ -7 \\ 1 \end{pmatrix} = \begin{pmatrix} 3.0\bar{5} \\ -1.9\bar{4} \\ 0.2\bar{7} \end{pmatrix}. \quad (2.65b)$$

This preferred solution has the smallest variance bound of $\|F_Y\|^2 = 13.1944$.

We find (counterintuitively) that even though the purple outcome itself cannot distinguish the marble color, the fact that one obtains a purple outcome at all still provides some useful information to the experimenter due to the *asymmetry* of the blue and yellow outcomes. Indeed, if for the red marble we instead found the symmetric detector response $P(b|r) = 0.3$, $P(y|r) = 0.5$, and $P(p|r) = 0.2$, the pseudoinverse would produce the preferred CVs $f_Y(b) = 5$, $f_Y(y) = -5$, and $f_Y(p) = 0$, indicating that the purple outcome was truly noninformative.

A less principled approach to solving (2.64) would be for the experimenter to assign a completely arbitrary value to one outcome, like $f_Y(b) = B$. The CV relation still produces a matrix equation,

$$\begin{pmatrix} +1 - BP(b|g) \\ -1 - BP(b|r) \end{pmatrix} = \begin{pmatrix} P(y|g) & P(p|g) \\ P(y|r) & P(p|r) \end{pmatrix} \begin{pmatrix} f_Y(y) \\ f_Y(p) \end{pmatrix}, \quad (2.66)$$

that can be solved to find,

$$f_Y(y) = B - 5, \quad f_Y(p) = 12.5 - 4B. \quad (2.67)$$

The bound for the variance of this solution is $\|F_Y\|^2 = 18B^2 - 110B + 181.25 \geq 13.1944$; the value of B that minimizes the bound is $B = 3.0\bar{5}$, which recovers the

pseudoinverse solution.

Although picking an arbitrary solution gives mathematically equivalent results, the experimenter will only increase the norm of the solution without any physical motivation. As such, the higher moments of the detector observable F_Y can be correspondingly larger, and more trials may be necessary for the estimated average of the system observable F_X to reach the desired precision.

2.5.4 Example: Continuous Detector

Consider the extreme example of a marble color detector that has a continuum of outcomes, such as the position of impact of a marble on a continuous screen. In such a case, the detector sample space Y is indexed by a real parameter $y \in \mathbb{R}$, and the relevant Boolean algebra Σ_Y can be chosen to be the set of all Borel subsets of the real line [77, 78].

After characterizing the detector, the experimenter finds that the detector displaces its initial probability distribution $dP_Y(y) = p_Y(y) dy$ by an amount z from the zero-point according to which marble-color is sent into the detector,

$$dP(y|g) = dP_Y(y - z), \quad dP(y|r) = dP_Y(y + z). \quad (2.68)$$

These probabilities define the probability observables,

$$dE(y) = g dP(y|g) + r dP(y|r), \quad (2.69)$$

such that $\int_{\mathbb{R}} dE(y) = 1_X$.

To infer information about the marble observable F_X using this detector, the experimenter must assign a continuum of CVs $f_Y(y)$ such that,

$$F_X = (+1)g + (-1)r = \int_{\mathbb{R}} f_Y(y) dE(y), \quad (2.70)$$

or in matrix form,

$$\begin{pmatrix} +1 \\ -1 \end{pmatrix} = \mathcal{S}[f_Y] = \begin{pmatrix} \int_{\mathbb{R}} f_Y(y) dP_Y(y - z) \\ \int_{\mathbb{R}} f_Y(y) dP_Y(y + z) \end{pmatrix}. \quad (2.71)$$

Since f_Y is a function, \mathcal{S} is a vector-valued functional, which is why we adopt the square-bracket notation.

In this case, the detector outcomes are overwhelmingly redundant. However, we can pick the least norm solution using the pseudoinverse of the map \mathcal{S} as before. To do so, we first calculate $\mathcal{S}\mathcal{S}^T$,

$$\mathcal{S}^T = \begin{pmatrix} p_Y(y - z) & p_Y(y + z) \end{pmatrix}, \quad (2.72a)$$

$$\mathcal{S}\mathcal{S}^T = \begin{pmatrix} a & b(z) \\ b(z) & a \end{pmatrix}, \quad (2.72b)$$

where,

$$a = \int_{\mathbb{R}} p_Y(y) dP_Y(y) = \int_{\mathbb{R}} p_Y^2(y) dy, \quad (2.73a)$$

$$b(z) = \int_{\mathbb{R}} p_Y(y + z)p_Y(y - z) dy, \quad (2.73b)$$

and we find its eigenvalues of $a + b(z)$ with corresponding normalized eigenvector $(1, 1)/\sqrt{2}$ and $a - b(z)$ with corresponding normalized eigenvector $(-1, 1)/\sqrt{2}$. We can then construct the orthogonal matrix \mathcal{U} composed of the normalized eigenvectors of $\mathcal{S}\mathcal{S}^T$ and the diagonal matrix Σ composed of the square roots of the

eigenvalues of $\mathcal{S}\mathcal{S}^T$,

$$\mathcal{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad (2.74)$$

$$\Sigma = \begin{pmatrix} \sqrt{a+b(z)} & 0 \\ 0 & \sqrt{a-b(z)} \end{pmatrix}. \quad (2.75)$$

Next we calculate the relevant eigenfunctions of $\mathcal{S}^T\mathcal{S}$ that correspond to the same nonzero eigenvalues $a \pm b(z)$ of $\mathcal{S}\mathcal{S}^T$; the remaining eigenfunctions belong to the nullspace of \mathcal{S} and do not contribute. Specifically, we have,

$$\begin{aligned} \mathcal{S}^T\mathcal{S}[h](y) &= p_Y(y-z) \int_{\mathbb{R}} h(y) dP_Y(y-z) \\ &\quad + p_Y(y+z) \int_{\mathbb{R}} h(y) dP_Y(y+z), \end{aligned} \quad (2.76)$$

where h is an arbitrary function. Then the equations,

$$\mathcal{S}^T\mathcal{S}[v_+](y) = (a+b(z))v_+(y), \quad (2.77a)$$

$$\mathcal{S}^T\mathcal{S}[v_-](y) = (a-b(z))v_-(y), \quad (2.77b)$$

define the normalized eigenfunctions,

$$v_+(y) = \frac{p_Y(y-z) + p_Y(y+z)}{\sqrt{2(a+b(z))}}, \quad (2.78a)$$

$$v_-(y) = -\frac{p_Y(y-z) - p_Y(y+z)}{\sqrt{2(a-b(z))}}, \quad (2.78b)$$

which allows us to construct the relevant part of the orthogonal map \mathcal{V}^T ,

$$\mathcal{V}^T[h] = \left(\int v_+(y)h(y) dy \quad \int v_-(y)h(y) dy \right), \quad (2.79)$$

completing the nonzero part of the singular value decomposition of $\mathcal{S} = \mathcal{U}\Sigma\mathcal{V}^T$.

Finally, we construct the pseudoinverse,

$$\begin{aligned} \mathcal{S}^+ &= \mathcal{V}\Sigma^+\mathcal{U}^T, \\ &= \left(\frac{v_+(y)}{\sqrt{2(a+b(z))}} - \frac{v_-(y)}{\sqrt{2(a-b(z))}} \quad \frac{v_+(y)}{\sqrt{2(a+b(z))}} + \frac{v_-(y)}{\sqrt{2(a-b(z))}} \right), \end{aligned} \quad (2.80)$$

and solve for the appropriate CV $f_Y(y)$ of the red/green observable F_X in (2.70),

$$f_Y(y) = \mathcal{S}^+ \begin{pmatrix} +1 \\ -1 \end{pmatrix} = \frac{p_Y(y-z) - p_Y(y+z)}{a-b(z)}, \quad (2.81)$$

where a and $b(z)$ are as defined in (2.73).

The pseudoinverse solution (2.81) contains only the physically relevant detector state density p_Y and provides direct physical intuition about the detection process. Namely, everything in the shifted distribution corresponding to the green marble $p_Y(y-z)$ is associated with the eigenvalue $+1$, while everything in the shifted distribution corresponding to the red marble $p_Y(y+z)$ is associated with the eigenvalue -1 . The overall amplification factor $a-b(z)$ indicates the discrepancy between the overlap of the shifted distributions and the distribution autocorrelation. The more the shifted distributions overlap, the more ambiguous the measurement will be, so the amplification factor makes the CVs larger to compensate. If the shifted distributions do not overlap, then $b(z) \rightarrow 0$ and the only amplification comes from the autocorrelation a that indicates the ambiguity of the intrinsic profile of the detector state. Moreover, the support of the CVs is equal to the support of both shifted detector distributions, which is physically satisfying.

The bound for the detector variance using the pseudoinverse solution is $\|f_Y\|^2 = 2/[a-b(z)]$, which depends solely on the amplification factor in the denominator. If the measurement is strong, such that $a-b(z) = 1$, then the variance bound reduces to the ideal variance bound of 2, as expected, leading to a maximum

RMS error of $\sqrt{2/n}$. Any additional ambiguity of the measurement stemming from distribution overlap or distributed autocorrelation amplifies the maximum RMS error by a factor of $\sqrt{1/[a - b(z)]}$.

Contrast these preferred values with the generic linear solution $f_Y(y) = y/z$, which also satisfies (2.70) when p_Y is symmetric about its mean [28, 34, 68]. While the generic solution could be argued to be simpler in form, it provides no information about the detector and provides no physical insight into the meaning or origin of the values themselves. It has nonzero support in areas where the detector has zero support and even gets progressively larger in regions that will not contribute to the average. Moreover, the bound for the detector variance diverges, indicating that the RMS error can in principle be unbounded. Hence, despite the mathematical equivalence, the linear solution is *physically* inferior as a solution when compared to the pseudoinverse (2.81).

3 Quantum Observable Measurement

We have to remember that what we observe is not nature herself, but nature exposed to our method of questioning.

Werner Heisenberg, (1958) [85]

To transition from the classical theory of observable measurement to the quantum theory we shall now treat the Boolean algebra $\Sigma_X^{\mathbb{R}}$ derived in Chapter 2 as a commutative subalgebra of the enveloping algebra $\mathcal{E}(\mathfrak{g})$ for a noncommutative Lie group \mathfrak{G} discussed in Appendix A. This approach serves to illustrate the myriad similarities between the quantum and classical probabilistic theories, while also highlighting their key differences. We shall see that the contextual-value formalism for indirect observable measurement is essentially unchanged, despite the modifications that must be made to the operational theory of measurement.

3.1 Sample Spaces and Observables

Quantum sample space.—The quantum theory of probability forms a superstructure on the classical theory of probability in the following sense: given a classical sample space X , the corresponding quantum sample space can be obtained as the orbit of X under the action of a simply connected Lie group. That is, the en-

the classical sample space X can be rotated to a different classical sample space $X' = \text{Ad}_U(X)$ under the action of some group element $U \in \mathfrak{G}$. We call each classical sample space generated in this fashion a *framework* to be consistent with other recent work [86]. The collection of all such continuously connected classical sample spaces is the quantum sample space, which we will notate as $\mathcal{Q}(X)$ to emphasize that it can be generated from X .

Representation.—If the sample space X is represented as a set of orthogonal rank-1 projections $\{|x\rangle\langle x|\}$ on a Hilbert space, the rotated sample space $X' = \text{Ad}_U(X)$ will be represented by a different set of orthogonal projections $\{\text{Ad}_U(|x\rangle\langle x|)\}$ on the same Hilbert space, as discussed in Appendix A.7. Any such group action Ad_U is a rotation in the Hilbert space and will have a spinor representation (see, e.g., [87–91]) as a two-sided product with a *unitary* rotor \hat{U} , such that $\hat{U}^\dagger\hat{U} = \hat{U}\hat{U}^\dagger = \hat{1}$, and $(\hat{U}^\dagger)^\dagger = \hat{U}$. The *involution* (\dagger) is the adjoint with respect to the inner product of the Hilbert space. While the projections $\{|x\rangle\langle x|\}$ correspond to subspaces spanned by vectors $\{|x\rangle\}$ in the Hilbert space, the rotated projections $\{U^\dagger|x\rangle\langle x|U\}$ correspond to subspaces spanned by rotated vectors $\{U^\dagger|x\rangle\}$. The representation of the quantum sample space $\mathcal{Q}(X)$ will therefore consist of all possible rank-1 projections on the complex Hilbert space in which the classical sample space X is represented.

Since the Hilbert space representation of a unitary rotor \hat{U} generally contains complex numbers in order to satisfy the group relations, the Hilbert space also becomes *complex*. However, it is important to note that the complex structure arises solely from the representation of the group and will not appear directly in any calculable quantity to follow¹. In what follows we shall tend to use the shorter algebraic notation x and adopt the equivalent Hilbert space notation $|x\rangle\langle x|$ as a projector onto a Hilbert space vector $|x\rangle$ only when it readily simplifies ex-

¹See [89–92] for discussion on purely real algebraic representations of the relativistic Dirac electron to drive home this point.

pressions. Similarly, we will omit the operator notation \hat{U} and \dagger in favor of the algebraic notation U and $*$ for the group elements. We will also abbreviate adjoint group actions Ad_U as the operations \mathcal{U} .

Quantum observables.—Each classical framework X has an associated Boolean algebra Σ_X and space of measurable observables $\Sigma_X^{\mathbb{R}}$ exactly as previously discussed. The space of measurable quantum observables is the collection of all measurable classical observables that are independently constructed in all the classical frameworks in $Q(X)$. We will denote this space as $\Sigma_{Q(X)}^{\mathbb{R}}$. Measurable quantum observables are therefore constructed entirely with real numbers that have empirical meaning for a laboratory setting; hence, their representations on a complex Hilbert space will be Hermitian operators.

For observables in the same framework $A, B \in \Sigma_X^{\mathbb{R}}$, we find that $\mathcal{U}(A)\mathcal{U}(B) = U^*AUU^*BU = U^*ABU = \mathcal{U}(AB)$, meaning that the group rotations preserve the algebraic product. As a corollary, all observables in $\Sigma_{Q(X)}^{\mathbb{R}}$ can be obtained by rotating observables constructed in a single framework $\Sigma_X^{\mathbb{R}}$; hence, our previous discussion of observables in Chapter 2 carries over to the quantum theory essentially unaltered. All measurable quantum observables are equivalent to classical observables in a particular framework.

Furthermore, the independence of the propositions in a framework X remains unaltered by unitary rotation, so every other framework X' has the same number of independent propositions. Thus, the number of independent propositions is an invariant known as the *quantum dimension*; for a representation it fixes the dimension of the Hilbert space. Similarly, the identity and zero observables are invariants, so are the same in every framework and unique in the quantum observable algebra. This feature is expected from the Lie group discussion in Appendix A.4.

Since each different framework forms a separate well-behaved classical sample space, the entire preceding discussion about classical probability theory applies

unaltered when restricted to a particular framework in the quantum theory. All observables constructed in a particular framework will commute with each other. We expect distinctly quantum features to appear only when comparing elements from different frameworks.

Noncommutativity.—The unitary rotations \mathcal{U} are generally *noncommutative* and so introduce noncommutativity into the quantum theory that is not present in the classical theory. Specifically, given $A, B \in \Sigma_X^{\mathbb{R}}$, $A' = \mathcal{U}(A)$, and $B' = \mathcal{V}(B)$, then $A'B' = U^\dagger AUV^\dagger BV \neq B'A'$, since U and V do not necessarily commute with each other or with A and B . That is, the noncommutativity of the observables stems directly from the noncommutativity of the Lie group. As a result, the Boolean algebras corresponding to different frameworks are *incompatible* with each other: propositions from one framework cannot form a Boolean logical AND with propositions from a different framework. We shall see in the next section, however, that the notion of *disturbance* followed by a logical AND *can* be generalized to the noncommutative setting in the form of the projection postulate.

Disturbance.—All nonconditioning disturbance operations \mathcal{D} in the quantum theory are postulated to be group rotation operations \mathcal{U} ; a classical disturbance (2.17) must then be a group rotation that happens to leave the framework invariant. Indeed, we shall see that the parallels between the quantum theory and the classical theory with disturbance are quite strong when one interprets all unitary rotations as a form of classical disturbance.

Time Evolution.—As an example, the continuous time-evolution of a closed quantum system is specified by a disturbance in the form of a one-parameter group rotation \mathcal{U}_t with corresponding rotor U_t , known as a *propagator*. This propagator is precisely the exponential map of a Hamiltonian generator discussed in Appendix A.2.

3.1.1 Example: Polarization

As an example quantum system we shall pick the simplest possible nontrivial system: a qubit. Specifically, we will consider the polarization degree of freedom of a laser beam. Suppose we are interested in measuring the linear polarization of the beam with respect to the surface of an optical table. We denote the polarization direction parallel to the table as “horizontal” (h) and the direction perpendicular to the table as “vertical” (v). Although we casually refer to the polarizations h and v as if they were properties of the light beam, the propositions h and v operationally refer to two independent outcomes of a polarization distinguishing device, such as a polarizing beam splitter, that can be implemented in the laboratory.

The two orthogonal polarizations form a classical sample space $X = \{h, v\}$ and a classical Boolean algebra $\Sigma_X = \{0, h, v, 1_X\}$, where $1_X = h + v$, similar to the classical sample space for the marble colors considered in Sec. 2.5.1. By extending the Boolean algebra over the reals to $\Sigma_X^{\mathbb{R}}$ as before we can define classical observables $F_X = ah + bv$ in this sample space, such as the Stokes observable $S_X = h - v$ that distinguishes the polarizations with a sign.

We can represent the commutative observable algebra $\Sigma_X^{\mathbb{R}}$ as diagonal 2×2 matrices,

$$h = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad v = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad F_X = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad (3.1)$$

which can also be understood as commuting Hermitian operators over a two-dimensional Hilbert space. The atomic propositions $h = |h\rangle\langle h|$ and $v = |v\rangle\langle v|$ are projectors that correspond to disjoint subspaces spanned by the orthonormal

Jones' polarization basis for the Hilbert space,

$$|h\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad |v\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (3.2)$$

To obtain the full quantum sample space $Q(X)$ from X , we introduce the group of possible polarization rotations. Algebraically, an arbitrary rotation $\mathcal{U}(F_X) = U^\dagger F_X U$ can be readily understood in terms of its rotor U , which is an element of the group $SU(2)$ and can be parametrized, for example, in terms of the Cartan decomposition $U_{\alpha,\beta,\gamma} = \exp(i\alpha\sigma_z/2) \exp(i\beta\sigma_y/2) \exp(i\gamma\sigma_z/2)$, which for a qubit happens to correspond to an Euler angle decomposition of a three-dimensional rotation². Here $i\sigma_z$ and $-i\sigma_y$ are two of the three generators of the Lie algebra for $SU(2)$ in terms of the standard Pauli matrices,

$$\sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \quad (3.3)$$

Since the group generators have been given a complex matrix representation, the unitary rotation $U_{\alpha,\beta,\gamma}$ will also have a complex matrix representation,

$$e^{i\frac{\alpha}{2}\sigma_z} = \begin{pmatrix} e^{i\frac{\alpha}{2}} & 0 \\ 0 & e^{-i\frac{\alpha}{2}} \end{pmatrix}, \quad (3.4a)$$

$$e^{i\frac{\beta}{2}\sigma_y} = \begin{pmatrix} \cos \frac{\beta}{2} & \sin \frac{\beta}{2} \\ -\sin \frac{\beta}{2} & \cos \frac{\beta}{2} \end{pmatrix}, \quad (3.4b)$$

$$U_{\alpha,\beta,\gamma} = \begin{pmatrix} e^{i(\alpha+\gamma)/2} \cos \frac{\beta}{2} & e^{i(\alpha-\gamma)/2} \sin \frac{\beta}{2} \\ -e^{-i(\alpha-\gamma)/2} \sin \frac{\beta}{2} & e^{-i(\alpha+\gamma)/2} \cos \frac{\beta}{2} \end{pmatrix}. \quad (3.4c)$$

The algebraic involution $U_{\alpha,\beta,\gamma}^\dagger$ is the complex transpose in the matrix represen-

²See Appendix A.8 for a discussion of how such a qubit arises from the constraints of space-time.

tation.

Physically, the factor $\exp(i\beta\sigma_y/2)$ corresponds to a rotation of the apparatus around the axis of the light beam by an angle $\beta/2$, while the factors $\exp(i\alpha\sigma_z/2)$ and $\exp(i\gamma\sigma_z/2)$ correspond to the action of phase plates that shift the relative phases of h and v by $\alpha/2$ and $\gamma/2$, respectively. Hence, the ubiquitous quantum phase also appears as a consequence of the unitary rotations.

Using the unitary rotations, we can generate other incompatible frameworks $\mathcal{U}_{\alpha,\beta,\gamma}(X) = \{\mathcal{U}_{\alpha,\beta,\gamma}(h), \mathcal{U}_{\alpha,\beta,\gamma}(v)\}$ in $Q(X)$,

$$\mathcal{U}_{\alpha,\beta,\gamma}(h) = U_{\alpha,\beta,\gamma}^\dagger h U_{\alpha,\beta,\gamma} = \begin{pmatrix} \cos^2 \frac{\beta}{2} & \frac{1}{2} e^{-i\gamma} \sin \beta \\ \frac{1}{2} e^{i\gamma} \sin \beta & \sin^2 \frac{\beta}{2} \end{pmatrix}, \quad (3.5a)$$

$$\mathcal{U}_{\alpha,\beta,\gamma}(v) = U_{\alpha,\beta,\gamma}^\dagger v U_{\alpha,\beta,\gamma} = \begin{pmatrix} \sin^2 \frac{\beta}{2} & -\frac{1}{2} e^{-i\gamma} \sin \beta \\ -\frac{1}{2} e^{i\gamma} \sin \beta & \cos^2 \frac{\beta}{2} \end{pmatrix}, \quad (3.5b)$$

which depend solely on the two parameters β and γ . The atomic propositions of such a rotated framework are projectors corresponding to each disjoint subspace spanned by an orthonormal Jones' polarization basis,

$$U_{\alpha,\beta,\gamma}^\dagger |h\rangle = \begin{pmatrix} e^{-i(\alpha+\gamma)/2} \cos \frac{\beta}{2} \\ e^{-i(\alpha-\gamma)/2} \sin \frac{\beta}{2} \end{pmatrix}, \quad (3.6a)$$

$$U_{\alpha,\beta,\gamma}^\dagger |v\rangle = \begin{pmatrix} -e^{i(\alpha-\gamma)/2} \sin \frac{\beta}{2} \\ e^{i(\alpha+\gamma)/2} \cos \frac{\beta}{2} \end{pmatrix}. \quad (3.6b)$$

that is a rotation of the Jones' basis (3.2).

Physically, one could in principle construct an apparatus corresponding to such a rotated framework using three laboratory elements: (1) attach a tunable phase plate to the incident port of a polarizing beam splitter with the fast axis aligned to the table, (2) rotate both the beam splitter and attached phase plate with respect to the table, and (3) attach a second tunable phase plate to the incident port of

the first phase plate with the fast axis aligned to the table. Of course this is only one possible parametrization for the unitary rotations; other parametrizations will correspond to other experimental implementations.

It follows that any observable in the full quantum observable space $\Sigma_{Q(X)}^{\mathbb{R}}$ can be obtained by rotating a classical observable $F_X = ah + bv$ to the appropriate framework,

$$\begin{aligned} F_{X'} &= \mathcal{U}_{\alpha,\beta,\gamma}(F_X) = a\mathcal{U}_{\alpha,\beta,\gamma}(h) + b\mathcal{U}_{\alpha,\beta,\gamma}(v), \\ &= \begin{pmatrix} \frac{a+b}{2} + \frac{a-b}{2} \cos \beta & \frac{a-b}{2} e^{-i\gamma} \sin \beta \\ \frac{a-b}{2} e^{i\gamma} \sin \beta & \frac{a+b}{2} - \frac{a-b}{2} \cos \beta \end{pmatrix}, \end{aligned} \quad (3.7)$$

where we have used the matrix representations in (3.5). We see that a general qubit observable depends on four parameters: the eigenvalues a and b , as well as the framework orientation angles β and γ . The complex representation of an observable stems solely from the unitary rotation of the atomic propositions h and v to a different relative framework. The observables no longer generally commute since the unitary rotations need not commute.

3.2 States, Densities, and Collapse

Quantum states.—A quantum state P is a classical state defined in a particular framework X that is then extended to apply to the entire quantum Boolean algebra $\Sigma_{Q(X)}$. The extension of a classical state P that has been defined in a framework X to a proposition $x' = \mathcal{U}(x) \in X' = \mathcal{U}(X)$ in a different framework can be accomplished by heuristically breaking down the state into a composition of the classical state in framework X and *transition probabilities* $D_x(x')$ that

connect the framework X to the different framework X' ,

$$P(x') = \sum_{x \in X} P(x) D_x(x'). \quad (3.8)$$

The transition probabilities characterize a *disturbance* (2.17) that connects the classical state P to propositions in incompatible frameworks.

To define the transition probabilities, we assume that atomic propositions in the framework X are undisturbed, so $D_x(x) = 1$. The only classical state with this property is the pure state which has a projector for a trace-density (2.12) $\rho = x$. Hence, we assume that we can consistently write the transition probability $D_x(x')$ in terms of the extension of the *trace* to the full Boolean algebra $\Sigma_{\mathcal{Q}(X)}$,

$$D_x(x') = \text{Tr}(x x'). \quad (3.9)$$

Notably, this definition makes the transition between frameworks symmetric.

Born rule.—We pick the trace extension to be the unique measure that satisfies the cyclic property $\text{Tr}(AB) = \text{Tr}(BA)$ for all $A, B \in \Sigma_{\mathcal{Q}(X)}$ and agrees with the classical trace (2.13) within any specific framework [93], which is exactly the trace defined in Appendix A.5. On a Hilbert space, (3.9) has the familiar form,

$$D_x(x') = \text{Tr}(|x\rangle\langle x| |x'\rangle\langle x'|) = |\langle x|x'\rangle|^2, \quad (3.10)$$

which we immediately recognize as the *Born rule* [94]. Hence, the complex square of the Hilbert space inner product can be seen as a disguised form of the natural extension of the trace to define transition probabilities between propositions in incompatible frameworks. If we recall that $x' = \mathcal{U}(x) = U^*xU$ we can also write the transition probability (3.10) in terms of the unitary rotor that connects the two propositions, $D_x(x') = \text{Tr}(|x\rangle\langle x| U^\dagger |x'\rangle\langle x'| U) = |\langle x|U|x'\rangle|^2$.

Density operator.—We can rewrite (3.8) in a more familiar form by using

the Born rule (3.9) and the full trace-density (2.12) of the original state $\rho = \sum_{x \in X} P(x) x$, which is traditionally known as the *density operator*,

$$P(x') = \sum_{x \in X} P(x) \text{Tr}(x x') = \text{Tr}(\rho x'). \quad (3.11)$$

This form of the probability functional conforms to *Gleason's theorem* [95]. We note, however, that it is the extension of the *trace* that extends the state to the noncommutative quantum setting since the trace-density ρ is identical to a classical trace-density in some particular framework X .

Moments.—Since the probabilities $P(x')$ are well-defined for a proposition in any framework $x' \in X'$, we can linearly extend P to an expectation functional $\langle \cdot \rangle$ on the entire quantum observable algebra $\Sigma_{Q(X)}^{\mathbb{R}}$,

$$\langle F_{X'} \rangle = \sum_{x' \in X'} f_{X'}(x') P(x') = \text{Tr}(\rho F_{X'}). \quad (3.12)$$

Similarly, observable moments will be well-defined by the expectation functional,

$$\langle (F_{X'})^n \rangle = \sum_{x' \in X'} f_{X'}^n(x') P(x') = \text{Tr}(\rho (F_{X'})^n). \quad (3.13)$$

Hence, the unitary rotations and resulting extension of the trace completely construct the quantum probability space from a single classical probability space and its associated observables.

Double-sided AND.—To be consistent with the assumptions made in (3.9), we must also ensure that conditioning a quantum state on an atomic proposition will collapse the state to a pure state with a trace-density equal to that atomic proposition. In other words, we must generalize the logical AND of the classical case to the noncommutative incompatible frameworks in the quantum case. The consistent way to do this is through a *double-sided product*, which

stems from the algebraic property (A.14) of primitive idempotents discussed in Appendix A.4. That is, given atomic propositions $x \in X$ and $x' \in X'$ then $x'x' = |x'\rangle\langle x'|x\rangle\langle x|x'\rangle\langle x'| = \text{Tr}(xx')x' = D_x(x')x'$, so the constant λ in (A.14) is identified with a *transition probability*.

The double-sided product with x' produces a transition probability $D_x(x')$ from x to x' as a proportionality factor in addition to collapsing the original proposition x to x' . In this sense, the double-sided product includes a form of *disturbance* in addition to the logical *and* of pure classical conditioning. If $X = X'$, so the frameworks coincide, then x and x' will commute; the disturbance will vanish, reducing the transition probability $D_x(x')$ to either 0 or 1; and, the classical AND will be recovered as a special case.

Lüders' rule.—Using the double-sided product as a disturbance followed by a logical AND, we find the quantum form of the *invasive conditioning rule* (2.19),

$$\langle \widetilde{F_X} \rangle_y = \frac{\langle yF_Xy \rangle}{P(y)} = \text{Tr}(\rho_y F_X), \quad (3.14a)$$

$$\rho_y = \frac{y\rho y}{\text{Tr}(y\rho y)}, \quad (3.14b)$$

for any Boolean proposition y in a framework algebra Σ_X measured prior to the observable F_X . As with the classical case, we use the tilde to indicate the intrinsic quantum invasiveness of the measurement process. If ρ and y commute, or if F_X and y commute, then the noninvasive classical conditioning rule (2.14) is properly recovered. This generalization of (2.19) is known as the projection postulate, or *Lüders' Rule* [96]. If y is an atomic proposition in X , then $\rho_y = y$ as in the classical case (2.14) and we consistently recover the assumption (3.9).

For contrast, Leifer and Spekkens [5] provide a careful quantum generalization of the *noninvasive* conditioning rule (2.14) using a formalism based around conditional density operators. They confirm that Lüder's Rule (3.14) cannot be obtained with pure conditioning, so it must imply additional disturbance from the

measurement process itself, as indicated here.

Aharonov-Bergmann-Lebowitz rule.—Just as with classical invasive conditioning, the order of conditioning will generally matter. Specifically, substituting a system proposition $z \in \Sigma_X$ into (3.14) yields $\langle \tilde{z} \rangle_y = P(yzy)/P(y)$; however, $P(yzy) \neq P(zyz)$, so the “joint probability” in the numerator is order-dependent unless y and z commute, just as in (2.32). That is, $\langle \tilde{z} \rangle_y$ explicitly describes the case when the conditioning proposition y is measured first as a *preselection*, followed by the proposition z .

To obtain the converse case when the conditioning proposition z is measured second as a *postselection*, we must derive the quantum form of (2.33). As in the classical case, we reinterpret the denominator of (3.14) as a marginalization $P(y) = \sum_z P(yzy)$ of the ordered joint probability that renormalizes the conditioning procedure; the identity $\sum_z z = 1_X$ permits the equality. With this interpretation, the postselected form of conditioning becomes straightforward,

$${}_z \langle \tilde{y} \rangle = \frac{P(yzy)}{\sum_{y' \in Y} P(y'zy')}. \quad (3.15)$$

As in the classical case, the different position of the subscript serves to distinguish the two conditioned expectations $\langle \tilde{\cdot} \rangle_z$ and ${}_z \langle \tilde{\cdot} \rangle$ corresponding to different measurement orderings.

For a pure state $\rho = x = |x\rangle\langle x|$, this postselected conditioning is known as the *Aharonov-Bergmann-Lebowitz (ABL) rule* [97], and has the form ${}_z \langle \tilde{y} \rangle_x = |z|y\rangle|^2|y|x\rangle|^2 / \sum_{y' \in Y} |z|y'\rangle|^2|y'|x\rangle|^2$. Unlike Lüders’ rule (3.14), the generalized ABL rule (3.15) does not perform a simple update to the trace-density ρ ; moreover, it depends on the entire disturbance of the first measurement via the normalization sum in the denominator. If y and z commute, then the disturbance vanishes and we again correctly recover the classical case (2.14) that is order-independent.

Bayes’ rule.—The two forms of quantum invasive conditioning also lead to a

modified form of Bayes' rule that relates the preselected conditioning of a sequence to the postselected conditioning of the same sequence, similarly to the classical case (2.35),

$${}_z\langle\tilde{y}\rangle = \langle\tilde{z}\rangle_y \frac{P(y)}{\sum_{y' \in Y} P(y'z y')}. \quad (3.16)$$

If y and z commute, then the disturbance vanishes and we correctly recover Bayes' rule (2.16).

The unusual form of (3.15) has led to postselected quantum conditioning being largely overlooked. The lack of symmetry in the density update under such postselected conditioning has even prompted works in multistate-density time-symmetric reformulations of quantum mechanics [28, 30–32, 35, 36, 98], which are outside the scope of this work. However, we see here that the form of the conditioning is the same as the classically *invasive* postselected conditioning (2.33). Later we shall use a fully generalized form of the ABL rule (3.15) together with CVs to consider the subtle case of postselected averages of observables in some detail, so we delay their consideration for now.

3.2.1 Example: Polarization State

A quantum state for a single system is a classical state in some particular framework. For a two-dimensional framework such as $\{h, v\}$, all probabilities for such a classical state can be completely specified by a mixing angle θ such that $P(h) = \cos^2(\theta/2)$ and $P(v) = \sin^2(\theta/2)$. Hence, after rotating the trace-density $\rho = P(h)h + P(v)v$ to an arbitrary framework according to (3.7), any quantum state

trace-density of polarization must have the form,

$$\begin{aligned} \rho_{\theta,\beta,\gamma} &= \cos^2(\theta/2) \mathcal{U}_{\alpha,\beta,\gamma}(h) + \sin^2(\theta/2) \mathcal{U}_{\alpha,\beta,\gamma}(v), \\ &= \frac{1}{2} \begin{pmatrix} 1 + \cos \beta \cos \theta & e^{-i\gamma} \sin \beta \cos \theta \\ e^{i\gamma} \sin \beta \cos \theta & 1 - \cos \beta \cos \theta \end{pmatrix}. \end{aligned} \quad (3.17)$$

The α parameter of the rotation disappears in favor of the θ parameter characterizing the classical state, leaving only three net parameters, in contrast to the four parameters of an arbitrary observable (3.7).

The expectation functional $\langle \cdot \rangle_{\theta,\beta,\gamma}$ is then defined from the trace-density $\rho_{\theta,\beta,\gamma}$ and the unique extension of the trace Tr to the whole observable algebra $\Sigma_{\mathcal{Q}(X)}^{\mathbb{R}}$ according to $\langle F_{X'} \rangle_{\theta,\beta,\gamma} = \text{Tr}(\rho_{\theta,\beta,\gamma} F_{X'})$. The trace extension is the sum of the diagonal matrix elements in the matrix representation. Hence for the expectation of an arbitrary observable (3.7) under an arbitrary state (3.17) we find,

$$\langle \mathcal{U}_{\alpha',\beta',\gamma'}(F_X) \rangle_{\theta,\beta,\gamma} = \frac{a+b}{2} + \frac{a-b}{2}(\cos \theta) \Xi, \quad (3.18a)$$

$$\Xi = \cos \beta \cos \beta' + \sin \beta \sin \beta' \cos(\gamma - \gamma'), \quad (3.18b)$$

where $\Xi \in [-1, 1]$ is an interference factor that depends only on the relative orientation between the state framework and the observable framework. If the frameworks coincide, then $\Xi = 1$ and the classical result is recovered.

3.3 Detectors and Probability Observables

Joint observable space.—As with the classical case, we can couple a system to a detector by enlarging the sample space to the product space XY of a particular pair of frameworks. We can then perform *local* unitary rotations on each space independently to form a joint quantum sample space from the classical joint observables $\mathcal{Q}(X)\mathcal{Q}(Y)$. However, the quantum observable space also admits *global*

unitary rotations on the classical joint observables to form a larger joint quantum sample space $\mathcal{Q}(XY)$. Just as with a single sample space, any two propositions in $\mathcal{Q}(XY)$ can be continuously connected with some global unitary rotation. This decomposition of a global Lie group action into *factors* that act independently on subgroups is precisely as discussed in Appendix A.4.

The full quantum observable space $\Sigma_{\mathcal{Q}(XY)}^{\mathbb{R}}$ is constructed from $\mathcal{Q}(XY)$ in the usual way. Product observables will maintain their product form under local unitary rotations, $\mathcal{U}_X(\mathcal{V}_Y(A_X B_Y)) = \mathcal{U}_X(A_X)\mathcal{V}_Y(B_Y)$. However, global unitary rotations can create unfactorable correlated joint observables in $\Sigma_{\mathcal{Q}(XY)}^{\mathbb{R}}$ even from product observables $\mathcal{U}(A_X B_Y)$.

Joint states.—Similarly, joint *states* on a classical product framework extend to joint quantum states on the quantum product observable space. Under local unitary rotations, product states remain product states and classically correlated states between two specific frameworks remain classically correlated. However, *global* unitary rotations performed on any state can also form *entangled* states that have no analog in the commutative classical theory [99]. Entangled states have some degree of *local-rotation-independent* correlation between frameworks, so display a stronger degree of correlation than can even be defined with a classically correlated state that is restricted to a single joint framework. As an extreme example, maximally entangled states are completely local-rotation-independent and perfectly correlated with respect to any joint framework.

Quantum operations.—The specifics of entanglement do not concern us here, since any type of correlation is sufficient to represent detector probabilities within the reduced system space. For the purposes of measurement, we only assume that the correlated state with density $\rho = \mathcal{U}^*(\rho_X \rho_Y) = U \rho_X \rho_Y U^*$ is connected to some initial product state with density $\rho_X \rho_Y$ via a unitary rotation \mathcal{U}^* . Since all quantum states can be continuously connected with some global unitary rotation that acts as a disturbance (2.27), this is always possible. Physically, the

unitary rotation couples the known detector state ρ_Y to an unknown system state ρ_X . Furthermore, we assume that the initial state of the detector has some (not necessarily unique) pure-state expansion that is meaningful with respect to the preparation procedure $\rho_Y = \sum_{y' \in Y'} P'(y') y'$.

It then follows that the numerator for the conditioning rules (3.14) and (3.15) becomes,

$$\begin{aligned} \langle y F_X y \rangle &= \text{Tr}(\rho y F_X y) = \text{Tr}_X(\text{Tr}_Y(U \rho_X \rho_Y U^* y F_X y)), \\ &= \langle \mathcal{E}_y(F_X) \rangle_X = \text{Tr}_X(\mathcal{E}_y^*(\rho_X) F_X), \end{aligned} \quad (3.19)$$

with the *operations* \mathcal{E}_y and \mathcal{E}_y^* defined as,

$$\mathcal{E}_y(F_X) = \langle U^* y F_X y U \rangle_Y = \sum_{y' \in Y'} P'(y') \text{Tr}_Y(y' U^* y F_X y U), \quad (3.20a)$$

$$= \sum_{y' \in Y'} M_{y,y'}^\dagger F_X M_{y,y'},$$

$$\mathcal{E}_y^*(\rho_X) = \text{Tr}_Y(y U \rho_X \rho_Y U^* y) = \sum_{y' \in Y'} P'(y') \text{Tr}_Y(y U \rho_X y' U^* y), \quad (3.20b)$$

$$= \sum_{y' \in Y'} M_{y,y'} \rho_X M_{y,y'}^\dagger,$$

$$M_{y,y'} = e^{i\phi_{y,y'}} \sqrt{P'(y')} \langle y | U | y' \rangle, \quad (3.20c)$$

$$M_{y,y'}^\dagger = e^{-i\phi_{y,y'}} \sqrt{P'(y')} \langle y' | U^\dagger | y \rangle. \quad (3.20d)$$

Here, the Hilbert space representations of the *Kraus operators* $\{M_{y,y'}\}$ have the form of partial matrix elements and are only well-defined up to the arbitrary phase factors $e^{i\phi_{y,y'}}$. We also stress that $\{M_{y,y'}\}$ depend not only on the measured detector outcome y , but also on a particular detector *preparation* y' .

As a result, we find the quantum versions of the probability observables (2.29),

$$P(y) = \langle \mathcal{E}_y(1_X) \rangle_X = \langle E_y \rangle_X, \quad (3.21)$$

$$E_y = \mathcal{E}_y(1_X) = \langle U^* y U \rangle_Y = \sum_{y' \in Y'} M_{y,y'}^\dagger M_{y,y'}, \quad (3.22)$$

and the general invasive measurement (2.30),

$$\left\langle \widetilde{F}_X \right\rangle_y = \frac{\langle \mathcal{E}_y(F_X) \rangle_X}{\langle \mathcal{E}_y(1_X) \rangle_X} = \frac{\sum_{y' \in Y'} \text{Tr}_X(\rho_X M_{y,y'}^\dagger F_X M_{y,y'})}{\text{Tr}_X(\rho_X E_y)}. \quad (3.23)$$

Similarly to the invasive classical case (2.31), the measurement of y on the detector must be described by a *quantum operation* \mathcal{E}_y in (3.19), which is a completely positive map [11, 17–25, 27, 100, 101] that performs a *generalized measurement* on the system state corresponding to the detector outcome y . The operation \mathcal{E}_y acting on the identity in (3.22) produces a positive operator known as a *quantum effect*, E_y . By construction, the set of operations $\{\mathcal{E}_y\}$ preserves the identity, $\sum_y \mathcal{E}_y(1_X) = 1_X$; hence, the effects form a partition of the identity, $\sum_y E_y = 1_X$, making them probability observables over a particular detector framework exactly as in (2.29).

Sequences of measurements emphasize the temporal ordering of operations, just as in the invasive classical case (2.32). Given two sets of quantum operations that define the sequential interaction of two detectors with the system and their subsequent conditioning, $\{\mathcal{E}_y\}$ and $\{\mathcal{E}'_z\}$, the joint probability of the ordered sequence of detector outcomes (y, z) is,

$$P(y)P(z|y) = P(yzy) = P(yz1_Xzy) = \langle \mathcal{E}_y(\mathcal{E}'_z(1_X)) \rangle_X = \langle \mathcal{E}_y(E'_z) \rangle_X, \quad (3.24)$$

where $E'_z = \mathcal{E}'_z(1_X)$. The proper sequential probability observable $\mathcal{E}_y(E'_z) = \sum_{y'} M_{y,y'}^\dagger E'_z M_{y,y'}$ is not a simple product of the individual probability observables

E_y and E'_z .

These sequence probabilities then give us the full generalization of the ABL rule (3.15),

$$\begin{aligned} {}_z\langle\widetilde{y}\rangle &= \frac{\langle\mathcal{E}_y(E'_z)\rangle_X}{\langle\mathcal{E}(E'_z)\rangle_X} = \frac{\langle\mathcal{E}_y(E'_z)\rangle_X}{\sum_{y''\in Y}\langle\mathcal{E}_{y''}(E'_z)\rangle_X}, \\ &= \frac{\sum_{y'\in Y'}\text{Tr}_X(\rho_X M_{y,y'}^\dagger E'_z M_{y,y'})}{\sum_{y''\in Y}\sum_{y'\in Y'}\text{Tr}_X(\rho_X M_{y'',y'}^\dagger E'_z M_{y'',y'})}, \end{aligned} \quad (3.25)$$

and the most general version of the invasive quantum Bayes' rule (3.16),

$${}_z\langle\widetilde{y}\rangle = \left\langle\widetilde{E}'_z\right\rangle_y \frac{\langle E_y\rangle_X}{\langle\mathcal{E}(E'_z)\rangle_X}, \quad (3.26)$$

As with (2.33) and (3.15), the postselected conditioning (3.25) depends on the entire disturbance of the first measurement via the *nonselective measurement* $\mathcal{E} = \sum_{y''\in Y}\mathcal{E}_{y''}$ in the denominator.

The noncommutativity of the detection operations \mathcal{E}_y emphasizes the fact that measurement is an active *process*: an experimenter alters the quantum state by coupling it to a detector and then conditioning on acquired information from the detector. Without some filtering process that completes the disturbance implied by (3.19), there is no measurement. The nonselective measurement \mathcal{E} also includes the active disturbance of the measurement process, but does not condition on a particular outcome. Furthermore, measuring a quantum state in a different order generally disturbs it differently. The state may also in certain conditions be probabilistically “uncollapsed” back to where it started by using the correct conditioning sequence [13–15]. In this sense, sequential quantum conditioning is analogous to a stochastic control process that guides the progressive disturbance of a state along some trajectory in the state space [27].

Measurement operators.—Since the quantum operation \mathcal{E}_y performs a measurement, we will refer to its Kraus operators $\{M_{y,y'}\}$ (3.20) as *measurement operators*.

However, a quantum operation generally has many equivalent double-sided product expansions like (3.20a) in terms of measurement operators. Each such set of measurement operators $\{M_{y,y'}\}$ corresponds to a specific choice of framework for the preparation of the detector state $\rho_Y = \sum_{y' \in Y'} P(y') y'$.

Given a specific set of measurement operators, the substitution $M_{y,y'} \rightarrow U_{y,y'} M_{y,y'}$ with unitary $U_{y,y'}$ will produce the same effect E_y according to (3.22) but will correspond to a different operation \mathcal{E}'_y . Hence, we conclude that many measurement operations can produce the same probability observables on the system space [102]. Therefore, *probability observables are not sufficient to completely specify a quantum measurement*: one needs to specify the full operations as in the classically invasive case (2.31).

Quantum process tomography.—Just as classical probability observables can be characterized via process tomography, operations can be characterized by *quantum process tomography*. One performs quantum process tomography by sending known states into a detector, measuring the detector, then measuring the resulting states to see how the state was changed by the detector. Since quantum operations contain information about disturbance as well as conditioning, quantum process tomography generally requires more characterization measurements than pure classical process tomography.

Pure operations.—An initially pure detector state with density y' produces a *pure operation* $\mathcal{E}_y(F_X) = M_y^\dagger F_X M_y$ with a single associated measurement operator $M_y = e^{i\phi_y} \langle y| U |y'\rangle$ that is unique up to the arbitrary phase factor $e^{i\phi_y}$. Most laboratory preparation procedures for the detector are designed to produce a pure initial state, so pure operations will be the typical case. A pure operation has the additional property of partially collapsing a pure state to another pure state. It is also most directly related to the probability observable $E_y = M_y^\dagger M_y$, since the single measurement operator has a polar decomposition $M_y = U_y E_y^{1/2}$ in terms of the positive root of the probability observable $E_y^{1/2}$.

Weak measurement.—If we wish for such a conditioning process to leave the state approximately unchanged, we must make a *weak measurement*, just as in the classical case (2.25). However, a quantum weak measurement requires a strict condition regarding the measurement operations and not just the probability observables due to the additional disturbance in the measurement. Formally, the measurement operations typically depend on a measurement strength parameter ϵ such that,

$$\forall y \in Y \lim_{\epsilon \rightarrow 0} \mathcal{E}_y(\epsilon; F_X) = P_Y(y)\mathcal{J}(F_X), \quad (3.27)$$

where \mathcal{J} is the identity operation and $P_Y(y)$ is the probability for obtaining the detector outcome y in the absence of interaction. As with the classical case, the limit as $\epsilon \rightarrow 0$ is an idealization known as the *weak measurement limit* and is not strictly achievable in the laboratory.

The definition (3.27) implies that subsequent measurements will be unaffected, $\forall y \in Y, \lim_{\epsilon \rightarrow 0} \langle \widetilde{F}_X \rangle_y = \langle F_X \rangle$, and that the probability observables are proportional to the identity in the weak limit, $\forall y \in Y, \lim_{\epsilon \rightarrow 0} E_y(\epsilon) = P_Y(y)1_X$, just as in the classical case (2.25). It also follows that any set of measurement operators $\{M_{y,y'}(\epsilon)\}$ that characterize $\mathcal{E}_y(\epsilon)$ must also be proportional to the identity in the weak limit $\forall y \in Y, y' \in Y', \lim_{\epsilon \rightarrow 0} M_{y,y'}(\epsilon) \propto 1_X$.

Weak measurements are more interesting in the quantum case than in the classical case due to the existence of incompatible frameworks. Since a weak measurement of an observable does not appreciably affect the quantum state, subsequent measurements on incompatible observables can be made that will probe approximately the same state. This technique allows (noisy) information about two incompatible frameworks to be gleaned from nearly the same quantum state in a single experiment, which is strictly impossible using strong measurements that collapse the state to a pure state in a particular framework after each measure-

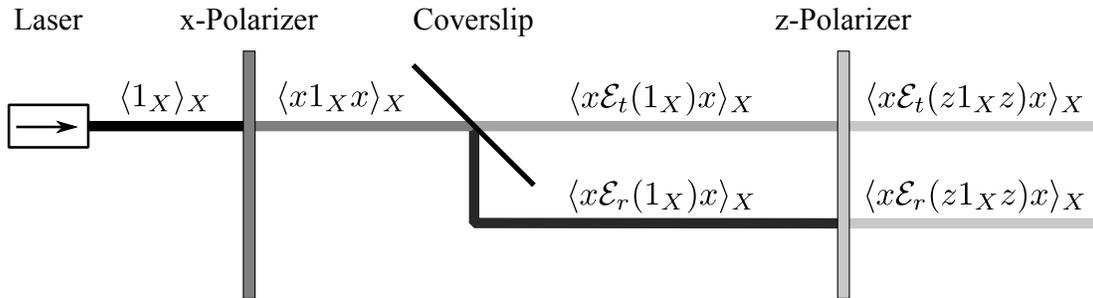


Figure 3.1: Coverslip polarization measurement. A laserbeam passes through a preselection x -polarizer, a glass microscope coverslip, and a postselection z -polarizer. The transmission probabilities for each segment of the apparatus are shown. By assigning appropriate contextual values $f_Y(t)$ and $f_Y(r)$ (3.46) to the output ports of the coverslip, the polarization observable $F_X = f_X(h)h + f_X(v)v$ can be measured using the equivalent expansion in terms of the appropriate measurement context $F_X = f_Y(t)\mathcal{E}_t(1_X) + f_Y(r)\mathcal{E}_r(1_X)$. Averaging the same contextual values with pre- and postselected conditional probabilities $\langle \tilde{t} \rangle_x = \langle x \mathcal{E}_t(z) x \rangle_X / (\langle x \mathcal{E}_t(z) x \rangle_X + \langle x \mathcal{E}_r(z) x \rangle_X)$ and ${}_z \langle \tilde{r} \rangle_x = \langle x \mathcal{E}_r(z) x \rangle_X / (\langle x \mathcal{E}_t(z) x \rangle_X + \langle x \mathcal{E}_r(z) x \rangle_X)$ produces the conditioned average (3.52) ${}_z \langle \tilde{F}_X \rangle_x = f_Y(t) {}_z \langle \tilde{t} \rangle_x + f_Y(r) {}_z \langle \tilde{r} \rangle_x$.

ment. The penalty for using weak measurements is that many more measurements are needed than in the strong measurement case to overcome the ambiguity of the measurement, as discussed in the classical case.

3.3.1 Example: Coverslip Polarization Detector

To cement these ideas, we consider the task of indirectly measuring polarization in a particular framework. For specificity, we will consider the passage of a laser beam with unknown polarization through a glass microscope coverslip, as shown in Fig. 3.1. Fresnel reflection off the coverslip leads to a disparity between transmission and reflection of the polarizations, so comparing transmitted to reflected light allows a generalized measurement of polarization. We will use exactly this measurement technique experimentally in Chapter 4.

The system sample space we wish to measure is the polarization with respect

to the table ($h = |h\rangle\langle h|$) and ($v = |v\rangle\langle v|$), which could in principle be measured ideally with a polarizing beam splitter. The detector sample space is the spatial degree of freedom of the transmitted ($t = |t\rangle\langle t|$) and reflected ($r = |r\rangle\langle r|$) ports of a coverslip rotated to some fixed angle with respect to the incident beam around an axis perpendicular to the table. The initial state of the detector is the pure state indicating that the beam enters a single incident port ($b = |b\rangle\langle b|$) of the coverslip with certainty. The rotation $U^*(\rho_X b) = U\rho_X bU^*$ that couples the system to the detector describes the interaction of the beam with the coverslip and has a unitary rotor U corresponding to the polarization-dependent scattering matrix of the coverslip. Assuming that the scattering preserves beams of pure polarization, so h remains h and v remains v , the rotor decouples into a direct sum of rotors that are specific to each polarization,

$$U = U_h \oplus U_v, \quad (3.28)$$

meaning that U has a block-diagonal structure when represented as a matrix.

Selecting each output port of the coverslip produces the two *measurement operators* according to (3.20),

$$M_t = \langle t| U |b\rangle = \begin{pmatrix} \langle t| U_h |b\rangle & 0 \\ 0 & \langle t| U_v |b\rangle \end{pmatrix}, \quad (3.29a)$$

$$M_r = \langle r| U |b\rangle = \begin{pmatrix} \langle r| U_h |b\rangle & 0 \\ 0 & \langle r| U_v |b\rangle \end{pmatrix}, \quad (3.29b)$$

which characterize the *pure measurement operations* that modify observables according to (3.20a),

$$\mathcal{E}_t(F_X) = M_t^\dagger F_X M_t, \quad (3.30a)$$

$$\mathcal{E}_r(F_X) = M_r^\dagger F_X M_r, \quad (3.30b)$$

and their adjoints that modify the state density according to (3.20b),

$$\mathcal{E}_t^*(\rho_X) = M_t \rho_X M_t^\dagger, \quad (3.31a)$$

$$\mathcal{E}_r^*(\rho_X) = M_r \rho_X M_r^\dagger. \quad (3.31b)$$

The pure measurement operations in turn produce *probability observables* according to (3.22),

$$E_t = \mathcal{E}_t(1_X) = M_t^\dagger M_t = \begin{pmatrix} |\langle t| U_h |b\rangle|^2 & 0 \\ 0 & |\langle t| U_v |b\rangle|^2 \end{pmatrix}, \quad (3.32a)$$

$$E_r = \mathcal{E}_r(1_X) = M_r^\dagger M_r = \begin{pmatrix} |\langle r| U_h |b\rangle|^2 & 0 \\ 0 & |\langle r| U_v |b\rangle|^2 \end{pmatrix}, \quad (3.32b)$$

in the same framework as h and v . These probability observables are therefore equivalent to classical probability observables (2.29) specified by the effective characterization probabilities $\tilde{P}(t|h) = |\langle t| U_h |b\rangle|^2$, $\tilde{P}(r|h) = |\langle r| U_h |b\rangle|^2$, $\tilde{P}(t|v) = |\langle t| U_v |b\rangle|^2$, and $\tilde{P}(r|v) = |\langle r| U_v |b\rangle|^2$.

The measurement operators (3.29) have a polar decomposition in terms of the roots of the probability observables and an extra unitary phase contribution,

$$M_t = \begin{pmatrix} e^{i\phi_{h,t}} \sqrt{\tilde{P}(t|h)} & 0 \\ 0 & e^{i\phi_{v,t}} \sqrt{\tilde{P}(t|v)} \end{pmatrix}, \quad (3.33a)$$

$$M_r = \begin{pmatrix} e^{i\phi_{h,r}} \sqrt{\tilde{P}(r|h)} & 0 \\ 0 & e^{i\phi_{v,r}} \sqrt{\tilde{P}(r|v)} \end{pmatrix}. \quad (3.33b)$$

Any nonzero relative phase, such as $\phi_{h,t} - \phi_{v,t}$, will affect the framework orientation for subsequent measurements; however, it will not contribute to the acquisition of information from the measurement since it does not contribute to the probability observables. Such relative phase is therefore part of the *disturbance* of the

measurement process.

Specifically, the initial state of polarization P_X will be conditioned by a selection of a particular port on the detector according to,

$$\langle \widetilde{F}_X \rangle_t = \frac{\langle \mathcal{E}_t(F_X) \rangle_X}{\langle \mathcal{E}_t(1_X) \rangle_X} = \frac{\text{Tr}_X(M_t \rho_X M_t^\dagger F_X)}{\text{Tr}_X(\rho_X E_t)}, \quad (3.34a)$$

$$\langle \widetilde{F}_X \rangle_r = \frac{\langle \mathcal{E}_r(F_X) \rangle_X}{\langle \mathcal{E}_r(1_X) \rangle_X} = \frac{\text{Tr}_X(M_r \rho_X M_r^\dagger F_X)}{\text{Tr}_X(\rho_X E_r)}. \quad (3.34b)$$

Although the probabilities in each denominator only depend on the probability observables, the altered states in each numerator depend on the measurement operations and will include effects from the relative phase in the measurement operators (3.33).

3.4 Contextual Values

Operation correspondence.—The introduction of contextual values in the quantum case proceeds identically to the classical case of invasive measurements (2.40). Since we must generally represent detector probabilities by *operations* $\{\mathcal{E}_y\}$ within the reduced system space according to (3.22) and (3.24), we must also generally represent detector observables by *weighted operations* within the reduced system space,

$$\langle F_Y \rangle = \sum_{y \in Y} f_Y(y) P(y) = \sum_{y \in Y} f_Y(y) \langle \mathcal{E}_y(1_X) \rangle_X = \langle \mathcal{F}_X(1_X) \rangle_X, \quad (3.35)$$

$$\mathcal{F}_X = \sum_{y \in Y} f_Y(y) \mathcal{E}_y. \quad (3.36)$$

If we are concerned with only a single measurement, or are working within a single framework as in the classical formalism, then for all practical purposes the operation \mathcal{F}_X reduces to its associated system observable $F_X = \mathcal{F}_X(1_X)$ as in the

classical definition (2.37).

Contextual values.—We observe a corollary exactly as in the classical case (2.38): if we can expand a *system* observable in terms of the probability observables generated by a particular measurement operation, then that observable can also be expressed as an equivalent *detector* observable,

$$F_X = \sum_{y \in Y} f_Y(y) E_y \implies F_Y = \sum_y f_Y(y) y, \quad (3.37)$$

which is the quantum form of our main result originally introduced in [68]. As in the classical case, we dub the required detector labels $f_Y(y)$ the **contextual values** (CV) of the quantum observable F_X with respect to the *context* of a specific detection scheme as represented in the system space by the measurement operations $\{\mathcal{E}_y\}$.

Unlike the classical case, the probability operators $\{E_y\}$ need not be diagonal, nor need they commute. All that is strictly necessary is for the equality (3.37) to hold in the sum. Moreover, since many measurement operations produce the same probability observables $\{\mathcal{E}_y(1_X) = E_y\}$, many detection schemes can use the same CVs to reproduce an observable average.

Moments.—As with classically invasive measurements (2.41), higher statistical moments of the observable require more care to measure. For instance, we require the following equality in order to accurately reproduce the n^{th} moment of an observable indirectly using the same CV,

$$\langle (F_X)^n \rangle_X = \sum_{y_1, \dots, y_n \in Y} f_Y(y_1) \cdots f_Y(y_n) \langle E_{y_1} \cdots E_{y_n} \rangle_X. \quad (3.38)$$

However, as indicated in (3.24), performing a sequence of n measurements produces the measurable probability $\langle \mathcal{E}_{y_1}(\cdots (E_{y_n}) \cdots) \rangle_X \neq \langle E_{y_1} \cdots E_{y_n} \rangle_X$. Indeed, $\langle E_{y_1} \cdots E_{y_n} \rangle_X$ will not generally be a well-formed probability. To obtain the equal-

ity (3.38) with a particular choice of CV, we need the additional constraint that *all the measurement operators must commute with each other*. As a result, they must be part of the same framework as the system observable and hence commute with that observable as well. We will call any detector with commuting measurement operators with respect to a particular observable a *fully compatible detector* for that observable.

Alternatively, as with the classical case, we can change the CVs to define new observables that correspond to powers of the original observable, such as $G_X = (F_X)^n = \sum_{y \in Y} g_Y(y) E_y$. These new observables can then be measured indirectly using the same experimental setup without the need for measurement sequences. The CVs $g_Y(y)$ for the n^{th} power of F_X will not be a simple power of the CVs $f_Y(y)$ for F_X unless the measurement is unambiguous.

Correlation functions.—If a time-evolution unitary rotation \mathcal{U}_t is inserted between different observable measurements, then we obtain a quantum *correlation function* instead,

$$\left\langle F_X(0) \widetilde{G_X}(t) \right\rangle = \langle \mathcal{F}_X(\mathcal{U}_t(\mathcal{G}_X(1_X))) \rangle_X, \quad (3.39)$$

which should be compared to the classical case (2.42). Similarly, n -time correlations can be defined with $n - 1$ time-evolutions between the observable measurements $\langle \mathcal{F}_1(\mathcal{U}_{t_1}(\mathcal{F}_2(\cdots \mathcal{U}_{t_{n-1}}(\mathcal{F}_n(1_X)) \cdots))) \rangle$.

Inversion.—Since the CVs depend only on the probability observables, which commute with the measured observable for a fully compatible detector, the procedure for determining the CVs will be identical to the classical case. That is, *the contextual values of a quantum observable exactly correspond to the detector labels for a classically ambiguous detector*. We shall refer the reader back to the classical inversion (2.46) for discussion on how to solve the relation (3.37). As a reminder, we advocate the pseudoinverse as a principled approach for picking the

CVs in the event of redundancy or course-graining.

Conditioned averages.—We can construct a general *postselected conditioned average* from the CVs and the fully generalized ABL rule (3.25) analogously to the classical case (2.43),

$$\begin{aligned} {}_z\langle \widetilde{F}_X \rangle &= \sum_y f_Y(y) {}_z\langle \widetilde{y} \rangle = \frac{\langle \mathcal{F}_X(E'_z) \rangle_X}{\langle \mathcal{E}(E'_z) \rangle_X}, \\ &= \frac{\sum_{y \in Y} \sum_{y' \in Y'} f_Y(y) \text{Tr} \left(\rho_X M_{y,y'}^\dagger E'_z M_{y,y'} \right)}{\sum_{y \in Y} \sum_{y' \in Y'} \text{Tr} \left(\rho_X M_{y,y'}^\dagger E'_z M_{y,y'} \right)}. \end{aligned} \quad (3.40)$$

We introduced this type of conditioned average in [68] for the typical case of pure operations $\{\mathcal{E}_y\}$ with single associated measurement operators $\{M_y\}$.

If the postselection is defined in the same framework as the measurement operation, then the nonselective measurement \mathcal{E} in the denominator will reduce to unity, leaving a classical conditioned average,

$$\langle F_X \rangle_z = \frac{\sum_{y \in Y} f_Y(y) \langle E_y E'_z \rangle_X}{\langle E'_z \rangle_X} = \frac{\langle F_X E'_z \rangle_X}{\langle E'_z \rangle_X}, \quad (3.41)$$

of the same form as (2.24). Similarly, the preselected conditioning (3.23) will also reduce to (3.41) for such a case. This special case cannot exceed the eigenvalue range of the observable: the observable F_X will always reduce to its eigenvalues since either the state or the postselection commute with it.

More generally, however, the combination of amplified CVs and the context-dependent probabilities in the general postselected average (3.40) can send it outside the eigenvalue range of the observable. We will see examples of this in Chapters 4, 5, and 6.

Strong-conditioned average.—There are two other important special cases of the conditioned average (3.40) worth mentioning: strong measurement and weak measurement. The strong measurement case is distinguished by being constrained

exclusively to the eigenvalue range of the observable. Specifically, (3.40) reduces to the form,

$${}_z \langle \widetilde{F}_X \rangle = \frac{\sum_{x \in X} f_X(x) P(x) D_x(z)}{\sum_{x \in X} P(x) D_x(z)} = \frac{\sum_{x \in X} f_X(x) \langle x | \rho | x \rangle |\langle x | z \rangle|^2}{\sum_{x \in X} \langle x | \rho | x \rangle |\langle x | z \rangle|^2}, \quad (3.42)$$

which contains only the eigenvalues $f_X(x)$ of the observable and factored probability products. However, it cannot be expressed solely in terms of the observable F_X and a conditioned state as in the classical case (2.43) due to the disturbances $D_x(z)$. Only when the state or postselection commutes with the observable does (3.42) reduce to a special case of (3.41) and become free from disturbance.

Weak values.—The weak measurement case is distinguished by being the only case of the quantum postselected conditioned average (3.40) that can become *context independent* for any state and postselection (under certain conditions, see Section 3.5). The context-independent weak limit of the conditioned average (3.40) is the *weak value* [28, 30–32, 35, 36, 68],

$${}_z \langle \widetilde{F}_X \rangle^w = \frac{\langle E'_z F_X + F_X E'_z \rangle_X}{2 \langle E'_z \rangle_X}, \quad (3.43)$$

and is expressed entirely in terms of the system expectation functional $\langle \cdot \rangle_X$, the postselection probability observable E'_z , and the observable F_X . Written in this form it is clear that it is a symmetrized version of the context-independent commuting case (3.41); however, unlike (3.41) the weak value (3.43) is not constrained to the eigenvalue range and can even diverge. For a pure initial state with trace-density x and pure postselection z , the weak value (3.43) takes the traditional form,

$${}_z \langle F_X \rangle_x^w \rightarrow \operatorname{Re} \frac{\langle z | F_X | x \rangle}{\langle z | x \rangle}. \quad (3.44)$$

We will have much more to say about weak values in Section 3.5 and in Chapter 6.

3.4.1 Example: Coverslip Detector Revisited

Continuing the example from Sec. 3.3.1 and Fig. 3.1, observables defined in the same framework as the probability observables may be expressed in terms of the probability observables according to (3.37) using *contextual values* (CVs), exactly as in the classical example (2.53),

$$F_X = f_X(h)h + f_X(v)v = f_Y(t)E_t + f_Y(r)E_r, \quad (3.45a)$$

$$\begin{pmatrix} f_X(h) \\ f_X(v) \end{pmatrix} = \begin{pmatrix} \tilde{P}(t|h) & \tilde{P}(r|h) \\ \tilde{P}(t|v) & \tilde{P}(r|v) \end{pmatrix} \begin{pmatrix} f_Y(t) \\ f_Y(r) \end{pmatrix}. \quad (3.45b)$$

Inverting this relation according to (2.46) produces the unique CVs,

$$f_Y(t) = \frac{\tilde{P}(r|v)f_X(h) - \tilde{P}(r|h)f_X(v)}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}, \quad (3.46a)$$

$$f_Y(r) = -\frac{\tilde{P}(t|v)f_X(h) - \tilde{P}(t|h)f_X(v)}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}. \quad (3.46b)$$

The denominator is unity when the output ports of the coverslip are perfectly correlated with the polarization. Otherwise, the denominator is less than one and serves to *amplify* the CVs to compensate for the ambiguity of the detection. The numerator contains cross-compensation factors that correct bias in the detector; that is, the eigenvalue $f_X(h)$ for h in the contextual value $f_Y(t)$ for t is weighted by the conditional probability $\tilde{P}(r|v)$ corresponding to the complementary quantities of v and r , and so forth.

The CVs define the detector observable that is actually being measured in the laboratory,

$$F_Y = f_Y(t)t + f_Y(r)r. \quad (3.47)$$

This detector observable corresponds to a detection *operation* on the system space

according to (3.36),

$$\mathcal{F}_X = f_Y(t)\mathcal{E}_t + f_Y(r)\mathcal{E}_r, \quad (3.48)$$

which fully describes the interaction with the detector, subsequent conditioning, and experimental convention for defining the observable. When no subsequent conditioning is performed on the system, this operation constructs the system observable $F_X = \mathcal{F}_X(1_X) = f_Y(t)E_t + f_Y(r)E_r$, as desired.

Since the pure measurement operations all belong to the same framework and commute with F_X , the operation \mathcal{F}_X is also *fully compatible* with the observable F_X , meaning it can measure any moment of that observable using the same CVs according to (3.38),

$$\langle \mathcal{F}_X^n(1_X) \rangle_X = \langle (F_X)^n \rangle_X = \sum_{i_1 \dots i_n} f_Y(i_1) \dots f_Y(i_n) \langle E_{i_1} \dots E_{i_n} \rangle_X. \quad (3.49)$$

The quantity $\mathcal{F}_X^n(1_X)$ indicates a sequence of n consecutive measurements made by the same coverslip on the beam to construct the observable $(F_X)^n$ for the n^{th} moment of F_X . That is, the output from each port of the coverslip is fed back into the coverslip to be measured again. There are 2^n possible outcome sequences (i_1, \dots, i_n) for n traversals through the coverslip, each with probability $\langle E_{i_1} \dots E_{i_n} \rangle_X$ of occurring. These probabilities are weighted with appropriate products of corresponding CVs and summed to correctly construct the n^{th} moment of F_X .

Alternatively, one can change the CVs to directly measure the observable $G_X = (F_X)^n = g_Y(t)E_t + g_Y(r)E_r$ from one traversal of the coverslip. The required

CVs for G_X ,

$$g_Y(t) = \frac{\tilde{P}(r|v)(f_X(h))^n - \tilde{P}(r|h)(f_X(v))^n}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}, \quad (3.50a)$$

$$g_Y(r) = -\frac{\tilde{P}(t|v)(f_X(h))^n - \tilde{P}(t|h)(f_X(v))^n}{\tilde{P}(t|h)\tilde{P}(r|v) - \tilde{P}(r|h)\tilde{P}(t|v)}, \quad (3.50b)$$

are not simple powers of the CVs (3.46) for F_X unless the measurement is unambiguous.

In addition to moments of F_X , we can obtain postselected *conditioned averages* of F_X by conditioning on a second measurement outcome characterized by a probability observable E'_z after the measurement by the coverslip according to (3.40),

$${}_z \langle \widetilde{F}_X \rangle = \frac{\langle \mathcal{F}_X(E'_z) \rangle_X}{\langle \mathcal{E}(E'_z) \rangle_X}, \quad (3.51)$$

where $\mathcal{E} = \mathcal{E}_t + \mathcal{E}_r$ is the nonselective measurement by the coverslip. The second measurement could be a polarizer, another coverslip, or any other method for measuring polarization a second time.

If the initial state is pure with a density $\rho = x = |x\rangle\langle x|$ and the final postselection is also pure $z = |z\rangle\langle z|$, then (3.51) simplifies to a pre- and postselected conditioned average,

$${}_z \langle \widetilde{F}_X \rangle_x = \frac{f_Y(t)|\langle z| M_t |x\rangle|^2 + f_Y(r)|\langle z| M_r |x\rangle|^2}{|\langle z| M_t |x\rangle|^2 + |\langle z| M_r |x\rangle|^2}. \quad (3.52)$$

If we relate both pure states to the reference state h via unitary rotations as defined in (3.4), $x = \mathcal{U}_{\alpha,\beta,\gamma}(h)$ and $z = \mathcal{U}_{\alpha',\beta',\gamma'}(h)$, then the probabilities take the

form,

$$|\langle z | M_t | x \rangle|^2 = \tilde{P}^h(t) \cos^2(\beta/2) \cos^2(\beta'/2) + \tilde{P}^v(t) \sin^2(\beta/2) \sin^2(\beta'/2) \quad (3.53a)$$

$$+ \frac{\sqrt{\tilde{P}^h(t)\tilde{P}^v(t)}}{2} \sin \beta \sin \beta' \cos(\gamma - \gamma' - \phi_{h,t} + \phi_{v,t}),$$

$$|\langle z | M_r | x \rangle|^2 = \tilde{P}^h(r) \cos^2(\beta/2) \cos^2(\beta'/2) + \tilde{P}^v(r) \sin^2(\beta/2) \sin^2(\beta'/2) \quad (3.53b)$$

$$+ \frac{\sqrt{\tilde{P}^h(r)\tilde{P}^v(r)}}{2} \sin \beta \sin \beta' \cos(\gamma - \gamma' - \phi_{h,r} + \phi_{v,r}).$$

We see that each probability possesses an interference term that stems from the relative orientations of the incompatible frameworks for the preparation, measurement, and postselection. In addition, the relative phases in the measurement operators (3.33) will affect the orientations of the frameworks and further disturb the measurement, as mentioned. For the classical case, the frameworks coincide, so $\beta, \beta' \in \{0, \pi\}$; the interference term vanishes; and, the probabilities reduce to the conditional probabilities that characterize the probability observables.

The combination of the expanded range of the CVs (3.46) and the interference term in the probabilities (3.53) can make the postselected conditioned averages (3.51) counter-intuitively exceed the eigenvalue range of the observable F_X . Such a violation of the eigenvalue range cannot occur from classical conditioning without disturbance as in Sec. 2.5.2.

3.4.2 Example: Calcite Polarization Detector

We can also measure polarization using a von Neumann measurement [11] that uses a detector with a continuous sample space detector, such as position³. For example, passing a beam of polarized light through a calcite crystal will continuously separate the polarizations h and v along a particular position axis. Measuring the

³In Chapter 6 we will solve such a von Neumann measurement exactly using a different method.

position profile of the resulting split beam along that axis allows information to be gained about the polarization.

For such a setup, measuring the position with a linear scale corresponds to measuring a detector observable $Q = \int_Y y d\varepsilon(y)$ for a continuous sample space of distinguishable positions. The observable Q has a conjugate D_Q that satisfies the Lie bracket $[Q, D_Q] = i1_Y$. The conjugate can thus generate translations in Q with a group action, $\exp(iqD_Q)Q\exp(-iqD_Q) = Q + [iqD_Q, Q] + [iqD_Q, [iqD_Q, Q]] + \dots = Q + q1_Y$. Hence, we can model the calcite crystal as a rotation governed by a unitary rotor of the form

$$U = \exp(-i(\epsilon_h h - \epsilon_v v)D_Q), \quad (3.54)$$

which will translate h polarization by some amount ϵ_h while simultaneously translating v polarization by some amount ϵ_v in the opposing direction. The parameters ϵ_h and ϵ_v will depend on the material of the crystal and the angle of the optical axis of the crystal to the incident beam⁴.

Suppose the light beam has an initially pure beam profile state described by a density $\rho = |\psi\rangle\langle\psi|$. The probability for obtaining a particular pure position $y = |y\rangle\langle y|$ in the profile would then be $dP_Y(y) = p_Y(y)dy = \text{Tr}(\rho y)dy = |\langle y|\psi\rangle|^2 dy$. Each complex factor $\langle y|\psi\rangle$ is the “wave function” of the transverse beam profile, whose complex square is the probability density with respect to the integral $p_Y(y) = |\langle y|\psi\rangle|^2$.

If we then pass the beam through the crystal described by the rotor (3.54) and measure its position in a pure position state $y = |y\rangle\langle y|$, we will have enacted a pure operation on the polarization of the beam that is characterized by a single

⁴For example, if h is aligned with the optic axis of the crystal then h is the ordinary ray with $\epsilon_h = 0$ while v is the extra-ordinary ray with $\epsilon_v \neq 0$.

measurement operator,

$$d\mathcal{E}_y(F_X) = M(y)^\dagger F_X M(y) dy, \quad (3.55a)$$

$$M(y) = \langle y | U | \psi \rangle = h \langle y - \epsilon_h | \psi \rangle + v \langle y + \epsilon_v | \psi \rangle, \quad (3.55b)$$

with components equal to the initial wave function of the detector profile shifted in position by an appropriate ϵ . The pure measurement operations define a continuous set of probability observables,

$$dE(y) = d\mathcal{E}_y(1_X) = M(y)^\dagger M(y) dy = h dP_Y(y - \epsilon_h) + v dP_Y(y + \epsilon_v), \quad (3.56)$$

with components equal to the initial transverse beam profile shifted in position by an appropriate ϵ . Unless the shifts become degenerate with $\epsilon_v = -\epsilon_h$ then these probability observables can be used to indirectly measure any observable in the framework of h and v .

Since the observable $\epsilon_h h - \epsilon_v v$ appears as a generator for the rotation U , it could be tempting to assert that the detector must specifically measure this observable. However, only the *framework* in which the generating observable is defined determines which observables can be measured. The choice of CV, which can be made in postprocessing, will calibrate the detector to measure specific observables in that framework.

We considered a classical version of similar probability observables in Section 2.5.4. Generalizing that derivation only slightly, we can find the preferred contextual values (CVs) $f_Y(y)$ for an arbitrary polarization observable $F_X = f_X(h)h + f_X(v)v$,

$$f_Y(y) = f_X(h) \frac{v_+(y) + v_-(y)}{2} + f_X(v) \frac{v_+(y) - v_-(y)}{2}, \quad (3.57a)$$

$$v_+(y) = \frac{p_Y(y - \epsilon_h) + p_Y(y + \epsilon_v)}{a + b(\epsilon_h, \epsilon_v)}, \quad (3.57b)$$

$$v_-(y) = \frac{p_Y(y - \epsilon_h) - p_Y(y + \epsilon_v)}{a - b(\epsilon_h, \epsilon_v)}, \quad (3.57c)$$

$$a = \int_Y p_Y^2(y) dy, \quad (3.57d)$$

$$b(\epsilon_h, \epsilon_v) = \int_Y p_Y(y - \epsilon_h) p_Y(y + \epsilon_v) dy. \quad (3.57e)$$

In particular, one can measure the orthogonal observables $h - v$ and 1_X using the expansions,

$$h - v = \int_Y v_-(y) dE(y), \quad (3.58)$$

$$1_X = h + v = \int_Y v_+(q) dE(y). \quad (3.59)$$

For the specific case of an initial Gaussian beam centered at zero, we have,

$$p(y) = \exp\left(-\frac{y^2}{2\sigma^2}\right) / \sigma\sqrt{2\pi}, \quad (3.60a)$$

$$\epsilon = (\epsilon_h + \epsilon_v)/2, \quad (3.60b)$$

$$\delta = (\epsilon_h - \epsilon_v)/2, \quad (3.60c)$$

$$a = \frac{1}{2\sigma\sqrt{\pi}}, \quad (3.60d)$$

$$b(\epsilon) = a \exp(-(\epsilon/\sigma)^2), \quad (3.60e)$$

$$v_-(y) = \sqrt{2} \frac{\exp(-\frac{(y-\delta)^2}{2\sigma^2}) \sinh(\frac{\epsilon(y-\delta)}{\sigma^2})}{\sinh(\frac{\epsilon^2}{2\sigma^2})}, \quad (3.60f)$$

$$v_+(y) = \sqrt{2} \frac{\exp(-\frac{(y-\delta)^2}{2\sigma^2}) \cosh(\frac{\epsilon(y-\delta)}{\sigma^2})}{\cosh(\frac{\epsilon^2}{2\sigma^2})}, \quad (3.60g)$$

What matters for the measurement is the average translation ϵ away from the midpoint $(y - \delta)$. The amplification of the CVs is controlled by the parameter ϵ/σ , which serves as an indicator for the ambiguity of the measurement. When the shift ϵ is large compared to the width of the Gaussian σ , then $\epsilon/\sigma \gg 1$; the shifted Gaussians for h and v are distinguishable; the CVs approach the eigenvalues of

the measurement; and, the measurement is unambiguous. When the shift is small compared to the width of the Gaussian, then $\epsilon/\sigma \ll 1$, the Gaussians for h and v largely overlap, the CVs diverge, and the measurement is ambiguous. Fig. 3.2 shows the CVs (3.60f) for the Gaussian initial beam profile, as well as for a Laplace and top-hat profile for comparison.

This sort of detection protocol was used in the original paper on weak values [28] in the form of a Stern-Gerlach apparatus that measures spin analogously to polarization using a continuous momentum displacement generated by a magnetic field. The initial Gaussian beam profile shifted an amount ϵ away from the midpoint of the initial beam profile in a direction corresponding to the value of the spin. Since the beam profile was symmetric about its mean, the generic CVs $f_Y(y) = y/\epsilon$ were implicitly assigned as a linear calibration of the detector, which targets a specific observable analogous to $h - v$. Motivating this implicit choice was the fact that when ϵ is sufficiently small, the two overlapping Gaussians produce to a good approximation a single resulting Gaussian with a shifted mean consistent with such a linear scaling, as shown in Fig. 3.3. That such a choice was being made was later pointed out explicitly in [34] before we identified the role of the CVs in [68] and derived the preferred form (3.60f). The proposed spin measurement protocol was adapted to a polarization measurement using a calcite crystal, as we have developed in this section, and then verified experimentally [29, 44].

To produce the weak value from the polarization measurement, we postselect on a second measurement to form a conditioned average. If the initial polarization state is pure with a density $\rho = x = |x\rangle\langle x|$ and the final postselection is also pure $z = |z\rangle\langle z|$, then we have the form,

$${}_z\langle\widetilde{F}_X\rangle_x = \frac{\int_Y f_Y(y) |\langle z | M(y) | x \rangle|^2 dy}{\int_Y |\langle z | M(y) | x \rangle|^2 dy}. \quad (3.61)$$

If we choose the symmetric Gaussian case (3.60) with $\delta = 0$ and take the form of $M(y)$ without additional unitary disturbance,

$$M(y) = \frac{1}{\sqrt{\sigma\sqrt{2\pi}}} \left[h \exp\left(-\frac{(y-\epsilon)^2}{4\sigma^2}\right) + v \exp\left(-\frac{(y+\epsilon)^2}{4\sigma^2}\right) \right], \quad (3.62)$$

and relate both pure states to the reference state h via unitary rotations as defined in (3.4), $x = \mathcal{U}_{\alpha,\beta,\gamma}(h)$ and $z = \mathcal{U}_{\alpha',\beta',\gamma'}(h)$, then the postselected probability density ${}_z\tilde{p}_x(y)$ takes the form,

$$\begin{aligned} |\langle z | M(y) | x \rangle|^2 &= \frac{\exp\left(-\frac{y^2+\epsilon^2}{2\sigma^2}\right)}{2\sigma\sqrt{2\pi}} \times \\ &\quad \left((1 + \cos\beta \cos\beta') \cosh\frac{y\epsilon}{\sigma^2} + (\cos\beta + \cos\beta') \sinh\frac{y\epsilon}{\sigma^2} \right. \\ &\quad \left. + \sin\beta \sin\beta' \cos(\gamma - \gamma') \right), \end{aligned} \quad (3.63)$$

Choosing the CVs (3.60f) to target the observable $h - v$, the conditioned average (3.61) then takes the form,

$${}_z \left\langle \widetilde{h - v} \right\rangle_x = \frac{\cos\beta + \cos\beta'}{1 + \cos\beta \cos\beta' + \Xi(\epsilon, \sigma)}, \quad (3.64a)$$

$$\Xi(\epsilon, \sigma) = \sin\beta \sin\beta' \cos(\gamma - \gamma') \exp\left(-\frac{\epsilon^2}{2\sigma^2}\right). \quad (3.64b)$$

The interference term $\Xi(\epsilon, \sigma)$ in the denominator is the only part of the conditioned average that depends on the details of the measurement context through the exponential dependence on ϵ/σ , which was also noted in [37, 103]. This conditioned average can exceed the eigenvalue range of the observable due to the combination of the amplified CVs and the disturbance linking the incompatible frameworks in the conditional probabilities. Fig. 3.4 shows the Gaussian measurement of the conditioned average (3.64), as well as top-hat and triangular measurements for comparison.

The conditioned average (3.64) has two limiting cases that eliminate the ex-

PLICIT context-dependence: (1) In the strong-measurement limit, $\epsilon/\sigma \rightarrow \infty$, the interference term vanishes, leaving a conditioned average of projective measurements that always stays in the eigenvalue range of the observable. (2) In the weak-measurement limit, $\epsilon/\sigma \rightarrow 0$, the conditioned average reduces to the *weak value*,

$${}_z\langle h - v \rangle_x^w = \operatorname{Re} \frac{\langle z | (h - v) | x \rangle}{\langle z | x \rangle} = \frac{\cos \beta + \cos \beta'}{1 + \cos \beta \cos \beta' + \sin \beta \sin \beta' \cos(\gamma - \gamma')}. \quad (3.65)$$

The weak value is distinguished by being the only case that can be written entirely in terms of the observable, the post-selection, and the pre-selected state without reference to the intermediate measurement. In this sense, it is the only context-independent form of the conditioned average. The derivation in the next section shows exactly under what conditions the weak value can be attained as such a limit point of a conditioned average.

3.5 The Weak Value as a Conditioned Average

As we have seen for the case of the calcite detector (3.65), the weak value (3.43) seems to arise naturally as the weak limit of post-selected conditioned averages. Indeed, much of the existing literature on weak values (e.g. [28, 30–32, 35, 36, 66]) operates under the assumption that it is the only weak limit of a conditioned average, or that it is a well-defined property of a pre- and postselected ensemble prior to the ensemble being measured. However, a conditioned average does not necessarily converge to the weak value in the weak measurement limit, as has been noted independently by several groups [33, 34, 37, 42, 68, 104], making its interpretation as a well-defined property worthy of more careful consideration. To obtain correct laboratory predictions for a conditioned average, the formula (3.40)

must be used, which generally requires the specification of the detection strategy and the protocol for assigning contextual values to target a specific observable.

Despite the interpretational controversy, the weak value (3.43) is distinguished by being a *context independent* weak limit of the conditioned average that is easy to compute theoretically and appears quite commonly in typical laboratory situations. The formal expression of the weak value can also appear in other measurement scenarios, such as in “modular values” [105], or even perturbative corrections to energy spectra⁵, which makes it an independently interesting quantity to study. We will examine it more closely in Chapter 6.

We will now demonstrate how the weak value (3.43) naturally appears from the general conditioned average (3.40) under a broad range of measurement conditions.

Preliminaries.—First we note from (3.20c) that each measurement operator has a polar decomposition, $M_{y,y'} = U_{y,y'}|M|_{y,y'}$, in terms of a unitary operator $U_{y,y'}$ and a positive operator $|M|_{y,y'}$. It then follows that,

$$\begin{aligned} M_{y,y'}^\dagger E'_z M_{y,y'} &= |M|_{y,y'} U_{y,y'}^\dagger E'_z U_{y,y'} |M|_{y,y'}, \\ &= \{|M|_{y,y'}^2, \mathcal{U}_{y,y'}(E'_z)\}/2 - [|M|_{y,y'}, [|M|_{y,y'}, \mathcal{U}_{y,y'}(E'_z)]]/2, \end{aligned} \quad (3.66)$$

where $\{A, B\} = AB + BA$ is the anticommutator, $[A, B] = AB - BA$ is the commutator, and $\mathcal{U}_{y,y'}(E'_z) = U_{y,y'}^\dagger E'_z U_{y,y'}$ is a unitary rotation of the postselection.

Next we make the following assumptions regarding the dependence of the relevant quantities on some strength parameter ϵ :

1. The measurement operators $M_{y,y'}$ are analytic functions of ϵ , and thus have

⁵If $|E\rangle$ is an eigenstate of a Hamiltonian H with energy E , and $|E'\rangle$ is an eigenstate of the perturbed Hamiltonian $\hat{H}' = \hat{H} + \hat{\Delta}$ with energy E' , then $\langle E' | \hat{H}' | E \rangle = E' \langle E' | E \rangle = E \langle E' | E \rangle + \langle E' | \hat{\Delta} | E \rangle$, so we can conclude that the perturbation in the eigenenergy corresponds to the (purely real) weak value of the perturbation $E' = E + \langle E' | \hat{\Delta} | E \rangle / \langle E' | E \rangle$.

well defined Taylor expansions around $\epsilon = 0$ such that they are proportional to the identity in the weak limit, $\forall y, y', \lim_{\epsilon \rightarrow 0} M_{y,y'} \propto 1_X$.

2. The unitary parts of the measurement operators $U_{y,y'} = \exp(iG_{y,y'}(\epsilon))$ are generated by Hermitian operators of order ϵ^k , $G_{y,y'}(\epsilon) = \epsilon^k G_{y,y'}^{(k)} + O(\epsilon^{k+1})$, for some integer $k \geq 1$. Furthermore, each $U_{y,y'}$ must commute with either the system state or the postselection, $\forall y, y', [U_{y,y'}, \rho_X] = 0$, or $\forall y, y', [U_{y,y'}, E'_z] = 0$.

Assertion.—Given the above sufficient conditions, we have the following observation: *in the weak limit $\epsilon \rightarrow 0$ the context dependence of the conditioned average (3.40) often vanishes, producing the weak value (3.43) as a limit point.*

Proof.—To prove our assertion, we expand (3.40) to the minimum necessary order of ϵ^n and then take the weak limit as $\epsilon \rightarrow 0$. First, we expand (3.66) to order ϵ^n using assumptions (1), and (4),

$$M_{y,y'}^\dagger E'_z M_{y,y'} = c_{y,y'}^2 \mathcal{U}_{y,y'}(E'_z) + c_{y,y'} \{ |M|_{y,y'}^{(n)}, \mathcal{U}_{y,y'}(E'_z) \} \epsilon^n + O(\epsilon^{n+1}). \quad (3.67)$$

Generally, the remaining unitary rotation of the postselection will disturb the weak limit. However, if $[U_{y,y'}, E'_z] = 0$ as in assumption (2), then $\mathcal{U}_{y,y'}(E'_z) = E'_z$ and the unitary disturbance disappears. If instead $[U_{y,y'}, \rho_X] = 0$, then we can apply the state to (3.67) and find,

$$\left\langle M_{y,y'}^\dagger E'_z M_{y,y'} \right\rangle_X = c_{y,y'}^2 \langle \mathcal{U}_{y,y'}(E'_z) \rangle_X + c_{y,y'} \left\langle \{ |M|_{y,y'}^{(n)}, \mathcal{U}_{y,y'}(E'_z) \} \right\rangle_X \epsilon^n + O(\epsilon^{n+1}). \quad (3.68)$$

Since $\langle \mathcal{U}_{y,y'}(E'_z) \rangle_X = \text{Tr}_X(\mathcal{U}^\dagger(\rho_X) E'_z) = \langle E'_z \rangle_X$, the first term simplifies. The unitary rotation in the second term expands to $\mathcal{U}_{y,y'}(E'_z) = E'_z + O(\epsilon^k)$, and the $O(\epsilon^k)$ correction can be absorbed into the overall $O(\epsilon^{n+1})$ correction.

Therefore, after summing over y' we find up to corrections of order ϵ^{n+1} ,

$$\sum_{y'} \left\langle M_{y,y'}^\dagger E'_z M_{y,y'} \right\rangle_X = \langle \{E_y(\epsilon), E'_z\} \rangle_X / 2, \quad (3.69)$$

where the probability observable has the expansion to order ϵ^n ,

$$E_y(\epsilon) = \sum_{y'} |M_{y,y'}|^2(\epsilon) = \sum_{y'} (c_{y,y'}^2 1_X + 2c_{y,y'} |M_{y,y'}|^{(n)} \epsilon^n + O(\epsilon^{n+1})). \quad (3.70)$$

Inserting (3.69) into (3.40), we find,

$${}_z \langle F_X \rangle = \frac{\langle \{F_X, E'_z\} / 2 \rangle_X + \sum_y f_Y(\epsilon; y) O(\epsilon^{n+1})}{\langle \{1_X, E'_z\} / 2 \rangle_X + O(\epsilon^{n+1})}, \quad (3.71)$$

where we have simplified $\sum_y f_Y(\epsilon; y) E_y(\epsilon) = F_X$ in the numerator, and $\sum_y E_y(\epsilon) = 1_X$ in the denominator. Hence, unless the CVs in the numerator have poles larger than $1/\epsilon^n$ the correction terms of order ϵ^{n+1} will vanish, producing the weak value (3.43) in the weak limit $\epsilon \rightarrow 0$, as claimed. The conditions for ensuring that these poles do not appear are explored in Appendix B for completeness.

Exceptions.—As the assertion indicates, the weak value will arise as the weak limit of a conditioned average in many common laboratory situations, which explains its stability in the literature. However, different weak limits are still possible. For example, an ϵ -dependent unitary disturbance in the measurement will effectively rotate the post-selection to a different framework for each measurement outcome, creating additional terms in the weak limit. Similarly, CVs that diverge more rapidly than $1/\epsilon^n$ produce additional terms in the weak limit. (See, for example, Ref. [45] or the semi-weak limit of Chapter 5.) This latter case can happen either from a pathological choice of CVs by the experimenter in the case of redundancy, or if the probability observables are insufficiently correlated to F_X when higher-order terms in ϵ are neglected.

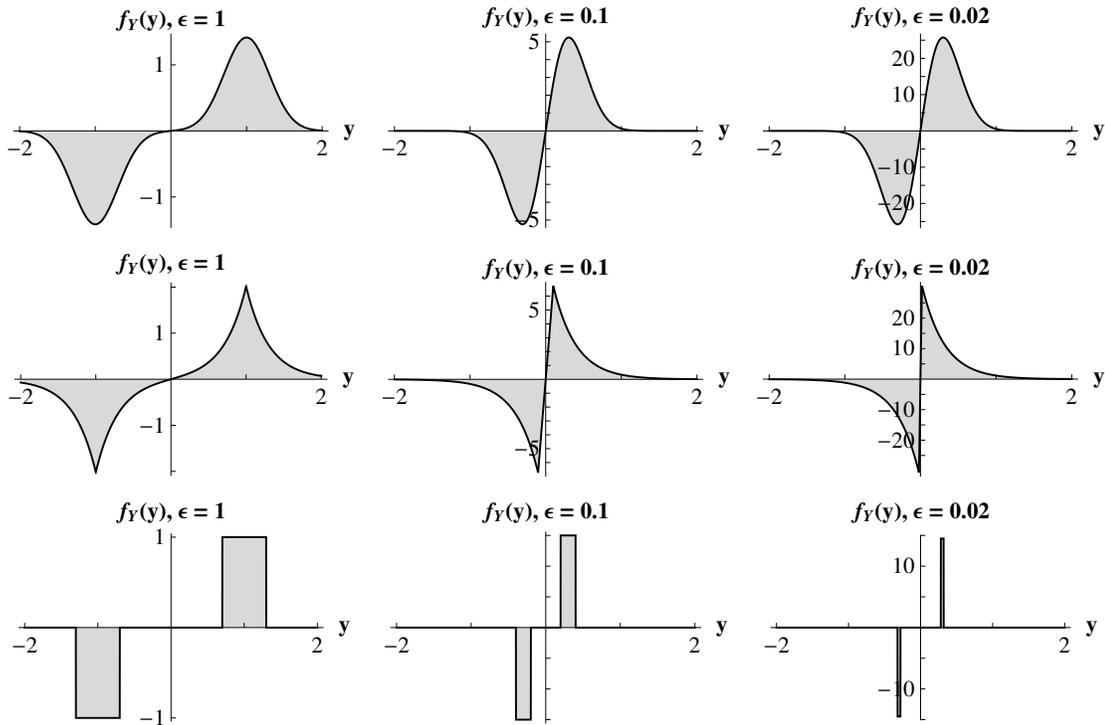


Figure 3.2: Preferred CVs $f_Y(y)$ given in (3.57c) for a calcite position measurement that targets the polarization observable $F_X = h - v$, shown for strong separation ($\epsilon = 1$), wimpy separation ($\epsilon = 0.1$), and weak separation ($\epsilon = 0.02$) of the polarizations. *Top Row*: Initial Gaussian beam profile. *Middle Row*: Initial Laplace beam profile. *Bottom Row*: Initial top-hat beam profile. Note that the top-hat CVs are the eigenvalues of ± 1 under strong separation, but become amplified as the distributions start to overlap; moreover, the top-hat CVs cancel out in the perfectly ambiguous overlapping region. The amplification and cancellation behavior of the CVs is more complicated for less definite detector profiles.

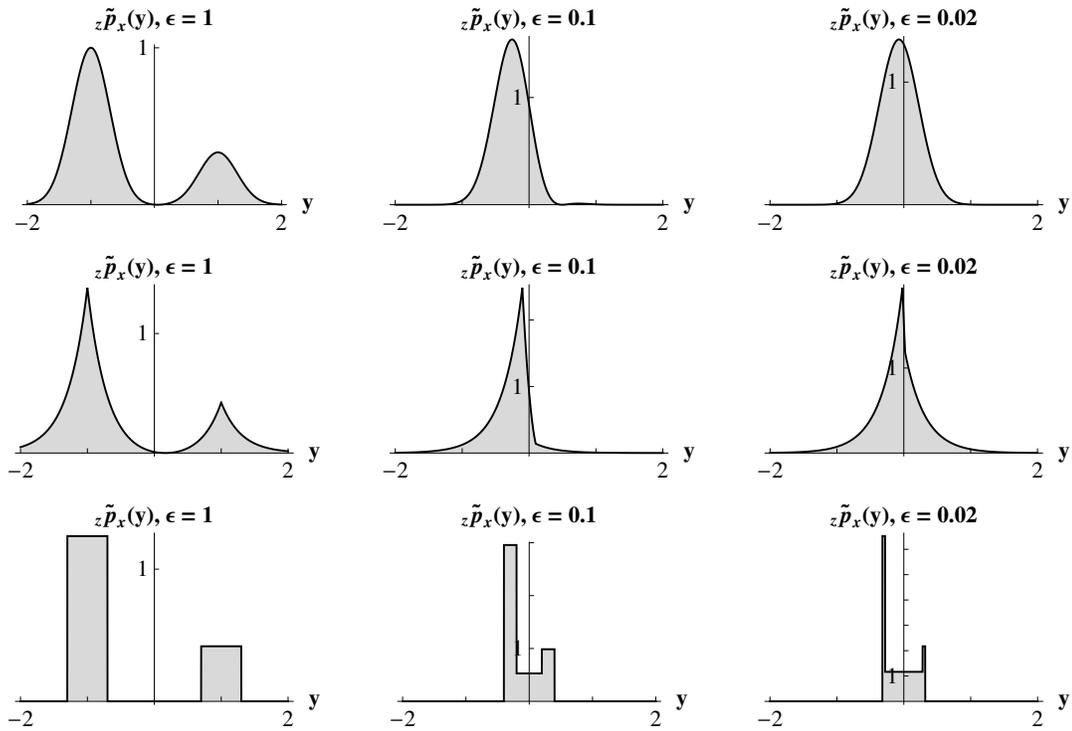


Figure 3.3: Pre- and postselected detector probability densities $z\tilde{p}_x(y)$ for the calcite position measurement (3.56), shown for strong separation ($\epsilon = 1$), wimpy separation ($\epsilon = 0.1$), and weak separation ($\epsilon = 0.02$) of the polarizations. The preselection is $x = |x\rangle\langle x|$ with associated vector $|x\rangle = \cos(4\pi/6)|h\rangle + \sin(4\pi/6)|v\rangle$. The postselection is $z = |z\rangle\langle z|$ with associated vector $|z\rangle = (|h\rangle + |v\rangle)/\sqrt{2}$. *Top Row*: Initial Gaussian beam profile. *Middle Row*: Initial Laplace beam profile. *Bottom Row*: Initial top-hat beam profile. Note that the Gaussian profile tilts to approximate a single shifted Gaussian under weak separation, as leveraged in the weak measurement protocol introduced in [28].

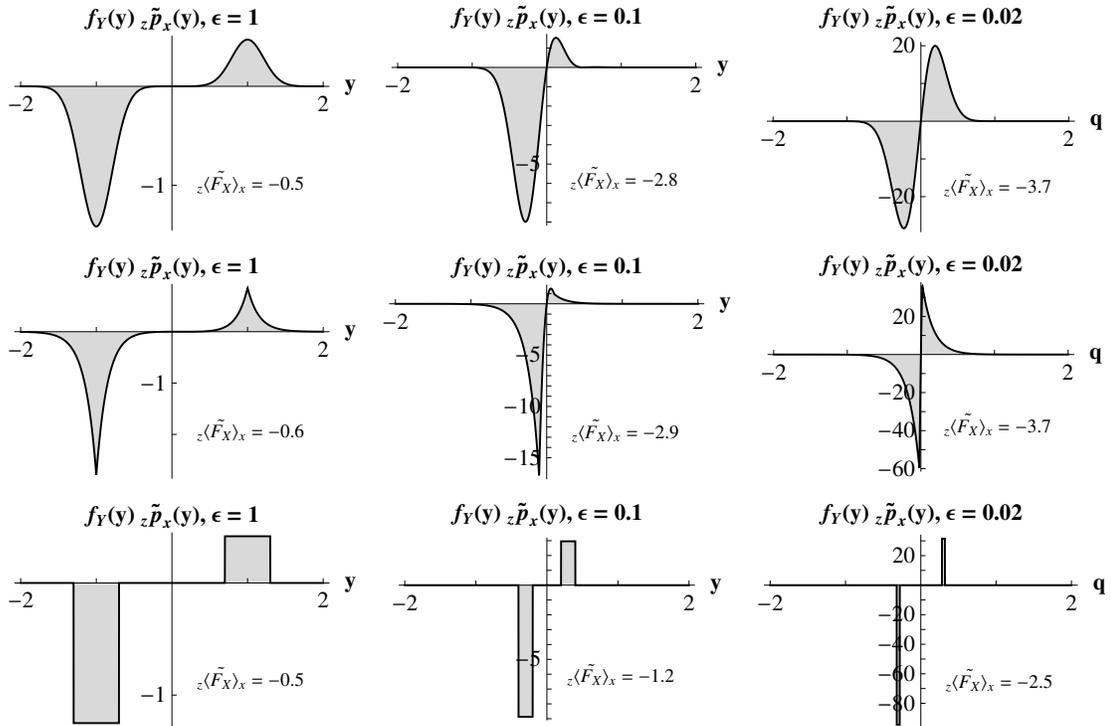


Figure 3.4: Pre- and postselected conditioned average densities $f_Y(y) z\tilde{p}_x(y)$ for a calcite position measurement targeting the observable $F_X = h - v$ with CVs as in Fig. 3.2, shown for strong separation ($\epsilon = 1$), wimpy separation ($\epsilon = 0.1$), and weak separation ($\epsilon = 0.02$) of the polarizations. The conditioned averages $z\langle\tilde{F}_X\rangle_x = \int_Y f_Y(y) z\tilde{p}_x(y) dy$ are the areas under the curves and are shown inset. As in Fig. 3.3, the preselection is $x = |x\rangle\langle x|$, where $|x\rangle = \cos(4\pi/6) |h\rangle + \sin(4\pi/6) |v\rangle$. The postselection is $z = |z\rangle\langle z|$, where $|z\rangle = (|h\rangle + |v\rangle)/\sqrt{2}$. *Top Row*: Initial Gaussian beam profile. *Middle Row*: Initial Laplace beam profile. *Bottom Row*: Initial top-hat beam profile. For sufficiently strong separation all three detector profiles will produce the strong conditioned average $z\langle\tilde{F}_X\rangle_x = -1/2$. For weak separation all three profiles approximate the weak value $z\langle F_X\rangle_x^w = -2 - \sqrt{3} \approx -3.73$. However, the different detector profiles converge to the weak value at different rates with decreasing ϵ .

4 Generalized Leggett-Garg Inequalities

Thus, in the present author's view, of the three major classes of 'resolution' of the quantum measurement paradox, the 'orthodox' one involves a major logical fallacy and the 'many-worlds' interpretation is simply a meaningless collage of words. The 'statistical' interpretation, if taken to its logical conclusion, is internally consistent but conflicts rather violently with the 'realistic' intuitions which most practising physicists probably find not only philosophically congenial, but almost essential, psychologically, in their everyday work. Thus, one is led to consider the possibility that the fundamental premise of the argument is wrong: that is, that the linear formalism of QM does not apply in unmodified form to macroscopic systems in the same way as it does to their microscopic constituents.

Anthony J. Leggett, (2002) [106]

To better understand and identify the apparent division between macroscopic and microscopic behavior, Leggett and Garg have distilled common implicit assumptions about the macroscopic world into a set of explicit postulates that they dub *macrorealism* (MR) [106, 107]. From these postulates, they construct inequalities analogous to Bell inequalities [61, 63, 108, 109] but involving multiple correlations in time. Such Leggett-Garg inequalities (LGIs) must be satisfied by any theory compatible with MR, but may be violated by quantum mechanics. As such, LGI violations have received increasing interest as signatures of distinctly quantum behavior in qubit implementations [16, 103, 110, 111], and have been recently confirmed experimentally in both solid-state [112] and optical systems [57].

Here we demonstrate a technique for systematically deriving generalized LGIs that admit multiple parties, invasive detection, and/or ambiguous detector results by considering a specific two-particle experimental setup with three measurements. We proceed to experimentally violate several such two-party LGIs simultaneously with a single data set produced from a setup using a *semi-weak* polarization measurement on an entangled biphoton state. The contextual values (CV) technique for observable measurement that we developed in Chapters 2 and 3 suggests a direct comparison between the classical and quantum treatments. Finally, we show that specific two-party LGIs are equivalent to constraints on convex sums of conditioned averages (CA), which, as we have shown, are the generalizations of the quantum weak value [28, 68] to an arbitrary measurement setup. The technique may be easily extended to check data from a setup with any number of measurements and parties.

4.1 Macro-realism

As introduced by Leggett, a *macro-realistic theory* consists of three key postulates [106, 107]:

1. If an object has several distinguishable states available to it, then at any given time it is in only one of those states.
2. One can *in principle* determine which state it is in without disturbing that state or its subsequent dynamics.
3. Its future state is determined causally by prior events.

In addition to these postulates, we acknowledge that physical detectors may be imperfect by being (a) *invasive* by altering the object state during the interaction, or (b) *ambiguous* by reporting results that only correlate probabilistically with the

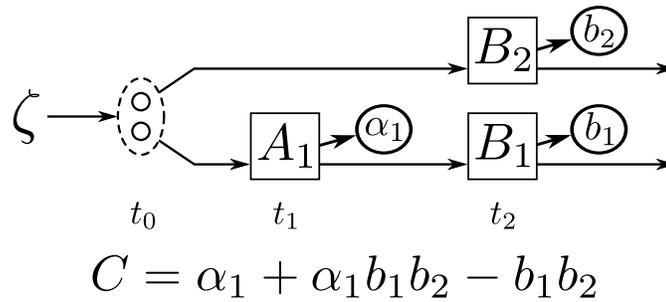


Figure 4.1: MR measurement schematic. An object pair is picked from an ensemble ζ at time t_0 . At t_1 object 1 of the pair interacts with an imperfect detector for the property A_1 , which reports a contextual value α_1 . At t_2 both objects interact with unambiguous detectors for the properties B_1 and B_2 that report values b_1 and b_2 . The two-party LG correlation C is constructed from the measured results.

object state due to inherent detector inefficiencies or errors. We will show that we can derive generalized Leggett-Garg inequalities by assuming that one or the other of these imperfections does not apply.

For convenience we consider dichotomic properties in what follows, though the discussion can be easily extended. Unambiguous detector outcomes will be assigned the (arbitrary) values $\{-1, 1\}$ corresponding to the two possible states of the property being measured. Ambiguous detectors will be calibrated to report the same ensemble average as an unambiguous detector for the same property. To do so, their outcomes must be assigned *contextual values* $\alpha \in S$ from an expanded set S , with $\min S \leq -1$ and $\max S \geq 1$, to compensate for the imperfect state correlation of the outcomes. The contextual values may be determined by measuring pure ensembles of either ± 1 , as discussed in Chapter 2.

4.2 Leggett-Garg Inequalities

We now derive a specific two-party generalized LGI for a particular experimental setup, keeping in mind that the method may be extended to any setup. Consider a

pair of MR objects that interacts with a sequence of detectors as shown in Fig. 4.1. At time t_0 the pair is picked from a known ensemble ζ . At time t_1 object 1 of the pair interacts with an imperfect detector for the dichotomic property A_1 , which reports a contextual value $\alpha_1 \in S_1$. Finally, at time t_2 objects 1 and 2 interact with unambiguous detectors for the dichotomic properties B_1 and B_2 , respectively, which report the values $b_1, b_2 \in \{-1, 1\}$.

For each object pair, we can keep all three results to construct the correlation product $\alpha_1 b_1 b_2$, or we can ignore some results as non-selective measurements [19] to construct the alternate quantities α_1 , b_1 , b_2 , $\alpha_1 b_1$, $\alpha_1 b_2$, or $b_1 b_2$. Since the latter terms involve voluntary loss of information after the measurement has been performed, we can compute them all from the *same data set*. Exploiting this freedom, we construct the correlation $C = \alpha_1 + \alpha_1 b_1 b_2 - b_1 b_2$ for each measured pair, which lies in the range, $-|1 - 2 \min S_1| \leq C \leq |2 \max S_1 - 1|$.

We repeat this procedure many times and average the results of C to obtain, $\langle C \rangle = \sum_{\alpha_1, b_1, b_2} P(\alpha_1|\zeta)P(b_1, b_2|\zeta, \alpha_1) (\alpha_1 + \alpha_1 b_1 b_2 - b_1 b_2)$, where $P(\alpha_1|\zeta)$ is the probability of detecting α_1 given the initial ensemble ζ , and $P(b_1, b_2|\zeta, \alpha_1)$ is the probability of detecting b_1 and b_2 given the initial ensemble ζ and the possibly invasive detection of α_1 .

Generally, we cannot separate the sums due to the α_1 -dependence of $P(b_1, b_2|\zeta, \alpha_1)$, so the best guaranteed bounds are $-|1 - 2 \min S_1| \leq \langle C \rangle \leq |2 \max S_1 - 1|$. As a special case, if the detector for A_1 is *unambiguous* then $\min S_1 = -1$, $\max S_1 = 1$, and we find the LGI,

$$-3 \leq \langle A_1 + A_1 B_1 B_2 - B_1 B_2 \rangle \leq 1. \quad (4.1)$$

Alternatively, if we assume that the detector is ambiguous but *noninvasive*¹

¹Note that our noninvasiveness assumption is stronger than the original MR requirement 2, which need not be satisfied by every measurement strategy.

so that $P(b_1, b_2 | \zeta, \alpha_1) = P(b_1, b_2 | \zeta)$ then the sums do separate and we can average A_1 first to find,

$$\langle C \rangle = \sum_{b_1, b_2} P(b_1, b_2 | \zeta) (\langle A_1 \rangle (1 + b_1 b_2) - b_1 b_2). \quad (4.2)$$

Since $-1 \leq \langle A_1 \rangle \leq 1$, each term can take only three possible values $\{-3, -1, 1\}$ and we again recover (4.1). Therefore, any violation of (4.1) will imply that at least one of the postulates (1-3) of MR does not hold, or that the detector for A_1 is *both* invasive and ambiguous.

We can construct many similar LGIs from the same set of data. For example, the three detectors in Fig. 4.1 allow the construction of the $2^3 - 1$ nontrivial correlation terms listed earlier, which can be combined with the three coefficients $\{-1, 0, 1\}$ ². Ignoring an overall sign, we can construct $(3^{2^3-1} - 1)/2 = 1093$ nonzero LGI correlations bounded in a similar manner to (4.1). The subset of $(3^{2^2-1} - 1)/2 = 13$ single-object LGIs can be obtained by ignoring the B_2 detector. Furthermore, if a fourth detector for A_2 were added before the detector for B_2 , we could test $(3^{2^4-1} - 1)/2 = 7174453$ such LGIs. One is formally identical to the CHSH-Bell inequality [109] (see also [114]), but tests MR and not Bell-locality.

For contrast, the original approach in [107] combines separate experiments for each correlation between *ideal* detectors to form a single LGI. Our approach uses a single experimental setup to determine all $2^M - 1$ correlations between M *general* sequential detectors to form a large number of LGIs. Hence we obtain an exponential improvement in experimental complexity for large M .

²Allowing other coefficients, as suggested in [113], produces even more possibilities.

4.3 Conditioned Averages

A single-object LGI, $-3 \leq \langle A_1 + A_1 B_1 - B_1 \rangle \leq 1$, was considered in [16, 103] and shown to have a one-to-one correspondence with an upper bound to the average of A_1 conditioned on the positive value of B_1 : ${}_{+1}\langle A \rangle \leq 1$. Three other LGIs similarly correspond to the bounds ${}_{-1}\langle A \rangle \geq -1$, and $-1 \leq {}_{-1}\langle A \rangle \leq 1$, as checked experimentally in [57].

We now extend these results to the two-object case using (4.1). First we define a marginal probability of measuring b_1 and b_2 given any result of A_1 as $P(b_1, b_2 | \zeta, A_1) = \sum_{\alpha_1} P(\alpha_1 | \zeta) P(b_1, b_2 | \zeta, \alpha_1)$. Then we define a conditional probability of measuring α_1 given the measurement of b_1 and b_2 as, $P(\alpha_1 | \zeta, b_1, b_2) = P(\alpha_1 | \zeta) P(b_1, b_2 | \zeta, \alpha_1) / P(b_1, b_2 | \zeta, A_1)$. Therefore, the average of A_1 conditioned on the measurements of b_1 and b_2 is ${}_{b_1, b_2}\langle A_1 \rangle = \sum_{\alpha_1} P(\alpha_1 | \zeta, b_1, b_2) \alpha_1$.

Using this definition, we rewrite the upper bound of (4.1) as,

$$\sum_{b_1, b_2} P(b_1, b_2 | \zeta, A_1) \left({}_{b_1, b_2}\langle A_1 \rangle (1 + b_1 b_2) - b_1 b_2 \right) \leq 1, \quad (4.3)$$

and insert the possible values for b_1 and b_2 to find the CA constraint,

$${}_{+1, +1}\langle A_1 \rangle p^+ + {}_{-1, -1}\langle A_1 \rangle p^- \leq 1, \quad (4.4)$$

where $p^\pm = P(\pm 1, \pm 1) / (P(1, 1) + P(-1, -1))$, and $P(i, j) = P(i, j | \zeta, A_1)$. The degeneracy of the product value $b_1 b_2$ results in an upper bound for a *convex sum* of CAs, in contrast to the single-object result in [16, 103]. A *sufficient* condition for violating (4.4) is for both CAs to exceed 1 simultaneously. Conversely, if all CAs were bounded by 1, then it would be impossible to violate (4.4) or (4.1).

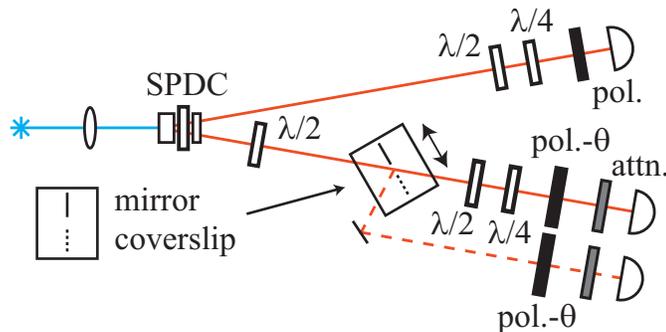


Figure 4.2: (color online) Experimental setup. A 488 nm laser produces degenerate down-converted photon pairs. The polarization of the photon in the lower arm is rotated by 45° with a half-wave plate, then undergoes semi-weak polarization measurement in the $\{h, v\}$ basis using Fresnel reflection (A_1) that encodes the information in the resulting spatial modes, and is finally projected into the $\{\theta, \theta_\perp\}$ basis with polarizers set at angle θ (B_1). The polarization of the photon in the upper arm is projected into the $\{h, v\}$ basis with another polarizer (B_2). The half and quarter waveplates prior to the polarizers are used for tomography of the input state; during data collection they are removed from the lower arm and used to switch between h and v polarization in the upper arm.

4.4 Quantum Formulation

Projective quantum measurements produce averages of eigenvalues analogous to the results of an unambiguous detector, but non-projective quantum measurements produce averages of *contextual values* as shown in Chapter 3 which need not lie in the eigenvalue range and are therefore analogous to the results of an ambiguous detector. By measuring A_1 weakly we can find quantum mechanical violations of (4.1) and (4.4).

Specifically, if we start with a 2-object density operator $\hat{\rho}$ and measure A_1 generally such that $\hat{A}_1 = \sum_{a_1} a_1 \Pi_{a_1} = \sum_{\alpha_1} \alpha_1 \hat{E}_{\alpha_1}$ (where $\{a_1\}$ are the eigenvalues corresponding to the projections $\{\Pi_{a_1}\}$ and $\{\alpha_1\}$ are the CV corresponding to the POVM $\{\hat{E}_{\alpha_1} = \hat{M}_{\alpha_1}^\dagger \hat{M}_{\alpha_1}\}$), and then measure $B_1 B_2$ projectively such that $\hat{B}_1 \otimes \hat{B}_2 = \sum_{b_1, b_2} b_1 b_2 \Pi_{b_1} \otimes \Pi_{b_2}$, we will find that the average correlation $\langle C \rangle =$

$\langle A_1 + A_1 B_1 B_2 - B_1 B_2 \rangle$ has the form,

$$\langle C \rangle = \sum_{\alpha_1, b_1, b_2} P(\alpha_1; b_1, b_2 | \hat{\rho}) (\alpha_1 + \alpha_1 b_1 b_2 - b_1 b_2), \quad (4.5)$$

where $P(\alpha_1; b_1, b_2 | \hat{\rho}) = \text{Tr} \left(\left(\hat{M}_{\alpha_1}^\dagger \Pi_{b_1} \hat{M}_{\alpha_1} \otimes \Pi_{b_2} \right) \hat{\rho} \right)$ is the probability of measuring outcome α_1 of the general measurement of A , followed by a joint projection of $b_1 b_2$. The appearance of the CV instead of the eigenvalues of \hat{A} in (4.5) combined with the non-separable probability $P(\alpha_1; b_1, b_2 | \hat{\rho})$ allows violations of the LGI (4.1).

The left side of (4.4) follows from (4.5), with the marginalized probability $P(b_1, b_2 | \hat{\rho}, A_1) = \sum_{\alpha_1} P(\alpha_1; b_1, b_2 | \hat{\rho})$ and with the quantum CA ${}_{b_1, b_2} \langle A_1 \rangle = \sum_{\alpha_1} \alpha_1 P(\alpha_1; b_1, b_2 | \hat{\rho}) / P(b_1, b_2 | \hat{\rho}, A_1)$ defined in (3.40), which converges to a *weak value* [28] (3.43) in the limit of minimal measurement disturbance, as shown in Section 3.5.

4.5 Experimental Setup

To implement Fig. 4.1 we use the polarization of an entangled biphoton with the setup shown in Fig. 4.2. A glass microscope coverslip measures a Stokes observable A_1 semi-weakly as described below (and in Section 3.3.1), and polarizers measure Stokes observables B_1 and B_2 projectively. We produce degenerate non-collinear type-II down-conversion by pumping a 2 mm walkoff-compensated BBO crystal [115] with a narrowband 488 nm laser. The down-converted light passes through automated polarization analyzers and 3 nm bandpass filters at 976 nm in each arm before being coupled into multimode fibers connected to single photon avalanche photodiodes (SPAD). We detect coincidences using a 3 ns window. We perform state tomography with maximum likelihood estimation [116], which gives the state shown in Fig. 4.3 with concurrence $C = 0.794$, and purity $\text{Tr}(\hat{\rho}^2) = 0.815$, and

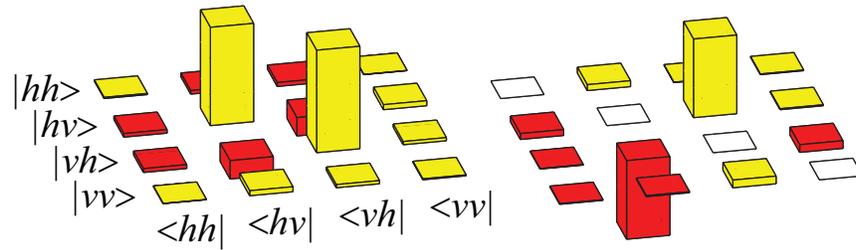


Figure 4.3: (color online) Real (left) and imaginary (right) parts of the reconstructed density matrix in the $\{h, v\}$ basis. Yellow and red represent positive and negative values, respectively.

which has a fidelity of 87% with the pure state vector $|\psi\rangle = (|hv\rangle - i|vh\rangle)/\sqrt{2}$.

After the state tomography, we remove the half- and quarter-wave plates from the lower arm and insert either a mirror or a coverslip using a computer-controlled translation stage. The reflected light passes through a polarization analyzer and couples into a third fiber and SPAD. We align the coverslip and the mirror to be parallel with an incidence angle of 40° relative to the incoming beam. Finally, we optimize the fiber incoupling and balance the collection efficiencies with attenuators so that the coincidences between the upper arm and either of the lower arms differ by only a few percent when the mirror is taken in and out of the beam path.

The coverslip acts as a polarization-dependent beamsplitter measuring $A_1 = \hat{\sigma}_z$. Averaging over the 3 nm bandwidth and the thickness variation ($\sim 150 \pm 0.6 \mu\text{m}$) produces an average Fresnel reflection similar to that of a single interface, with horizontal (h) polarization relative to the table exhibiting zero reflection near Brewster's angle and vertical (v) polarization exhibiting increasing reflection with incident angle.

For a pure state of polarization $|\psi\rangle = \alpha|h\rangle + \beta|v\rangle$ with $|\alpha|^2 + |\beta|^2 = 1$, the resulting state after passing through the coverslip is $|\psi'\rangle = (\gamma\alpha|h\rangle + \bar{\eta}\beta|v\rangle)|r\rangle - (\bar{\gamma}\alpha|h\rangle + \eta\beta|v\rangle)|t\rangle$, where $|j\rangle$, $j \in \{t, r\}$, specify the transmitted and reflected spatial modes of the coverslip, and the reflection and transmission probabilities

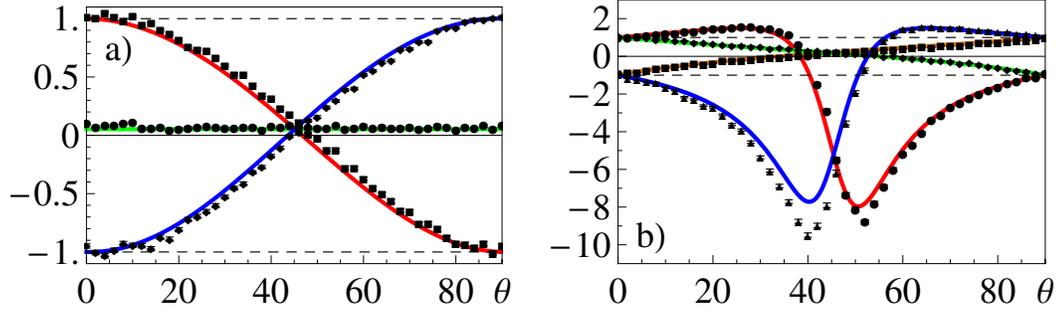


Figure 4.4: (color online) In all data plots, solid lines indicate theory and points indicate experimental data. a) $\langle \hat{\sigma}_z^{(1)} \rangle$ (green, flat), $\langle_{\theta} \hat{\sigma}_z^{(1)} \rangle$ (red, decreasing), and $\langle_{\theta_{\perp}} \hat{\sigma}_z^{(1)} \rangle$ (blue, increasing). b) $\langle_{\theta, h} \hat{\sigma}_z^{(1)} \rangle$ (red, bottom right), and $\langle_{\theta_{\perp}, v} \hat{\sigma}_z^{(1)} \rangle$ (blue, bottom left), violating negative bounds, unlike $\langle_{\theta_{\perp}, h} \hat{\sigma}_z^{(1)} \rangle$ (orange, increasing), and $\langle_{\theta, v} \hat{\sigma}_z^{(1)} \rangle$ (green, decreasing).

for h - and v -polarized light are $R_h = \gamma^2$, $R_v = \bar{\eta}^2$, $T_h = \bar{\gamma}^2$, and $T_v = \eta^2$, such that $R_i + T_i = 1$. Written this way, the coverslip reflection can be viewed as a generalization of the weak measurement in [48] and discussed in [68].

From $|\psi'\rangle$, we find the measurement operators for the back-action of the coverslip outcomes to be $\hat{M}_r = \gamma\Pi_h + \bar{\eta}\Pi_v$ and $\hat{M}_t = \bar{\gamma}\Pi_h + \eta\Pi_v$, where Π_i , $i \in \{h, v\}$, are polarization projectors. The corresponding POVM elements are $\hat{E}_r = R_h\Pi_h + R_v\Pi_v$ and $\hat{E}_t = T_h\Pi_h + T_v\Pi_v$, with which we can expand the polarization Stokes operator as $\hat{\sigma}_z = \Pi_h - \Pi_v = \alpha_r\hat{E}_r + \alpha_t\hat{E}_t$, as discussed before (4.5), where $\alpha_r = (T_h + T_v)/(R_h - R_v)$ and $\alpha_t = -(R_h + R_v)/(R_h - R_v)$ are the CV.

We determine the values of R_h and R_v with calibration polarizers before the coverslip, yielding $R_h = 0.0390 \pm 0.0007$ and $R_v = 0.175 \pm 0.001$. The reflected arm is largely projected to v , while the transmitted arm is only weakly perturbed, making the total coverslip effect a *semi-weak measurement*. The CV, $\alpha_r = -13.1 \pm 0.1$ and $\alpha_t = 1.57 \pm 0.01$, are correspondingly amplified from the eigenvalues of $\hat{\sigma}_z$.

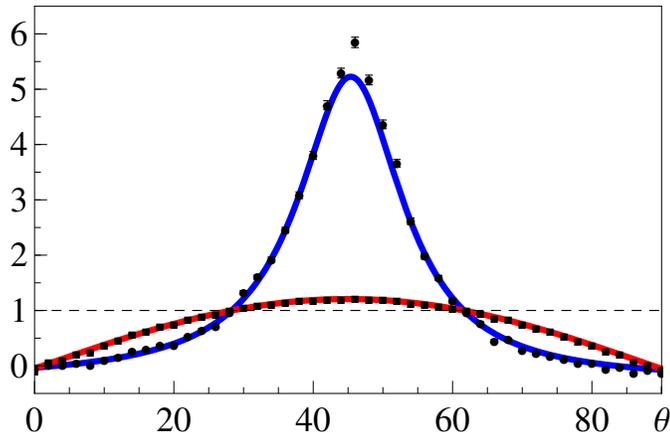


Figure 4.5: (color online) LGI correlation $\langle -\sigma_z^{(1)} - \sigma_z^{(1)} \sigma_\theta^{(1)} \sigma_z^{(2)} - \sigma_\theta^{(1)} \sigma_z^{(2)} \rangle$ (red, squares) and the corresponding convex sum of the CAs $\langle -\hat{\sigma}_z^{(1)} \rangle_{\theta,h}$ and $\langle -\hat{\sigma}_z^{(1)} \rangle_{\theta_\perp,v}$ (blue, circles), both violating their upper bounds of 1 in the same domain of θ . Compare to Fig. 4.4 (b) and note that the LGI violation includes the region where the two CAs both exceed their bounds.

4.6 Results

To complete the state preparation, we place a half-wave plate before the coverslip in the lower arm and rotate the polarization by 45° to produce a state similar to $|\psi''\rangle = (|ha\rangle + i|vd\rangle)/\sqrt{2}$. We then measure (4.1) by choosing the observables A_1 , B_1 , and B_2 to be the Stokes observables $\hat{\sigma}_z^{(1)}$, $\hat{\sigma}_\theta^{(1)}$ and $\hat{\sigma}_z^{(2)}$, respectively, where $\hat{\sigma}_\theta$ is the $\hat{\sigma}_z$ operator rotated to the $\{\theta, \theta_\perp\}$ basis (e.g. $\hat{\sigma}_{0^\circ} = \hat{\sigma}_z$ and $\hat{\sigma}_{45^\circ} = \hat{\sigma}_x$). By changing the single parameter, θ , we can explore a range of observables.

Fig. 4.4 shows the various averages of $\hat{\sigma}_z^{(1)}$. Averaging all results for orthogonal settings on $\hat{\sigma}_\theta^{(1)}$ and $\hat{\sigma}_z^{(2)}$ gives the expectation value $\langle \hat{\sigma}_z^{(1)} \rangle$, which is properly constant and near zero for all θ since the reduced density operator is almost fully mixed. Averaging only the results for the orthogonal settings of $\hat{\sigma}_z^{(2)}$ gives the single CAs $\langle \hat{\sigma}_z^{(1)} \rangle_\theta$ and $\langle \hat{\sigma}_z^{(1)} \rangle_{\theta_\perp}$, which are also well-behaved. Finally, averaging only the results for specific settings gives the double CAs $\langle \hat{\sigma}_z^{(1)} \rangle_{\theta,v}$, $\langle \hat{\sigma}_z^{(1)} \rangle_{\theta_\perp,h}$,

$_{\theta,v}\langle\hat{\sigma}_z^{(1)}\rangle$, and $_{\theta_{\perp},v}\langle\hat{\sigma}_z^{(1)}\rangle$, which can exceed the eigenvalue range for some range of θ due to the non-local correlations in the entangled biphoton state.

Using the same set of data, Fig. 4.5 shows the upper bound of the LGI $-3 \leq \langle -\sigma_z^{(1)} - \sigma_z^{(1)}\sigma_{\theta}^{(1)}\sigma_z^{(2)} - \sigma_{\theta}^{(1)}\sigma_z^{(2)} \rangle \leq 1$ being violated in the same range of θ that the appropriate convex sum of $_{\theta,h}\langle-\hat{\sigma}_z^{(1)}\rangle$ and $_{\theta_{\perp},v}\langle-\hat{\sigma}_z^{(1)}\rangle$ violates its upper bound according to (4.4).

We can violate several more LGIs using the same set of data as well. Fig. 4.6 shows two such correlations, $\langle \sigma_z^{(1)}\sigma_z^{(2)} + \sigma_z^{(2)}\sigma_{\theta}^{(1)} - \sigma_z^{(1)}\sigma_{\theta}^{(1)} \rangle$, and $\langle -\sigma_z^{(1)}\sigma_z^{(2)} + \sigma_z^{(2)}\sigma_{\theta}^{(1)} + \sigma_z^{(1)}\sigma_{\theta}^{(1)} \rangle$ that between them violate an upper bound over nearly the whole range of θ , for illustration.

All solid curves in Figures 4.4, 4.5, and 4.6 are quantum predictions analogous to (4.5) using the measurement operators, CV, and the reconstructed initial state. They also include compensation for a few percent deviation in the thickness of the half-wave plate in the upper arm. The points indicate experimental data and include Poissonian error bars. The small discrepancies between theory and data can be explained by sensitivity to the state reconstruction and additional equipment imperfections. The violations indicate either that MR is inconsistent with experiment or that the semi-weak measurement device is *both* invasive and ambiguous in the MR sense.

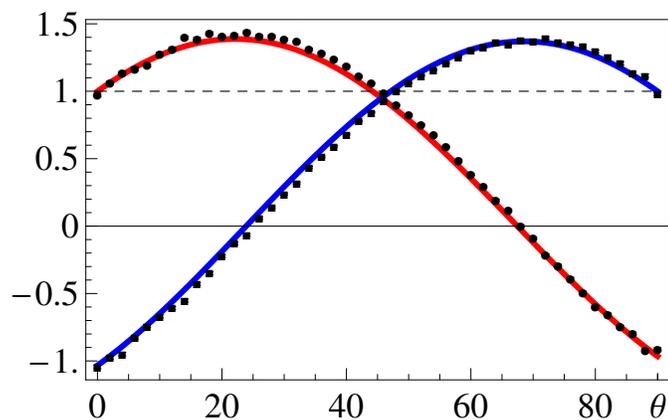


Figure 4.6: (color online) LGI correlations $\langle \sigma_z^{(1)} \sigma_z^{(2)} + \sigma_z^{(2)} \sigma_\theta^{(1)} - \sigma_z^{(1)} \sigma_\theta^{(1)} \rangle$ (red, circles), and $\langle -\sigma_z^{(1)} \sigma_z^{(2)} + \sigma_z^{(2)} \sigma_\theta^{(1)} + \sigma_z^{(1)} \sigma_\theta^{(1)} \rangle$ (blue, squares) violating their upper bounds of 1 for nearly the entire θ domain.

5 Electronic Mach-Zehnder Interferometry

... present quantum theory not only does not use—it does not even dare to mention—the notion of a “real physical situation.” Defenders of the theory say that this notion is philosophically naive, a throwback to outmoded ways of thinking, and that recognition of this constitutes deep new wisdom about the nature of human knowledge. I say that it constitutes a violent irrationality, that somewhere in this theory the distinction between reality and our knowledge of reality has become lost, and the result has more the character of medieval necromancy than of science.

Edwin T. Jaynes, (1980) [117]

The construction of electronic Mach-Zehnder interferometers (MZIs) in the solid state is a recent innovation in the fabrication and control of coherent mesoscopic systems. The first experiment of this kind, published by the Heiblum group [118], used the edge states [119] of an integer quantum Hall Corbino geometry as the electronic analog of light beams and quantum point contacts (QPCs) [120–122] as the electronic analogs of optical beam splitters to construct an interferometer with a visibility of 62%. Other interferometer designs have since been similarly implemented as electronic interferometers in the integer quantum Hall regime [123–126].

The electronic MZI differs from its optical counterpart in several respects. The arms of the MZI accumulate relative phase differences not only due to kinetic propagation of electrons along the arms, but also because the electrons

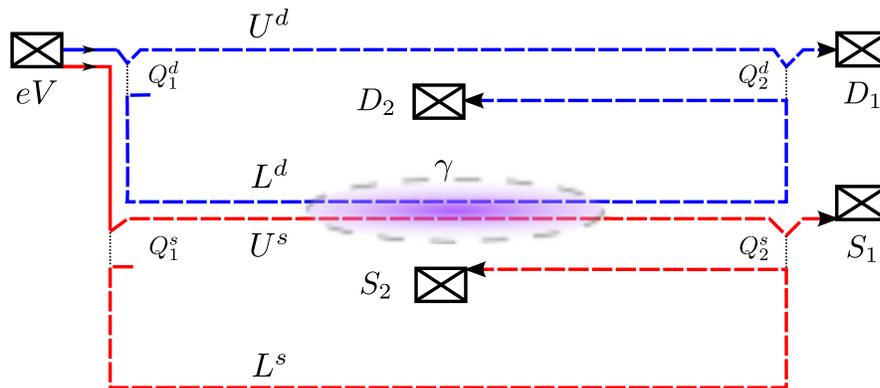


Figure 5.1: Schematic of coupled electronic MZIs. An ohmic source in a quantum Hall system at filling factor $\nu = 2$ injects chiral excitation pairs biased at energy eV relative to the ohmic reference drains D_1 , D_2 , S_1 , and S_2 into independent edge channels. The bias is kept low enough to allow only one excitation per channel on average. The outer (red) channel is transmitted entirely through the QPC Q_1^d and partially transmitted through Q_1^s and Q_2^s , forming the system MZI. The inner (blue) channel is reflected entirely from Q_1^s and Q_2^s and partially transmitted through Q_1^d and Q_2^d , forming a separate detector MZI. The Coulomb interaction between the copropagating arms L^d and U^s induces an average relative phase shift γ between each excitation pair that couples the interferometers.

are charged particles and can thus acquire a geometric Aharonov-Bohm phase [127, 128] when the arms enclose a magnetic flux. This charge can also lead to strong electron-electron interactions, giving rise to a variety of effects that have no counterpart in an optical MZI. For example, the interactions can produce differences in the counting statistics [129, 130], can induce temperature-dependent decoherence [131–136], can be used to detect external charges [137, 138], and can even lead to lobe structure in the visibility at high voltage bias [139–142]. Such lobe structure was unexpected and has generated numerous theoretical explanations [143–150], some hypothesizing Luttinger liquid physics as the cause.

Here we take a more modest theoretical approach for describing electronic MZIs in a quantum Hall system that focuses on the low bias regime within a single-particle edge-state model. We consider a configuration of two such single-particle electronic MZIs coupled together by the Coulomb interaction, as shown

in Figure 5.1. This geometry has similarities to Hardy’s paradox [52, 151–153] and has been considered previously at various levels of detail by several authors [154–157] to explore such phenomena as quantum erasure [49, 158–162] and Bell inequality violations. In our treatment, we include interactions between the MZIs via a minimal phenomenological model that adds a relative interaction-induced phase shift between a pair of electrons that occupy adjacent edge states simultaneously. The relative phase shift has the effect of entangling [163–165] the states, mixing the path information of the two MZIs. This kind of interaction has been experimentally shown to be capable of producing a π -phase shift on a single electron—perhaps the most dramatic difference from the optical analog [133]. Thus, all elements of our theoretical analysis are based on currently available technology.

Our work considers the task of detecting which-path information in one MZI by using the second MZI as the detector. Since the system and the detector are identical devices, this arrangement has several appealing features. First, the symmetry of the geometry indicates that there should be a duality between “which-path” information in one MZI versus “which-fringe” information in the other. We will show that this is indeed the case, which relates this work to earlier “controlled dephasing” experiments [166–169]. We apply the contextual values formalism [68, 69] for generalized measurements [20, 22, 27] to show how even with inefficient detection [170, 171] and low visibility, the which-path information may be extracted from the detector currents systematically. Next, the fact that both the system and the detector have their own inputs, outputs and coherence allows the effects of measurement to be explored in detail. In particular, the correlations between them can be experimentally measured and analyzed, taking various forms such as joint counting statistics, or even conditioned measurements (see Sukhorukov et al. [172] for an example with incoherent electrons). The ability to condition (or post-select) measurements performed on a system with quantum coherence

also allows the possibility of measuring weak values [28, 32, 44, 48, 68]. Weak values, in addition to being of interest in their own right, have been shown to be useful as an amplification technique for measuring small variations of a system parameter [50, 51, 53, 54, 173], as well as for tests of bona fide quantum behavior [66, 69, 103, 107, 152].

5.1 Coupled MZIs

We consider a pair of electronic MZIs embedded in a two-dimensional electron gas in the integer quantum Hall regime at filling factor $\nu = 2$ as illustrated in Figure 5.1. An ohmic source with a small DC-bias eV above the Fermi energy E_F injects chiral electron-like edge excitations of charge e into the sample that propagate uni-directionally along two independent edge channels [119]. Each edge channel forms an interferometer from two appropriately tuned quantum point contacts (QPCs) that coherently split and then recombine the possible paths. The relative phase between the arms of each interferometer is determined not only by a local dynamical phase accumulated during kinetic propagation along each arm, but also by a global geometric phase [128] (in the form of the Aharonov-Bohm (AB) effect) [127] arising from the closed paths. After the paths interfere, the charges are collected at ohmic reference drains held at the Fermi energy, producing fluctuating output currents that can be temporally averaged.

The two interferometers accrue an additional relative phase shift due to the Coulomb interaction where the charges copropagate. Intuitively, the mutual repulsion affects the dynamical phases by effectively warping the propagation paths, which also affects the geometrical phases by changing the areas enclosed by the paths. A more careful microscopic analysis of the joint interaction phase starting from the screened Coulomb interaction is provided in Appendix C. Such additional relative phase has the effect of *entangling* the joint state of the two interferom-

eters, mixing the which-path information. Due to the entanglement, extracting information from the drains of one interferometer allows one to infer correlated which-path information about the other interferometer. That is, one interferometer can be used as a detector to *indirectly measure* [20, 22, 27, 68] the which-path information of the other. As we shall see, the characteristics of the measurement will depend on the tuning of the detector interferometer as well as on the coupling phase.

We model the coupled MZI system using the elastic scattering approach of Landauer and Büttiker [120–122] for coherent charge transport. As the transport is largely ballistic in the integer quantum Hall regime, the formalism directly relates the average currents I_l collected at each ohmic lead $l \in \{D_1, D_2, S_1, S_2\}$ to the *transmission probabilities* $P_l(E) \in [0, 1]$ for plane waves of fixed energy E to traverse the sample successfully. Treating the ohmic leads as thermal reservoirs, the average currents from spinless single-channel transport are,

$$I_l = e \int_0^{eV} \frac{dE}{h} (f(E + eV) - f(E)) P_l(E) \quad (5.1a)$$

$$\approx \frac{e^2 V}{h} P_l(E_F). \quad (5.1b)$$

Here, $f(E) = (\exp((E - E_F)/k_B T) + 1)^{-1}$ is the equilibrium Fermi distribution relative to the Fermi energy E_F at a temperature T ; h is Planck's constant; and, k_B is Boltzmann's constant.

The approximate equality (5.1b) holds in the low-bias regime when $E_F \gg eV \gg k_B T$ and the transmission probabilities P_l are constant with respect to the small variations in energy. We assume that the source operates in such a regime. Due to the small spectral width of the source, the fermionic excitations will then be well approximated as plane waves at a fixed energy on the scale of the sample; hence, on average only one excitation per channel will occupy the sample and intrachannel interactions can be ignored. In particular, we avoid the anomalous

lobe structure in the interference that appears at higher bias [139–142].

We also assume for simplicity of discussion that the source only injects spinless excitation *pairs* with one excitation per channel so that the coupling interaction between the channels is fixed; the results will be averaged over a more realistic source distribution in Section 5.2.7. With these approximations, the initial joint scattering state for an excitation pair can be written in second-quantized notation as,

$$|\Psi\rangle = \hat{a}^{d\dagger} \hat{a}^{s\dagger} |0\rangle, \quad (5.2)$$

where $|0\rangle$ is the filled Fermi sea of the edge channels and $\hat{a}^{d\dagger}$ and $\hat{a}^{s\dagger}$ are creation operators for plane waves of a fixed energy injected into the inner and outer channels, respectively. Operators corresponding to different edge channels commute due to the independence of the channels.

The inner channel will form a Mach-Zehnder interferometer as shown in Figure 5.1, which we refer to as the upper MZI, or the *detector* MZI. Similarly, the outer channel will form an identical interferometer, which we refer to as the lower MZI, or the *system* MZI. We will use the lowercase superscripts d and s throughout to distinguish quantities specific to the detector and the system, respectively, and to avoid confusion with the detector and system drains that we denote with capital letters D_1 , D_2 , S_1 and S_2 .

The QPCs Q_1^d , Q_2^d , Q_1^s , and Q_2^s shown in Figure 5.1 each elastically scatter the plane waves, affecting only the complex amplitudes of the joint scattering state. Hence, for $m \in \{d, s\}$, $i \in \{1, 2\}$ we can represent the effect of each QPC as a unitary scattering matrix,

$$\hat{U}_i^m = \begin{pmatrix} e^{i\chi_i^m} t_i^m & e^{i\xi_i^m} r_i^m \\ e^{i\chi_i^m} r_i^m & e^{i\xi_i^m} t_i^m \end{pmatrix}, \quad (5.3)$$

where $t_i^m = \sqrt{T_i^m}$ and $r_i^m = i\sqrt{R_i^m}$ are given in terms of the transmission and

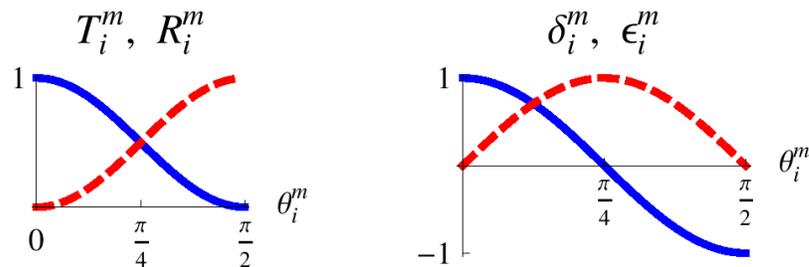


Figure 5.2: Complementary QPC balance parameters (5.4) for $m \in \{d, s\}$ and $i \in \{1, 2\}$ as parametrized by the balance angle θ_i^m . Left: transmission probability T_i^m (solid, blue) and reflection probability R_i^m (dashed, red). Right: particle-like parameter δ_i^m (solid, blue) and wave-like parameter ϵ_i^m (dashed, red).

reflection probabilities $T_i^m \in [0, 1]$ and $R_i^m = 1 - T_i^m$ through Q_i^m . The additional scattering phases χ_i^m and ξ_i^m may arise from QPC asymmetry.

The QPCs are kept tunable subject to the constraints that the outer channel is fully transmitted through Q_1^d and Q_2^d and the inner channel is fully reflected from Q_1^s and Q_2^s to create the two separate interfering paths. There is an additional QPC near drain S_1 not shown in Figure 5.1 that is kept fixed to allow full transmission of the outer channel and full reflection of the inner channel in order to divert the outer channel for collection at the drain S_1 .

For later convenience we also introduce the complementary QPC balance parameters,

$$\delta_i^m = T_i^m - R_i^m \in [-1, 1], \quad (5.4a)$$

$$\epsilon_i^m = 2\sqrt{T_i^m R_i^m} \in [0, 1], \quad (5.4b)$$

for $m \in \{d, s\}$ and $i \in \{1, 2\}$ that satisfy $(\epsilon_i^m)^2 + (\delta_i^m)^2 = 1$. All such QPC parameters can be related by a QPC balance angle $\theta_i^m \in [0, \pi/2]$ such that, $T_i^m = \cos^2 \theta_i^m$, $R_i^m = \sin^2 \theta_i^m$, $\delta_i^m = \cos 2\theta_i^m$, and $\epsilon_i^m = |\sin 2\theta_i^m|$ as illustrated in Figure 5.2. We shall see that the parameters δ_i^m control the particle-like path-bias of the excita-

tion after a QPC, while the parameters ϵ_i^m control the complementary wave-like interference visibility.

The joint state (5.2) can be scattered through Q_1^d and Q_1^s using (5.3) into the basis of the MZI paths, yielding the replacements,

$$\hat{a}^{d\dagger} = e^{i\chi_1^d} t_1^d \hat{a}_{L^d}^\dagger + e^{i\xi_1^d} r_1^d \hat{a}_{U^d}^\dagger, \quad (5.5a)$$

$$\hat{a}^{s\dagger} = e^{i\chi_1^s} t_1^s \hat{a}_{L^s}^\dagger + e^{i\xi_1^s} r_1^s \hat{a}_{U^s}^\dagger. \quad (5.5b)$$

During propagation to the second pair of QPCs, each path $p \in \{L^d, U^d, L^s, U^s\}$ accumulates an additional dynamical phase ϕ_p that depends on the excitation energy and the path-length. When the paths recombine, the difference between the dynamical phases contributes to the interference. Closing the paths for MZI $m \in \{s, d\}$ also contributes a relative geometric Aharonov-Bohm (AB) phase ϕ_{AB}^m that depends on the magnetic flux enclosed by the path.

We can compactly account for the various phase effects contributing to the interference by defining *tuning phases* for each MZI,

$$\phi^d = \phi_{AB}^d + \phi_{L^d} - \phi_{U^d} + \chi_1^d - \xi_1^d, \quad (5.6a)$$

$$\phi^s = \phi_{AB}^s + \phi_{L^s} - \phi_{U^s} + \chi_1^s - \xi_1^s. \quad (5.6b)$$

Finally, the joint scattering amplitude corresponding to co-occupation of L^d and U^s acquires a effective Coulomb interaction phase γ that couples the two interferometers. (See Appendix C for discussion about how the Coulomb effect can produce such a phase shift.) This interaction phase compactly encodes opposing shifts in the combined dynamical and geometric phases of each MZI due to the Coulomb repulsion of the charge pair. For simplicity, we assume for now that the relative phase is constant; we will allow it to fluctuate for consecutive pairs in Section 5.2.7. We also note that any additional Coulomb phase acquired during

copropagation after QPC Q_1^d and before QPC Q_1^s will only contribute to the tuning phase ϕ^d and can therefore be ignored.

After adding the phenomenological phases, the scattered joint state just before the second pair of QPCs is,

$$|\Psi'\rangle = \left(t_1^d t_1^s e^{i(\phi^d + \phi^s)} \hat{a}_{L^d}^\dagger \hat{a}_{L^s}^\dagger + r_1^d r_1^s \hat{a}_{U^d}^\dagger \hat{a}_{U^s}^\dagger + r_1^d t_1^s e^{i\phi^s} \hat{a}_{U^d}^\dagger \hat{a}_{L^s}^\dagger + t_1^d r_1^s e^{i(\phi^d + \gamma)} \hat{a}_{L^d}^\dagger \hat{a}_{U^s}^\dagger \right) |0\rangle, \quad (5.7)$$

up to a global phase of $\exp(i(\phi_{U^d} + \phi_{U^s} + \xi_1^d + \xi_1^s))$ not written.

The interaction phase γ has the effect of *entangling* the two interferometers, which we can show by computing the concurrence [163],

$$\mathcal{C} \left[|\Psi'\rangle \right] = \epsilon_1^d \epsilon_1^s \left| \sin \frac{\gamma}{2} \right| \in [0, 1]. \quad (5.8)$$

We see that the entanglement reaches a maximum when the phase $\gamma \rightarrow \pi$ and vanishes as $\gamma \rightarrow 0$. Furthermore, the entanglement directly depends on the QPCs Q_1^d and Q_1^s preparing interfering wave-like excitations in each MZI, which is measured by $\epsilon_1^d \epsilon_1^s$; maximum entanglement can only occur for balanced QPCs with $T_1^d = T_1^s = 1/2$, or $\epsilon_1^d \epsilon_1^s = 1$.

At this point, we conceptually break the symmetry between the two interferometers to treat one as a detector for information about the other. We will treat the upper MZI as the *detector* and the lower MZI as the *system* being measured, though obviously we could exchange those roles by the symmetry of the geometry. To do this we finish scattering the detector MZI through Q_2^d into the basis of the ohmic detector drains $\{D_1, D_2\}$ using (5.3),

$$\begin{pmatrix} \hat{a}_{L^d}^\dagger \\ \hat{a}_{U^d}^\dagger \end{pmatrix} = \hat{U}_2^d \begin{pmatrix} \hat{a}_{D_1}^\dagger \\ \hat{a}_{D_2}^\dagger \end{pmatrix} \quad (5.9)$$

yielding,

$$|\Psi''\rangle = \left(\hat{a}_{D_1}^\dagger \left[C_{D_1, L^s} e^{i\phi^s} t_1^s \hat{a}_{L^s}^\dagger + C_{D_1, U^s} r_1^s \hat{a}_{U^s}^\dagger \right] \right. \\ \left. + \hat{a}_{D_2}^\dagger \left[C_{D_2, L^s} e^{i\phi^s} t_1^s \hat{a}_{L^s}^\dagger + C_{D_2, U^s} r_1^s \hat{a}_{U^s}^\dagger \right] \right) |0\rangle, \quad (5.10)$$

up to the same global phase as in (5.7). For later convenience, we have defined the detector scattering amplitudes,

$$C_{D_1, L^s} = e^{i\chi_2^d} \left(t_1^d t_2^d e^{i\phi^d} + r_1^d r_2^d \right), \quad (5.11a)$$

$$C_{D_1, U^s} = e^{i\chi_2^d} \left(t_1^d t_2^d e^{i(\phi^d + \gamma)} + r_1^d r_2^d \right), \quad (5.11b)$$

$$C_{D_2, L^s} = e^{i\xi_2^d} \left(t_1^d r_2^d e^{i\phi^d} + r_1^d t_2^d \right), \quad (5.11c)$$

$$C_{D_2, U^s} = e^{i\xi_2^d} \left(t_1^d r_2^d e^{i(\phi^d + \gamma)} + r_1^d t_2^d \right). \quad (5.11d)$$

5.2 Measurement Interpretation

The joint scattering model is useful for computing probabilities and average currents, but it does not provide direct insight into the measurement being performed by one interferometer on the other. To make the connection to measurement more apparent, we will use the *contextual values* formalism [68, 69] that links the detector drain probabilities directly to the “which-path” operator for the system. We will see that we can understand the various subtleties of the measurement quite transparently using this technique.

5.2.1 POVM

To facilitate the interpretation of the distinguishable detector drains as the *outcomes* of a measurement being performed on the system, we define the single

particle state kets,

$$|D_1\rangle = \hat{a}_{D_1}^\dagger |0\rangle, \quad |D_2\rangle = \hat{a}_{D_2}^\dagger |0\rangle, \quad (5.12a)$$

$$|L^s\rangle = \hat{a}_{L^s}^\dagger |0\rangle, \quad |U^s\rangle = \hat{a}_{U^s}^\dagger |0\rangle, \quad (5.12b)$$

define the reduced system state in absence of interaction,

$$|\psi^s\rangle = e^{i\phi^s} t_1^s |L^s\rangle + r_1^s |U^s\rangle, \quad (5.13)$$

and write (5.10) in the form,

$$|\Psi''\rangle = |D_1\rangle \otimes \hat{M}_{D_1} |\psi^s\rangle + |D_2\rangle \otimes \hat{M}_{D_2} |\psi^s\rangle. \quad (5.14)$$

The interaction with the detector in (5.14) is entirely represented by operators acting on the reduced *system* state (5.13) that contain all the scattering information of the detector,

$$\hat{M}_{D_1} = C_{D_1, L^s} |L^s\rangle \langle L^s| + C_{D_1, U^s} |U^s\rangle \langle U^s|, \quad (5.15a)$$

$$\hat{M}_{D_2} = C_{D_2, L^s} |L^s\rangle \langle L^s| + C_{D_2, U^s} |U^s\rangle \langle U^s|. \quad (5.15b)$$

The operator \hat{M}_{D_1} encodes the interaction followed by the absorption of the detector excitation at the drain D_1 . Similarly, the operator \hat{M}_{D_2} encodes the interaction followed by absorption at D_2 . We refer to $\{\hat{M}_{D_1}, \hat{M}_{D_2}\}$ as *measurement operators* [20, 22, 27].

As the coupling phase $\gamma \rightarrow 0$ the measurement operators (5.15) become nearly proportional to the identity. We call $\gamma \rightarrow 0$ the *weak coupling limit* since the reduced system state is only weakly perturbed for small γ . Conversely the limit $\gamma \rightarrow \pi$ is called the *strong coupling limit* since the measurement operators are maximally different from the identity and maximally perturb the reduced system

state.

The measurement operators also form a *Positive Operator-Valued Measure* (POVM) on the system,

$$\hat{E}_{D_1} = \hat{M}_{D_1}^\dagger \hat{M}_{D_1}, \quad \hat{E}_{D_2} = \hat{M}_{D_2}^\dagger \hat{M}_{D_2}, \quad (5.16)$$

such that $\hat{E}_{D_1} + \hat{E}_{D_2} = \hat{1}$. The POVM elements $\{\hat{E}_{D_1}, \hat{E}_{D_2}\}$ act as *probability operators* for the measurement outcomes.

Hence, the probability of absorbing the detector excitation at a drain $D \in \{D_1, D_2\}$ can be expressed either as an expectation of the projection operator of the detector drain under the joint state (5.10) *or*, equivalently, as an expectation of the probability operator (5.16) under the unperturbed system state (5.13),

$$P_D = |\langle D | \Psi'' \rangle|^2 = \langle \psi^s | \hat{E}_D | \psi^s \rangle = |C_{D,L} t_2^d|^2 + |C_{D,U} r_2^d|^2. \quad (5.17)$$

By working with the reduced state (5.13), the measurement operators (5.15), and the probability operators (5.16), we treat the detector as an abstract entity whose sole purpose is to measure the system. Such abstraction allows us to more clearly examine the measurement being made upon the system.

5.2.2 Contextual Values

In order to relate the measurement on the system to observable information that we can interpret, we will use *contextual values* [68, 69] to formally construct system observables from the probability operators (5.16). This formalism acknowledges that the only quantities to which we have experimental access are the detector drain probabilities, so all observations we wish to make about the system must be contained somehow in those probabilities. Generally, the correspondence between the detector drains and a particular system observable will be imperfect, but

we can compensate for such *ambiguity* of the detection by weighting the drain probabilities with appropriate values for the particular measurement setup.

Generally, we cannot construct information about just any system observable from a particular measurement. To find which observables we *can* measure, it is useful to decompose the probability operators (5.16) into an orthonormal basis for the observable space. In our case, the system state space is two-dimensional, so any Hermitian operator can be spanned by the four basis operators,

$$\hat{\sigma}_0^s = \hat{1} = |L^s\rangle \langle L^s| + |U^s\rangle \langle U^s|, \quad (5.18a)$$

$$\hat{\sigma}_1^s = \hat{\sigma}_x^s = |L^s\rangle \langle U^s| + |U^s\rangle \langle L^s|, \quad (5.18b)$$

$$\hat{\sigma}_2^s = \hat{\sigma}_y^s = -i(|L^s\rangle \langle U^s| - |U^s\rangle \langle L^s|), \quad (5.18c)$$

$$\hat{\sigma}_3^s = \hat{\sigma}_z^s = |L^s\rangle \langle L^s| - |U^s\rangle \langle U^s|, \quad (5.18d)$$

which are equivalent to the identity operator and the Pauli spin operators. To find the real components of an observable in this basis we introduce the normalized Hilbert-Schmidt inner product between operators,

$$\langle \hat{A}, \hat{B} \rangle = \frac{\text{Tr}(\hat{A}^\dagger \hat{B})}{\text{Tr}(\hat{1})}, \quad (5.19)$$

under which the operator basis is orthonormal,

$$\langle \hat{\sigma}_\mu^s, \hat{\sigma}_\nu^s \rangle = \delta_{\mu\nu}. \quad (5.20)$$

Here $\mu, \nu \in 0, 1, 2, 3$ and $\delta_{\mu\nu}$ is the Kronecker delta that is 1 if $\mu = \nu$ and 0

otherwise. Using this basis, any system observable can be written,

$$\hat{A} = \sum_{\mu} a_{\mu} \hat{\sigma}_{\mu}^s, \quad (5.21a)$$

$$a_{\mu} = \langle \hat{A}, \hat{\sigma}_{\mu}^s \rangle = \text{Tr} \left(\hat{A} \hat{\sigma}_{\mu}^s \right) / 2, \quad (5.21b)$$

where $\{a_{\mu}\}$ are real-valued components of the observable.

Using (5.21), we can expand the probability operators on the system (5.16) in the basis (5.18) to determine their structure,

$$\hat{E}_{D_1} = \frac{1}{2} (\beta_+^d - V^d \Delta^d) \hat{\sigma}_0^s - \frac{1}{2} V^d \Gamma^d \hat{\sigma}_3^s, \quad (5.22a)$$

$$\hat{E}_{D_2} = \frac{1}{2} (\beta_-^d + V^d \Delta^d) \hat{\sigma}_0^s + \frac{1}{2} V^d \Gamma^d \hat{\sigma}_3^s, \quad (5.22b)$$

where we see that the measurement is characterized by the detector parameters,

$$\beta_+^d = 2 (T_1^d T_2^d + R_1^d R_2^d) = 1 + \delta_1^d \delta_2^d, \quad (5.23a)$$

$$\beta_-^d = 2 (T_1^d R_2^d + R_1^d T_2^d) = 1 - \delta_1^d \delta_2^d, \quad (5.23b)$$

$$V^d = 4 \sqrt{T_1^d R_1^d T_2^d R_2^d} = \epsilon_1^d \epsilon_2^d, \quad (5.23c)$$

$$\Gamma^d = \sin \frac{\gamma}{2} \sin \left(\frac{\gamma}{2} + \phi^d \right), \quad (5.23d)$$

$$\Delta^d = \cos \phi^d - \Gamma^d. \quad (5.23e)$$

defined in terms of the QPC balance parameters (5.4), the tuning phases (5.6), and the coupling phase γ . We will describe these parameters in detail in the next section.

The probability operators only contain components in the subspace spanned by $\{\hat{\sigma}_0^s, \hat{\sigma}_3^s\}$; therefore, *we can only construct observables that are contained within that subspace*. That is, we can construct any observable of the form $\hat{A} = a_0 \hat{\sigma}_0^s + a_3 \hat{\sigma}_3^s$. We denote observables of this form as being *compatible* with the measurement

(5.22). Other observables are *incompatible* with the measurement.

To construct such a compatible system observable from the measurement, we expand its operator directly in terms of the probability operators (5.16),

$$\hat{A} = a_0 \hat{\sigma}_0^s + a_3 \hat{\sigma}_3^s = \alpha_{D_1} \hat{E}_{D_1} + \alpha_{D_2} \hat{E}_{D_2}. \quad (5.24)$$

The required expansion coefficients α_{D_1} and α_{D_2} are generalized eigenvalues, or *contextual values* [68, 69] of the operator, as developed in Chapters 2 and 3. Using this expansion, we can recover the same information *on average* as a projective measurement by using only the drain probabilities, $\langle \hat{A} \rangle = \alpha_{D_1} P_{D_1} + \alpha_{D_2} P_{D_2}$.

To determine the appropriate contextual values to assign in order to construct \hat{A} , we insert (5.22) into (5.24) and solve it as a standard matrix equation using the orthonormal basis, which yields the unique contextual values,

$$\alpha_{D_1} = a_0 - \frac{a_3}{\Gamma^d} \left(\frac{\beta_-^d}{V^d} + \Delta^d \right), \quad (5.25a)$$

$$\alpha_{D_2} = a_0 + \frac{a_3}{\Gamma^d} \left(\frac{\beta_+^d}{V^d} - \Delta^d \right). \quad (5.25b)$$

As long as the contextual values do not diverge, the expansion (5.24) of the compatible operator \hat{A} is well defined, and we can perfectly recover its average,

$$\langle \hat{A} \rangle = \langle \psi^s | \hat{A} | \psi^s \rangle = a_0 + a_3 \delta_1^s. \quad (5.26)$$

The observable parameter a_0 sets the reference point for the average, so contributes no information about the system; we will set it to zero in what follows without loss of generality. Similarly, the remaining parameter a_3 sets the scale of the average; we will set it to one in what follows.

We formally conclude that the detector drains perform a generalized measurement of the *which-path operator* $\hat{\sigma}_3^s$, as might be intuitively expected from the

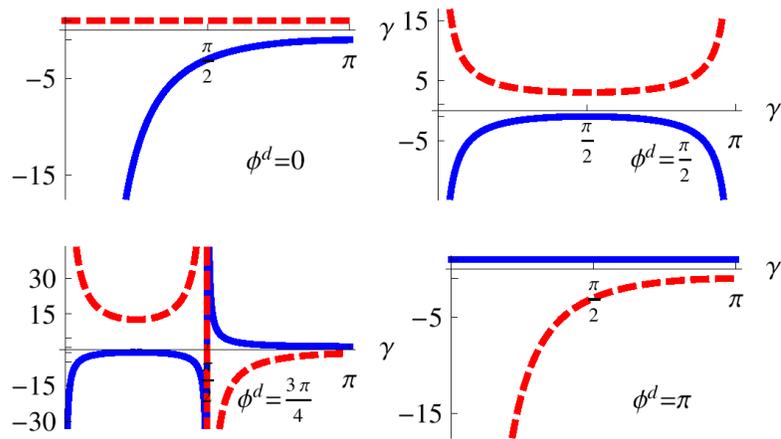


Figure 5.3: The contextual values (5.25) of the which-path operator $\hat{\sigma}_3^s$: α_{D_1} (solid, blue) and α_{D_2} (dashed, red), as a function of the coupling phase γ . The curves are shown for efficient detection $V^d = 1$ and detector tunings $\phi^d = \{0, \pi/2, 3\pi/4, \pi\}$. The tuning strongly affects the ambiguity of the measurement; moreover, the roles of the detector drains flip as the tuning varies from $\phi^d = 0$ to $\phi^d = \pi$.

path-dependent interaction. Moreover, the QPC parameter δ_1^s defined in (5.4a) determines the particle-like which-path behavior on average. No other information about the system can be inferred from the measurement.

The contextual values (5.25) are shown in Figure 5.3 for a few parameter choices. If they are equal to the eigenvalues of the which-path operator, $\alpha_{D_1}, \alpha_{D_2} = \pm 1$, then the measurement is *unambiguous*: one obtains perfect knowledge about the path information with every drain detection, and the system state is projected to a pure path state. If the contextual values diverge, $\alpha_{D_1}, \alpha_{D_2} \rightarrow \pm\infty$, then the measurement is *completely ambiguous*: no knowledge about the path information can be obtained, and the system state is unprojected; however, we shall see in Section 5.2.4 that the system state may still be unitarily perturbed by the coupling. In between these extremes the measurement is *partially ambiguous*: partial knowledge is obtained about the path information with each drain detection, and the system state is partially projected toward a particular path state.

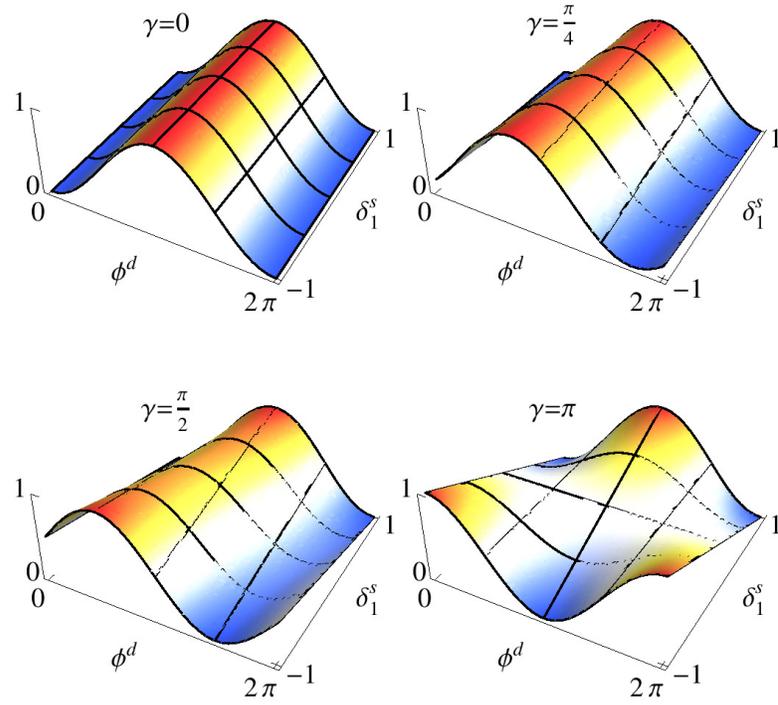


Figure 5.4: The drain probability P_{D_1} (5.27) as a function of the detector tuning ϕ^d and the which-path information δ_1^s , shown for efficient detection $V^d = 1$ and coupling phases $\gamma = \{0, \pi/4, \pi/2, \pi\}$. For zero coupling the interference is independent of the which-path information; for strong coupling $\gamma = \pi$ the interference maximally corresponds to the which-path information.

5.2.3 Parameters

To better understand the parameters (5.23) we write the drain probabilities (5.17) explicitly,

$$P_{D_1} = \frac{1}{2} (\beta_+^d - V^d (\Delta^d + \delta_1^s \Gamma^d)), \quad (5.27a)$$

$$P_{D_2} = \frac{1}{2} (\beta_-^d + V^d (\Delta^d + \delta_1^s \Gamma^d)). \quad (5.27b)$$

The probability P_{D_1} is illustrated in Figure 5.4 for several values of the coupling strength.

The particle-like parameters $\beta_+^d, \beta_-^d \in [0, 1]$ are determined entirely by the path-bias parameters δ_1^d and δ_2^d ; they indicate the average background signal of each detector drain and satisfy $(\beta_+^d + \beta_-^d)/2 = 1$. The wave-like parameter $V^d \in [0, 1]$ is determined entirely by the path-uncertainty parameters ϵ_1^d and ϵ_2^d ; it indicates the *visibility* of the interference. The parameter $\Gamma^d \in [-1, 1]$ indicates the deviation in the interference caused by the coupling phase γ , which is the only effect of the charge coupling. The parameter $\Delta^d \in [-1, 1]$ indicates the interference unrelated to the path information of the system. As the coupling $\gamma \rightarrow 0$, then $\Gamma^d \rightarrow 0$ and $\Delta^d \rightarrow \cos \phi^d$, which recovers the signal for an isolated interferometer [118]. As the coupling $\gamma \rightarrow \pi$, then $\Gamma^d \rightarrow \cos \phi^d$ and $\Delta^d \rightarrow 0$, and the interference maximally corresponds to the path information.

The parameters (5.23) also give insight into the nature of the *measurement* by the role they play in the *contextual values* (5.25). The parameter Γ^d indicates the correlation between the detector drains and the which-path information. Its magnitude $|\Gamma^d| \in [0, 1]$ denotes the *correlation strength*, with 1 indicating perfect correlation and 0 indicating no correlation; due to the inverse dependence in (5.25), any imperfect correlation will *amplify* the contextual values to compensate for the resulting measurement ambiguity. The sign of Γ^d indicates the correspondence of the detector drains to the which path information, with $-$ denoting the mapping $\{D_1, D_2\} \leftrightarrow \{L^s, U^s\}$ and $+$ denoting the mapping $\{D_2, D_1\} \leftrightarrow \{L^s, U^s\}$. Note that the correlation strength depends not only on the coupling phase γ , but also on the tuning phase ϕ^d ; hence, it is possible for the detector drains to be uncorrelated with the system paths even under strong coupling (e.g. examine $\phi^d = \pi/2$ in Figure 5.4 when $\gamma = \pi$).

The parameters β_+^d and β_-^d in (5.25) counterbalance the bias in the average drain background caused by a preferred particle-like path. For instance, if $\beta_+^d > \beta_-^d$ then the signal at drain D_1 is stronger on average in (5.27); hence, the contextual value (5.25a) assigned to D_1 is proportional to the smaller value β_-^d to compensate.

The visibility parameter V^d controls the wave-like interference produced by Q_1^d and Q_2^d . The transmission of each QPC should be balanced in order to provide the interaction phase with an equal-amplitude reference phase for later interference. Any imbalance leads to *inefficiency* of the measurement [170] by reducing the interference visibility, effectively hiding the correlations. Such inefficiency increases the measurement *ambiguity* and results in an amplification of the contextual values. All correlations are hidden at zero interference visibility when either Q_1^d or Q_2^d is fully transmissive or reflective, $T_1^d, T_2^d \in \{0, 1\}$, which leads to divergent contextual values. Maximum interference visibility occurs for balanced transmission with $V^d = 1$. We see that to optimally measure the particle-like which-path information for the system, the detector must itself exhibit maximal wave-like interference; *the detector and system behaviors are therefore complementary*.

The parameter Δ^d is the portion of the interference not affected by the coupling, meaning $\Delta^d + \Gamma^d = \cos \phi^d$. It indicates an additional bias in the drain correspondence caused by the interference not pertinent to the which-path measurement. The contextual values naturally subtract the contribution from this irrelevant background interference to retrieve the measurement information. In the limit of strong coupling $\gamma \rightarrow \pi$, *all* the detector interference encodes the measurement result: $\Gamma^d \rightarrow \cos \phi^d$ and $\Delta^d \rightarrow 0$.

Practically speaking, the detector must be *calibrated* in the laboratory before it can be used to probe an unknown system state. That is, the detector parameters (5.23) must be predetermined by examining the drain outputs of the detector under known system configurations. For example, pinching off QPC Q_1^s to prevent any interactions allows most of the parameters to be set directly by tuning the detector QPCs and the magnetic field. The remaining interaction parameter γ can be inferred from an additional reference system state. Therefore, the process of detector calibration can be viewed as the experimental determination of the appropriate *contextual values* to assign to the detection apparatus.

5.2.4 Measurement Disturbance

The measurement necessarily disturbs the system state by extracting information. We can see the effect of such disturbance by characterizing the system interferometer with analogous parameters to (5.23),

$$\beta_+^s = 2(T_1^s T_2^s + R_1^s R_2^s) = 1 + \delta_1^s \delta_2^s, \quad (5.28a)$$

$$\beta_-^s = 2(T_1^s R_2^s + R_1^s T_2^s) = 1 - \delta_1^s \delta_2^s, \quad (5.28b)$$

$$V^s = 4\sqrt{T_1^s R_1^s T_2^s R_2^s} = \epsilon_1^s \epsilon_2^s, \quad (5.28c)$$

$$\Gamma^s = \sin \frac{\gamma}{2} \sin \left(\frac{\gamma}{2} - \phi^s \right), \quad (5.28d)$$

$$\Delta^s = \cos \phi^s - \Gamma^s. \quad (5.28e)$$

Using these parameters the absorption probabilities for the system drain take the simple form similar to (5.27),

$$P_{S_1} = \frac{1}{2} (\beta_+^s - V^s (\Delta^s - \delta_1^d \Gamma^s)), \quad (5.29a)$$

$$P_{S_2} = \frac{1}{2} (\beta_-^s + V^s (\Delta^s - \delta_1^d \Gamma^s)). \quad (5.29b)$$

With efficient detection $V^d = 1$ and strong coupling $\gamma \rightarrow \pi$, then $\delta_1^d \rightarrow 0$ and $\Delta^s \rightarrow 0$, so the system drain probabilities display no interference, $P_{S_1} \rightarrow \beta_+^s/2$ and $P_{S_2} \rightarrow \beta_-^s/2$; that is, a strongly coupled, efficient which-path measurement will force particle-like statistics in the system [133, 154].

The measurement disturbance may be analyzed more explicitly by rewriting

the measurement operators (5.15) for the case of efficient detection $V^d = 1$,

$$\hat{M}_{D_1} = i e^{i\chi_2^d} e^{i\phi^d/2} \hat{U}_\gamma \hat{E}_{D_1}^{1/2}, \quad (5.30a)$$

$$\hat{M}_{D_2} = i e^{i\xi_2^d} e^{i\phi^d/2} \hat{U}_\gamma \hat{E}_{D_2}^{1/2}, \quad (5.30b)$$

$$\hat{E}_{D_1}^{1/2} = \sin \frac{\phi^d}{2} |L^s\rangle \langle L^s| + \sin \frac{\phi^d + \gamma}{2} |U^s\rangle \langle U^s|, \quad (5.30c)$$

$$\hat{E}_{D_2}^{1/2} = \cos \frac{\phi^d}{2} |L^s\rangle \langle L^s| + \cos \frac{\phi^d + \gamma}{2} |U^s\rangle \langle U^s|, \quad (5.30d)$$

$$\hat{U}_\gamma = \exp\left(i\frac{\gamma}{2} |U^s\rangle \langle U^s|\right). \quad (5.30e)$$

The disturbance manifests itself as two distinct processes. First, the positive roots of the POVM (5.16) $\{\hat{E}_{D_1}^{1/2}, \hat{E}_{D_2}^{1/2}\}$ perform the information extraction necessary for the measurement, partially projecting the reduced system state toward a particular path. Second, the coupling-dependent unitary factor \hat{U}_γ contributes an additional evolution of the system that is unrelated to the extraction of information. The remaining phase factors contribute only to the global phase of the measured state and do not alter the subsequent measurement statistics.

Unambiguous measurements extract maximal information from the system and thus project the system state to a definite path; they are frequently known as *projective* or *strong* measurements. Ambiguous measurements extract partial information from the system and thus partially project the system state toward a particular path. Completely ambiguous measurements extract no information from the system and thus are completely unitary. When the system state is nearly unperturbed up to a global phase, the measurement is called *weak*, which corresponds to the case of a nearly completely ambiguous measurement with a negligible unitary evolution.

5.2.5 Strong Coupling

An unambiguous measurement can only be obtained in the limits of efficient detection $V^d \rightarrow 1$ and strong coupling $\gamma \rightarrow \pi$. In this situation, the ambiguity will be determined only by the tuning phase of the detector ϕ^d , and the POVM will have the most symmetric dependence on the which-path operator,

$$\alpha_{D_1} \rightarrow \frac{-1}{\cos \phi^d}, \quad (5.31a)$$

$$\alpha_{D_2} \rightarrow \frac{1}{\cos \phi^d}, \quad (5.31b)$$

$$\hat{E}_{D_1} \rightarrow \frac{1}{2} (\hat{1} - \hat{\sigma}_3^s \cos \phi^d), \quad (5.31c)$$

$$\hat{E}_{D_2} \rightarrow \frac{1}{2} (\hat{1} + \hat{\sigma}_3^s \cos \phi^d). \quad (5.31d)$$

As the tuning phase ϕ^d varies, the POVM elements oscillate between pure path projections and the identity, despite the strong coupling. The tuning-dependent drain ambiguity contributes to the inefficiency of the measurement by erasing the potentially extractable which-path information from the detector state. Indeed, we shall see in Section 5.3.2 that such ambiguity in the measurement allows the system interference to be recovered by conditioning the system results on specific detector outcomes: Such a phenomenon is known as *quantum erasure* [49, 158–162].

In a laboratory quantum Hall system the AB phase will precess due to slow decay of the transverse magnetic field, so the tuning phase ϕ^d will also precess slowly. Hence, the ambiguity of the measurement will generally oscillate between extremes, while also flipping the correspondence of the drains to the which-path information. Despite any ambiguity in the measurement, however, the system will always be perturbed by the additional unitary evolution (5.30e) that induces a relative phase shift of $\pi/2$ between the arms. Since the system state will be appreciably altered by the strong coupling, the measurement will not be weak

even when completely ambiguous. Hence, *ambiguity of the measurement need not indicate weakness of the measurement.*

The measurement becomes *unambiguous* when the tuning is held fixed at $\cos \phi^d = \pm 1$. In this situation, the detector drains are perfect “bright” and “dark” ports: detection at the dark port will occur deterministically when the system excitation is in the upper arm. The detector drains are perfectly correlated to the which-path information of the system, so the system state is projected to a definite path, and the measurement is *strong*.

5.2.6 Weak Coupling Limit

The weak coupling limit is the limit as the coupling phase $\gamma \rightarrow 0$ and the system and detector become nearly uncoupled. Since at zero coupling the measurement operators (5.30) must either be zero or be proportional to the identity, the weak coupling limit of a measurement must have outcomes that are inherently ambiguous. Hence, we expect the contextual values (5.25) to diverge. However, since $\Gamma^d = \sin \phi^d(\gamma/2) + \cos \phi^d(\gamma/2)^2 + O(\gamma^3)$ the nature of the divergence will also depend upon the tuning.

If the tuning is not an integer multiple of π , then both measurement operators (5.30) will approach the identity as $\gamma \rightarrow 0$ and the measurement will be *weak* for all outcomes. That is, the system state will be nearly unperturbed for any outcome of the measurement. In this case, $\Gamma^d = \sin \phi^d(\gamma/2) + O(\gamma^2)$ and the divergence of the contextual values (5.25) will be linear in γ . For an efficient

detector with $V^d = 1$, we find to $O(\gamma^2)$,

$$\alpha_{D_1} \rightarrow 1 - \frac{2}{\gamma} \frac{1 + \cos \phi^d}{\sin \phi^d}, \quad (5.32a)$$

$$\alpha_{D_2} \rightarrow 1 + \frac{2}{\gamma} \frac{1 - \cos \phi^d}{\sin \phi^d}, \quad (5.32b)$$

$$\hat{E}_{D_1} \rightarrow \frac{1 - \cos \phi^d}{2} \hat{1} + \frac{\gamma}{2} \sin \phi^d |U^s\rangle \langle U^s|, \quad (5.32c)$$

$$\hat{E}_{D_2} \rightarrow \frac{1 + \cos \phi^d}{2} \hat{1} - \frac{\gamma}{2} \sin \phi^d |U^s\rangle \langle U^s|. \quad (5.32d)$$

The POVM has simple dependence on the projection to the upper path, which can also be written in terms of the which-path operator, $|U^s\rangle \langle U^s| = (\hat{1} - \hat{\sigma}_3^s)/2$. The most symmetric case of $\phi^d = \pi/2$ is shown in the upper-right of Figure 5.3.

However, if the tuning $\phi^d = n\pi$ with integer n , then *only one* of the measurement operators will approach the identity as $\gamma \rightarrow 0$. The remaining outcome remains proportional to a projector with a vanishing coefficient and will thus strongly perturb the system state. Hence, only one contextual value diverges while the other remains a constant eigenvalue. In this case, $\Gamma^d = (-1)^n \sin^2(\gamma/2)$ so the divergence will be quadratic in γ . We call such a measurement a *semi-weak* measurement [69] since only a subset of outcomes are weak. For an efficient detector with $V^d = 1$, we find,

$$\alpha_{D_1} \rightarrow \frac{-1}{\sin^2 \frac{\gamma}{2}} \left((-1)^n + \cos^2 \frac{\gamma}{2} \right), \quad (5.33a)$$

$$\alpha_{D_2} \rightarrow \frac{1}{\sin^2 \frac{\gamma}{2}} \left((-1)^n - \cos^2 \frac{\gamma}{2} \right), \quad (5.33b)$$

$$\hat{E}_{D_1} \rightarrow \frac{1}{2} (1 - (-1)^n) \hat{1} + (-1)^n \sin^2 \frac{\gamma}{2} |U^s\rangle \langle U^s|, \quad (5.33c)$$

$$\hat{E}_{D_2} \rightarrow \frac{1}{2} (1 + (-1)^n) \hat{1} - (-1)^n \sin^2 \frac{\gamma}{2} |U^s\rangle \langle U^s|. \quad (5.33d)$$

The POVM retains the simple dependence on the projection to the upper path. The cases for $n = 0$ and $n = 1$ are shown in the upper-left and lower-right of

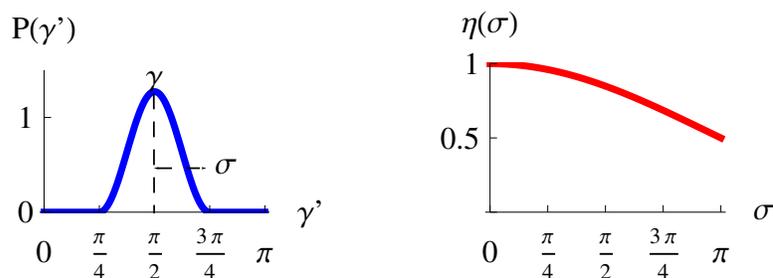


Figure 5.5: Left: The raised cosine distribution showing a spread in the coupling phase centered at $\gamma = \pi/2$ by a half-width $\sigma = \pi/4$. Right: The inefficiency factor $\eta(\sigma)$ defined in (5.36) as a function of the half-width σ .

Figure 5.3, respectively.

For the semi-weak measurement, the effect of absorption at one of the drains is projective. The projective drain outcome unambiguously indicates that the system excitation took the upper U^s path; therefore, the contextual value assigned to the *complementary* drain is an eigenvalue. In contrast, the effect of absorption at the complementary drain only weakly perturbs the system state. Its outcome only ambiguously corresponds to which-path information; therefore, the contextual value assigned to the *projective* drain must be amplified. Such complementary behavior of the contextual value amplification can be counter-intuitive, but it emphasizes that the function of the amplification is to compensate for the ambiguity of the measurement.

We shall see in Section 5.3.6 that while conditioned averages of the weak measurements (5.32) will lead to *weak values*, the conditioned averages of the semi-weak measurements (5.33) have different limiting behavior and lead to different values. The two limiting cases will compete depending on the relative magnitudes of γ and ϕ^d .

5.2.7 Fluctuating Coupling

If the coupling between excitation pairs is not a constant relative phase γ , but instead can fluctuate within a finite uncertainty width σ around an average γ , then the average measurement will be correspondingly more ambiguous. We could quantify this effect by averaging the joint state (5.10) over a range of coupling phases to create a mixed state represented by a density operator; the measurement operators (5.15) and resulting POVM (5.16) could then be generalized to an *averaged* measurement from that density operator. However, that procedure would be completely equivalent to the simpler procedure of averaging the probability operators (5.16) over the coupling width directly, which we choose to do here.

For simplicity, we consider as a coupling distribution the *raised cosine distribution*, which is Gaussian-like, but has compact support. We center the distribution around $\gamma \in [0, 2\pi]$, and give it the half-width $\sigma \in [0, \pi]$. The density for the distribution is nonzero in the domain $\gamma' \in [\gamma - \sigma, \gamma + \sigma]$ and has the form,

$$P(\gamma') = \frac{1}{2\sigma} \left(1 + \cos \left(\frac{\pi}{\sigma} (\gamma' - \gamma) \right) \right). \quad (5.34)$$

An example of the distribution is shown in Figure 5.5 centered at $\gamma = \pi/2$ and with half-width $\sigma = \pi/4$.

Averaging the probability operators (5.16) only affects the constant Γ^d , which is replaced by the averaged version,

$$\Gamma^d(\gamma) \rightarrow \int_{\gamma-\sigma}^{\gamma+\sigma} d\gamma' P(\gamma') \Gamma^d(\gamma') = \eta(\sigma) \Gamma^d(\gamma), \quad (5.35)$$

$$\eta(\sigma) = \left(\frac{\pi^2}{\pi^2 - \sigma^2} \right) \left(\frac{\sin \sigma}{\sigma} \right). \quad (5.36)$$

A plot of the damping factor $\eta(\sigma)$ is shown in Figure 5.5.

We assumed in (5.2) that the source emits only excitation pairs. However,

any contribution of *unpaired* excitations in the initial joint state is equivalent to a contribution of joint states with $\gamma = 0$. The net effect of the source emitting such unpaired excitations is thus to modify Γ^d by an additional probability factor P_p that denotes the likelihood of pair emission. Hence, the only effect of an imperfect source is to introduce a net inefficiency factor $\eta' = P_p \eta(\sigma) \in [0, 1]$ in the parameter Γ^d .

Since the contextual values (5.25) inversely depend on Γ^d , any such inefficiency will introduce an overall amplification factor of $1/\eta'$. In other words, any uncertainty in the coupling strength will lead to additional ambiguity in the average measurement by degrading the portion of the detector interference that is coupled to the system.

5.2.8 Observation Time

Since ambiguous measurements provide less information about an observable per measurement, more measurements will be required to achieve a desired precision for an observable average. We can characterize the necessary increase in observation time as follows. The total detector current is $I = (e^2VP)/h = e/\tau_m$ according to (5.1b), where P is the total probability for excitations to traverse the sample; hence, we can infer that the average time per detector absorption is $\tau_m = h/eVP$. For our single-particle model to apply we wish for the voltage bias V to be low enough that the interferometers contain less than one excitation per channel on average. The characteristic measurement time τ_m will then be on the order of the time-of-flight $\tau_m \approx \ell/v_F$ of an excitation pair through the sample, where ℓ is the average path length of the interferometers and v_F is the Fermi velocity of the ballistic excitations. An observation time of T at the drains D_1 and D_2 therefore roughly corresponds to $n \approx T/\tau_m$ individual measurement events.

The contextual values can be used to provide an upper bound for the number of

measurement events for a desired root-mean-square (RMS) error in the estimation of the average. Specifically, to estimate the average $\langle \hat{\sigma}_3^s \rangle$ from a sequence of n random drain absorptions (d_1, d_2, \dots, d_n) , where $d_i \in \{D_1, D_2\}$, one can use an *unbiased estimator* for the average,

$$\mathbb{E}[\hat{\sigma}_z] = \frac{1}{n} \sum_i^n \alpha_{d_i}, \quad (5.37)$$

that is defined in terms of the *contextual values* assigned to each measurement realization. As $n \rightarrow \infty$, the estimator (5.37) converges to $\langle \hat{\sigma}_3^s \rangle = \alpha_{D_1} P_{D_1} + \alpha_{D_2} P_{D_2}$. The *mean squared error* (MSE) of this estimator is given by the variance of the contextual values over the number of measurements,

$$\text{MSE}[\mathbb{E}[\hat{\sigma}_z]] = \frac{\alpha_{D_1}^2 P_{D_1} + \alpha_{D_2}^2 P_{D_2} - \langle \hat{\sigma}_3^s \rangle^2}{n}. \quad (5.38)$$

Hence, the RMS error $\sqrt{\text{MSE}[\mathbb{E}]}$ scales as $1/\sqrt{n}$ and improves with an increasing number of measurements.

Without prior knowledge of the state, a reasonable upper bound one can make for the MSE is the norm-squared of the contextual values over the number of measurements,

$$\text{MSE}[\mathbb{E}[\hat{\sigma}_z]] \leq \frac{\alpha_{D_1}^2 + \alpha_{D_2}^2}{n}. \quad (5.39)$$

It then follows that to guarantee a maximum desired RMS error ϵ one needs an observation time on the order of,

$$T \approx \tau_m \frac{\alpha_{D_1}^2 + \alpha_{D_2}^2}{\epsilon^2}. \quad (5.40)$$

As the measurement becomes more ambiguous the contextual values become more amplified and so lengthen the observation time necessary to achieve the RMS

error of ϵ . For a strong measurement the upper bound on the observation time is $T \approx 2\tau_m/\epsilon^2$.

5.3 Conditioned Measurements

To gain further insight into the which-path information, we can *condition* the measurement on the subsequent absorption of the system excitation at a specific system drain. To do this we must obtain the joint transmission probabilities for pairs of detector and system drains. Conditional probabilities can then be defined in terms of the joint and single transmission probabilities.

As pointed out by Kang [154] these probabilities are experimentally accessible in the low-bias regime through the zero-frequency cross-correlation noise power between a detector drain $D \in \{D_1, D_2\}$ and a system drain $S \in \{S_1, S_2\}$,

$$S_{D,S} \approx 2 \frac{e^3 V}{h} (P_{S,D}(E_F) - P_S(E_F)P_D(E_F)). \quad (5.41)$$

Hence, knowledge of both the average currents (5.1) and the noise power (5.41) allows the determination of both the joint and single transmission probabilities.

5.3.1 Joint Scattering

We can determine the joint probabilities directly in the scattering model by rewriting (5.10) in the basis of the system drains using (5.3),

$$\begin{pmatrix} \hat{a}_{L^s}^\dagger \\ \hat{a}_{U^s}^\dagger \end{pmatrix} = \hat{U}_2^s \begin{pmatrix} \hat{a}_{S_1}^\dagger \\ \hat{a}_{S_2}^\dagger \end{pmatrix} \quad (5.42)$$

yielding,

$$|\Psi'''\rangle = \left(C_{D_1, S_1} \hat{a}_{D_1}^\dagger \hat{a}_{S_1}^\dagger + C_{D_1, S_2} \hat{a}_{D_1}^\dagger \hat{a}_{S_2}^\dagger + C_{D_2, S_1} \hat{a}_{D_2}^\dagger \hat{a}_{S_1}^\dagger + C_{D_2, S_2} \hat{a}_{D_2}^\dagger \hat{a}_{S_2}^\dagger \right) |0\rangle, \quad (5.43)$$

up to the same global phase as in (5.10).

The relevant joint scattering amplitudes are,

$$C_{D_1, S_1} = e^{i(\chi_2^d + \chi_2^s)} \times \quad (5.44a)$$

$$\left[r_1^d r_2^d r_1^s r_2^s + t_1^d t_2^d r_1^s r_2^s e^{i(\phi^d + \gamma)} + r_1^d r_2^d t_1^s t_2^s e^{i\phi^s} + t_1^d t_2^d t_1^s t_2^s e^{i(\phi^d + \phi^s)} \right],$$

$$C_{D_1, S_2} = e^{i(\chi_2^d + \xi_2^s)} \times \quad (5.44b)$$

$$\left[r_1^d r_2^d r_1^s t_2^s + t_1^d t_2^d r_1^s t_2^s e^{i(\phi^d + \gamma)} + r_1^d r_2^d t_1^s r_2^s e^{i\phi^s} + t_1^d t_2^d t_1^s r_2^s e^{i(\phi^d + \phi^s)} \right],$$

$$C_{D_2, S_1} = e^{i(\xi_2^d + \chi_2^s)} \times \quad (5.44c)$$

$$\left[r_1^d t_2^d r_1^s r_2^s + t_1^d r_2^d r_1^s r_2^s e^{i(\phi^d + \gamma)} + r_1^d t_2^d t_1^s t_2^s e^{i\phi^s} + t_1^d r_2^d t_1^s t_2^s e^{i(\phi^d + \phi^s)} \right],$$

$$C_{D_2, S_2} = e^{i(\xi_2^d + \xi_2^s)} \times \quad (5.44d)$$

$$\left[r_1^d t_2^d r_1^s t_2^s + t_1^d r_2^d r_1^s t_2^s e^{i(\phi^d + \gamma)} + r_1^d t_2^d t_1^s r_2^s e^{i\phi^s} + t_1^d r_2^d t_1^s r_2^s e^{i(\phi^d + \phi^s)} \right].$$

The joint probabilities for absorption in detector drain $D \in \{D_1, D_2\}$ and system drain $S \in \{S_1, S_2\}$ can then be understood as either expectations of joint projections $|S\rangle \langle S| \otimes |D\rangle \langle D|$ under the joint state (5.43), *or*, equivalently, as expectations of system projections $|S\rangle \langle S|$ under the *measured* reduced system state $\hat{M}_D |\psi^s\rangle$,

$$P_{D,S} = |\langle S, D | \Psi''' \rangle|^2 = |\langle S | \hat{M}_D |\psi^s\rangle|^2 = |C_{D,S}|^2. \quad (5.45)$$

These joint probabilities can be written explicitly in terms of the parameters

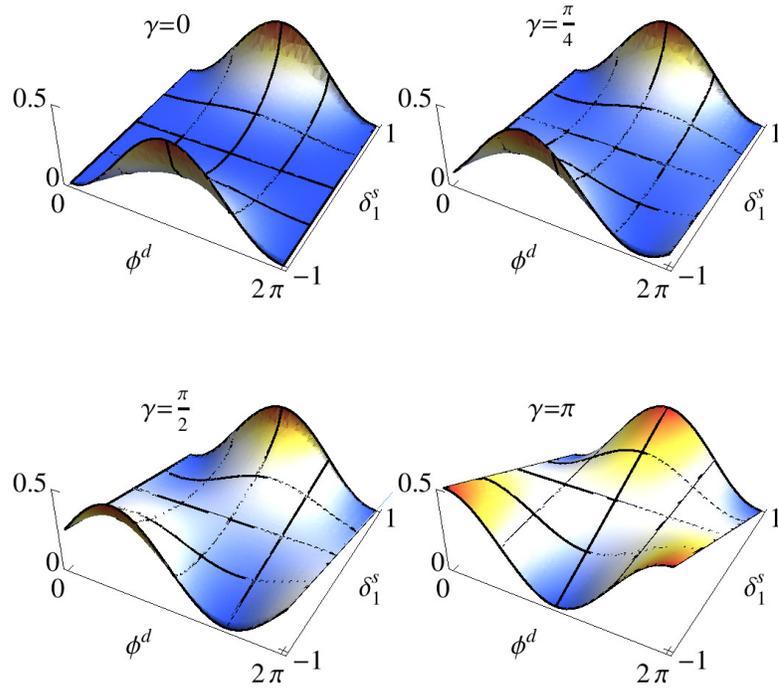


Figure 5.6: The joint probability P_{D_1, S_1} (5.46a) with system tuning $\phi^s = 0$ as a function of the detector tuning ϕ^d and the which-path information δ_1^s , shown for efficient detection $V^d = 1$, balanced system drains $\epsilon_2^s = 1$, and the coupling phases $\gamma = \{0, \pi/4, \pi/2, \pi\}$.

(5.4), (5.23), and (5.28) as,

$$P_{D_1, S_1} = \frac{1}{4} \left(\beta_+^d \beta_+^s + V^d V^s \Delta^{ds} \right. \quad (5.46a)$$

$$\begin{aligned} & - V^d (\Delta^d \beta_+^s + \Gamma^d (\delta_1^s + \delta_2^s)) \\ & \left. - V^s (\Delta^s \beta_+^d - \Gamma^s (\delta_1^d + \delta_2^d)) \right), \end{aligned}$$

$$P_{D_1, S_2} = \frac{1}{4} \left(\beta_+^d \beta_-^s - V^d V^s \Delta^{ds} \right. \quad (5.46b)$$

$$\begin{aligned} & - V^d (\Delta^d \beta_-^s + \Gamma^d (\delta_1^s - \delta_2^s)) \\ & \left. + V^s (\Delta^s \beta_+^d - \Gamma^s (\delta_1^d + \delta_2^d)) \right), \end{aligned}$$

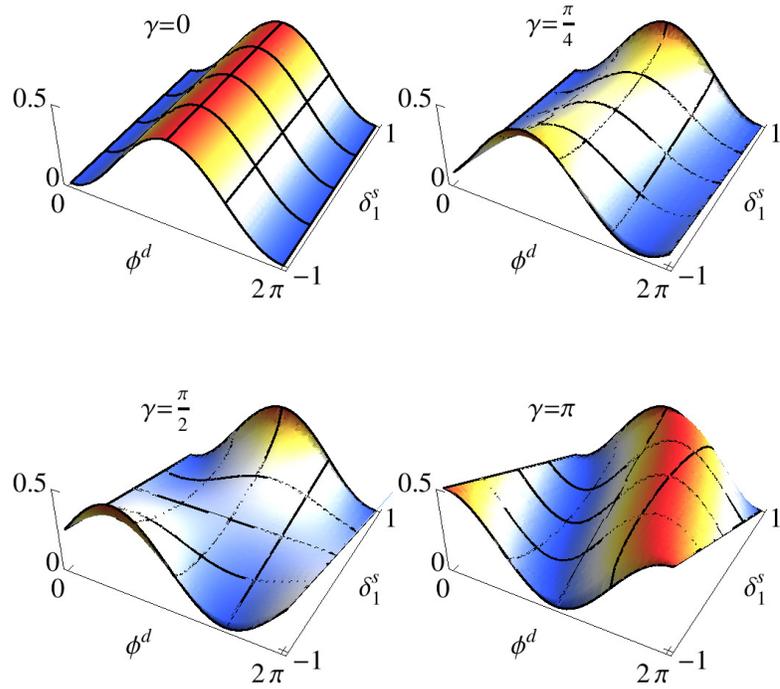


Figure 5.7: The joint probability P_{D_1, S_1} (5.46a) with system tuning $\phi^s = \pi/2$ as a function of the detector tuning ϕ^d and the which-path information δ_1^s , shown for efficient detection $V^d = 1$, balanced system drains $\epsilon_2^s = 1$, and the coupling phases $\gamma = \{0, \pi/4, \pi/2, \pi\}$.

$$P_{D_2, S_1} = \frac{1}{4} \left(\beta_-^d \beta_+^s - V^d V^s \Delta^{ds} \right. \quad (5.46c)$$

$$\left. + V^d (\Delta^d \beta_+^s + \Gamma^d (\delta_1^s + \delta_2^s)) \right. \\ \left. - V^s (\Delta^s \beta_-^d - \Gamma^s (\delta_1^d - \delta_2^d)) \right),$$

$$P_{D_2, S_2} = \frac{1}{4} \left(\beta_-^d \beta_-^s + V^d V^s \Delta^{ds} \right. \quad (5.46d)$$

$$\left. + V^d (\Delta^d \beta_-^s + \Gamma^d (\delta_1^s - \delta_2^s)) \right. \\ \left. + V^s (\Delta^s \beta_-^d - \Gamma^s (\delta_1^d - \delta_2^d)) \right),$$

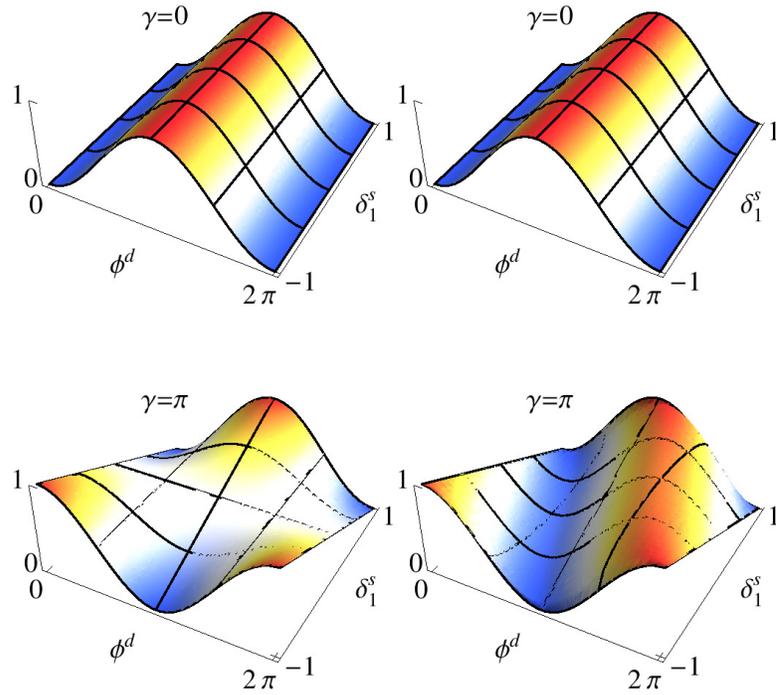


Figure 5.8: The conditional probability $P_{D_1|S_1} = P_{D_1, S_1}/P_{S_1}$ with system tunings $\phi^s = 0$ (left) and $\phi^s = \pi/2$ (right) as a function of the detector tuning ϕ^d and the which-path information δ_1^s , shown for efficient detection $V^d = 1$, balanced system drains $\epsilon_2^s = 1$, and the coupling phases $\gamma = \{0, \pi\}$.

where we have defined the additional parameter Δ^{ds} as,

$$\Delta^{ds} = \cos \phi^d \cos \phi^s - \Gamma^{ds}, \quad (5.47a)$$

$$\Gamma^{ds} = \sin \frac{\gamma}{2} \sin \left(\frac{\gamma}{2} + \phi^{ds} \right), \quad (5.47b)$$

$$\phi^{ds} = \phi^d - \phi^s. \quad (5.47c)$$

For illustration purposes, we have plotted the joint probability P_{D_1, S_1} (5.46a) for several parameter choices in Figure 5.6 and Figure 5.7.

The parameter $\Delta^{ds} \in [-1, 1]$ represents the joint interference between the system and detector. The parameter $\Gamma^{ds} \in [-1, 1]$ is the portion of the joint

interference that depends explicitly on the coupling phase γ and the difference between the tuning phases ϕ^{ds} . As the coupling $\gamma \rightarrow 0$, then $\Gamma^{ds} \rightarrow 0$ and the joint interference reduces to a decoupled interference product $\Delta^{ds} \rightarrow \cos \phi^d \cos \phi^s$. As the coupling $\gamma \rightarrow \pi$, then $\Gamma^{ds} \rightarrow \cos \phi^{ds}$ and the joint interference will be maximally coupled $\Delta^{ds} \rightarrow -\sin \phi^d \sin \phi^s$.

We can marginalize the joint probabilities (5.46) to obtain both the detector probabilities (5.27) as $P_D = \sum_S P_{D,S}$ and the system probabilities (5.29) as $P_S = \sum_D P_{D,S}$. Furthermore, we can construct the *conditional probabilities* $P_{D|S}$ for absorption in a detector drain D given an absorption in a system drain S , as well as the conditional probabilities $P_{S|D}$ for absorption in a system drain S given an absorption in a detector drain D ,

$$P_{D|S} = \frac{P_{D,S}}{P_S}, \quad (5.48a)$$

$$P_{S|D} = \frac{P_{D,S}}{P_D}. \quad (5.48b)$$

For comparison with the joint probabilities, we have plotted the conditional detector probability $P_{D_1|S_1}$ in Figure 5.8 for several parameter choices.

5.3.2 Quantum Erasure

We can use the conditional system probabilities $P_{S|D}$ to clarify the phenomenon of *quantum erasure* [49, 158–162], which has also been explored in this system by Kang [154]. Any wave-like interference patterns in the system drain probabilities will degrade with the coupling phase γ , as shown in the upper-left of Figure 5.9 for P_{S_1} and maximum system visibility $V^s = 1$. At strong coupling $\gamma = \pi$ the wave-like interference will be completely destroyed. However, parts of the interference may be restored by *conditioning* the drain on an appropriate detector measurement.

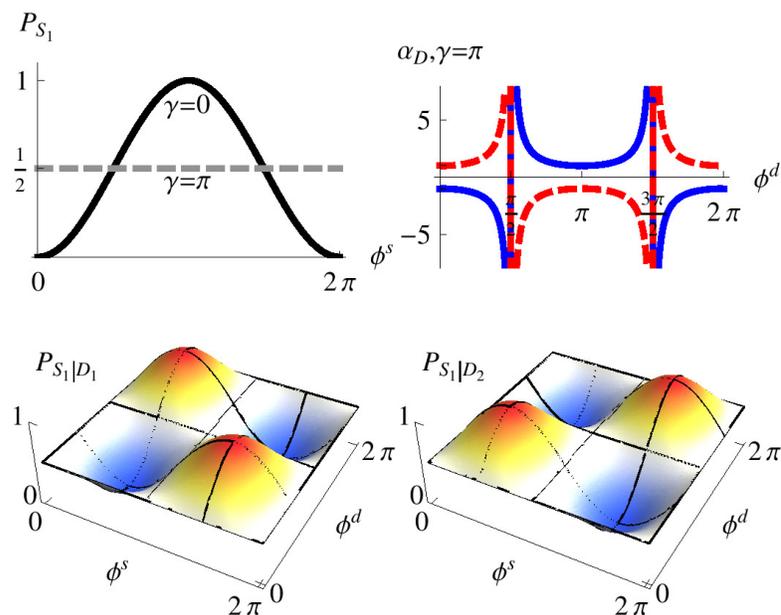


Figure 5.9: Quantum erasure. Any system interference in the system drain probability P_{S_1} (upper-left, black) is completely destroyed at strong coupling $\gamma = \pi$ (gray, dashed). Conditioning on the detector drains recovers phase-shifted interference in the conditional probabilities $P_{S_1|D_1}$ and $P_{S_1|D_2}$ (bottom), but with a visibility that is dependent on the measurement ambiguity as controlled by the detector tuning phase ϕ^d . The contextual values $\{\alpha_{D_1}, \alpha_{D_2}\}$ (upper-right) diverge for maximum ambiguity and reduce to the eigenvalues of the which-path operator $\hat{\sigma}_3^s$ for zero ambiguity. The plots are shown for efficient detection $V^d = 1$, strong coupling $\gamma = \pi$, and maximum system visibility $V^s = 1$.

To restore the interference in the system statistics, the detector must make an *ambiguous* measurement, as we shall see. Strong coupling destroys the interference in the unconditioned system statistics by recording the which-path information in the detector state via the coupling phase γ . As the which-path information is available in the detector state for later collection, at least in principle, the total reduced system statistics must reflect the degree of *potential* information acquisition. However, such information in the detector state has not yet been extracted classically since the detector drains have not yet been probed; hence, the information in the state only indicates the *potential* for later extraction of which-

path information at the detector drains. A partially ambiguous measurement extracts some of that potential information and erases the rest; a completely ambiguous measurement extracts no information and erases all of the potential for later extraction in the process. The recovered interference in the conditioned system statistics reflects the erasure of the information acquisition potential by the ambiguous measurement, even though the total statistics of the reduced system are unchanged by probing the detector.

As discussed in Section 5.2.5 the detector phase ϕ^d determines the ambiguity of the measurement under such strong coupling, so the degree of possible erasure will also depend on the detector phase. We can see the dependence of the interference recovery on ϕ^d in Figure 5.9 in the lower two plots. As the detector phase ϕ^d varies from 0 to 2π , the conditional probabilities $P_{S_1|D_1}$ and $P_{S_1|D_2}$ continuously vary from flat particle-like statistics to complementary wave-like interference patterns. The visibilities of the complementary interference patterns directly depend on the measurement ambiguity, as can be seen in the plot of the contextual values in the upper-right of Figure 5.9. Maximum visibility corresponds to maximum measurement ambiguity where the contextual values diverge; zero visibility corresponds to zero measurement ambiguity where the contextual values reduce to the eigenvalues of the which-path operator. We also note that the effect of the additional coupling evolution (5.30e) creates a $\pi/2$ phase shift in the interference pattern that *cannot be erased since it is not part of the information extraction of the measurement*.

Such erasure under strong coupling has an intuitive analogy to an optical double-slit experiment (that you can even try at home!) [162] as shown in Figure 5.10. In the optical equivalent, a coherent beam of light passes through two slits and displays an interference pattern on a remote screen. However, if an experimenter tags each slit with horizontal and vertical polarizing filters, then the which-path information of the light can later be extracted from the polarization

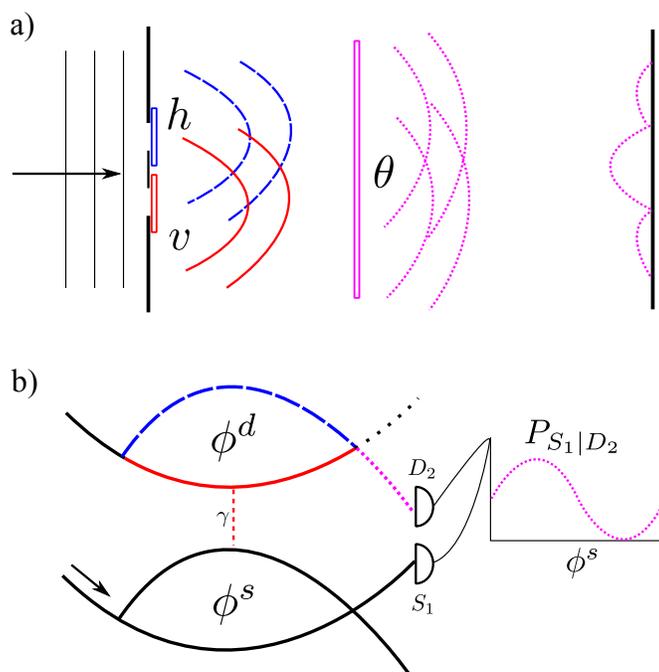


Figure 5.10: Analogy to optical quantum erasure of two-slit interference. a) A coherent light beam passes through two slits tagged with horizontal (h) and vertical (v) polarization. After conditioning on a subsequent polarizer oriented at an angle θ to horizontal, an interference pattern can be recovered with a visibility that depends upon θ . b) A coherent electron passes through a QPC and is tagged by a detector electron via an interaction phase shift γ . After conditioning on the drain D_2 after a subsequent detector QPC that forms an MZI with tuning phase ϕ^d , an interference pattern can be recovered in the system drain S_1 as the system tuning ϕ^s is varied with a visibility that depends upon ϕ^d .

degree of freedom; hence, the total interference pattern on the remote screen will be destroyed. The experimenter can subsequently condition the measurement on a particular polarization by placing another polarizer after the two slits oriented at some angle θ relative to horizontal. If the conditioning polarizer is oriented horizontally or vertically, then the path measurement will be unambiguous and extract all which-path information, so no conditioned system interference will be recovered. However, if the conditioning polarizer is oriented diagonally, then the path measurement will be completely ambiguous and the potential which-path

information will be erased, recovering all of the interference in the conditioned statistics. In the electronic version, the system excitation plays the role of the light beam, the relative coupling phase γ records the potential path information, the conditioning on a particular detector drain plays the role of the polarizer, and the tuning phase ϕ^d selects the conditioning basis analogously to the angle θ , controlling the ambiguity of the measurement and degree of erasure.

Furthermore, one could in principle implement a delayed-choice [158, 159, 161] version of the quantum erasure by placing the detector drains much further away in the sample than the system drains. The interaction phase γ could be recorded and the system excitations collected, upon which a controlled change in the magnetic field could set the tuning phase of the detector. Upon conditioning the data, the interference would reappear according to which tuning phase had been chosen after the system excitation had already been collected.

We stress that the erasure of the potential which-path information and recovery of the system interference will be apparent only when conditioning the collected data. Without conditioning, even completely ambiguous measurements under strong coupling will destroy the system interference. The interference patterns recovered by conditioning on the detector drains will be complementary to each other in such a case and thus cancel in the unconditioned statistics.

5.3.3 Conditioned Averages

We can also use the conditional probabilities to condition the averages of the which-path measurement on a subsequent *system* drain absorption. To do this, we weight the conditional detector probabilities (5.48) with the contextual values for the measurement (5.25) [68],

$${}_S\langle\hat{\sigma}_3^s\rangle = \sum_D \alpha_D P_{D|S}. \quad (5.49)$$

In terms of the various characterization parameters, these take the explicit form,

$$s_1 \langle \hat{\sigma}_3^s \rangle = \frac{1}{2P_{S_1}} \left(\delta_1^s + \delta_2^s - V^s \frac{\Xi^{ds}}{\Gamma^d} \right), \quad (5.50a)$$

$$s_2 \langle \hat{\sigma}_3^s \rangle = \frac{1}{2P_{S_2}} \left(\delta_1^s - \delta_2^s + V^s \frac{\Xi^{ds}}{\Gamma^d} \right), \quad (5.50b)$$

where there is a joint interference contribution to the conditioned average,

$$\Xi^{ds} = \Delta^{ds} - (\Delta^s - \delta_1^d \Gamma^s) + \frac{\delta_1^d \delta_2^d}{V^d} (\Delta^s \beta_-^d - \Gamma^s (\delta_1^d + \delta_2^d)). \quad (5.51)$$

The joint interference simplifies considerably in the special case of efficient detection $V^d = 1$,

$$\frac{\Xi^{ds}}{\Gamma^d} \rightarrow 2 \sin \frac{\gamma}{2} \cot \left(\frac{\gamma}{2} + \phi^d \right) \cos \left(\frac{\gamma}{2} - \phi^s \right). \quad (5.52)$$

This case is plotted in Figure 5.11.

The interference term is scaled by the visibility of the *system* interference V^s , which measures the wave-like behavior of the system excitation. Any wave-like behavior of the system leads to a contribution to the conditioned averages that depends on properties of the correlated *detector* as well, due to the joint interference. Hence, *the which-path information of the system cannot be decoupled from the detector that is measuring it in general.*

The conditioned averages properly obey the consistency relation,

$$\langle \hat{\sigma}_3^s \rangle = \sum_S s \langle \hat{\sigma}_3^s \rangle P_S, \quad (5.53)$$

and are bounded by the *contextual values* (5.25). Since the contextual values are usually larger than the eigenvalues of $\hat{\sigma}_3^s$ due to amplification from measurement ambiguity, the conditioned averages can counter-intuitively lie outside the eigenvalue range. In the weak coupling limit $\gamma \rightarrow 0$, such conditioned averages

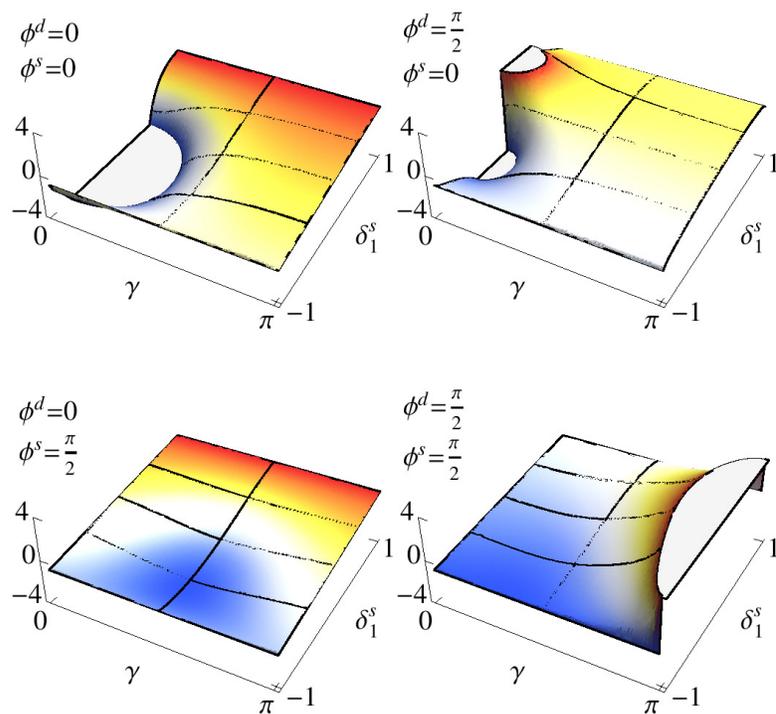


Figure 5.11: The conditioned average $s_1 \langle \hat{\sigma}_3^s \rangle$ (5.50a) as a function of the coupling phase γ and the which-path information δ_1^s , shown for efficient detection $V^d = 1$, balanced system drains $\epsilon_2^s = 1$, and tuning phases $\phi^d, \phi^s = \{0, \pi/2\}$.

can become detector-independent and converge to *weak values* [28, 68], as we will show later.

However, for any macroscopic property such conditioned averages will always lie inside the eigenvalue range, even when measured ambiguously. As shown in Chapter 4, the eigenvalue range constraint for conditioned averages is equivalent to a generalized Leggett-Garg constraint [69, 103, 107] that must be satisfied for any non-invasively measured, realistic property. As such, any violation of the eigenvalue range in a conditioned average can be seen as a signature of *nonclassical* behavior stemming from quantum interference.

5.3.4 Deterministic Measurement

If either of the system QPCs is fully transmissive or reflective then $V^s = 0$, the system interference vanishes, and the conditioned averages (5.50) reduce to ± 1 for any coupling strength. In such a case the excitation path is deterministic and the system displays purely particle-like behavior. The post-selection perfectly determines the path, and the which-path measurement made by the detector will always agree with the post-selected value.

5.3.5 Strong Coupling

For the case of strong coupling $\gamma = \pi$ and an efficient detector $V^d = 1$, the conditioned averages (5.50) reduce to,

$$s_1 \langle \hat{\sigma}_3^s \rangle \rightarrow \frac{\delta_1^s + \delta_2^s}{\beta_+^s} + \frac{V^s}{\beta_+^s} \tan \phi^d \sin \phi^s, \quad (5.54a)$$

$$s_2 \langle \hat{\sigma}_3^s \rangle \rightarrow \frac{\delta_1^s - \delta_2^s}{\beta_-^s} - \frac{V^s}{\beta_-^s} \tan \phi^d \sin \phi^s. \quad (5.54b)$$

The tuning phase ϕ^d is the sole detector parameter that specifies the ambiguity of the measurement.

If the measurement is also unambiguous $\phi^d \rightarrow n\pi$, then the interference contribution vanishes. The conditioned averages become the *detector-independent* quantities $s_1 \langle \hat{\sigma}_3^s \rangle \rightarrow (\delta_1^s + \delta_2^s)/\beta_+^s$ and $s_2 \langle \hat{\sigma}_3^s \rangle \rightarrow (\delta_1^s - \delta_2^s)/\beta_-^s$ that always lie in the eigenvalue range. A strong which-path measurement made by the detector therefore forces the system excitation to display particle-like conditioned statistics.

However, even with strong coupling any ambiguity introduced in the detector will lead to quantum erasure that recovers interference in the conditioned statistics of the system, as discussed in Section 5.3.2. Such recovered interference can take the conditioned averages of the which-path information outside the eigenvalue range. For an almost completely ambiguous measurement the tuning phase will

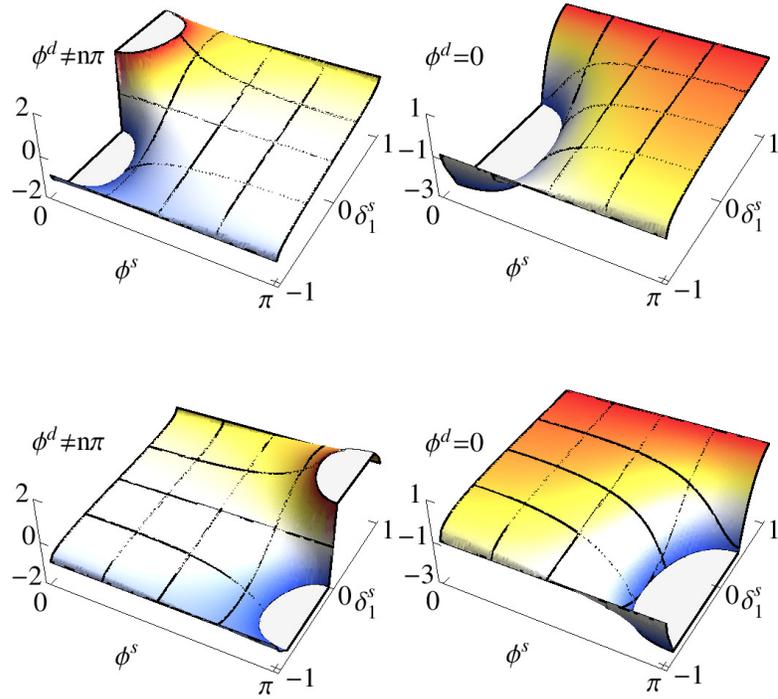


Figure 5.12: The weak limit $\gamma \rightarrow 0$ of the conditioned averages $s_1 \langle \hat{\sigma}_3^s \rangle$ (top) and $s_2 \langle \hat{\sigma}_3^s \rangle$ (bottom) for the distinct cases of weak measurement $\phi^d \neq 0$ (left) given in (5.55) and semi-weak measurement $\phi^d = 0$ (right) given in (5.57) as a function of the system tuning ϕ^s and the which-path information δ_1^s . The values are shown for efficient detection $V^d = 1$ and balanced system drains $\epsilon_2^s = 1$.

deviate from $\pi/2$ only by a small angle $\delta\phi^d$. Since $\tan(\pi/2 + \delta\phi^d) = -1/\delta\phi^d + O(\delta\phi^d)$, the interference contribution will dominate, and the conditioned averages will *diverge*.

5.3.6 Weak Coupling Limit

The weak coupling limit $\gamma \rightarrow 0$ of an efficient detector $V^d = 1$ leads to conditioned averages (5.50) that generally depend on the detector tuning ϕ^d , as anticipated during the discussion in Section 5.2.6. For the weak measurement case when the detector tuning is not an integer multiple of π , then as $\gamma \rightarrow 0$ the joint interference

term in the numerators of (5.50) vanishes, $\sin(\gamma/2) \cot(\gamma/2 + \phi^d) \rightarrow 0$, yielding the *detector-independent* quantities,

$$s_1 \langle \hat{\sigma}_3^s \rangle \rightarrow \frac{\delta_1^s + \delta_2^s}{\beta_+^s - V^s \cos \phi^s}, \quad (5.55a)$$

$$s_2 \langle \hat{\sigma}_3^s \rangle \rightarrow \frac{\delta_1^s - \delta_2^s}{\beta_-^s + V^s \cos \phi^s}. \quad (5.55b)$$

In contrast with the unambiguous case, system interference remains in the denominators.

These expressions match the real parts of the *weak value* expressions defined in Aharonov et al. [28],

$$s_1 \langle \hat{\sigma}_3^s \rangle^w = \frac{\langle S_1 | \hat{\sigma}_3^s | \psi^s \rangle}{\langle S_1 | \psi^s \rangle} = \frac{\delta_1^s + \delta_2^s}{\beta_+^s - V^s \cos \phi^s} - i \frac{V^s \sin \phi^s}{\beta_+^s - V^s \cos \phi^s}, \quad (5.56a)$$

$$s_2 \langle \hat{\sigma}_3^s \rangle^w = \frac{\langle S_2 | \hat{\sigma}_3^s | \psi^s \rangle}{\langle S_2 | \psi^s \rangle} = \frac{\delta_1^s - \delta_2^s}{\beta_+^s + V^s \cos \phi^s} + i \frac{V^s \sin \phi^s}{\beta_+^s + V^s \cos \phi^s}. \quad (5.56b)$$

As pointed out in Section 3.5 any additional unitary coupling evolution such as (5.30e) could in principle affect the convergence of the conditioned averages to these detector-independent weak value expressions. However, in this case the limit is unaffected and the standard expressions are recovered.

For the semi-weak measurement case when $\phi^d = n\pi$ with n an integer, then as $\gamma \rightarrow 0$ the joint interference term in the numerators of (5.50) reduces to $\sin(\gamma/2) \cot(\gamma/2 + n\pi) \rightarrow (-1)^n$; therefore, system interference remains in *both* the numerators and denominators, yielding the modified expressions,

$$s_1 \langle \hat{\sigma}_3^s \rangle \rightarrow \frac{\delta_1^s + \delta_2^s - (-1)^n V^s \cos \phi^s}{\beta_+^s - V^s \cos \phi^s}, \quad (5.57a)$$

$$s_2 \langle \hat{\sigma}_3^s \rangle \rightarrow \frac{\delta_1^s - \delta_2^s + (-1)^n V^s \cos \phi^s}{\beta_-^s + V^s \cos \phi^s}. \quad (5.57b)$$

The integer n selects which detector drain is projective. Hence, unlike the weak values (5.56), these “semi-weak values” do not conform to general detector-independent

expressions, but explicitly depend on the details of the measurement operators (5.30).

We emphasize that the detector tunings $\phi^d = n\pi$ are critical points around which the $\gamma \rightarrow 0$ limit is *unstable*, meaning that any laboratory approach to the weak coupling limit of the conditioned averages (5.50) can approximate either the weak values (5.55) or the semi-weak values (5.57) depending on the relative magnitudes of γ and ϕ^d . Hence, the limiting values will compete with each other as γ becomes small and ϕ^d approaches a critical point. The difference between these limiting cases is plotted in Figure 5.12 using the choice $n = 0$.

6 The Complex Weak Value

Strangeness by itself is not a problem; self-consistency is the real issue. In this sense the logic of the weak values is similar to the logic of special relativity: That light has the same velocity in all reference frames is certainly highly unusual, but everything works in a self consistent way, and because of this special relativity is rather easy to understand. We are convinced that, due to its self-consistency, the weak measurements logic will lead to a deeper understanding of the nature of quantum mechanics.

Yakir Aharonov *et al.*, (2002) [152]

In their seminal Letter, Aharonov *et al.* [28] claimed that they could consistently assign a particular value to an observable that was being weakly measured in a pre- and post-selected ensemble. To illustrate their technique, they weakly coupled an observable \hat{A} to a continuous detector with an initial Gaussian wave-function. Normally, such a weak von Neumann coupling [11] would approximately shift the mean of the Gaussian detector wave-function by the expectation value $\langle \psi_i | \hat{A} | \psi_i \rangle$ of \hat{A} in the initial state $|\psi_i\rangle$, which would effectively measure \hat{A} ; however, they showed that by post-selecting a final state $|\psi_f\rangle$ after the weak coupling, the mean of the Gaussian detector wave-function could be made to approximately shift by a *complex* quantity that they dubbed the *weak value* of the observable,

$$A_w = \frac{\langle \psi_f | \hat{A} | \psi_i \rangle}{\langle \psi_f | \psi_i \rangle}. \quad (6.1)$$

Notably, the weak value expression is not constrained to the eigenvalue range for the observable \hat{A} , so it can become arbitrarily large for nearly orthogonal pre- and post-selections.

This complex shift in the mean of the Gaussian detector wave-function was only approximate under weak von Neumann coupling and not directly observable, so its significance was not overtly clear; however, the Letter [28] also showed that both the real and imaginary parts of (6.1) could be *operationally* obtained from the linear response of the detector under separate conjugate observable measurements. The practical benefit of this observation was that one could amplify the response of the detector by making a clever choice of post-selection, which potentially allowed for the sensitive determination of other small parameters contributing to the evolution.

After theoretical clarifications of the derivation in [29], experimental confirmation of such amplified detector response soon followed in optical systems [44, 174]. The amplification has since been used successfully to sensitively measure a variety of phenomena [50, 51, 55, 56, 175, 176] to remarkable precision, using both the real and imaginary parts of (6.1) as amplification parameters. Several theoretical extensions of the original derivation of the amplification [33, 34, 37, 39, 40, 42, 43, 53, 177–185] and several proposals for other amplification measurements have also appeared [41, 173, 186–188]. In particular, it has been noted that how the amplification effect arises in such a continuous wave-function detector is not intrinsically quantum mechanical, but can also occur in classical wave mechanics [54], which has prompted recent study into the mathematical phenomenon of *superoscillations* (e.g. [189, 190]).

Conceptually, however, the weak value expression (6.1) has remained quite controversial: since it is generally complex and not constrained to the spectrum of \hat{A} , how should it be interpreted? Its primary interpretation in the literature has rested somewhat loosely upon the observation that despite its anomalous behavior

one can still decompose an expectation value through the insertion of the identity into an average of weak values,

$$\langle \psi_i | \hat{A} | \psi_i \rangle = \sum_f |\langle \psi_f | \psi_i \rangle|^2 \frac{\langle \psi_f | \hat{A} | \psi_i \rangle}{\langle \psi_f | \psi_i \rangle}, \quad (6.2)$$

which has the same form as decomposing a classical expectation value $E(X|i)$ into an average of *conditioned expectation values* $E(X|i) = \sum_f P(f|i)E(X|i, f)$. This observation, together with its approximate appearance operationally in weak conditioned measurements, make it tempting to interpret the weak value as a disturbance-free counter-factual conditioned average that can be assigned to the observable within the context of a pre- and post-selected ensemble even when it is not strictly measured [30, 31, 66, 191].

Supporting this point of view is the fact that when the *real part* of (6.1) is bounded by the eigenvalue range of \hat{A} , it agrees with the classical conditioned expectation value for the observable [31]. Moreover, even when the *real part* is outside the normal eigenvalue range, it still obeys a self-consistent logic [32] and seems to indicate oddly sensible information regarding the operator \hat{A} . As such, it has been used quite successfully to analyze and interpret many quantum-mechanical paradoxes both theoretically and experimentally, such as tunneling time [192–195], vacuum Cherenkov radiation [196], cavity QED correlations [45], double-slit complementarity [46, 49], superluminal group velocities [197], the N-box paradox [198, 199], phase singularities [200], Hardy’s paradox [38, 52, 152, 153], photon arrival time [201], Bohmian trajectories [58, 202–204], and Leggett-Garg inequality violations [57, 69, 103] (as shown in Chapter 4).

Arguably more important for its status as a quantity pertaining to the measurement of \hat{A} , however, is the fact that the *real part* of (6.1) appears as a stable weak limit point for conditioned measurements even when the detector is not a von Neumann-coupled continuous wave that can experience superoscillatory in-

terference (e.g. [48, 57, 69, 72, 205]). As a result, we can infer that at least the real part of (6.1) must have some operational significance specifically pertaining to the measurement of \hat{A} that extends beyond the scope of the original derivation.

Indeed, a principled contextual value treatment of a *general conditioned average* of an observable can in fact converge in the weak measurement limit to a generalized expression for the *real part* of (6.1),

$$\text{Re}A_w = \frac{\text{Tr}\left(\hat{P}_f\{\hat{A}, \hat{\rho}_i\}\right)}{2\text{Tr}\left(\hat{P}_f\hat{\rho}_i\right)}, \quad (6.3)$$

where $\{\hat{A}, \hat{\rho}_i\} = \hat{A}\hat{\rho}_i + \hat{\rho}_i\hat{A}$ is the *anti-commutator* between the observable operator and an arbitrary initial state $\hat{\rho}_i$ represented by a density operator, and where \hat{P}_f is an arbitrary post-selection represented by an element from a positive operator-valued measure (POVM)¹. The general conditioned average converges to (6.3) provided that the manner in which \hat{A} is measured satisfies reasonable sufficiency conditions [68, 70, 71] that ensure that the disturbance intrinsic to the measurement process does not persist in the weak limit (see Section 3.5).

It is in this precise restricted sense that we can operationally interpret the *real part* of the weak value (6.3) as an *idealized conditioned average of \hat{A} in the limit of zero measurement disturbance*. Since it is also the only apparent limiting value of the general conditioned average that no longer depends on *how* the measurement of \hat{A} is being made, it is also distinguished as a *measurement context-independent* conditioned average, as anticipated in the discussion around (3.43). These observations provide strong justification for the treatment of the real part of the weak value (6.3) as a form of value assignment [30, 31, 63, 64, 66, 191] for the observable \hat{A} that depends only upon the preparation and post-selection².

¹This equation is a restatement of (3.43) in an operator form, proved in Section 3.5.

²Note that such a value assignment does not violate the Bell-Kochen-Specker theorem [63, 64, 66] since (6.3) does not generally obey the product rule, $(AB)_w \neq A_w B_w$.

However, we are still left with a mystery: what is the significance of the *imaginary part* of (6.1) that appears in the von Neumann measurement, and how does it relate to the operator \hat{A} ? We can find a partial answer to this question in existing literature (e.g. [31, 33, 192, 193]) that has associated the appearance of the imaginary part of (6.1) in the response of the detector with the intrinsic *disturbance*, or *back-action*, of the measurement process. For example, regarding continuous von Neumann detectors Aharonov and Botero [31, p.8] note that “the imaginary part of the complex weak value can be interpreted as a ‘bias function’ for the posterior sampling point [of the detector].” Furthermore, they note that “the weak value of an observable \hat{A} is tied to the role of \hat{A} as a generator for infinitesimal unitary transformations” [31, p.11]. Similarly, while discussing measurements of tunneling time Steinberg [192] states that the imaginary part is a “measure of the back-action on the particle due to the measurement interaction itself” and that the detector shift corresponding to the imaginary part “is sensitive to the details of the measurement apparatus (in particular, to the initial uncertainty in momentum), unlike the [shift corresponding to the real part].”

In Section 6.3, we will augment these observations in the literature by providing a precise operational interpretation of the following generalized expression for the imaginary part of (6.1),

$$\text{Im}A_w = \frac{\text{Tr}\left(\hat{P}_f[-i\hat{A}, \hat{\rho}_i]\right)}{2\text{Tr}\left(\hat{P}_f\hat{\rho}_i\right)}, \quad (6.4)$$

where $[\hat{A}, \hat{\rho}_i] = \hat{A}\hat{\rho}_i - \hat{\rho}_i\hat{A}$ is the *commutator* between \hat{A} and the initial state. We will see that the imaginary part of the weak value does not pertain to the measurement of \hat{A} as an observable. Instead, we will interpret it as half the *logarithmic directional derivative of the post-selection probability along the flow of the group action generated by the operator \hat{A}* . As such, it provides an explicit measure for the idealized disturbance that the coupling to \hat{A} would have induced upon the

initial state in the limit that the detector was not measured, which resembles the suggestion by Steinberg [192]; however, we shall see that the measurement of the detector can strongly alter the state evolution away from that ideal. The explicit commutator in (6.4) also indicates that the imaginary part of the weak value involves the operator \hat{A} in its role as a generator for unitary group transformations as suggested by Aharonov and Botero [31], in contrast to the real part (6.3) that involves the operator \hat{A} in its role as a measurable observable.

Before we provide these interpretations, however, we wish to make it clear how the generalized weak value expressions (6.3) and (6.4) and their interpretations arise within a traditional von Neumann detector. Hence, we will first provide an exact treatment of a von Neumann measurement in Section 6.1 using the formalism of quantum operations (e.g. [22, 25, 27]). In addition to augmenting existing derivations in the literature that are concerned largely with understanding the detector response (e.g. [33, 34, 37, 42, 43, 178–182, 184, 185]), our exact approach serves to connect the standard treatment of weak values to our more general contextual values analysis from Chapter 3 that produces the real part [68, 70, 71] more explicitly. Our exact solutions will also show that the generalized form of the weak value is a universal feature in von Neumann measurements, extending its validity well beyond the weak measurement limit originally conceived by Aharonov *et al.*

We will also briefly comment on the time-symmetry of the weak value in Section 6.4 and provide several explicit examples in Section 6.5 that specialize our exact solutions to typically investigated cases: a particular momentum weak value, a qubit observable measurement, a Gaussian detector, and a detector composed of arbitrary Hermite-Gauss modes. As a consequence, we will show that the Gaussian detector is notable since it induces measurement disturbance that purely *decoheres* the system state into the eigenbasis of \hat{A} in the Lindblad sense with increasing measurement strength. Surprisingly, the pure decoherence allows the

shifts in a Gaussian detector to be completely parametrized by a single complex weak value to all orders in the coupling strength, which allows those shifts to be completely understood using our interpretations of that weak value. For the more general Hermite-Gauss modes, the imaginary part of the weak value must be augmented by an additional correction, further supporting our interpretation of it as a disturbance term.

6.1 von Neumann Measurement

The traditional approach for obtaining a complex weak value [28] for a system observable is to post-select a weak Gaussian von Neumann measurement [11]. The real and imaginary parts of the complex weak value then appear as scaled shifts in the conditioned expectations of conjugate detector observables to linear order in the coupling strength. To clarify how these shifts occur and how the weak value can be interpreted, we shall solve the von Neumann measurement model exactly in the presence of post-selection.

6.1.1 Traditional Analysis

A von Neumann measurement [11, 28] unitarily couples an operator \hat{A} on a *system* Hilbert space \mathcal{H}_s to a momentum operator \hat{p} on a continuous *detector* Hilbert space \mathcal{H}_d via a time-dependent interaction Hamiltonian of the form,

$$\hat{H}_I(t) = g(t)\hat{A} \otimes \hat{p}. \quad (6.5)$$

The interaction profile $g(t)$ is assumed to be a function that is only nonzero over some interaction time interval $t \in [0, T]$. The interaction is also assumed to be *impulsive* with respect to the natural evolution of the initial joint state $\hat{\rho}$ of

the system and detector; i.e., the interaction Hamiltonian (6.5) acts as the total Hamiltonian during the entire interaction time interval.

Solving the Schrödinger equation,

$$i\hbar\partial_t\hat{U} = \hat{H}_I\hat{U}, \quad (6.6)$$

with the initial condition $\hat{U}_0 = \hat{1}$ produces a unitary operator,

$$\hat{U}_T = \exp\left(\frac{g}{i\hbar}\hat{A} \otimes \hat{p}\right), \quad (6.7)$$

$$g = \int_0^T dt g(t), \quad (6.8)$$

that describes the full interaction over the time interval T . The constant g acts as an effective coupling parameter for the impulsive interaction. If the interaction is *weakly coupled* then g is sufficiently small so that $\hat{U}_T \approx \hat{1}$ and the effect of the interaction will be approximately negligible; however, we will make no assumptions about the weakness of the coupling *a priori*.

The unitary interaction (6.7) will entangle the system with the detector so that performing a direct measurement on the detector will lead to an *indirect* measurement being performed on the system. Specifically, we note that the position operator \hat{x} of the detector satisfies the canonical commutation relation $[\hat{x}, \hat{p}] = i\hbar\hat{1}_d$, and thus will evolve in the Heisenberg picture of the interaction according to,

$$(\hat{1}_s \otimes \hat{x})_T = \hat{U}_T^\dagger(\hat{1}_s \otimes \hat{x})\hat{U}_T = \hat{1}_s \otimes \hat{x} + g\hat{A} \otimes \hat{1}_d. \quad (6.9)$$

As a result, measuring the mean of the detector position after the interaction $\langle x \rangle_T = \text{Tr}((\hat{1}_s \otimes \hat{x})_T \hat{\rho})$ will produce,

$$\langle x \rangle_T = \langle x \rangle_0 + g \langle A \rangle_0. \quad (6.10)$$

Hence, the mean of the detector position will be shifted from its initial mean by the mean of the system observable \hat{A} in the *initial reduced system state*, linearly scaled by the coupling strength g . For this reason we say that directly measuring the average of the detector position \hat{x} results in an indirect measurement of the average of the system observable \hat{A} .

The detector momentum \hat{p} , on the other hand, does not evolve in the Heisenberg picture since $[\hat{U}_T, \hat{1}_s \otimes \hat{p}] = 0$. Hence, we expect that measuring the average detector momentum will provide no information about the system observable \hat{A} .

As discussed in the introduction, however, when one conditions such a von Neumann measurement of the detector upon the outcome of a second measurement made only upon the system, then the conditioned average of *both* the position and the momentum of the detector can experience a shift. To see why this is so, we will find it useful to switch to the language of *quantum operations* (e.g. [22, 25, 27]) in order to dissect the measurement in more detail.

6.1.2 Quantum Operations

Unconditioned Measurement

As before, we will assume an impulsive interaction in what follows so that any natural time evolution in the joint system and detector state will be negligible on the time scale of the measurement. (For considerations of the detector dynamics, see [34].) We will also assume for simplicity of discussion that the initial joint state of the system and detector before the interaction is a product state and that the detector state is pure,

$$\hat{\rho} = \hat{\rho}_i \otimes |\psi\rangle\langle\psi|, \quad (6.11)$$

though we will be able to relax this assumption in our final results. Conceptually, this assumption states that a typical detector will be initially well-calibrated and uncorrelated with the unknown system state that is being probed via the interaction.

Evolving the initial state with the interaction unitary \hat{U}_T (6.7) will *entangle* the system with the detector. Hence, subsequently measuring a particular detector position will be equivalent to performing an *operation* \mathcal{M}_x upon the reduced system state, as illustrated in Figure 6.1,

$$\mathcal{M}_x(\hat{\rho}_i) = \text{Tr}_D \left((\hat{1}_s \otimes |x\rangle \langle x|) \hat{U}_T \hat{\rho} \hat{U}_T^\dagger \right) = \hat{M}_x \hat{\rho}_i \hat{M}_x^\dagger, \quad (6.12)$$

$$\hat{M}_x = \langle x| \hat{U}_T |\psi\rangle. \quad (6.13)$$

where $\text{Tr}_D(\cdot)$ is the partial trace over the detector Hilbert space, and \hat{M}_x is the *Kraus operator* associated with the operation \mathcal{M}_x . Furthermore, since $\langle x|\psi\rangle = \psi(x)$ is the initial detector position wave-function we find,

$$\hat{M}_x = \int da \exp(-ga\partial_x) \psi(x) |a\rangle \langle a| = \int da \psi(x - ga) |a\rangle \langle a|, \quad (6.14)$$

or, more compactly, $\hat{M}_x = \psi(x - g\hat{A})$.

If we do *not* perform a subsequent post-selection on the system state, then we trace out the system to find the total probability density for detecting the position x ,

$$p(x) = \text{Tr}_S (\mathcal{M}_x(\hat{\rho}_i)) = \text{Tr}_S \left(\hat{E}_x \hat{\rho}_i \right), \quad (6.15)$$

$$\hat{E}_x = \hat{M}_x^\dagger \hat{M}_x = \langle \psi| \hat{U}_T^\dagger (\hat{1}_s \otimes |x\rangle \langle x|) \hat{U}_T |\psi\rangle, \quad (6.16)$$

where $\text{Tr}_S(\cdot)$ is the partial trace over the system Hilbert space. The *probability operator* \hat{E}_x is a positive system operator that encodes the probability of measuring a particular detector position x , and can also be written in terms of the

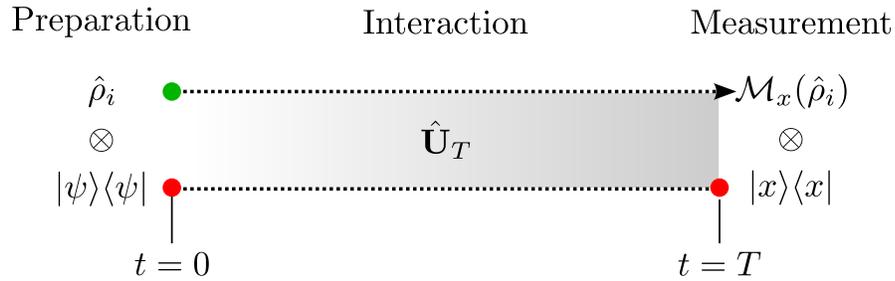


Figure 6.1: (color online) Schematic for a von Neumann measurement. An initially prepared system state $\hat{\rho}_i$ and detector state $|\psi\rangle\langle\psi|$ become entangled with the von Neumann unitary interaction \hat{U}_T (6.7) over a time interval T . Measuring a particular detector position x after the interaction updates the detector state to $|x\rangle\langle x|$ and also updates the system state to $\mathcal{M}_x(\hat{\rho}_i)$, where \mathcal{M}_x (6.12) is an effective measurement operation that encodes the entanglement with and subsequent measurement of the detector.

initial detector position wave-function as $\hat{E}_x = |\psi(x - g\hat{A})|^2$. To conserve probability it satisfies the condition, $\int dx \hat{E}_x = \hat{1}_s$, making the operators \hat{E}_x a *positive operator-valued measure* (POVM) on the system space.

Consequently, averaging the position of the detector will effectively average a system observable with the *initial system state*,

$$\langle x \rangle_T = \int_{-\infty}^{\infty} dx x p(x) = \text{Tr}_S \left(\hat{O} \hat{\rho}_i \right), \quad (6.17)$$

$$\hat{O} = \int_{-\infty}^{\infty} dx x \hat{E}_x = \langle \psi | \hat{U}_T^\dagger (\hat{1}_s \otimes \hat{x}) \hat{U}_T | \psi \rangle = \langle x \rangle_0 \hat{1}_s + g\hat{A}, \quad (6.18)$$

where we see the Heisenberg evolved position operator (6.9) naturally appear.

Since the probability operators \hat{E}_x are diagonal in the basis of \hat{A} , then the effective system operator \hat{O} will also be diagonal in the same basis. Hence, by modifying the values that we assign to the position measurements, we can arrange an indirect measurement of any system observable spanned by $\{\hat{E}_x\}$ in the basis

of \hat{A} , including \hat{A} itself,

$$\hat{A} = \int_{-\infty}^{\infty} dx \left(\frac{x - \langle x \rangle_0}{g} \right) \hat{E}_x, \quad (6.19)$$

The chosen set of values $(x - \langle x \rangle_0)/g$ are *contextual values* for \hat{A} , which, as discussed in Chapters 2 and 3 can be thought of as a generalized spectrum that relates \hat{A} to the specific POVM $\{\hat{E}_x\}$ associated with the *measurement context* $\{\mathcal{M}_x\}$ [68, 70, 71]. They are not the only values that we could assign to the position measurement in order to obtain the equality (6.19) (see, e.g., (3.57)), but they are arguably the simplest to obtain and compute, as well as the most frequently used in the literature. It is in this precise sense that we can say that the von Neumann coupling leads to an indirect measurement of the average of \hat{A} in the absence of post-selection.

The measurement of \hat{A} comes at a cost, however, since the system state is necessarily *disturbed* by the operations \mathcal{M}_x in order to obtain the probability operators \hat{E}_x . The state may even be disturbed more than is strictly required to make the measurement of \hat{A} , which can be seen by rewriting the measurement operators in polar form, $\hat{M}_x = \hat{U}_x |\hat{E}_x|^{1/2}$, with the positive root of the probability operator $|\hat{E}_x|^{1/2}$ and an additional unitary operator \hat{U}_x . This decomposition implies that \mathcal{M}_x splits into an effective composition of two distinct operations,

$$\mathcal{M}_x(\hat{\rho}_i) = \mathcal{U}_x(\mathcal{E}_x(\hat{\rho}_i)), \quad (6.20a)$$

$$\mathcal{E}_x(\hat{\rho}_i) = |\hat{E}_x|^{1/2} \hat{\rho}_i |\hat{E}_x|^{1/2}, \quad (6.20b)$$

$$\mathcal{U}_x(\hat{\rho}'_i) = \hat{U}_x \hat{\rho}'_i \hat{U}_x^\dagger. \quad (6.20c)$$

We can interpret the operation \mathcal{E}_x that involves only the roots of the probability operator $|\hat{E}_x|^{1/2}$ as the *pure measurement operation* producing \hat{E}_x . That is, it represents the *minimum necessary disturbance* that one must make to the initial

state in order to extract a measurable probability. The second operation \mathcal{U}_x unitarily disturbs the initial state, but does not contribute to \hat{E}_x . Since only \hat{E}_x can be used to infer information about \hat{A} through the identity (6.19), we conclude that the disturbance from \mathcal{U}_x is superfluous.

To identify the condition for eliminating \mathcal{U}_x , we can rewrite the Kraus operator (6.13) using the polar form of the initial detector position wave-function $\psi(x) = \exp(i\psi_s(x))\psi_r(x)$,

$$\hat{M}_x = \exp(i\psi_s(x - g\hat{A}))\psi_r(x - g\hat{A}). \quad (6.21)$$

The phase factor becomes the unitary operator $\hat{U}_x = \exp(i\psi_s(x - g\hat{A}))$ for \mathcal{U}_x , while the magnitude becomes the required positive root $|\hat{E}_x|^{1/2} = \psi_r(x - g\hat{A})$ for \mathcal{E}_x . Hence, to eliminate the superfluous operation \mathcal{U}_x from a von Neumann measurement with coupling Hamiltonian (6.5), one must use a *purely real* initial detector wave-function in position.

For contrast, measuring only a particular detector momentum p will be equivalent to performing a different operation \mathcal{N}_p upon the reduced system state,

$$\mathcal{N}_p(\hat{\rho}_i) = \text{Tr}_D \left((\hat{1}_s \otimes |p\rangle \langle p|) \hat{U}_T \hat{\rho}_i \hat{U}_T^\dagger \right) = \hat{N}_p \hat{\rho}_i \hat{N}_p^\dagger, \quad (6.22)$$

$$\hat{N}_p = \langle p | \hat{U}_T | \psi \rangle = \exp \left(\frac{gp}{i\hbar} \hat{A} \right) \langle p | \psi \rangle. \quad (6.23)$$

The Kraus operator \hat{N}_p has a purely unitary factor containing \hat{A} that will disturb the system, regardless of the form of the initial momentum wave-function $\langle p | \psi \rangle$. Moreover, the probability operator associated with the momentum measurement has the form,

$$\hat{F}_p = \hat{N}_p^\dagger \hat{N}_p = |\langle p | \psi \rangle|^2 \hat{1}_s, \quad (6.24)$$

which can only be used to measure the identity $\hat{1}_s$.

For completeness we also briefly note that the conjugate Kraus operators \hat{M}_x and \hat{N}_p are related through a Fourier transform,

$$\hat{N}_p = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dx e^{-ipx/\hbar} \hat{M}_x, \quad (6.25a)$$

$$\hat{M}_x = \frac{1}{\sqrt{2\pi\hbar}} \int_{-\infty}^{\infty} dp e^{ipx/\hbar} \hat{N}_p, \quad (6.25b)$$

and that both detector probability operators can be obtained as marginals of a *Wigner quasi-probability operator* on the system Hilbert space,

$$\hat{W}_{x,p} = \frac{1}{\pi\hbar} \int_{-\infty}^{\infty} dy e^{2ipy/\hbar} \hat{M}_{x+y}^\dagger \hat{M}_{x-y}, \quad (6.26a)$$

$$\hat{E}_x = \int_{-\infty}^{\infty} dp \hat{W}_{x,p}, \quad (6.26b)$$

$$\hat{F}_p = \int_{-\infty}^{\infty} dx \hat{W}_{x,p}. \quad (6.26c)$$

In the absence of interaction, then the Wigner quasi-probability operator reduces to the Wigner quasi-probability distribution $W(x, p)$ for the initial *detector* state, $\hat{W}_{x,p} \xrightarrow{g=0} W(x, p) \hat{1}_s$.

Conditioned Measurement

To post-select the system, an experimenter must perform a second measurement after the von Neumann measurement and filter the two-measurement event space based on the outcomes for the second measurement. In other words, the experimenter keeps only those pairs of outcomes for which the second outcome satisfies some constraint. The remaining measurement pairs can then be averaged to produce *conditioned averages* of the first measurement.

If we represent the second measurement as a set of probability operators $\{\hat{P}_f\}$ indexed by some parameter f that can be derived analogously to (6.16) from a set of operations $\{\mathcal{P}_f\}$ as illustrated in Figure 6.2, then the total joint probability

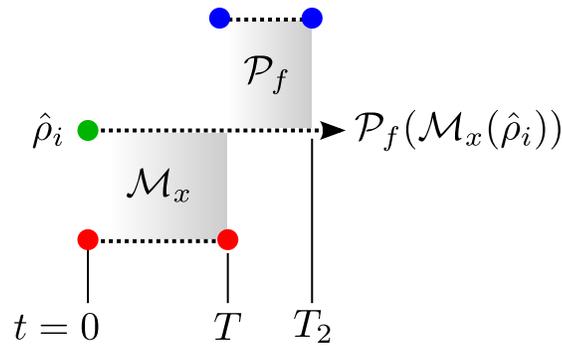


Figure 6.2: (color online) Schematic for a sequence of two indirect measurements. After the von Neumann interaction and measurement of x illustrated in Figure 6.1 that produces the effective measurement operation \mathcal{M}_x upon the initial system state, a second detector interacts impulsively with the system for a time interval $T_2 - T$. The second detector is then measured to have a particular outcome f , which updates the system state to $\mathcal{P}_f(\mathcal{M}_x(\hat{\rho}_i))$, where \mathcal{P}_f is another measurement operation. Taking the trace of the final system state will then produce the joint probability densities (6.27).

densities for the ordered sequences of measurement outcomes (x, f) and (p, f) will be,

$$p(x, f) = \text{Tr}_S \left(\hat{P}_f \mathcal{M}_x(\hat{\rho}_i) \right) = \text{Tr}_S \left(\hat{E}_{x,f} \hat{\rho}_i \right), \quad (6.27a)$$

$$p(p, f) = \text{Tr}_S \left(\hat{P}_f \mathcal{N}_p(\hat{\rho}_i) \right) = \text{Tr}_S \left(\hat{F}_{p,f} \hat{\rho}_i \right), \quad (6.27b)$$

where the joint probability operators,

$$\hat{E}_{x,f} = \hat{M}_x^\dagger \hat{P}_f \hat{M}_x, \quad (6.28a)$$

$$\hat{F}_{p,f} = \hat{N}_p^\dagger \hat{P}_f \hat{N}_p. \quad (6.28b)$$

are not simple products of the post-selection \hat{P}_f and the probability operators (6.16) or (6.24). Those operators can be recovered, however, by marginalizing over the index f , since the post-selection probability operators must satisfy a POVM condition $\sum_f \hat{P}_f = \hat{1}_s$.

The joint probabilities (6.27) will contain information not only about the first measurement and the initial system state, but also about the second measurement and any *disturbance* to the initial state that occurred due to the first measurement. In particular, the joint probability operators (6.28) can no longer satisfy the identity (6.19) due to the second measurement, so averaging the probabilities (6.27) must reveal more information about the measurement process than can be obtained solely from the operator \hat{A} , the initial state $\hat{\rho}_i$, and the post-selection \hat{P}_f . As a poignant example, the unitary disturbance \mathcal{U}_x in (6.20) that did not contribute to the operator identity (6.19) *will* contribute to the joint probability operators, $\hat{E}_{x,f} = |\hat{E}_x|^{1/2} \hat{U}_x^\dagger \hat{P}_f \hat{U}_x |\hat{E}_x|^{1/2}$.

The total probability for obtaining the post-selection outcome f can be obtained by marginalizing over either x or p in the joint probabilities,

$$p(f) = \int_{-\infty}^{\infty} dx p(x, f) = \int_{-\infty}^{\infty} dp p(p, f) = \text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right), \quad (6.29)$$

$$\mathcal{E}(\hat{\rho}_i) = \text{Tr}_D \left(\hat{U}_T (\hat{\rho}_i \otimes |\psi\rangle \langle \psi|) \hat{U}_T^\dagger \right), \quad (6.30)$$

where the operation \mathcal{E} is the total *non-selective* measurement that has been performed on $\hat{\rho}_i$. Since \mathcal{E} is not the identity operation, it represents the total *disturbance* intrinsic to the measurement process. It includes unitary evolution of the reduced system state due to the interaction Hamiltonian (6.5), as well as *decoherence* stemming from entanglement with the measured detector.

By experimentally filtering the event pairs to keep only a particular outcome f of the second measurement, an experimenter can obtain the conditional proba-

bilities,

$$p(x|f) = \frac{p(x, f)}{p(f)} = \frac{\text{Tr}_S \left(\hat{P}_f \mathcal{M}_x(\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)}, \quad (6.31a)$$

$$p(p|f) = \frac{p(p, f)}{p(f)} = \frac{\text{Tr}_S \left(\hat{P}_f \mathcal{M}_p(\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)}, \quad (6.31b)$$

which can then be averaged to find the exact conditioned averages for the detector position and momentum,

$${}_f \langle x \rangle_T = \int_{-\infty}^{\infty} dx x p(x|f) = \frac{\text{Tr} \left((\hat{P}_f \otimes \hat{x}) \hat{\rho}_T \right)}{\text{Tr} \left((\hat{P}_f \otimes \hat{1}_d) \hat{\rho}_T \right)}, \quad (6.32a)$$

$${}_f \langle p \rangle_T = \int_{-\infty}^{\infty} dp p p(p|f) = \frac{\text{Tr} \left((\hat{P}_f \otimes \hat{p}) \hat{\rho}_T \right)}{\text{Tr} \left((\hat{P}_f \otimes \hat{1}_d) \hat{\rho}_T \right)}, \quad (6.32b)$$

where we have written out the operations explicitly and where $\hat{\rho}_T = \hat{U}_T(\hat{\rho}_i \otimes |\psi\rangle \langle\psi|) \hat{U}_T^\dagger$ is the post-interaction joint state. It is worth noting at this point that we can relax the assumption (6.11) made about the initial state in the exact expressions (6.32). Similarly, if different contextual values are used to average the conditional probabilities in (6.32), then corresponding detector observables with the same spectra will appear in (6.32) in place of \hat{x} or \hat{p} ; for example, averaging the values $\alpha(x) = (x - \langle x \rangle_0)/g$ used in (6.19) will replace the detector observable \hat{x} with $\hat{\alpha} = \int_{-\infty}^{\infty} dx \alpha(x) |x\rangle \langle x|$.

6.2 Joint Weak Values

As written, Eqs. (6.32) show that the joint observables $\hat{P}_f \otimes \hat{x}$ and $\hat{P}_f \otimes \hat{p}$ are averaged with respect to the final joint state $\hat{\rho}_T$. However, we can also express these averages in terms of the *initial* joint state by commuting the detector observables

symmetrically past the evolution operators \hat{U}_T using the canonical commutation relations as in (6.9),

$${}_f\langle x \rangle_T = \text{Re} \langle x \rangle^w + g \text{Re} \langle A \rangle^w, \quad (6.33a)$$

$${}_f\langle p \rangle_T = \text{Re} \langle p \rangle^w. \quad (6.33b)$$

The averages are exactly characterized by the real parts of three *generalized weak values* [45, 68, 70, 71, 73] that are defined in the *joint* Hilbert space of the system and detector,

$$\langle A \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{A} \otimes \hat{1}_d) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (6.34a)$$

$$\langle x \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{1}_s \otimes \hat{x}) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (6.34b)$$

$$\langle p \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{1}_s \otimes \hat{p}) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}. \quad (6.34c)$$

The pre-selection for each weak value is equal to the initial joint state $\hat{\rho}$, while the post-selection is equal to the Heisenberg-evolved joint post-selection operator, $\hat{P}_T = \hat{U}_T^\dagger (\hat{P}_f \otimes \hat{1}_d) \hat{U}_T$. Higher-order detector moments are provided in Appendix D for completeness, and all have similar expansions into joint weak values.

The form of the equations (6.34) clearly illustrates how the post-selection will affect the measurement. If the post-selection is the identity operator, $\hat{P}_f = \hat{1}_s$, then the unitary operators \hat{U}_T causing the total disturbance of the initial state will cancel, leaving the averages in the *initial* states that were previously obtained,

$$\langle x \rangle_T = \langle x \rangle_0 + g \langle A \rangle_0, \quad (6.35a)$$

$$\langle p \rangle_T = \langle p \rangle_0. \quad (6.35b)$$

In this sense, commuting the detector operators \hat{x} and \hat{p} in (6.32) through the unitary operators to arrive at (6.33) is equivalent to evolving them in the Heisenberg picture back from the time of measurement T to the initial time 0 in order to compare them with the initial states. However, the presence of the post-selection operator \hat{P}_f will now generally spoil the cancellation of the unitary operators that is implicit in the Heisenberg picture, leading to corrections from the disturbance between the pre- and post-selection.

Importantly, these relations hold for any coupling strength g , any (possibly entangled) initial joint state $\hat{\rho}$, and any generalized post-selection \hat{P}_f ; that is, *all von Neumann detector (conditioned) averages are exactly described by generalized weak values*. This important result seems to have been missed in the existing literature due to the fact that the generalized weak values (6.34) cannot be written in a form with projective pre- and post-selections as defined originally by AAV [28]. Moreover, they explicitly include the detector information, so are not solely system quantities.

6.2.1 Reduced State Expressions

If we prepare a product initial state $\hat{\rho} = \hat{\rho}_i \otimes \hat{\rho}_d$, where $\hat{\rho}_i$ ($\hat{\rho}_d$) is the initial state of the system (detector), then we can exploit the product form of the observables to further simplify Eqs. (6.34). Notably, since $[\hat{A}, \hat{U}_T] = 0$ we can express Eq. (6.34a) as a weak value only on the system Hilbert space,

$$\langle A \rangle^w = \frac{\text{Tr}_S \left(\hat{P}_f \hat{A} \mathcal{E}(\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)}, \quad (6.36)$$

where the pre-selection state,

$$\mathcal{E}(\hat{\rho}_i) = \text{Tr}_D \left(\hat{U}_T (\hat{\rho}_i \otimes \hat{\rho}_d) \hat{U}_T^\dagger \right), \quad (6.37)$$

is the reduced system state *after* the interaction, which has the same form as a non-selective measurement (6.30). All detector information has been absorbed into an effective preparation of the reduced system state $\mathcal{E}(\hat{\rho}_i)$.

To directly compare the joint weak values Eqs. (6.34b) and (6.34c) with (6.36), we also express them within the system Hilbert space,

$$\text{Re} \langle x \rangle^w = \frac{\text{Tr}_S \left(\hat{P}_f \mathcal{X}(\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)}, \quad (6.38a)$$

$$\text{Re} \langle p \rangle^w = \frac{\text{Tr}_S \left(\hat{P}_f \mathcal{P}(\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)}, \quad (6.38b)$$

by introducing the operations,

$$\mathcal{X}(\hat{\rho}_i) = \text{Tr}_D \left(\hat{U}_T \left(\hat{\rho}_i \otimes \left(\frac{\hat{x}\hat{\rho}_d + \hat{\rho}_d\hat{x}}{2} \right) \right) \hat{U}_T^\dagger \right), \quad (6.39a)$$

$$\mathcal{P}(\hat{\rho}_i) = \text{Tr}_D \left(\hat{U}_T \left(\hat{\rho}_i \otimes \left(\frac{\hat{p}\hat{\rho}_d + \hat{\rho}_d\hat{p}}{2} \right) \right) \hat{U}_T^\dagger \right), \quad (6.39b)$$

that act upon the initial system state.

The Weyl-ordered operator products with the initial detector state that appear motivate us to define the Wigner distribution of the detector and its Fourier transform,

$$W_d(x, p) = \frac{1}{2\pi\hbar} \int dy \langle x - y/2 | \hat{\rho}_d | x + y/2 \rangle e^{ipy/\hbar}, \quad (6.40a)$$

$$\widetilde{W}_d(x, y) = \int dp W_d(x, p) e^{-ipy/\hbar} = \langle x - y/2 | \hat{\rho}_d | x + y/2 \rangle. \quad (6.40b)$$

With this distribution, we can express the exact reduced system state (6.37) and

the operations (6.39) in a useful and compact form,

$$\mathcal{E}(\hat{\rho}_i) = \int dx \widetilde{W}_d(x, g \text{ad}^*[\hat{A}])(\hat{\rho}_i), \quad (6.41a)$$

$$\mathcal{X}(\hat{\rho}_i) = \int dx x \widetilde{W}_d(x, g \text{ad}^*[\hat{A}])(\hat{\rho}_i), \quad (6.41b)$$

$$\mathcal{P}(\hat{\rho}_i) = i\hbar \left[\partial_z \int dx \widetilde{W}_d(x, z) \right]_{z \rightarrow g \text{ad}^*[\hat{A}]}(\hat{\rho}_i) = i\hbar \partial_{g \text{ad}^*[\hat{A}]} \mathcal{E}(\hat{\rho}_i). \quad (6.41c)$$

Notice that the left adjoint action (A.11) of \hat{A} , $\text{ad}^*[\hat{A}](\hat{B}) = \hat{A}\hat{B} - \hat{B}\hat{A}$, discussed in Appendix A.3 naturally appears, and that Eq. (6.41c) allows one to simply obtain the momentum response once the functional form of reduced system state (6.41a) is known. Derivations of these expressions, as well as generalizations to higher-order detector moments are provided in Appendix D. We will return to examples of these general expressions in Section 6.5.4.

6.3 Linear Response Regime

To see how the operations \mathcal{E} , \mathcal{X} , and \mathcal{P} defined in (6.37) and (6.39) depend on the coupling strength g , we can also expand them perturbatively,

$$\mathcal{E}(\hat{\rho}_i) = \sum_{n=0} \frac{1}{n!} \left(\frac{g}{\hbar}\right)^n \langle p^n \rangle_0 (\text{ad}^*[-i\hat{A}])^n(\hat{\rho}_i), \quad (6.42a)$$

$$\mathcal{X}(\hat{\rho}_i) = \sum_{n=0} \frac{1}{n!} \left(\frac{g}{\hbar}\right)^n \frac{\langle \{p^n, x\} \rangle_0}{2} (\text{ad}^*[-i\hat{A}])^n(\hat{\rho}_i), \quad (6.42b)$$

$$\mathcal{P}(\hat{\rho}_i) = \sum_{n=0} \frac{1}{n!} \left(\frac{g}{\hbar}\right)^n \langle p^{n+1} \rangle_0 (\text{ad}^*[-i\hat{A}])^n(\hat{\rho}_i), \quad (6.42c)$$

where $\{p^n, x\} = p^n x + x p^n$ is the anti-commutator. Again, the left adjoint action (A.11) of \hat{A} as a Lie algebraic element explicitly appears and describes how \hat{A} *disturbs* the initial state due to the interaction that measures it.

The initial detector state plays a critical role in (6.42) by determining the

various moments, $\langle p^n \rangle_0$, $\langle p^{n+1} \rangle_0$ and $\langle \{p^n, x\}/2 \rangle_0$ that appear in the series expansions. Notably, if we make the initial detector wave-function purely real so that it minimally disturbs the system state then all moments containing odd powers of \hat{p} will vanish. We conclude that those moments of the disturbance operations are superfluous for obtaining the measurable probabilities that allow the measurement of \hat{A} , while the moments with even powers of \hat{p} are necessary.

After expanding the exact conditioned averages (6.33) to first order in g , we obtain the linear response of the conditioned detector means due to the interaction,

$${}_f\langle x \rangle_T \rightarrow \langle x \rangle_0 + \frac{g}{\hbar} \frac{\langle \{p, x\} \rangle_0}{2} \frac{\text{Tr}_S \left(\hat{P}_f \text{ad}^*[-i\hat{A}](\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \hat{\rho}_i \right)} + g \frac{\text{Tr}_S \left(\hat{P}_f \{ \hat{A}, \hat{\rho}_i \} \right)}{2 \text{Tr}_S \left(\hat{P}_f \hat{\rho}_i \right)}, \quad (6.43a)$$

$${}_f\langle p \rangle_T \rightarrow \langle p \rangle_0 + \frac{g}{\hbar} \langle p^2 \rangle_0 \frac{\text{Tr}_S \left(\hat{P}_f \text{ad}^*[-i\hat{A}](\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \hat{\rho}_i \right)}. \quad (6.43b)$$

Measurements for which this linear response is a good approximation are known as (AAV) *weak measurements*.

By introducing the generalized weak value with no measurement disturbance as a complex parameter,

$$A^w = \frac{\text{Tr}_S \left(\hat{P}_f \hat{A} \hat{\rho}_i \right)}{\text{Tr}_S \left(\hat{P}_f \hat{\rho}_i \right)}, \quad (6.44)$$

we thus have the following approximate replacements to linear order in g ,

$$g \text{Re} \langle A \rangle^w \approx g \text{Re} A^w, \quad (6.45a)$$

$$\text{Re} \langle x \rangle^w \approx \langle x \rangle_0 + \frac{g}{\hbar} \frac{\langle \{p, x\} \rangle_0}{2} (2 \text{Im} A^w), \quad (6.45b)$$

$$\text{Re} \langle p \rangle^w \approx \langle p \rangle_0 + \frac{g}{\hbar} \langle p^2 \rangle_0 (2 \text{Im} A^w), \quad (6.45c)$$

in terms of not only its real part in the system weak value, but also *twice its imaginary part* in the *detector* weak values. The linear response formulas (6.43) acquire the compact form,

$${}_f\langle x \rangle_T \approx \langle x \rangle_0 + \frac{g}{\hbar} \frac{\langle \{p, x\} \rangle_0}{2} (2\text{Im}A^w) + g \text{Re}A^w, \quad (6.46a)$$

$${}_f\langle p \rangle_T \approx \langle p \rangle_0 + \frac{g}{\hbar} \langle p^2 \rangle_0 (2\text{Im}A^w). \quad (6.46b)$$

If the initial detector state $\hat{\rho}_d = |\psi\rangle\langle\psi|$ consists of a purely real position wavefunction $\langle x|\psi\rangle = \psi(x)$, so that the measurement is minimally disturbing, then $\langle \{p, x\}/2 \rangle_0$ will vanish, leaving only $\text{Re}A^w$ in ${}_f\langle x \rangle_T$ as a linear approximation to the full conditioned average $\text{Re}\langle A \rangle^w$ defined in (6.36). However, the term proportional to $2\text{Im}A^w$ will not vanish in ${}_f\langle p \rangle_T$ to linear order in g , making it an element of measurement disturbance that persists even for minimally disturbing weak measurements.

The linear response formulas (6.46) for the von Neumann measurement have also been obtained and discussed in the literature with varying degrees of generality and rigor (e.g. [28, 29, 32–34, 37, 42, 43, 178, 180–182, 184, 185]). However, our derivation has a conceptual advantage in that we see explicitly how the origins of the real and imaginary parts of the weak value differ with respect to the measurement of \hat{A} . We are therefore in a position to give concrete interpretations for each part.

The real part (6.3) of the complex weak value parameter $\text{Re}A^w$ stems directly from the part of the conditioned shift of the detector pointer that corresponds to the measurement of \hat{A} and does not contain any further perturbation induced by the measurement coupling that would be indicated by factors of $\text{ad}^*[-i\hat{A}]$. As a result, it can be interpreted as an idealized limit point for the average of \hat{A} in the initial state $\hat{\rho}_i$ that has been conditioned on the post-selection \hat{P}_f without any appreciable intermediate measurement disturbance. We expect this interpretation

due to our analysis in Section 3.5 that shows $\text{Re}A^w$ appears naturally as a limit point even for minimally disturbing measurements that are not of von Neumann type.

The imaginary part (6.4) of the complex weak value parameter $\text{Im}A^w$, on the other hand, stems directly from the *disturbance* of the measurement and explicitly contains $\text{ad}^*[-i\hat{A}]$, which is the action of \hat{A} as a generator for unitary group evolution due to the specific Hamiltonian (6.5). The factor $2\text{Im}A^w$ appears in (6.46) along with information about the initial detector momentum that is being coupled to \hat{A} in the Hamiltonian (6.5), as well as factors of \hbar , in stark contrast to the real part.

The significance of $2\text{Im}A^w$ becomes more clear once we identify the left adjoint action of \hat{A} that appears in its numerator as a *directional derivative*,

$$\delta_A(\cdot) = \text{ad}^*[-i\hat{A}](\cdot). \quad (6.47)$$

That is, $\delta_A(\hat{\rho}_i)$ indicates the *rate of change* of the initial state $\hat{\rho}_i$ along a *flow* in state-space generated by \hat{A} .

As mentioned in Appendix A.3, this directional derivative should be familiar from the Heisenberg equation of motion (A.12), or from the equivalent Liouville-Schrödinger equation written in the form $\partial_t \hat{\rho} = [\hat{H}, \hat{\rho}]/i\hbar = \delta_\Omega(\hat{\rho})$, where the scaled Hamiltonian $\hat{\Omega} = \hat{H}/\hbar$ is a characteristic frequency operator. The integration of this equation is an adjoint group operation in exponential form $\hat{\rho}(t) = \exp(t\delta_\Omega)(\hat{\rho}(0)) = \exp(-it\hat{\Omega})\hat{\rho}(0)\exp(it\hat{\Omega})$ that specifies a *flow* in state space, which is a collection of curves that is parametrized both by a time parameter t and by the initial condition $\hat{\rho}(0)$. Specifying the initial condition $\hat{\rho}(0) = \hat{\rho}_i$, picks out the specific curve from the flow that contains $\hat{\rho}_i$. The directional derivative of the initial state along that specific curve is then defined in the standard way, $\partial_t \hat{\rho}(t)|_{t=0} = \delta_\Omega(\hat{\rho}_i)$.

The fact that the quantum state space is always a continuous manifold of states built fundamentally from a Lie group, as discussed more generally in Appendix A, allows such a flow to be defined using any Hermitian operator, such as \hat{A} , as a generator. Analogously to time evolution, such a flow has the form of a unitary operation, $\hat{\rho}(\epsilon) = \exp(\epsilon\delta_A)(\hat{\rho}(0))$, where the real parameter ϵ for the flow has units inverse to \hat{A} . Therefore, taking the directional derivative of $\hat{\rho}_i$ along the specific curve of this flow that passes through $\hat{\rho}_i$ will produce (6.47). For an explicit example that we will detail in Section 6.5.2, the state-space of a qubit can be parametrized as the continuous volume of points inside the unit Bloch sphere; the derivative (6.47) produces the vector field illustrated in Figure 6.3 tangent to the flow corresponding to Rabi oscillations of the qubit.

With this intuition in mind, we define the post-selection probability for measuring \hat{P}_f given an initial state $\hat{\rho}_i(\epsilon) = \exp(\epsilon\delta_A)(\hat{\rho}_i)$ that is changing along the flow generated by \hat{A} ,

$$p_f(\epsilon) = \text{Tr}_S \left(\hat{P}_f \hat{\rho}_i(\epsilon) \right). \quad (6.48)$$

The logarithmic directional derivative of this post-selection probability then produces the factor $2\text{Im}A_w$ that appears in (6.46),

$$2\text{Im}A^w = \partial_\epsilon \ln p_f(\epsilon) \Big|_{\epsilon=0}. \quad (6.49)$$

Thus, the imaginary part of the complex weak value parameter is half the logarithmic directional derivative of the post-selection probability along the natural group flow generated by \hat{A} . It does not provide any information about the measurement of \hat{A} as an observable, but rather *indicates an instantaneous rate of multiplicative change to the post-selection probability* due to disturbance of the initial state caused by \hat{A} in its role as a generator for unitary transformations.

Specifically, for small ϵ we have the approximate relation,

$$\frac{p_f(\epsilon)}{p_f(0)} \approx 1 + (2\text{Im}A^w)\epsilon. \quad (6.50)$$

6.4 Time Symmetry

As noted in [32, 97], a quantum system that has been pre- and post-selected exhibits time symmetry. We can make the time symmetry more apparent in our operational treatment by introducing the *retrodictive state*,

$$\hat{\rho}_f = \frac{\hat{P}_f}{\text{Tr}_S(\hat{P}_f)}, \quad (6.51)$$

associated with the post-selection (see, e.g., [206, 207]) and rewriting our main results in the time-reversed retrodictive picture.

After cancelling normalization factors, the joint weak values (6.36) and (6.38) that characterize the response (6.33) for a system retrodictively prepared in the final state $\hat{\rho}_f$ and then conditioned on the *pre-selection* measurement producing the initial system state $\hat{\rho}_i$ have the equivalent form,

$$\langle A \rangle^w = \frac{\text{Tr}_S(\mathcal{E}^*(\hat{\rho}_f)\hat{A}\hat{\rho}_i)}{\text{Tr}_S(\mathcal{E}^*(\hat{\rho}_f)\hat{\rho}_i)}, \quad (6.52a)$$

$$\text{Re} \langle x \rangle^w = \frac{\text{Tr}_S(\mathcal{X}^*(\hat{\rho}_f)\hat{\rho}_i)}{\text{Tr}_S(\mathcal{E}^*(\hat{\rho}_f)\hat{\rho}_i)}, \quad (6.52b)$$

$$\text{Re} \langle p \rangle^w = \frac{\text{Tr}_S(\mathcal{P}^*(\hat{\rho}_f)\hat{\rho}_i)}{\text{Tr}_S(\mathcal{E}^*(\hat{\rho}_f)\hat{\rho}_i)}, \quad (6.52c)$$

where the retrodictive operations \mathcal{E}^* , \mathcal{X}^* , and \mathcal{P}^* are the adjoints with respect to

the trace of the predictive operations in (6.37) and (6.39),

$$\mathcal{E}^*(\hat{\rho}_f) = \text{Tr}_D \left(\hat{U}_T^\dagger(\hat{\rho}_f \otimes \hat{1}_d) \hat{U}_T(\hat{1}_s \otimes \hat{\rho}_d) \right), \quad (6.53a)$$

$$\mathcal{X}^*(\hat{\rho}_f) = \text{Tr}_D \left(\hat{U}_T^\dagger(\hat{\rho}_f \otimes \hat{1}_d) \hat{U}_T \left(\hat{1}_s \otimes \left(\frac{\hat{x}\hat{\rho}_d + \hat{\rho}_d\hat{x}}{2} \right) \right) \right), \quad (6.53b)$$

$$\mathcal{P}^*(\hat{\rho}_f) = \text{Tr}_D \left(\hat{U}_T^\dagger(\hat{\rho}_f \otimes \hat{1}_d) \hat{U}_T \left(\hat{1}_s \otimes \left(\frac{\hat{p}\hat{\rho}_d + \hat{\rho}_d\hat{p}}{2} \right) \right) \right). \quad (6.53c)$$

Notably, the retrodictive state propagates backwards in time. As discussed in [207], it describes the measurement apparatus outcome for the post-selection, and not the system itself; in particular, the retrodictive state does not equal the predictive state after the post-selection unless the post-selection is projective. One can thus envision each state as corresponding to a boundary condition specified by a measurement apparatus.

After expanding the retrodictive operations perturbatively as in (6.42),

$$\mathcal{E}^*(\hat{\rho}_f) = \sum_{n=0} \frac{1}{n!} \left(\frac{g}{\hbar} \right)^n \langle p^n \rangle_0 (\text{ad}[-i\hat{A}])^n(\hat{\rho}_f), \quad (6.54a)$$

$$\mathcal{X}^*(\hat{\rho}_f) = \sum_{n=0} \frac{1}{n!} \left(\frac{g}{\hbar} \right)^n \frac{\langle \{p^n, x\} \rangle_0}{2} (\text{ad}[-i\hat{A}])^n(\hat{\rho}_f), \quad (6.54b)$$

$$\mathcal{P}^*(\hat{\rho}_f) = \sum_{n=0} \frac{1}{n!} \left(\frac{g}{\hbar} \right)^n \langle p^{n+1} \rangle_0 (\text{ad}[-i\hat{A}])^n(\hat{\rho}_f), \quad (6.54c)$$

where $\text{ad}[-i\hat{A}](\cdot) = -\text{ad}^*[-i\hat{A}](\cdot) = [\cdot, -i\hat{A}]$ is the right adjoint action of \hat{A} defined in (A.11) in Appendix A.3, then the linear response of the detector (6.46) can be written in terms of the retrodictive forms of the real and imaginary parts of the complex weak value parameter,

$$\text{Re}A^w = \frac{\text{Tr}_S \left(\{\hat{\rho}_f, \hat{A}\} \hat{\rho}_i \right)}{2\text{Tr}_S(\hat{\rho}_f \hat{\rho}_i)}, \quad (6.55)$$

$$2\text{Im}A^w = \frac{\text{Tr}_S \left([\hat{\rho}_f, -i\hat{A}] \hat{\rho}_i \right)}{\text{Tr}_S(\hat{\rho}_f \hat{\rho}_i)}, \quad (6.56)$$

which can be compared with (6.3), and (6.4).

6.5 Examples

6.5.1 Bohmian Mechanics

To make the preceding abstract discussion of the weak value more concrete, let us consider a special case that has been recently discussed by Leavens [202], Wiseman [203], and Hiley [204], where the operator $\hat{A} = \hat{p}$ being measured is the momentum operator of the system particle. Since the wave-number operator $\hat{k} = -\hat{p}/\hbar$ generates a flow that is parametrized by the position x , then we expect from the discussion surrounding (6.49) that the imaginary part of a momentum weak value will give information about how the post-selection probability will change along changes in position.

If we restrict our initial system state to be a pure state $\hat{\rho}_i = |\phi\rangle\langle\phi|$, and post-select the measurement of the momentum on a particular position $\hat{P}_f = |x\rangle\langle x|$, then the detector will have the linear response relations (6.46) with the complex weak value given by,

$$p_w = \frac{\langle x | \hat{p} | \phi \rangle}{\langle x | \phi \rangle} = \frac{-i\hbar\partial_x\phi(x)}{\phi(x)}. \quad (6.57)$$

We can split this value naturally into its real and imaginary parts by considering the polar decomposition of the initial system state $\phi(x) = r(x)\exp(iS(x))$,

$$p_w = \hbar\partial_x S(x) - i\hbar\partial_x \ln r(x). \quad (6.58)$$

The real part of the weak value $\text{Re } p_w = \hbar\partial_x S(x)$ is the phase gradient, or *Bohmian momentum* for the initial state, which we can now interpret operationally as the *average momentum conditioned on the subsequent measurement of*

a particular x in the ideal limit of no measurement disturbance. Indeed, Hiley has recently identified the Bohmian momentum as the local momentum component of the energy-momentum tensor [91]. The connection between the real part of a weak value and the Bohmian momentum that was pointed out in [202, 203] has recently allowed Kocsis *et al.* [58] to experimentally reconstruct the averaged Bohmian trajectories in an optical two-slit interference experiment using such a von Neumann measurement.

The imaginary part of the weak value, $\text{Im } p_w = -\hbar \partial_x \ln r(x)$, on the other hand, is the logarithmic gradient of the root of the probability density $\rho(x) = |\phi(x)|^2 = r^2(x)$ for the particle at the point x . Written in the form,

$$2\text{Im } p_w = -\hbar \partial_x \ln \rho(x), \quad (6.59)$$

it describes the instantaneous exponential rate of positional change of the probability density with respect to the particular post-selection point x , as expected. This quantity, scaled by an inverse mass $1/m$, was introduced under the name “osmotic velocity” in the context of a stochastic interpretation of quantum mechanics developed by Nelson [208], where it produced a diffusion term in the stochastic equations of motion for a classical point particle with diffusion coefficient $\hbar/2m$. Nelson’s interpretation was carefully contrasted with a stochastic interpretation for the Bohmian pilot wave by Bohm and Hiley [209], and the connection of the osmotic velocity with a weak value was recently emphasized by Hiley [204].

Hence, the imaginary part of the momentum weak value does not provide information about a measurement of the momentum in the initial state. Instead, it indicates the logarithmic directional derivative of the probability density for measuring x along the flow generated by \hat{p} . The scaled derivative $-\hbar \partial_x$ appears since $\hat{p} = -\hbar \hat{k}$ and \hat{k} generates flow along the position x .

6.5.2 Qubit Observable

To make the full von Neumann measurement process more concrete, let us also consider a simple example where \hat{A} operates on the two-dimensional Hilbert space of a qubit. (See also [33, 34, 37, 42, 43, 178–182, 184, 185].) We can in such a case simplify the perturbative expansions (6.42) using the following identities,

$$\hat{A} = A\hat{\sigma}_3, \quad (6.60a)$$

$$\hat{\rho}_i = \frac{1}{2} \left(\hat{1}_s + \sum_k r_k \hat{\sigma}_k \right), \quad (6.60b)$$

$$[\hat{\sigma}_j, \hat{\sigma}_k] = 2i\epsilon_{jkl}\hat{\sigma}_l, \quad (6.60c)$$

$$\{\hat{\sigma}_j, \hat{\sigma}_k\} = 2\delta_{jk}\hat{1}_s, \quad (6.60d)$$

where $\{\hat{\sigma}_k\}_{k=1}^3$ are the usual Pauli operators, the components of the initial system state $\{r_k\}_{k=1}^3$ are real and satisfy the inequality $0 \leq \sum_k r_k^2 \leq 1$, ϵ_{jkl} is the completely antisymmetric Levi-Civita pseudotensor, and δ_{jk} is the Kronecker delta. We have defined $\hat{\sigma}_3$ to be diagonal in the eigenbasis of \hat{A} and have rescaled the spectrum of \hat{A} for simplicity to zero out its maximally mixed mean $\text{Tr}_S(\hat{A}\hat{1}_s/2) = 0$. As a result, $\langle A \rangle_0 = Ar_3$.

It follows that for positive integer n the repeated actions of \hat{A} on the various qubit operators have the forms,

$$\text{ad}^*[\hat{A}]^n(\hat{1}_s) = 0, \quad (6.61a)$$

$$\text{ad}^*[\hat{A}]^{2n-1}(\hat{\sigma}_1) = i\hat{\sigma}_2(2A)^{2n-1}, \quad (6.61b)$$

$$\text{ad}^*[\hat{A}]^{2n}(\hat{\sigma}_1) = \hat{\sigma}_1(2A)^{2n}, \quad (6.61c)$$

$$\text{ad}^*[\hat{A}]^{2n-1}(\hat{\sigma}_2) = -i\hat{\sigma}_1(2A)^{2n-1}, \quad (6.61d)$$

$$\text{ad}^*[\hat{A}]^{2n}(\hat{\sigma}_2) = \hat{\sigma}_2(2A)^{2n}, \quad (6.61e)$$

$$\text{ad}^*[\hat{A}]^n(\hat{\sigma}_3) = 0, \quad (6.61f)$$

which collectively imply that,

$$\text{ad}^*[\hat{A}]^{2n-1}(\hat{\rho}_i) = \frac{i}{2}(2A)^{2n-1}(r_1\hat{\sigma}_2 - r_2\hat{\sigma}_1), \quad (6.62a)$$

$$\text{ad}^*[\hat{A}]^{2n}(\hat{\rho}_i) = \frac{1}{2}(2A)^{2n}(r_1\hat{\sigma}_1 + r_2\hat{\sigma}_2), \quad (6.62b)$$

and hence that the nonselective measurement operation has the exact form,

$$\mathcal{E}(\hat{\rho}_i) = \hat{\rho}_i + \frac{c(g)r_1 - s(g)r_2}{2}\hat{\sigma}_1 + \frac{c(g)r_2 + s(g)r_1}{2}\hat{\sigma}_2, \quad (6.63a)$$

$$c(g) = \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} \left(\frac{2Ag}{\hbar}\right)^{2n} \langle p^{2n} \rangle_0, \quad (6.63b)$$

$$s(g) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)!} \left(\frac{2Ag}{\hbar}\right)^{2n-1} \langle p^{2n-1} \rangle_0. \quad (6.63c)$$

The correction term can be interpreted as a Rabi oscillation of the qubit that has been perturbed by the coupling to the detector. Indeed, if the detector operator \hat{p} were replaced with a constant p , then the interaction Hamiltonian (6.5) would constitute an evolution term for the qubit that would induce Rabi oscillations around the $\hat{\sigma}_3$ axis of the Bloch sphere, which would be the natural flow in state space generated by the action of \hat{A} . With the substitution $\hat{p} \rightarrow p$ then $\langle p^n \rangle_0 \rightarrow p^n$, so $c(g) \rightarrow \cos(2gAp/\hbar) - 1$ and $s(g) \rightarrow \sin(2gAp/\hbar)$, which restores the unperturbed Rabi oscillations.

Similarly, we find that the averaging operations for the detector position and momentum (6.39) have the exact forms,

$$\mathcal{X}(\hat{\rho}_i) = \langle x \rangle_0 \hat{\rho}_i + \frac{c_x(g)r_1 - s_x(g)r_2}{2}\hat{\sigma}_1 + \frac{c_x(g)r_2 + s_x(g)r_1}{2}\hat{\sigma}_2, \quad (6.64a)$$

$$c_x(g) = \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} \left(\frac{2Ag}{\hbar}\right)^{2n} \frac{\langle \{p^{2n}, x\} \rangle_0}{2}, \quad (6.64b)$$

$$s_x(g) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)!} \left(\frac{2Ag}{\hbar}\right)^{2n-1} \frac{\langle \{p^{2n-1}, x\} \rangle_0}{2}, \quad (6.64c)$$

and,

$$\mathcal{P}(\hat{\rho}_i) = \langle p \rangle_0 \hat{\rho}_i + \frac{c_p(g)r_1 - s_p(g)r_2}{2} \hat{\sigma}_1 + \frac{c_p(g)r_2 + s_p(g)r_1}{2} \hat{\sigma}_2, \quad (6.65a)$$

$$c_p(g) = \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n)!} \left(\frac{2Ag}{\hbar} \right)^{2n} \langle p^{2n+1} \rangle_0, \quad (6.65b)$$

$$s_p(g) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{(2n-1)!} \left(\frac{2Ag}{\hbar} \right)^{2n-1} \langle p^{2n} \rangle_0. \quad (6.65c)$$

These operations differ from \mathcal{E} only in how the various moments of the initial detector distribution weight the series for the Rabi oscillation. In particular, given the substitutions $\hat{p} \rightarrow p$ and $\hat{x} \rightarrow x$, then $\langle \{p^n, x\}/2 \rangle_0 \rightarrow p^n x$ and $\langle p^{n+1} \rangle_0 \rightarrow p^{n+1}$, so $c_x(g) \rightarrow x (\cos(2gAp/\hbar) - 1)$, $s_x(g) \rightarrow x \sin(2gAp/\hbar)$, $c_p(g) \rightarrow p (\cos(2gAp/\hbar) - 1)$, and $s_p(g) \rightarrow p \sin(2gAp/\hbar)$. Therefore, if the detector remained uncorrelated with the system the averaging operations would reduce to $\mathcal{X}(\hat{\rho}_i) \rightarrow x \mathcal{E}(\hat{\rho}_i)$ and $\mathcal{P}(\hat{\rho}_i) \rightarrow p \mathcal{E}(\hat{\rho}_i)$, which are the decoupled initial detector means scaling the Rabi-oscillating qubit state.

Since we have assumed that \hat{A} does not have a component proportional to the identity, the symmetric product $\{\hat{A}, \hat{\rho}_i\}/2 = Ar_3 \hat{1}_s/2 = \langle A \rangle_0 (\hat{1}_s/2)$ for a qubit will act effectively as an inner product that extracts the part of the initial state proportional to \hat{A} . Therefore, the correction to $\langle A \rangle_0$ in the full position response ${}_f \langle x \rangle_T$ that appears in (6.33) has the simple form,

$$g \frac{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\{\hat{A}, \hat{\rho}_i\}/2) \right)}{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)} = \frac{g \langle A \rangle_0}{\tilde{p}(f)}. \quad (6.66)$$

where the conditioning factor,

$$\begin{aligned} \tilde{p}(f) &= 2 \text{Tr}_S (\hat{\rho}_f \mathcal{E}(\hat{\rho}_i)) \\ &= 2 \text{Tr}_S (\hat{\rho}_f \hat{\rho}_i) + (c(g)r_1 - s(g)r_2) \text{Tr}_S (\hat{\rho}_f \sigma_1) + (c(g)r_2 + s(g)r_1) \text{Tr}_S (\hat{\rho}_f \sigma_2), \end{aligned} \quad (6.67)$$

is $(2/\text{Tr}_S(\hat{P}_f))$ times the total probability of obtaining the post-selection. We have expressed $\tilde{p}(f)$ more compactly in terms of the retrodictive state (6.51) to show how the deviations from the initial state that are induced by \hat{A} become effectively averaged by the post-selection state. In the absence of post-selection, the retrodictive state will be maximally mixed $\hat{\rho}_f = \hat{1}_s/2$ and $\tilde{p}(f) \rightarrow 1$, recovering the unconditioned average $\langle A \rangle_0$.

The correction to the detector mean position $\langle x \rangle_0$ in (6.33) can be expressed in a similar way,

$$\begin{aligned} \frac{\text{Tr}_S(\hat{P}_f \mathcal{X}(\hat{\rho}_i))}{\text{Tr}_S(\hat{P}_f \mathcal{E}(\hat{\rho}_i))} &= \frac{1}{\tilde{p}(f)} \left(2 \langle x \rangle_0 \text{Tr}_S(\hat{\rho}_f \hat{\rho}_i) \right. \\ &\quad \left. + (c_x(g)r_1 - s_x(g)r_2) \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_1) \right. \\ &\quad \left. + (c_x(g)r_2 + s_x(g)r_1) \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_2) \right), \end{aligned} \quad (6.68)$$

as can the correction to the detector mean momentum $\langle p \rangle_0$ in (6.33),

$$\begin{aligned} \frac{\text{Tr}_S(\hat{P}_f \mathcal{P}(\hat{\rho}_i))}{\text{Tr}_S(\hat{P}_f \mathcal{E}(\hat{\rho}_i))} &= \frac{1}{\tilde{p}(f)} \left(2 \langle p \rangle_0 \text{Tr}_S(\hat{\rho}_f \hat{\rho}_i) \right. \\ &\quad \left. + (c_p(g)r_1 - s_p(g)r_2) \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_1) \right. \\ &\quad \left. + (c_p(g)r_2 + s_p(g)r_1) \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_2) \right). \end{aligned} \quad (6.69)$$

Expanding the full solution (6.33) using (6.66), (6.68), and (6.69) to linear order in g , we find the linear response (6.46) in terms of the real and imaginary

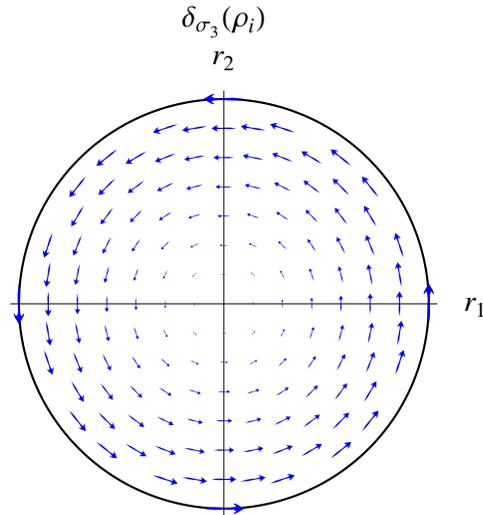


Figure 6.3: (color online) The projection onto the plane $r_3 = 0$ of the qubit Bloch sphere, showing the vector field $\delta_{\sigma_3}(\hat{\rho}_i) = -r_2\hat{\sigma}_1 + r_1\hat{\sigma}_2$ for arbitrary initial states $\hat{\rho}_i = (\hat{1} + r_1\hat{\sigma}_1 + r_2\hat{\sigma}_2 + r_3\hat{\sigma}_3)/2$. The curves of the flow through this vector field are the Rabi oscillations around the r_3 axis that are generated by the unitary action of $\hat{\sigma}_3$. The quantity $2\text{Im}A_w$ (6.70) is the logarithmic rate of change of the post-selection probability (6.49) along this vector field.

parts of the qubit weak value,

$$\text{Re}A_w = \frac{\langle A \rangle_0}{2\text{Tr}_S(\hat{\rho}_f \hat{\rho}_i)}, \quad (6.70a)$$

$$2\text{Im}A_w = \frac{\text{Tr}_S(\hat{\rho}_f \delta_A(\hat{\rho}_i))}{\text{Tr}_S(\hat{\rho}_f \hat{\rho}_i)}, \quad (6.70b)$$

$$\delta_A(\hat{\rho}_i) = A(-r_2\hat{\sigma}_1 + r_1\hat{\sigma}_2). \quad (6.70c)$$

As expected, the real part contains information regarding the measurement of \hat{A} as an observable in the initial state, conditioned by the disturbance-free overlap between the predictive and retrodictive states. The imaginary part, on the other hand, contains $\delta_A(\hat{\rho}_i)$, which is a tangent vector field on the Bloch sphere—illustrated in Figure 6.3—that corresponds to an infinitesimal portion of the Rabi oscillation being generated by \hat{A} . This tangent vector field contains only the

components r_1 and r_2 from bases *orthogonal* to \hat{A} in the initial state $\hat{\rho}_i$, so $2\text{Im}A_w$ contains only the retrodictive averages of corrections to bases *orthogonal* to \hat{A} , and thus contains no information about the measurement of \hat{A} as an observable. As discussed in (6.49), $2\text{Im}A_w$ is the logarithmic rate of change of the post-selection probability along the vector field $\delta_A(\hat{\rho}_i)$. Scaling it by a small factor with units inverse to A will produce a probability correction to linear order. In the absence of post-selection, then $\hat{\rho}_f \rightarrow \hat{1}_s/2$, $\text{Re}A_w \rightarrow \langle A \rangle_0$, and $\text{Im}A_w \rightarrow 0$.

6.5.3 Gaussian Detector

We can also apply our general results to the traditional case when the initial detector state in (6.11) is a zero-mean Gaussian in position,

$$\langle x|\psi\rangle = \frac{1}{(2\pi\sigma^2)^{1/4}} e^{-x^2/4\sigma^2}, \quad (6.71)$$

Then the measurement operators for position detection (6.13) have the initial Gaussian form shifted by $g\hat{A}$,

$$\hat{M}_x = \frac{1}{(2\pi\sigma^2)^{1/4}} e^{-(x-g\hat{A})^2/4\sigma^2}, \quad (6.72)$$

while the conjugate measurement operators for momentum detection (6.23) have the initial Gaussian modified by a unitary factor containing \hat{A} ,

$$\hat{N}_p = \left(\frac{2\sigma^2}{\pi\hbar^2}\right)^{1/4} e^{-p^2\sigma^2/\hbar^2} e^{gp\hat{A}/i\hbar}. \quad (6.73)$$

The Wigner quasi-probability operator (6.26a) correspondingly decouples into a product of Gaussian distributions, with only the position shifted by the system

operator,

$$\hat{W}_{x,p} = \frac{1}{\pi\hbar} e^{-(x-g\hat{A})^2/2\sigma^2} e^{-2p^2\sigma^2/\hbar^2}. \quad (6.74)$$

Marginalizing the Wigner operator over momentum and position separately produces the probability operators (6.16) and (6.24),

$$\hat{E}_x = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-g\hat{A})^2/2\sigma^2}, \quad (6.75a)$$

$$\hat{F}_p = \frac{\sigma}{\hbar} \sqrt{\frac{2}{\pi}} e^{-2p^2\sigma^2/\hbar^2} \hat{1}_s. \quad (6.75b)$$

As anticipated, the probability operator for momentum no longer contains any information about the system operator \hat{A} and is proportional to the identity, so measuring the momentum provides zero information about any system operator not proportional to the identity.

In the presence of post-selection we can also exactly compute the perturbative expansions of the disturbance operations (6.42) using the following identities for the Gaussian detector moments,

$$\langle p^{2n} \rangle_0 = \left(\frac{\hbar}{2\sigma} \right)^{2n} (2n-1)!!, \quad (6.76a)$$

$$\langle p^{2n-1} \rangle_0 = 0, \quad (6.76b)$$

$$\langle \{p^n, x\}/2 \rangle_0 = 0, \quad (6.76c)$$

$$\frac{(2n-1)!!}{(2n)!} = \frac{1}{2^n n!}, \quad (6.76d)$$

which hold for positive integer n . We find the simple results,

$$\mathcal{E}(\hat{\rho}_i) = \exp\left(\left(\frac{g}{2\sigma}\right)^2 \frac{1}{2} \text{ad}^*[-i\hat{A}]^2\right)(\hat{\rho}_i), \quad (6.77a)$$

$$\mathcal{X}(\hat{\rho}_i) = 0, \quad (6.77b)$$

$$\mathcal{P}(\hat{\rho}_i) = \frac{g}{\hbar} \frac{\hbar^2}{4\sigma^2} \text{ad}^*[-i\hat{A}](\mathcal{E}(\hat{\rho}_i)). \quad (6.77c)$$

Note that this independent derivation via perturbative methods has reproduced the Gaussian case of our more general equations (6.41).

The quantity $\epsilon = (g/2\sigma)^2$ with units inverse to \hat{A}^2 emerges as the natural decoherence parameter, which we can see more clearly by rewriting the non-selective measurement operation in (6.77) as,

$$\hat{\rho}_i(\epsilon) = \mathcal{E}(\hat{\rho}_i) = \exp\left(\epsilon \mathcal{L}[\hat{A}]\right)(\hat{\rho}_i), \quad (6.78a)$$

$$\mathcal{L}[\hat{A}](\hat{\rho}_i) = \hat{A}\hat{\rho}_i\hat{A}^\dagger - \frac{1}{2}\{\hat{\rho}_i, \hat{A}^\dagger\hat{A}\} = \frac{1}{2}\text{ad}^*[-i\hat{A}]^2 = \frac{1}{2}\delta_A^2. \quad (6.78b)$$

The operation $\mathcal{L}[\hat{A}](\hat{\rho}_i)$ is the *Lindblad operation* [25, 27, 101] that produces *decoherence* in continuous dynamical systems, with \hat{A} playing the role of the Lindblad operator that decoheres the system³. Note that for a Lie algebraic element like $-i\hat{A}$ the Lindblad operation takes the intuitive form of the second-order directional derivative. Since $\partial_\epsilon \hat{\rho}_i(\epsilon) = \mathcal{L}[\hat{A}](\hat{\rho}_i(\epsilon))$, the Gaussian measurement acts as an effective Lindblad evolution that *decoheres* the system state with increasing ϵ via the action of \hat{A} , but does not cause unitary disturbance along the natural flow of \hat{A} .

The exact expressions for the conditioned Gaussian detector means follow from

³That von Neumann coupling can lead to such Lindblad evolution in the case of *continuous* indirect non-selective measurements with zero-mean Gaussian detectors was also noted by Breuer and Petruccione [25, §3.5.2, p.162]; however, here we make a single Gaussian detection after a duration of time, producing an effective flow parameter $(g/2\sigma)^2$.

(6.33) and (6.77),

$${}_f\langle x \rangle_T = g \frac{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\{\hat{A}, \hat{\rho}_i\}) \right)}{2 \text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)}, \quad (6.79a)$$

$${}_f\langle p \rangle_T = \frac{g}{\hbar} \frac{\hbar^2}{4\sigma^2} \frac{\text{Tr}_S \left(\hat{P}_f \text{ad}^*[-i\hat{A}](\mathcal{E}(\hat{\rho}_i)) \right)}{\text{Tr}_S \left(\hat{P}_f \mathcal{E}(\hat{\rho}_i) \right)}. \quad (6.79b)$$

Surprisingly, the special properties of the Gaussian moments (6.76) allow (6.79) to be written in a form proportional to the real and imaginary part of a complex weak-value parameter involving the *decohered* system state (6.78) to all orders in the coupling strength g ,

$$A^w(\epsilon) = \frac{\text{Tr}_S \left(\hat{P}_f \hat{A} \hat{\rho}_i(\epsilon) \right)}{\text{Tr}_S \left(\hat{P}_f \hat{\rho}_i(\epsilon) \right)}, \quad (6.80a)$$

$${}_f\langle x \rangle_T = g \text{Re} A^w(\epsilon), \quad (6.80b)$$

$${}_f\langle p \rangle_T = \frac{g}{\hbar} \frac{\hbar^2}{4\sigma^2} (2 \text{Im} A^w(\epsilon)). \quad (6.80c)$$

Following the interpretations outlined in this paper we can therefore understand the position shift $\text{Re} A^w(\epsilon)$ to all orders in g as the *average of the observable \hat{A} in the decohered initial system state $\hat{\rho}_i(\epsilon)$ conditioned on the post-selection \hat{P}_f* . Similarly, we can understand the factor $2 \text{Im} A^w(\epsilon)$ in the momentum shift to all orders in g as the *logarithmic directional derivative of the probability of post-selecting \hat{P}_f given the decohered initial system state $\hat{\rho}_i(\epsilon)$ along the unitary flow generated by \hat{A}* .

If the measured operator is the qubit operator $\hat{A} = A\hat{\sigma}_3$ as in (6.60), then we

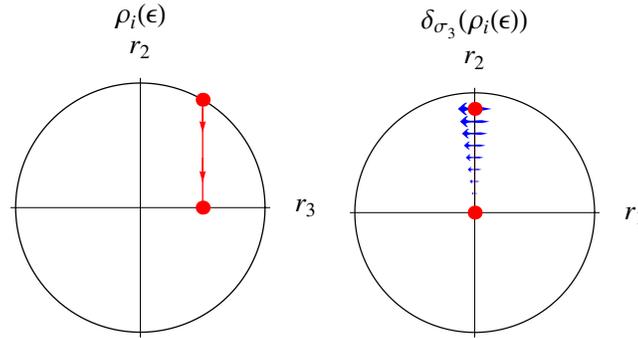


Figure 6.4: (color online) Two projections of the Bloch sphere showing the pure decoherence of the specific state $\hat{\rho}_i(\epsilon) = \exp(\epsilon\mathcal{L}[\hat{\sigma}_3])(\hat{\rho}_i) = (\hat{1} + \exp(-2\epsilon)\sqrt{3}\hat{\sigma}_2/2 + \hat{\sigma}_3/2)/2$ due to the Gaussian detector (6.81). (left) The projection onto the plane $r_1 = 0$ showing the progressive collapse of $\hat{\rho}_i(\epsilon)$ onto the r_3 axis with increasing ϵ . (right) The projection onto the plane $r_3 = 0$ showing the vector field $\delta_{\sigma_3}(\hat{\rho}_i(\epsilon))$ during the progressive collapse. Notably the quantity $2\text{Im}A^w(\epsilon)$ (6.80) is the rate of change of the post-selection probability (6.49) along this vector field for all ϵ , but not along the purely decohering trajectory that $\hat{\rho}_i(\epsilon)$ actually follows.

can further simplify the expression (6.77) using the identities (6.62) to find,

$$\begin{aligned} \mathcal{E}(\hat{\rho}_i) &= \hat{\rho}_i + \frac{1}{2}(e^{-(Ag/\sigma)^2/2} - 1)(r_1\hat{\sigma}_1 + r_2\hat{\sigma}_2), \\ &= \frac{1}{2}\left(\hat{1} + r_3\hat{\sigma}_3 + e^{-(Ag/\sigma)^2/2}(r_1\hat{\sigma}_1 + r_2\hat{\sigma}_2)\right), \end{aligned} \quad (6.81)$$

which shows how the measurement decoheres the bases orthogonal to \hat{A} in the initial state with an increase in the dimensionless flow parameter $(Ag/\sigma)^2$ ⁴. This decoherence is illustrated in Figure 6.4. The conditioned means (6.79) of a Gaus-

⁴The continuous measurement of a double quantum dot that is discussed in [68, 103] can be mapped onto this problem so that $(Ag/\sigma)^2 \rightarrow \gamma t$ where $\gamma = 1/T_m$ is an inverse characteristic measurement time that acts as a dephasing rate due to the continuous non-selective measurement. Hence, the results obtained therein are special cases of the exact solution (6.82).

sian qubit detector consequently have the exact form,

$${}_f\langle x \rangle_T = g \frac{\langle A \rangle_0}{\tilde{p}(f)}, \quad (6.82a)$$

$${}_f\langle p \rangle_T = \frac{g}{\hbar} \frac{\hbar^2}{4\sigma^2} \frac{2A}{\tilde{p}(f)} e^{-(Ag/\sigma)^2/2} (r_1 \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_2) - r_2 \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_1)),$$

$$\tilde{p}(f) = 1 + r_3 \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_3) + e^{-(Ag/\sigma)^2/2} (r_1 \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_1) + r_2 \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_2)),$$

to all orders in the coupling strength g . When expanded to linear order in g , (6.82) reduces to (6.46) with the real and imaginary parts of the qubit weak value (6.70), as expected.

For contrast, as g becomes large the unconditioned measurement of \hat{A} becomes essentially projective and the operation \mathcal{E} almost completely decoheres the initial state (6.81) into the basis of \hat{A} as the pointer basis,

$$\mathcal{E}(\hat{\rho}_i) \approx \frac{1}{2} (\hat{1}_s + r_3 \hat{\sigma}_3). \quad (6.83)$$

Hence, in this *strong measurement* regime, the conditioned means (6.82) approximate,

$${}_f\langle x \rangle_T \approx g \frac{\langle A \rangle_0}{1 + r_3 \text{Tr}_S(\hat{\rho}_f \hat{\sigma}_3)}, \quad (6.84a)$$

$${}_f\langle p \rangle_T \approx 0. \quad (6.84b)$$

The position shift contains the average of \hat{A} in the *decohered* initial system state $\mathcal{E}(\hat{\rho}_i)$, conditioned by the post-selection. Moreover, since the decohered initial system state $\mathcal{E}(\hat{\rho}_i)$ is essentially diagonal in the basis of \hat{A} , it will no longer Rabi oscillate, so the directional derivative along the flow generated by \hat{A} will be essentially zero. Hence, the probability correction factor represented by $2\text{Im}A^w(\epsilon)$ vanishes.

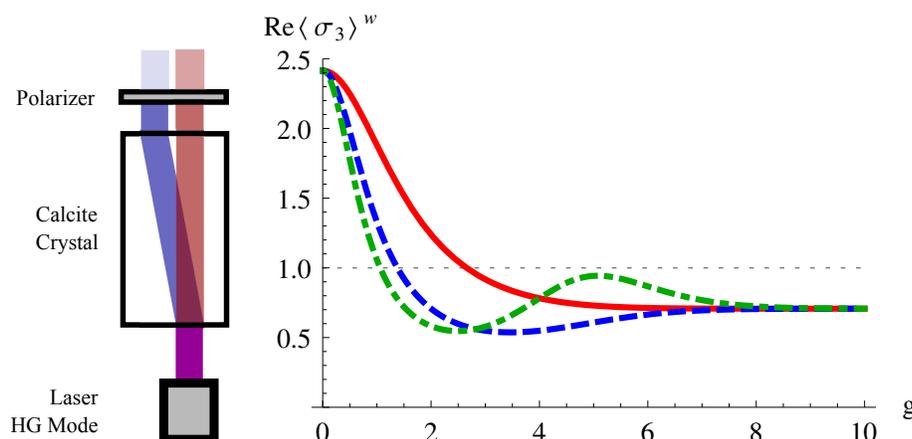


Figure 6.5: (color online) (left) A possible implementation of a conditioned polarization measurement similar to [44], where the length of a birefringent crystal determines the coupling strength g . (right) The weak value $\text{Re}\langle\sigma_3\rangle^w$ corresponding to the Hermite-Gauss detector profiles in Fig. 6.6 and reduced states in (6.89) with $m = 0$ (solid, red), $m = 1$ (dashed, blue), and $m = 2$ (dot-dashed, green), obtained by averaging according to Eq. (6.87). The weak limit $g \rightarrow 0$ is identical for all detectors, as is the strong limit $g \rightarrow \infty$ of a classical conditioned average, but the specifics of the transition depend on how the detector decoheres the state. The dotted horizontal line is the eigenvalue bound of 1.

6.5.4 Hermite-Gauss Modes

To show how our general Wigner function expressions in Eqs.(6.41) can be applied beyond the Gaussian case, we now consider the more general Hermite-Gauss modes $\{|h_m\rangle\}$, which are a widely used complete set of transverse modes naturally generated in laser cavities that can describe an initial zero-mean and collimated detecting beam. The Gaussian state in the last section is a special case for $m = 0$. The Wigner distribution for a Hermite-Gauss mode of order $m \in \{0, 1, 2, \dots\}$ has the form [210],

$$W_m^{\text{HG}}(x, p) = \frac{(-1)^m}{\pi\hbar} L_m[2G(x, p)] e^{-G(x, p)}, \quad (6.85a)$$

$$G(x, p) = \frac{x^2}{2\sigma^2} + \frac{2\sigma^2 p^2}{\hbar^2}, \quad (6.85b)$$

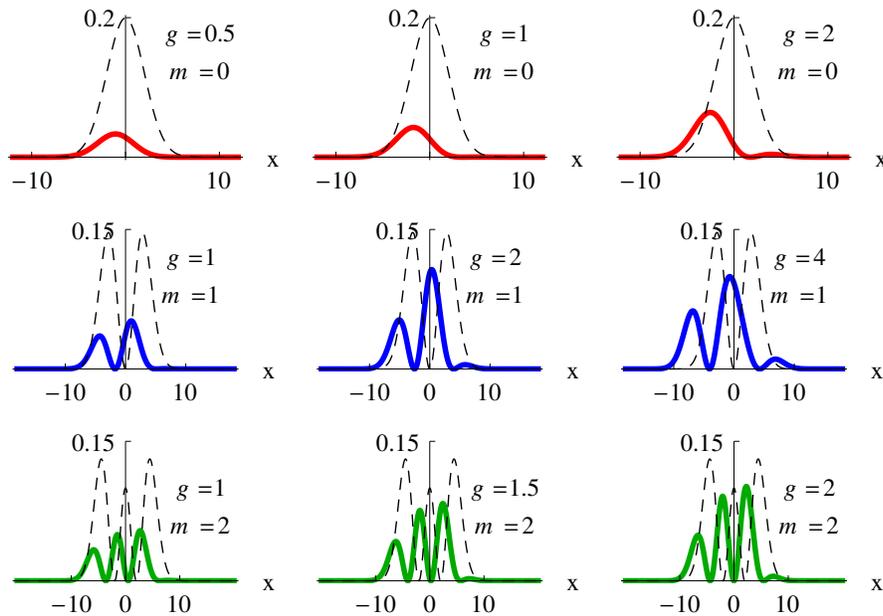


Figure 6.6: (color online) Post-selected detector intensities for the initial polarization state $|\psi_i\rangle = (\cos(7\pi/8), \sin(7\pi/8))$ and final post-selection $|\psi_f\rangle = (1, 1)/\sqrt{2}$, using the first three Hermite-Gauss detector modes with $\sigma = 2$. Averaging these profiles produces weak values according to Eq. (6.87) and shown in Fig. 6.5. The dashed line indicates the initial detector intensity.

where L_m is a Laguerre polynomial of order m . The first few such polynomials are shown in Table 6.1 for reference.

After Fourier-transforming Eq. (6.85) and integrating according to Eq. (6.41a), we obtain a compact expression for the exact post-interaction reduced system state for any coupling strength and initial detector mode m ,

$$\hat{\rho}_i^{(m)}(\epsilon) = \mathcal{E}_m(\hat{\rho}_i) = L_m \left[-2\epsilon \mathcal{L}[\hat{A}] \right] \exp \left[\epsilon \mathcal{L}[\hat{A}] \right] (\hat{\rho}_i), \quad (6.86)$$

which generalizes (6.78) to any mode number m . As with (6.78), a measurement strength parameter $\epsilon = (g/2\sigma)^2$ naturally appears for all modes along with the Lindblad operation $\mathcal{L}[\hat{A}] = \text{ad}^*[-i\hat{A}]^2/2$ that decoheres bases orthogonal to the eigenbasis of \hat{A} [25, 73]. Furthermore, the functional form of (6.86) is the same

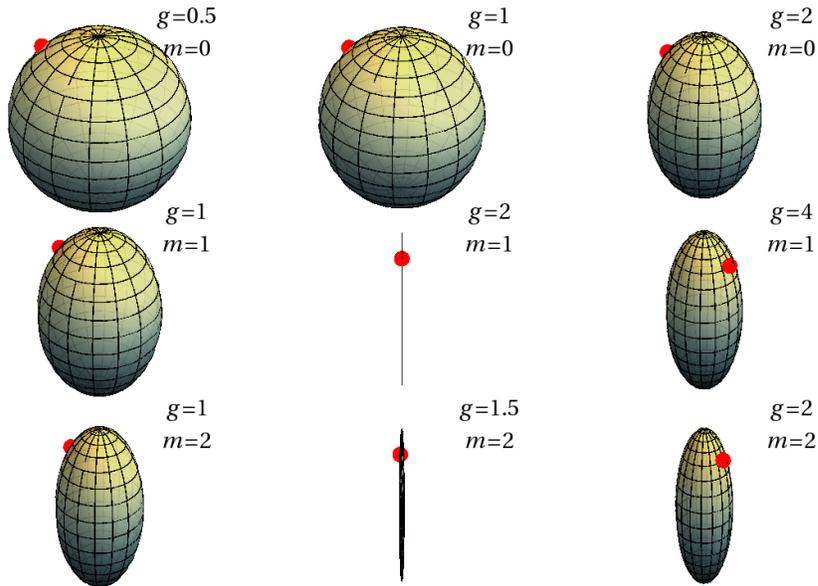


Figure 6.7: (color online) Reduced polarization states (6.89) corresponding to the detector responses in Fig. 6.6. Bloch sphere distortions are shown with the σ_3 axis aligned vertically; the red dot tracks the initial state chosen in Fig. 6.6. For $m > 0$ any initial state will experience decoherence oscillations and pass directly through the σ_3 axis before partially recohering.

as the Wigner distribution (6.85) up to normalization, but with the function $G(x, p)$ replaced by the Lindblad operation $-\epsilon\mathcal{L}[\hat{A}]$. Superpositions of modes are considered in Appendix D.

Using Eqs. (6.86), (6.41), and (6.33) we also obtain the following compact results for the exact detector averages for any initial Hermite-Gauss detector mode of order m ,

$${}_f\langle x \rangle = g \operatorname{Re} A^w(\epsilon), \quad (6.87a)$$

$${}_f\langle p \rangle = g \frac{\hbar}{(2\sigma)^2} 2 \operatorname{Im}(A^w(\epsilon) + \Delta_m(\epsilon)). \quad (6.87b)$$

Perhaps surprisingly, they are still parametrized for all orders in ϵ by a complex weak value parameter (6.80) with pre-selection equal to the reduced post-

m	$L_m(x)$	$-2 L'_m(x)$
0	1	0
1	$1 - x$	2
2	$1 - 2x + x^2/2$	$4 - 2x$
3	$1 - 3x + 3x^2/2 - x^3/6$	$6 - 6x + x^2$

Table 6.1: Laguerre polynomials $L_m(x)$ and their derivatives for the first few m . These polynomials appear naturally for Hermite-Gauss modes in their Wigner distribution (6.85), as well as the resulting system operations (6.86) and (6.88).

interaction system state $\hat{\rho}_i^{(m)}(\epsilon)$ given in Eq. (6.86), and one additional weak-value-like correction term for the higher mode numbers $m \geq 1$,

$$\Delta_m(\epsilon) = \frac{\text{Tr}_S \left(\hat{P}_f \hat{A} \mathcal{M}_m(\hat{\rho}_i) \right)}{\text{Tr}_S \left(\hat{P}_f \hat{\rho}_i^{(m)}(\epsilon) \right)}, \quad (6.88a)$$

$$\mathcal{M}_m(\hat{\rho}_i) = -2 L'_m \left[-2 \epsilon \mathcal{L}[\hat{A}] \right] \exp \left[\epsilon \mathcal{L}[\hat{A}] \right] (\hat{\rho}_i). \quad (6.88b)$$

The first few polynomials $-2 L'_m(x)$ in \mathcal{M}_m that contain the derivatives of Laguerre polynomials are shown in Table 6.1 for reference.

The appearance of a correction to $\text{Im}A^w$ in Eq. (6.87b) further strengthens the observation in that $\text{Im}A^w$ pertains solely to the rate of change of the post-selection probability and not to the measurement of \hat{A} itself. Indeed, for $m = 0$ Eqs. (6.86) and (6.87) correctly reproduce the exact Gaussian detector case that we derived using the perturbative method in the last section.

We stress that these are general results for any system observable \hat{A} . Figs. 6.5, 6.6, and 6.7 show the special case of an optical application, where $\hat{A} = \hat{\sigma}_3$ is a polarization observable being measured by a Hermite-Gaussian beam. Given an initial state $\hat{\rho}_i = (\hat{1} + \sum_k r_k \hat{\sigma}_k)/2$ with Pauli operators $\hat{\sigma}_k$ as in (6.60) and a measurement of $\hat{A} = \hat{\sigma}_3$ being made with Hermite-Gauss mode m , the post-interaction state from Eq. (6.86) can be computed with the qubit identities (6.61)

to produce,

$$\hat{\rho}_i^{(m)}(g) = \frac{1}{2} \left[\hat{1} + r_3 \hat{\sigma}_3 + L_m \left[(g/\sigma)^2 \right] e^{-(g/\sigma)^2/2} (r_1 \hat{\sigma}_1 + r_2 \hat{\sigma}_2) \right]. \quad (6.89)$$

Fig. 6.5 shows a possible implementation of this example that is analogous to the experiment performed in [44], as well as how the generalized weak value (6.36) continuously changes into a classical conditioned average as the initial state decoheres into the reduced state (6.89). Fig. 6.6 shows the post-interaction spatial intensity profiles for the detector, while Fig. 6.7 shows the corresponding reduced polarization states (6.89). Notably, these results for the qubit show that the decoherence generally undergoes oscillations due to the lobed structure of the detector profile; it need not be a simple exponential decay.

7 Concluding remarks

I know of nothing more terrible than the poor creatures who have learned too much. Instead of the sound powerful judgement which would probably have grown up if they had learned nothing, their thoughts creep timidly and hypnotically after words, principles and formulae, constantly by the same paths. What they have acquired is a spider's web of thoughts too weak to furnish sure supports, but complicated enough to provide confusion.

Ernst Mach, (1886)

As we were trained classically, we physicists speak of physical objects like ‘point particles’ and ‘fields’ that possess tangible *observable* properties like ‘energy,’ ‘momentum,’ ‘charge,’ and ‘spin.’ We describe these objects as statically existing extended curves or surfaces contained in a self-consistently warped yet unchanging geometric background of ‘spacetime.’ We speak of experience in this description as a meta-rule: as proper time elapses, a point or hyperplane traverses a pre-determined curve in the static geometry at a fixed rate, steadily uncovering a pre-determined future. Such a description of changeless form embodies the deprecated philosophy of realistic determinism; it is neat and tidy, and ultimately an averaged approximation.

Irreducible stochasticity, such as that present in quantum measurements, undermines the entire concept of a physical spacetime. A stochastic event transforms the uncertain potentiality of the future to certainty in the past, which necessitates

irreversible change and distinguishes the present as the center of that change. Put another way, a stochastic event implies an active *stochastic process*. A physical spacetime is a static structure with future, past, and present on equal footing, so it does not admit this sort of change. Since stochastic events cannot be determined *a priori*, a physical spacetime can be constructed only as a retrospective history of events that have already occurred. It will become real only after it is measured, in a manner of speaking. However, the reality will be that of the correlations between past events recorded in a history book¹.

Mathematical objects defined in terms of spacetime—e.g., particle and field states, and their observable properties—must suffer a similar disillusionment of their physical status in any stochastic theory. Those objects may give us insight into correlations between the physical events of our experience, but they cannot be the reality to which those events correspond. Any advocacy of the reality of spacetime and its associated objects would require the denial of intrinsic stochasticity, and thus inexorably lead to quantum interpretations that contain no stochastic measurement events (such as the many-worlds interpretation). However, forcing a deterministic philosophy simply to allow the reification of mathematical abstractions like spacetime and its associated objects is at best a “bad habit” [211]. At worst, this bad habit creates a delusion that obstructs our ability to reason honestly about our models of the physical world.

If we wish to avoid delusions of this sort, then we must change the language we use and embrace the inherent stochasticity observed in the laboratory. Spacetime cannot be understood as fundamentally physical in a stochastic theory, so we must describe it in a manner that better reflects its significance: it is a set of *symmetries* that collections of recorded stochastic events are observed to obey. It specifies how relatively moving, rotated, and translated observers will record the same set of

¹It is worth noting that this tension between stochasticity and the intrinsic physicality of spacetime underpins both the quantum measurement problem and the deep incompatibility between general relativity and the quantum theory.

events, and thus specifies how one must transform different event records so that they agree on the same content. These symmetries are embodied by a continuous *Lie group* known as the Poincaré group, which constitutes a powerful constraint upon the form of the possible events that can occur. The events themselves are physical; spacetime indicates the symmetries that those events satisfy and thus provides a structural constraint that potential events must obey. We outlined how such a Lie group will lead to a natural algebraic structure in Appendix A. The possible events themselves take the algebraic form of spectral idempotents. Working algebraically thus provides a direct connection between the symmetry constraints and the physical events.

As a corollary, states in the form of spacetime fields cannot be understood as fundamentally physical: they are *probability functionals* over a set of potential events constrained by group symmetries that produce collections of likelihoods for events to occur, and thus govern how realized events will be correlated. Hence, states must correspond solely to undetermined events—events that are known to have occurred are not described directly by such a state, but instead provide *boundary conditions* that serve to further constrain the state. States are thus manifestly *counter-factual* fictions, existing only as abstract constructions for deriving inferences that satisfy the constraints imposed by symmetries (such as spacetime) and other measured boundary conditions. Their associated probabilistic predictions depend upon the assumed symmetries, the assumed prior information about past events, and the presumed set of future events that may occur. This probability calculus provides a form of *logic* that is suitable for manipulating uncertainty, as we explored in Chapters 2 and 3.

The measurable observables have a peculiar dual role in this stochastic structure. On one hand they are defined abstractly as group generators in a Lie algebra that dictate the possible physical symmetries, as discussed in Appendix A. On the other hand, they can be constructed from spectral idempotents or other proba-

bility observables that correspond directly to measurable events, which allows them to be assigned measurable values associated with a counter-factual state or an experimentally realized ensemble that spans an equivalence class of measured spacetime points. We detailed this correspondence by introducing generalized spectra known as *contextual values* in Chapters 2 and 3. These contextual values are assigned by an experimenter to each event recorded with an imperfectly correlated detector in order to construct the average value of an observable that can be assigned to an ensemble of events. We then showed how to condition these averages in a general way by partitioning the possible ensembles of recorded detector events into classes labeled by subsequently recorded events.

The deterministic notion of a classically intrinsic objective property manifests only in a restricted sense when examining large collections of simultaneously recorded measurement events that begin to approximate a mean frequency field over a realized spacetime; associating such an emergent mean field with the assigned average values of the group generators produces a classical spacetime field. However, the intrinsic stochasticity present in a *sequence* of measurement events—such as those considered with conditioned averages—can expose the effective disturbance that breaks the identification of deterministic properties compatible with the measured sequence. We showed this behavior explicitly by deriving and experimentally violating generalized Leggett-Garg inequalities using an optical setup in Chapter 4. Notably, the violations could be understood as a form of classically invasive disturbance occurring between each recorded event.

We also showed nonclassical behavior explicitly in Chapter 5 by considering the which-path information of an electronic Mach-Zehnder interferometer as measured by an equivalent coupled interferometer. Examination of measurement event sequences in the form of conditioned observable averages shows similar violation of classical spectral bounds, exposing the effective disturbance between the measurements due to intrinsic stochasticity. We also showed that when the recorded events

are unbiased with respect to the which-path observable then quantum erasure can be attained by partitioning the detector events into complementary sets, restoring in a conditioned sense the original counter-factual state on each sub-ensemble.

In Chapters 3, 4, and 5, we saw that the real part of a peculiar complex quantity known as the *quantum weak value* kept appearing as a limiting value for a weakly measured conditioned observable average. To further understand the complex weak value, we carefully explored how it has been traditionally defined in the literature by exactly solving the von Neumann measurement protocol algebraically in Chapter 6. We found that while the real part of the weak value may indeed be interpreted as a conditioned average of an observable, the imaginary part does not pertain to the measurement of an observable. Instead, the imaginary part describes the role of the observable as the generator of a Lie group, as expected from the discussion in Appendix A, and thus indicates the symmetry group foundation of the quantum theory.

The significance of quantum observables has become more clear through the investigations presented in this work. While they are defined as dynamical entities that generate continuous group transformations, observables can acquire an emergent description as physical properties given a sufficiently large ensemble of recorded stochastic events. Provided that a set of realized events are tightly correlated to a locally emergent spacetime region, one can approximately associate observable properties to that region as *classical mean fields*; however, distributed ensembles of events cannot produce such a local mean field description *a priori*. These distributed ensembles, however, can be used to *inferentially* assign mean observable values to a local spacetime region using the contextual values technique. Hence, an extended notion of a classical mean field can be constructed *a posteriori* over an emergent spacetime of measured past events. Such a technique provides an intriguing window into the structure of the irreducible stochasticity present in the physical world.

Bibliography

- [1] I. Prigogine, *The end of certainty: time, chaos, and the new laws of nature* (The Free Press, New York, 1997).
- [2] E. T. Jaynes, *Probability Theory: The Logic of Science* (Cambridge University Press, Cambridge, 2003).
- [3] C. M. Caves, C. A. Fuchs, and R. Schack, *Phys. Rev. A* **65**, 22305 (2002).
- [4] C. A. Fuchs, “QBism, the Perimeter of Quantum Bayesianism,” (2010), arXiv:1003.5209 .
- [5] M. S. Leifer and R. W. Spekkens, “Formulating Quantum Theory as a Causally Neutral Theory of Bayesian Inference,” (2011), arXiv:1107.5849 .
- [6] H. Everett, *Rev. Mod. Phys.* **29**, 454 (1957).
- [7] B. S. DeWitt, *Physics Today* **23**, 30 (1970).
- [8] D. Deutsch, *Proc. Royal Soc. Lond.* **A455**, 3129 (1999).
- [9] W. H. Zurek, *Phys. Rev. A* **71**, 052105 (2005).
- [10] P. A. M. Dirac, *Principles of Quantum Mechanics* (Oxford University Press, Oxford, 1930).
- [11] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer, 1932).
- [12] A. N. Korotkov, *Phys. Rev. B* **60**, 5737 (1999).
- [13] A. N. Korotkov and A. N. Jordan, *Phys. Rev. Lett.* **97**, 166805 (2006).
- [14] N. Katz, M. Neeley, M. Ansmann, R. Bialczak, M. Hofheinz, E. Lucero, A. OConnell, H. Wang, A. Cleland, J. Martinis, and A. Korotkov, *Phys. Rev. Lett.* **101**, 200401 (2008).

- [15] Y.-S. Kim, Y.-W. Cho, Y.-S. Ra, and Y.-H. Kim, *Opt. Exp.* **17**, 11978 (2009).
- [16] A. N. Jordan and A. N. Korotkov, *Phys. Rev. B* **74**, 085307 (2006).
- [17] E. B. Davies and J. T. Lewis, *Comm. Math. Phys.* **17**, 239 (1970).
- [18] K. Kraus, *Annals of Physics* **64**, 311 (1971).
- [19] K. Kraus, *States, effects and operations: fundamental notions of quantum theory* (Springer-Verlag, Berlin, 1983).
- [20] V. Braginski and F. Khalili, *Quantum Measurement* (Cambridge University Press, Cambridge, 1992).
- [21] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics* (Berlin: Springer-Verlag, 1995).
- [22] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information* (Cambridge University Press, Cambridge, 2000).
- [23] R. Alicki and M. Fannes, *Quantum dynamical systems* (Oxford University Press, Oxford, 2001).
- [24] M. Keyl, *Physics Reports* **369**, 431 (2002).
- [25] H. Breuer and F. Petruccione, *The Theory of Open Quantum Systems* (Oxford University Press, Oxford, 2007).
- [26] S. M. Barnett, *Quantum Information* (Oxford University Press, Oxford, 2009).
- [27] H. M. Wiseman and G. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2009).
- [28] Y. Aharonov, D. Z. Albert, and L. Vaidman, *Phys. Rev. Lett.* **60**, 1351 (1988).
- [29] I. M. Duck, P. M. Stevenson, and E. C. G. Sudarshan, *Phys. Rev. D* **40**, 2112 (1989).
- [30] Y. Aharonov and L. Vaidman, *Phys. Rev. A* **41**, 11 (1990).
- [31] Y. Aharonov and A. Botero, *Phys. Rev. A* **72**, 052111 (2005).
- [32] Y. Aharonov and L. Vaidman, *Lect. Notes Phys.* **734**, 399 (2008).
- [33] R. Jozsa, *Phys. Rev. A* **76**, 044103 (2007).

- [34] A. Di Lorenzo and J. C. Egues, *Phys. Rev. A* **77**, 042108 (2008).
- [35] Y. Aharonov, S. Popescu, J. Tollaksen, and L. Vaidman, *Phys. Rev. A* **79**, 052110 (2009).
- [36] Y. Aharonov, S. Popescu, and J. Tollaksen, *Physics Today* **63**, 27 (2010).
- [37] T. Geszti, *Phys. Rev. A* **81**, 044102 (2010).
- [38] A. Hosoya and Y. Shikano, *J. Phys. A* **43**, 385307 (2010).
- [39] Y. Shikano and A. Hosoya, *J. Phys. A* **43**, 025304 (2010).
- [40] Y. Shikano and A. Hosoya, *Physica E* **43**, 776 (2011).
- [41] S. Kagami, Y. Shikano, and K. Asahi, *Physica E* **43**, 761 (2011).
- [42] S. Wu and Y. Li, *Phys. Rev. A* **83**, 052106 (2011).
- [43] E. Haapasalo, P. Lahti, and J. Schultz, *Phys. Rev. A* **84**, 052107 (2011).
- [44] N. W. M. Ritchie, J. G. Story, and R. G. Hulet, *Phys. Rev. Lett.* **66**, 1107 (1991).
- [45] H. M. Wiseman, *Phys. Rev. A* **65**, 032111 (2002).
- [46] H. M. Wiseman, *Phys. Lett. A* **311**, 285 (2003).
- [47] N. Brunner, V. Scarani, M. Wegmuller, M. Legre, and N. Gisin, *Phys. Rev. Lett.* **93**, 203902 (2004).
- [48] G. J. Pryde, J. L. O'Brien, A. G. White, T. C. Ralph, and H. M. Wiseman, *Phys. Rev. Lett.* **94**, 220405 (2005).
- [49] R. Mir, J. S. Lundeen, M. W. Mitchell, A. M. Steinberg, J. L. Garretson, and H. M. Wiseman, *New J. Phys.* **9**, 287 (2007).
- [50] O. Hosten and P. Kwiat, *Science* **319**, 787 (2008).
- [51] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell, *Phys. Rev. Lett.* **102**, 173601 (2009).
- [52] J. S. Lundeen and A. M. Steinberg, *Phys. Rev. Lett.* **102**, 020404 (2009).
- [53] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, *Phys. Rev. A* **80**, 041803 (2009).
- [54] J. C. Howell, D. J. Starling, P. B. Dixon, P. K. Vudyaletu, and A. N. Jordan, *Phys. Rev. A* **81**, 033813 (2010).

- [55] D. J. Starling, P. B. Dixon, A. N. Jordan, and J. C. Howell, *Phys. Rev. A* **82**, 063822 (2010).
- [56] D. J. Starling, P. B. Dixon, N. S. Williams, A. N. Jordan, and J. C. Howell, *Phys. Rev. A* **82**, 011802(R) (2010).
- [57] M. E. Goggin, M. P. Almeida, M. Barbieri, B. P. Lanyon, J. L. O'Brien, A. G. White, and G. J. Pryde, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1256 (2011).
- [58] S. Kocsis, B. Braverman, S. Ravets, M. J. Stevens, R. P. Mirin, L. K. Shalm, and A. M. Steinberg, *Science* **332**, 1170 (2011).
- [59] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber, *Nature* **474**, 188 (2011).
- [60] J. S. Lundeen and C. Bamber, *Phys. Rev. Lett.* **108**, 070402 (2012).
- [61] J. Bell, *Rev. Mod. Phys.* **38**, 447 (1966).
- [62] S. Kochen and E. P. Specker, *J. Math. Mech.* **17**, 59 (1967).
- [63] N. D. Mermin, *Rev. Mod. Phys.* **65**, 803 (1993).
- [64] R. W. Spekkens, *Phys. Rev. A* **71**, 052108 (2005).
- [65] M. S. Leifer and R. W. Spekkens, *Phys. Rev. Lett.* **95**, 200405 (2005).
- [66] J. Tollaksen, *J. Phys. A* **40**, 9033 (2007).
- [67] R. W. Spekkens, *Phys. Rev. Lett.* **101**, 020401 (2008).
- [68] J. Dressel, S. Agarwal, and A. N. Jordan, *Phys. Rev. Lett.* **104**, 240401 (2010).
- [69] J. Dressel, C. J. Broadbent, J. C. Howell, and A. N. Jordan, *Phys. Rev. Lett.* **106**, 040402 (2011).
- [70] J. Dressel and A. N. Jordan, *J. Phys. A: Math. Theor.* **45**, 015304 (2012).
- [71] J. Dressel and A. N. Jordan, *Phys. Rev. A* **85**, 022123 (2012).
- [72] J. Dressel, Y. Choi, and A. N. Jordan, *Phys. Rev. B* **85**, 045320 (2012).
- [73] J. Dressel and A. N. Jordan, *Phys. Rev. A* **85**, 012107 (2012).
- [74] J. Dressel and A. N. Jordan, *Phys. Rev. Lett.* **109**, 230402 (2012).
- [75] R. T. Cox, *Am. J. Phys.* **14**, 1 (1946).

- [76] R. T. Cox, *The Algebra of Probable Inference* (John Hopkins University Press, Baltimore, 1961).
- [77] W. Rudin, *Real and Complex Analysis* (McGraw Hill, 1987).
- [78] P. Szekeres, *A Course in Modern Mathematical Physics: Groups, Hilbert Spaces, and Differential Geometry* (Cambridge University Press, Cambridge, 2004).
- [79] A. Robinson, *Non-standard Analysis* (Princeton University Press, 1966).
- [80] E. Nelson, *Radically Elementary Probability Theory* (Princeton University Press, 1987).
- [81] S. Sakai, Bull. Amer. Math. Soc. **71**, 149 (1965).
- [82] H. Kosaki, Pub. Res. Inst. Math. Sci. **18**, 379 (1982).
- [83] M. Raginsky, J. Math. Phys. **44**, 5003 (2003).
- [84] R. C. Jeffrey, *The logic of decision* (The University of Chicago Press, 1965).
- [85] W. Heisenberg, *Physics and Philosophy: The Revolution in Modern Science* (Harper and Row, New York, 1958).
- [86] R. B. Griffiths, "A Consistent Quantum Ontology," (2011), arXiv:1105.3932 .
- [87] D. Hestenes and G. Sobczyk, *Clifford Algebra to Geometric Calculus: A Unified Language for Mathematics and Physics* (D. Riedel Publishing, Dordrecht, 1987).
- [88] D. Hestenes, *New Foundations for Classical Mechanics* (Kluwer Academic Publishers, Dordrecht, 1999).
- [89] A. Crumeyrolle, *Orthogonal and Symplectic Clifford Algebras* (Kluwer Academic Publishers, Dordrecht, 1990).
- [90] C. Doran and A. Lasenby, *Geometric Algebra for Physicists* (Cambridge University Press, Cambridge, 2007).
- [91] B. J. Hiley and R. Callaghan, Found. Phys. **42**, 192 (2012).
- [92] D. Hestenes, *Space-Time Algebra* (Routledge, 1966).
- [93] M. Takesaki, *Theory of Operator Algebras I, II, and III* (Springer, 1979).
- [94] M. Born, Zeitschrift für Physik **37**, 863 (1926).

- [95] A. Gleason, *Indiana Univ. Math. J.* **6**, 885 (1957).
- [96] G. Lüders, *Ann. Phys.* **8**, 322 (1951).
- [97] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz, *Phys. Rev* **134**, B1410 (1964).
- [98] S. Massar and S. Popescu, *Phys. Rev. A* **84**, 052106 (2011).
- [99] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, *Rev. Mod. Phys.* **81**, 865 (2009).
- [100] J. M. Jauch, *Foundations of quantum mechanics* (Addison-Wesley Pub. Co., Reading, 1968).
- [101] G. Lindblad, *Comm. Math. Phys.* **48**, 119 (1976).
- [102] S. L. Braunstein and C. M. Caves, *Found. Phys. Lett.* **1**, 3 (1988).
- [103] N. S. Williams and A. N. Jordan, *Phys. Rev. Lett.* **100**, 026804 (2008).
- [104] S. Parrott, “Quantum weak values are not unique,” (2009), arXiv:0909.0295v2 .
- [105] Y. Kedem and L. Vaidman, *Phys. Rev. Lett.* **105**, 230401 (2010).
- [106] A. J. Leggett, *J. Phys.: Condens. Matter* **14**, R415 (2002).
- [107] A. J. Leggett and A. Garg, *Phys. Rev. Lett.* **54**, 857 (1985).
- [108] J. S. Bell, *Physics* **1**, 195 (1965).
- [109] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
- [110] R. Ruskov, A. N. Korotkov, and A. Mizel, *Phys. Rev. Lett.* **96**, 200404 (2006).
- [111] M. W. Wilde and A. Mizel, *Found. Phys.* **42**, 256 (2012).
- [112] A. Palacios-Laloy, F. Mallet, F. Nguyen, P. Bertet, D. Vion, D. Esteve, and A. N. Korotkov, *Nature Phys.* **6**, 442 (2010).
- [113] M. Barbieri, *Phys. Rev. A* **80**, 034102 (2009).
- [114] S. Marcovitch and B. Reznik, “Testing Bell inequalities with weak measurements,” (2010), arXiv:1005.3236 .

- [115] P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih, *Phys. Rev. Lett.* **75**, 4337 (1995).
- [116] D. F. V. James, P. G. Kwiat, W. J. Munro, and A. G. White, *Phys. Rev. A* **64**, 052312 (2001).
- [117] E. T. Jaynes (Plenum Press, New York, 1980) pp. 37–43.
- [118] Y. Ji, Y. Chung, D. Sprinzak, M. Heiblum, D. Mahalu, and H. Shtrikman, *Nature* **422**, 415 (2003).
- [119] E. V. Devyatov, *Physics-Uspekhi* **50**, 197 (2007).
- [120] M. Büttiker, *Phys. Rev. B* **46**, 12485 (1992).
- [121] T. Martin and R. Landauer, *Phys. Rev. B* **45**, 1742 (1992).
- [122] Y. Blanter and M. Büttiker, *Phys. Rep.* **336** (2000).
- [123] I. Neder, N. Ofek, Y. Chung, M. Heiblum, D. Mahalu, and V. Umansky, *Nature* **448**, 333 (2007).
- [124] V. Giovannetti, F. Taddei, D. Frustaglia, and R. Fazio, *Phys. Rev. B* **77**, 155320 (2008).
- [125] P. V. Lin, F. E. Camino, and V. J. Goldman, *Phys. Rev. B* **80**, 125310 (2009).
- [126] L. Chirulli, E. Strambini, V. Giovannetti, F. Taddei, V. Piazza, R. Fazio, F. Beltram, and G. Burkard, *Phys. Rev. B* **82**, 045403 (2010).
- [127] Y. Aharonov and D. Bohm, *Phys. Rev.* **115**, 485 (1959).
- [128] J. Anandan, *Nature* **360**, 307 (1992).
- [129] V. S.-W. Chung, P. Samuelsson, and M. Büttiker, *Phys. Rev. B* **72**, 125320 (2005).
- [130] H. Förster, S. Pilgram, and M. Büttiker, *Phys. Rev. B* **72**, 075301 (2005).
- [131] J. T. Chalker, Y. Gefen, and M. Y. Veillette, *Phys. Rev. B* **76**, 085320 (2007).
- [132] L. V. Litvin, H.-P. Tranitz, W. Wegscheider, and C. Strunk, *Phys. Rev. B* **75**, 33315 (2007).
- [133] I. Neder, M. Heiblum, D. Mahalu, and V. Umansky, *Phys. Rev. Lett.* **98**, 36803 (2007).

- [134] P. Roulleau, F. Portier, D. C. Glattli, P. Roche, A. Cavanna, G. Faini, U. Gennser, and D. Mailly, *Phys. Rev. Lett.* **100**, 126802 (2008).
- [135] S.-C. Youn, H.-W. Lee, and H.-S. Sim, *Phys. Rev. B* **80**, 113307 (2009).
- [136] M. Hashisaka, A. Helzel, S. Nakamura, L. Litvin, Y. Yamauchi, K. Kobayashi, T. Ono, H.-P. Tranitz, W. Wegscheider, and C. Strunk, *Physica E* **42**, 1091 (2010).
- [137] G. L. Khym and K. Kang, *Phys. Rev. B* **79**, 195306 (2009).
- [138] G. Haack, H. Förster, and M. Büttiker, *Phys. Rev. B* **82**, 155303 (2010).
- [139] I. Neder, M. Heiblum, Y. Levinson, D. Mahalu, and V. Umansky, *Phys. Rev. Lett.* **96**, 16804 (2006).
- [140] P. Roulleau, F. Portier, D. C. Glattli, P. Roche, A. Cavanna, G. Faini, U. Gennser, and D. Mailly, *Phys. Rev. B* **76**, 161309 (2007).
- [141] L. V. Litvin, A. Helzel, H. P. Tranitz, W. Wegscheider, and C. Strunk, *Phys. Rev. B* **78**, 075303 (2008).
- [142] E. Bieri, M. Weiss, O. Göktas, M. Hauser, C. Schönenberger, and S. Oberholzer, *Phys. Rev. B* **79**, 245324 (2009).
- [143] E. V. Sukhorukov and V. V. Cheianov, *Phys. Rev. Lett.* **99**, 156801 (2007).
- [144] I. P. Levkivskyi and E. V. Sukhorukov, *Phys. Rev. B* **78**, 045322 (2008).
- [145] I. Neder and E. Ginossar, *Phys. Rev. Lett.* **100**, 196806 (2008).
- [146] C. Neuenhahn and F. Marquardt, *New J. Phys.* **10**, 115018 (2008).
- [147] S.-C. Youn, H.-W. Lee, and H.-S. Sim, *Phys. Rev. Lett.* **100**, 196807 (2008).
- [148] I. P. Levkivskyi and E. V. Sukhorukov, *Phys. Rev. Lett.* **103**, 036801 (2009).
- [149] D. L. Kovrizhin and J. T. Chalker, *Phys. Rev. B* **80**, 161306 (2009).
- [150] D. L. Kovrizhin and J. T. Chalker, *Phys. Rev. B* **81**, 155318 (2010).
- [151] L. Hardy, *Phys. Rev. Lett.* **68**, 2981 (1992).
- [152] Y. Aharonov, A. Botero, S. Popescu, B. Reznik, and J. Tollaksen, *Phys. Lett. A* **301**, 130 (2002).
- [153] K. Yokota, T. Yamamoto, M. Koashi, and N. Imoto, *New J. Phys.* **11**, 033011 (2009).

- [154] K. Kang, Phys. Rev. B **75**, 125326 (2007).
- [155] K. Kang and K. Ho Lee, Physica E **40**, 1395 (2008).
- [156] D. Braun and M.-S. Choi, Phys. Rev. A **78**, 032114 (2008).
- [157] V. Shpitalnik, Y. Gefen, and A. Romito, Phys. Rev. Lett. **101**, 226802 (2008).
- [158] J. A. Wheeler, *Problems in the Formulation of Physics* (North Holland, Amsterdam, 1979).
- [159] M. O. Scully and K. Drühl, Phys. Rev. A **25**, 2208 (1982).
- [160] M. Scully, B. Englert, and H. Walther, Nature **351**, 111 (1991).
- [161] T. J. Herzog, P. G. Kwiat, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. **75**, 3034 (1995).
- [162] R. Hilmer and P. Kwiat, Sci. Am. **296**, 90 (2007).
- [163] W. K. Wootters, Phys. Rev. Lett. **80**, 2245 (1998).
- [164] P. Samuelsson, E. V. Sukhorukov, and M. Büttiker, New J. Phys. **7**, 176 (2005).
- [165] D. Frustaglia and A. Cabello, Phys. Rev. B **80**, 201312 (2009).
- [166] E. Buks, R. Schuster, M. Heiblum, D. Mahalu, and V. Umansky, Nature **391**, 871 (1998).
- [167] D. Sprinzak, E. Buks, M. Heiblum, and H. Shtrikman, Phys. Rev. Lett. **84**, 5820 (2000).
- [168] I. Neder, F. Marquardt, M. Heiblum, D. Mahalu, and V. Umansky, Nature Phys. **3**, 524 (2007).
- [169] P. Roulleau, F. Portier, P. Roche, A. Cavanna, G. Faini, U. Gennser, and D. Mailly, Phys. Rev. Lett. **102**, 236802 (2009).
- [170] S. Pilgram and M. Büttiker, Phys. Rev. Lett. **89**, 200401 (2002).
- [171] D. V. Averin and E. V. Sukhorukov, Phys. Rev. Lett. **95**, 126803 (2005).
- [172] E. V. Sukhorukov, A. N. Jordan, S. Gustavsson, R. Leturcq, T. Ihn, and K. Ensslin, Nat. Phys. **3**, 243 (2007).
- [173] O. Zilberberg, A. Romito, and Y. Gefen, Phys. Rev. Lett. **106**, 080405 (2011).

- [174] A. D. Parks, D. W. Cullin, and D. C. Stoudt, *Proc. R. Soc. Lond.* **454**, 2997 (1998).
- [175] M. D. Turner, C. A. Hagedorn, S. Schlamming, and J. H. Gundlach, *Opt. Lett.* **36**, 1479 (2011).
- [176] J. M. Hogan, J. Hammer, S.-W. Chiow, S. Dickerson, D. M. S. Johnson, T. Kovachy, A. Sugarbaker, and M. A. Kasevich, *Opt. Lett.* **36**, 1698 (2011).
- [177] Y.-W. Cho, H.-T. Lim, Y.-S. Ra, and Y.-H. Kim, *New J. Phys.* **12**, 023036 (2010).
- [178] A. D. Parks and J. E. Gray, *Phys. Rev. A* **84**, 012116 (2011).
- [179] X. Zhu, Y. Zhang, S. Pang, C. Qiao, Q. Liu, and S. Wu, *Phys. Rev. A* **84**, 052111 (2011).
- [180] A. G. Kofman, S. Ashkab, and F. Nori, *Phys. Rep.* **520**, 43 (2012).
- [181] K. Nakamura, A. Nishizawa, and M.-K. Fujimoto, *Phys. Rev. A* **85**, 012113 (2012).
- [182] T. Koike and S. Tanaka, *Phys. Rev. A* **84**, 062106 (2011).
- [183] C. Simon and E. S. Polzik, *Phys. Rev. A* **83**, 040101(R) (2011).
- [184] A. Di Lorenzo, *Phys. Rev. A* **85**, 032106 (2012).
- [185] A. K. Pan and A. Matzkin, *Phys. Rev. A* **85**, 022122 (2012).
- [186] A. Romito, Y. Gefen, and Y. M. Blanter, *Phys. Rev. Lett.* **100**, 056801 (2008).
- [187] N. Brunner and C. Simon, *Phys. Rev. Lett.* **105**, 010405 (2010).
- [188] C.-F. Li, X.-Y. Xu, J.-S. Tang, J.-S. Xu, and G.-C. Guo, *Phys. Rev. A* **83**, 044102 (2011).
- [189] Y. Aharonov, F. Columbo, I. Sabadini, D. C. Struppa, and J. Tollaksen, *J. Phys. A: Math. Theor.* **44**, 365304 (2011).
- [190] M. V. Berry and P. Shukla, *J. Phys. A: Math. Theor.* **45**, 015301 (2011).
- [191] A. Hosoya and M. Koga, *J. Phys. A: Math. Theor.* **44**, 415303 (2011).
- [192] A. M. Steinberg, *Phys. Rev. Lett.* **74**, 2405 (1995).
- [193] A. M. Steinberg, *Phys. Rev. A* **52**, 32 (1995).

- [194] Y. Aharonov, N. Erez, and B. Reznik, *Phys. Rev. A* **65**, 052124 (2002).
- [195] Y. Aharonov, N. Erez, and B. Reznik, *J. Mod. Opt.* **50**, 1139 (2003).
- [196] D. Rohrlich and Y. Aharonov, *Phys. Rev. A* **66**, 042102 (2002).
- [197] N. Brunner, V. Scarani, M. Wegmuller, M. Legre, and N. Gisin, *Phys. Rev. Lett.* **93**, 203902 (2004).
- [198] Y. Aharonov and L. Vaidman, *J. Phys. A: Math. Gen.* **24**, 2315 (1991).
- [199] K. J. Resch, J. S. Lundeen, and A. M. Steinberg, *Phys. Lett. A* **324**, 125 (2004).
- [200] D. R. Solli, C. F. McCormick, R. Y. Chiao, S. Popescu, and J. M. Hickmann, *Phys. Rev. Lett.* **92**, 043601 (2004).
- [201] Q. Wang, F. W. Sun, Y. S. Zhang, Jian-Li, Y. F. Huang, and G. C. Guo, *Phys. Rev. A* **73**, 023814 (2006).
- [202] C. R. Leavens, *Found. Phys.* **35**, 469 (2005).
- [203] H. M. Wiseman, *New J. Phys.* **9**, 165 (2007).
- [204] B. J. Hiley, “Weak Values: Approach through the Clifford and Moyal Algebras,” (2011), arXiv:1111.6536 .
- [205] M. Iinuma, Y. Suzuki, G. Taguchi, Y. Kadoya, and H. F. Hofmann, *New J. Phys.* **13**, 033041 (2011).
- [206] D. Pegg, S. Barnett, and J. Jeffers, *J. Mod. Opt.* **49**, 913 (2002).
- [207] T. Amri, J. Laurat, and C. Fabre, *Phys. Rev. Lett.* **106**, 020502 (2011).
- [208] E. Nelson, *Phys. Rev.* **150**, 1079 (1966).
- [209] D. Bohm and B. J. Hiley, *Phys. Reports* **172**, 93 (1989).
- [210] R. Simon and G. Agarwal, *Opt. Lett.* **25**, 1313 (2000).
- [211] N. D. Mermin, *Physics Today* **62**, 8 (2009).
- [212] P. A. M. Dirac, *Proc. Roy. Soc. A* **133**, 60 (1931).
- [213] H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).
- [214] E. P. Wigner, *Group Theory* (Academic Press Inc., New York, 1959).

- [215] D. Hestenes, *Found. Phys.* **35**, 903 (2005).
- [216] D. Hestenes, *Am. J. Phys.* **71**, 104 (2003).
- [217] D. Hestenes, *Am. J. Phys.* **71**, 1 (2003).
- [218] B. J. Hiley, *Proc. of the Int. Meeting of ANPA* **22**, 107 (2001).
- [219] G. Sobczyk, *The College Math. J.* , 27 (1997).
- [220] G. Sobczyk, *Amer. Math. Month.* **108**, 289 (2001).
- [221] D. Bohm, P. G. Davies, and B. J. Hiley, *AIP Conf. Proc.* **810**, 314 (2005).
- [222] E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
- [223] A. S. Holevo, *J. Math. Phys.* **39**, 1373 (1998).
- [224] A. Connes, *Noncommutative Geometry* (Boston, MA: Academic Press, 1994).
- [225] M. Rédei, *Stud. Hist. Phil. Mod. Phys.* **21**, 493 (1996).
- [226] P. A. M. Dirac, *Phys. Rev.* **139**, B684 (1965).
- [227] M. Gadella and F. Gómez, *Found. Phys.* **32**, 815 (2002).
- [228] M. Gadella and F. Gómez, *Int. J. Theor. Phys.* **42**, 2225 (2003).
- [229] R. de la Madrid, *Eur. J. Phys.* **26**, 287 (2005).
- [230] T. Petrosky and I. Prigogine, *Phys. Lett. A* **182**, 5 (1993).
- [231] A. R. Bohm, *Phys. Rev. A* **60**, 861 (1999).
- [232] O. Civitarese and M. Gadella, *Phys. Rep.* **396**, 41 (2004).
- [233] E. Karpov, I. Prigogine, T. Petrosky, and G. Pronko, *J. Math. Phys.* **41**, 118 (2000).
- [234] G. Ordonez, T. Petrosky, and I. Prigogine, *Phys. Rev. A* **63**, 052106 (2001).
- [235] T. Petrosky, G. Ordonez, and I. Prigogine, *Phys. Rev. A* **64**, 062101 (2001).
- [236] R. Nakano, N. Hatano, and T. Petrosky, *Int. J. Theor. Phys.* **50**, 1134 (2011).
- [237] I. Antoniou, S. A. Shkarin, and Z. Suchanecki, *J. Math. Phys.* **40**, 4106 (1999).

- [238] M. Courbage, *Lett. Math. Phys.* **4**, 425 (1980).
- [239] G. Ordóñez, T. Petrosky, E. Karpov, and I. Prigogine, *Chaos, Solitons and Fractals* **12**, 2591 (2001).
- [240] M. Courbage, *Int. J. Theor. Phys.* **46**, 1881 (2007).
- [241] G. Ordonez and N. Hatano, *Phys. Rev. A* **79**, 042102 (2009).

A Algebraic Quantum Mechanics

The steady progress of physics requires for its theoretical formulation a mathematics which get continually more advanced. This is only natural and to be expected. What however was not expected by the scientific workers of the last century was the particular form that the line of advancement of mathematics would take, namely it was expected that mathematics would get more and more complicated, but would rest on a permanent basis of axioms and definitions, while actually the modern physical developments have required a mathematics that continually shifts its foundation and gets more abstract. Non-euclidean geometry and noncommutative algebra, which were at one time were considered to be purely fictions of the mind and pastimes of logical thinkers, have now been found to be very necessary for the description of general facts of the physical world. It seems likely that this process of increasing abstraction will continue in the future and the advance in physics is to be associated with continual modification and generalisation of the axioms at the base of mathematics rather than with a logical development of any one mathematical scheme on a fixed foundation.

Paul A. M. Dirac (1931) [212]

To complement the discussion in the main text, we briefly review how observables in the quantum theory can be defined from a slightly more abstract point of view advocated by Weyl [213] and Wigner [214]. Namely, we delay the introduction of operators and Hilbert space in favor of a purely algebraic approach that makes transparent the structural foundation of the theory. To accomplish this, we review the essential features of continuous Lie groups, their generating Lie algebras, the universal enveloping $*$ -algebra that contains both, and the

possible embedding of the enveloping algebra as a subalgebra of an even richer algebra. Traditional quantum observables will be identified as the elements of a non-commutative Lie algebra that satisfy the associative algebraic product of its universal enveloping algebra. We will see that all the essential aspects of the quantum theory will appear within this enveloping algebra.

Of particular interest for the main text will be the appearance of idempotent-valued spectral measures for the Lie algebraic elements and their algebraic powers. These measures will lead to a natural definition of the trace and the Hilbert-Schmidt inner product on the algebra, both of which play an important role in the main text. We deliberately omit the stochastic content of the quantum theory here, since it is developed more systematically in Chapters 2 and 3.

A.1 Lie Groups

A *group* \mathfrak{G} is a set of algebraic elements that is closed under an associative binary product, has a multiplicative identity, and has an involution that specifies a unique inverse for each element. These properties are summarized in Table A.1. A group is a highly abstract—and thus very general—mathematical concept. However, we will be most interested in its foundational significance for quantum physics, where it naturally specifies physical *symmetries* [213].

The elements of a *symmetry group* specify the possible transformations for a physical system in an abstract way, without specifying exactly what is being transformed. The identity element is a trivial transformation that leaves the system unchanged. The group product indicates a composition of transformations. Hence, any combination of transformations is another valid transformation, and

Closed:	$gh \in \mathfrak{G}$
Associative:	$g(hk) = (gh)k = ghk$
Unital:	$g1 = 1g = g$
Inverse:	$g^{**} = g$
	$gg^* = g^*g = 1$

Table A.1: Algebraic properties of a group \mathfrak{G} . Here, $g, h, k \in \mathfrak{G}$ and $*$: $\mathfrak{G} \rightarrow \mathfrak{G}$ is the group involution producing the inverse.

transformations can be reduced pairwise provided that their order is unchanged.

The inverse indicates that any transformation can be undone.

The most important type of symmetry group for specifying physical transformations is a *continuous* symmetry group—known as a *Lie group*—which consists of a continuous differentiable manifold of group elements such that the inverse and product are continuous operations on that manifold. The groups of spatial rotations, spacetime-translations, and Lorentz boosts are poignant examples of Lie groups. We shall assume that such a Lie group manifold is simply connected for reasons that will soon become clear.

A.1.1 Lie Algebra

The manifold structure of a Lie group implies that any local region of the manifold around a particular group element $g \in \mathfrak{G}$ can be parametrized by a set of real local coordinates $\vec{x} = (x^1, \dots, x^N) \in \mathbb{R}^N$ called a *coordinate chart*, where N is the dimension of the manifold. Each point $g(\vec{x}) \in \mathfrak{G}$ of the group manifold has a *tangent vector space* $T_{g(\vec{x})}\mathfrak{G}$ with basis elements that can be determined by taking partial derivatives with respect to a chosen set of local coordinates $\vec{D}_i(\vec{x}) = \partial_{x^i}g(\vec{x})$. Any vector in the tangent space at the point $g(\vec{x})$ of the manifold is a linear combination of these basis vectors, $\vec{Y}(\vec{x}) = \sum_i y^i \vec{D}_i(\vec{x})$, where the

components $y^i \in \mathbb{R}$ are real numbers.

The tangent vector space $\mathfrak{g} = T_1\mathfrak{G}$ at the identity element $g(0) = 1$ of a Lie group is particularly important for determining the structure of the group. This tangent vector space is also a non-associative *algebra* under a binary product known as the *Lie bracket*, and is known as the *Lie algebra* of the group. The Lie bracket relations between a set of basis elements \vec{D}_i of the vector space,

$$[\vec{D}_i, \vec{D}_j] = c_{ij}^k \vec{D}_k, \quad (\text{A.1})$$

specify the *structure constants* c_{ij}^k that completely determine the Lie algebra. We will return to the significance of this bracket shortly.

The Lie algebra completely determines the local structure of a Lie group manifold that is simply connected to the identity, which is why we restricted ourselves to simply connected group manifolds. That is, the smoothness of the manifold implies that points near to the identity $g(0) = 1$ along a straight coordinate curve $g(\epsilon\vec{y})$ parametrized by a real parameter $\epsilon \in \mathbb{R}$ can be formally determined by expanding it into a Taylor series around the identity,

$$g(\epsilon\vec{y}) = \sum_{n=0}^{\infty} \frac{1}{n!} (\partial_{\epsilon})^n g(\epsilon\vec{y})|_{\epsilon=0} = \sum_n \frac{1}{n!} (\epsilon\vec{Y})^n = \exp(\epsilon\vec{Y}), \quad (\text{A.2})$$

where $\vec{Y} = \sum_i y^i \vec{D}_i(0)$ is the vector in the Lie algebra that directly corresponds to the local coordinate tangent vector $\vec{y} = (y^i)$. Inverting the direction of this tangent vector produces the inverse group element $g^*(\epsilon\vec{y}) = g(-\epsilon\vec{y}) = \exp(-\epsilon\vec{Y})$. Thus, *exponentiating* the Lie algebra directly translates the group identity to other elements of the Lie group without having to specify a local coordinate chart on the group manifold *a priori*. We will henceforth dispense with these coordinate

charts as superfluous, and characterize the group directly with its Lie algebra. To move our discussion forward, we briefly delay the justification of the vector products that appear in the sum (A.2) until Section A.1.3.

Generally, Lie groups do not commute, $gh \neq hg$ (e.g., the Lie group of spatial rotations). We can measure the local structure of group commutation by considering the quantity ghg^*h^* for group elements close to the identity, $g = \exp(\epsilon\vec{Y})$, and $h = \exp(\epsilon\vec{Z})$. Computing this to linear order in ϵ produces the Lie bracket by definition, $\exp(\epsilon\vec{Y})\exp(\epsilon\vec{Z})\exp(-\epsilon\vec{Y})\exp(-\epsilon\vec{Z}) = 1 + \epsilon[\vec{Y}, \vec{Z}] + O(\epsilon^2)$, which has the explicit algebraic form of a *commutator*,

$$[\vec{Y}, \vec{Z}] = \vec{Y}\vec{Z} - \vec{Z}\vec{Y}. \quad (\text{A.3})$$

Hence, the Lie bracket specifies the degree to which associated group actions do not locally commute.

A.1.2 Lie Exponential Map

The *Lie exponential map* $\exp_\epsilon(\vec{Y})$ is formally defined to uniquely specify the one-parameter subgroup of the manifold that has the tangent vector \vec{Y} at the identity. That is, $\exp_\epsilon(\vec{Y}) \in \mathfrak{G}$ is the unique curve parametrized by ϵ such that $\exp_0(\vec{Y}) = 1$, $\partial_\epsilon \exp_\epsilon(\vec{Y})|_{\epsilon=0} = \vec{Y}$, and $\exp_{\epsilon_1}(\vec{Y})\exp_{\epsilon_2}(\vec{Y}) = \exp_{\epsilon_1+\epsilon_2}(\vec{Y})$. The exponential series expansion we used in (A.2) is equivalent to this map provided that the tangent vectors along the curve of the subgroup point in the same direction so that they can be identified, $\vec{D}_i(\epsilon) = \vec{D}_i(0)$.

More generally, we can consider specifying a Lie group element by translating along an arbitrary curve of the group manifold specified by a *tangent vector field*

such that the tangent vector rotates as it is transported along the curve. This can be described by a rotation field \mathcal{R}_ϵ along the curve that specifies the appropriate local correction to the tangent vectors $\vec{D}_i(\epsilon) = \mathcal{R}_\epsilon(\vec{D}_i(0))$ with \mathcal{R}_0 being the identity operation [215]. In such a case the series expansion (A.2) of the Lie exponential map for this curve will only formally hold to first order in ϵ , while the full map will be the integration of the differential equation,

$$\partial_\epsilon \exp_\epsilon(\vec{Y}) = \vec{Y}(\epsilon) \exp_\epsilon(\vec{Y}), \quad (\text{A.4})$$

where $\vec{Y}(\epsilon) = \mathcal{R}_\epsilon(\vec{Y})$ is the rotating, and thus ϵ -dependent, generator for the curved one-dimensional subgroup. This equation has a formal solution as the exponential series of an integral,

$$\exp_\epsilon(\vec{Y}) = \sum_{n=0}^{\infty} \frac{1}{n!} \left[\int_0^\epsilon d\epsilon' \vec{Y}(\epsilon') \right]^n = \exp \left(\int_0^\epsilon d\epsilon' \vec{Y}(\epsilon') \right). \quad (\text{A.5})$$

A.1.3 Universal Enveloping Algebra

Strictly speaking, for the formal algebraic expressions used in (A.2), (A.3), (A.4), and (A.5) to make sense we must extend the applicability of the algebraic product. That is, we must work within the *universal enveloping algebra* $\mathcal{E}(\mathfrak{g})$ that contains the real numbers \mathbb{R} , the Lie group elements \mathfrak{G} , their Lie algebra \mathfrak{g} , and all the direct algebraic powers of the Lie algebra \mathfrak{g}^n , such that the group product and the Lie bracket are faithfully represented by the product in the enveloping algebra.

This enveloping algebra can be constructed as a quotient ring from the most general graded *tensor algebra* generated by the Lie algebra vector space, $\mathcal{T}(\mathfrak{g}) = \bigoplus_{n=0}^{\infty} T^n \mathfrak{g}$, where the grade-0 elements $T^0 \mathfrak{g} = \mathbb{R}$ are the real numbers, and the

grade- k elements $T^k \mathfrak{g} = \bigotimes_{n=1}^k \mathfrak{g}$ are the k^{th} -order tensor products of the Lie algebra. The enveloping algebra is the quotient $\mathcal{E}(\mathfrak{g}) = \mathcal{T}(\mathfrak{g})/\mathcal{J}_L$ of the tensor algebra with a two-sided ideal chosen to represent the Lie bracket as a commutator to match (A.3),

$$\mathcal{J}_L = \left\{ C \left(\vec{A} \otimes \vec{B} - \vec{B} \otimes \vec{A} - [\vec{A}, \vec{B}] \right) D \mid \vec{A}, \vec{B} \in \mathfrak{g}, C, D \in \mathcal{T}(\mathfrak{g}) \right\}. \quad (\text{A.6})$$

This quotient creates an equivalence relation between all elements satisfying the form of the ideal \mathcal{J}_L . With this construction of the universal enveloping algebra, the exponential sums in (A.2) and (A.5) makes algebraic sense and are well defined. Moreover these sums ensure that the algebraic product of the enveloping algebra $\mathcal{E}(\mathfrak{g})$ will be a faithful extension of the group product of \mathfrak{G} .

The constructed universal enveloping algebra $\mathcal{E}(\mathfrak{g})$ has two associative binary operations: a (generally non-commutative) product inherited from the Lie group, and a commutative sum inherited from the Lie algebra. It is closed under both these operations, obeys the standard distributive property between product and sum, and contains both the additive identity (0), and the multiplicative identity (1). The sum always has an inverse obtained by taking the negative of an element. The product is not generally invertible.

The group involution can be extended to the enveloping algebra so that it becomes a $*$ -algebra. This is done by applying the involution to the exponential in (A.2), which implies that it must be equivalent to the additive inverse when applied to the Lie algebra $\vec{Y}^* = -\vec{Y}$, making the Lie algebra *anti-Hermitian* under the algebraic involution. The involution then extends naturally to all the algebraic products of the Lie algebra by reversing the product order in the same manner as a transpose $(\vec{A}\vec{B})^* = \vec{B}^*\vec{A}^*$, and distributing over sums $(\vec{A} + \vec{B})^* = \vec{A}^* + \vec{B}^*$.

Closed:	$AB \in \mathcal{E}(\mathfrak{g})$ $A + B \in \mathcal{E}(\mathfrak{g})$
Associative:	$A(BC) = (AB)C = ABC$ $A + (B + C) = (A + B) + C = A + B + C$
Multiplicative identity:	$A1 = 1A = A$
Additive identity:	$A + 0 = 0 + A = A$
Scalar multiplication:	$(\alpha A)(\beta B) = (\alpha\beta)(AB)$ $\alpha(A + B) = \alpha A + \alpha B$
Distributivity:	$A(B + C) = AB + AC$ $(A + B)C = AC + BC$
Involution:	$A^{**} = A$ $(AB)^* = B^*A^*$ $(A + B)^* = A^* + B^*$

Table A.2: Properties of the universal enveloping algebra $\mathcal{E}(\mathfrak{g})$ that faithfully contains the Lie group \mathfrak{G} , its Lie algebra \mathfrak{g} , and the scalars \mathbb{R} . Here, $A, B, C \in \mathcal{E}(\mathfrak{g})$ and $\alpha, \beta \in \mathbb{R}$. Complexifying this algebra substitutes \mathbb{C} for \mathbb{R} .

These properties are all summarized in Table A.2.

This universal enveloping algebra is the minimal algebra that faithfully contains the Lie group and its Lie algebra. However, it may in turn be embedded as a closed subalgebra of an even larger $*$ -algebra $\mathcal{E}(\mathfrak{g}) \subset \mathfrak{A}$ that can contain additional structure not specified solely by a Lie group, but which may still be physically relevant or convenient. For a poignant example, the graded enveloping algebra $\mathcal{E}(\mathfrak{g})$ may be embedded as the even-graded subalgebra of a larger super-algebra such as a Clifford algebra that can have other helpful properties or physical relevance [87–92, 215–217]. We will illustrate how the Lie group of quantum spin can be readily described using such an embedding in Section A.8. In what follows, however, we need only assume that we are working in some appropriate algebra \mathfrak{A} that contains the universal enveloping algebra $\mathcal{E}(\mathfrak{g})$ of the group as a closed subalgebra.

A.2 Quantum Observables

Traditional quantum observables are fundamentally defined to be the elements of a non-commutative Lie algebra contained within its universal enveloping algebra. Hence, the structure of a set of observables is completely specified by their commutation relations. Indeed, we will find that all aspects of the traditional quantum theory are contained in the enveloping algebra for the Lie algebra of the observables without the introduction of any auxiliary Hilbert space.

Conventionally in physics we *complexify* the algebra (i.e., replace the scalars of $\mathcal{E}(\mathfrak{g})$ with the complex numbers \mathbb{C}) and formally write the anti-Hermitian Lie algebra elements \vec{Y} using a scalar imaginary, $\vec{Y} = -i\vec{G}$, where $\vec{G}^* = \vec{G}$ are then *Hermitian* generators under the involution, and where the involution is extended to the complex numbers as the complex conjugate $i^* = -i$. The structure constants (A.1) specified by a basis $\vec{D}_i = -i\vec{G}_i$ acquire a scalar imaginary factor, $[\vec{G}_i, \vec{G}_j] = (ic_{ij}^k) \vec{G}_k$, when specified in terms of the Hermitian generators.

With the introduction of the scalar imaginary, we can write (A.4) and its solution (A.5) in the more familiar form,

$$i\partial_\epsilon U_\epsilon = \vec{G}(\epsilon)U_\epsilon, \quad (\text{A.7})$$

$$U_\epsilon = \exp\left(-i \int_0^\epsilon d\epsilon' \vec{G}(\epsilon')\right), \quad (\text{A.8})$$

where $U_\epsilon = \exp_\epsilon(-i\vec{G})$ is the exponential map specifying the curve through the group manifold generated by a (possibly ϵ -dependent) Hermitian algebraic element $\vec{G}(\epsilon)$ acting as a vector field..

Equation (A.7) is a generalization of the *Schrödinger equation* expressed directly at the level of its *propagator*, where the flow parameter ϵ is usually a proper

time, and the generator $\vec{G} = \vec{H}/\hbar$ is usually a scaled *Hamiltonian* that generates the translations along that proper time. We can therefore understand such a differential equation as specifying the orbit of a particular one-dimensional subgroup through a larger Lie group manifold.

To match standard quantum conventions for clarity, we will adopt the convention of using a scalar imaginary and Hermitian generators $-i\vec{G}$ in what follows. We will use the term *observable* to specifically denote these Hermitian generators. We will also use the term *propagator* to denote a one-parameter subgroup orbit U_ϵ specified by an equation like (A.7).

A.3 Adjoint Maps

A Lie group can act directly on itself as a transformation. The form of the action of the group element g on another element h can take three distinct forms: a left product $h \mapsto g^*h$, a right product $h \mapsto hg$, and a double-sided product $h \mapsto g^*hg$. Of particular interest in this work will be the double-sided product, since it preserves the group structure, $hk \mapsto g^*hgg^*kg = g^*hkg$. Such a structure-preserving transformation that uses the group elements to map the group onto itself is an *inner automorphism*.

The *adjoint map* lifts a group element g to its inner automorphism,

$$\text{Ad}_g(h) = g^*hg. \tag{A.9}$$

The inverse of the adjoint map flips the product order, $\text{Ad}_g^*(h) = \text{Ad}_{g^*}(h) = ghg^*$, so that $\text{Ad}_g^*\text{Ad}_g(h) = \text{Ad}_g\text{Ad}_g^*(h) = h$ for any $h \in \mathfrak{G}$.

A Lie group \mathfrak{G} can also act directly on its Lie algebra \mathfrak{g} and its enveloping

algebra $\mathcal{E}(\mathfrak{g})$ via this adjoint map. Notably, if $h = \exp_\epsilon(-i\vec{G})$, then $\text{Ad}_g(h) = g^* \exp_\epsilon(-i\vec{G})g = \exp_\epsilon(-i\text{Ad}_g(\vec{G}))$, since the automorphism preserves the algebraic products in the exponential sum.

As another notable example, a propagator $g = U_\epsilon = \exp_\epsilon(-i\vec{G})$ as in (A.7) acts directly on an observable \vec{X} according to,

$$\text{Ad}_{U_\epsilon}(\vec{X}) = U_\epsilon^* \vec{X} U_\epsilon = \exp_\epsilon(i\vec{G}) \vec{X} \exp_\epsilon(-i\vec{G}). \quad (\text{A.10})$$

The adjoint group action on an observable thus generalizes the *Heisenberg picture* of the evolution of an observable, for which $\vec{G} = \vec{H}/\hbar$ is usually the scaled Hamiltonian and ϵ a proper time.

The adjoint map for the Lie group also defines an induced adjoint map for its Lie algebra by taking a derivative,

$$\text{ad}[-i\vec{G}](-i\vec{X}) = \partial_\epsilon \text{Ad}_{\exp_\epsilon(-i\vec{G})}(-i\vec{X})|_{\epsilon=0} = [-i\vec{X}, -i\vec{G}]. \quad (\text{A.11})$$

This induced map lifts an element of the Lie algebra to its associated right action on the Lie algebra itself, which has the form of a commutator. Similarly, the induced map that corresponds to the inverse adjoint map is the reversed commutator, $\text{ad}^*[-i\vec{G}](-i\vec{X}) = [-i\vec{G}, -i\vec{X}]$, or left action on the Lie algebra.

Notably, the induced adjoint map can specify the evolution of an observable under a propagator directly,

$$\partial_\epsilon \vec{X}_\epsilon = \text{ad}[-i\vec{G}(\epsilon)](\vec{X}) = [\vec{X}, -i\vec{G}(\epsilon)], \quad (\text{A.12})$$

where $\vec{X}_\epsilon = \text{Ad}_{U_\epsilon}(\vec{X})$. This equation generalizes *Heisenberg's equation of motion*

for an observable by specifying how to transport it along a particular curve of the group manifold; it has the same information content as the Schrödinger equation for the propagator itself (A.7). Its solution relates the adjoint map of a Lie group to the induced adjoint map of its algebra,

$$\text{Ad}_{\exp_{\epsilon}(-i\vec{G})}(\vec{X}) = \exp_{\epsilon}\left(\text{ad}[-i\vec{G}]\right)(\vec{X}), \quad (\text{A.13})$$

and takes the form of an exponential sum of nested commutators that act as *directional derivatives* along the flow of the group action.

A.4 Idempotents, Spectra, and Irreducible Factors

Since the universal enveloping algebra $\mathcal{E}(\mathfrak{g})$ is a closed graded algebra, it will have a graded basis. For example, 1 is the sole basis element of the grade-0 part of $\mathcal{E}(\mathfrak{g})$, while the Lie algebra basis $\{-i\vec{G}_j\}$ will be a basis for the grade-1 part of $\mathcal{E}(\mathfrak{g})$. The bases of higher grades in $\mathcal{E}(\mathfrak{g})$ will be specified by the algebraic products of the grade-1 basis. As an alternative to this graded basis, however, algebraic elements may also be expanded in terms of characteristic spectral *idempotents* ε , which square to themselves $\varepsilon^2 = \varepsilon$. These spectral idempotents will play a fundamental role in the development of the statistical aspect of the theory in Chapters 2 and 3.

Idempotents can be given a natural partial order, such that $\varepsilon_1 \leq \varepsilon_2 \iff \varepsilon_1\varepsilon_2 = \varepsilon_2\varepsilon_1 = \varepsilon_1$. Hence, if $\varepsilon_1 < \varepsilon_2$ then $\varepsilon_2 = \varepsilon_1 + \bar{\varepsilon}_1$ for some other idempotent $\bar{\varepsilon}_1$ that is *disjoint* to ε_1 in the product sense $\varepsilon_1\bar{\varepsilon}_1 = \bar{\varepsilon}_1\varepsilon_1 = 0$. According to this partial

ordering, 0 is the smallest idempotent, while 1 is the largest; these always exist, so are called the *trivial* idempotents. A nontrivial idempotent that is strictly greater than zero according to this ordering but that cannot be further decomposed is called *primitive*. Primitive idempotents have the important algebraic property that for any $A \in \mathcal{E}(\mathfrak{g})$,

$$\varepsilon A \varepsilon = \lambda \varepsilon, \tag{A.14}$$

for some λ that commutes with the entire algebra, which for $\mathcal{E}(\mathfrak{g})$ must be a scalar [89].

The idempotent partial ordering implies that the identity may be decomposed into a sum of smaller disjoint idempotents, $1 = \varepsilon_1 + \bar{\varepsilon}_1$. Furthermore, if there exist primitive idempotents then any such *partition of unity* may be further split until it is resolved into a sum of primitive disjoint idempotents,

$$1 = \sum_n \varepsilon_n. \tag{A.15}$$

If a countable set of primitive idempotents do not exist, then this splitting procedure may be continued indefinitely. In such a case, there is a continuum of idempotents and we can consider this splitting procedure to define a *measure space* of disjoint idempotent intervals [218] such that the identity may be decomposed using an integral,

$$1 = \int d\varepsilon. \tag{A.16}$$

The *idempotent-valued measure* $d\varepsilon$ takes a measurable set from some indexing measurable space, such as the Borel sets on the real line, and uniquely assigns it

a corresponding idempotent.

Given such a partition of unity, other inequivalent partitions may be specified through group automorphisms, $1 = \text{Ad}_g(1) = \int \text{Ad}_g(d\varepsilon)$. Each new measure $\text{Ad}_g(d\varepsilon)$ will specify a different characteristic set of idempotents for the algebra.

The *generalized spectral theorem* for associative algebraic elements provides the link between these idempotent partitions of unity and the enveloping algebra $\mathcal{E}(\mathfrak{g})$. Specifically, any algebraic element can be decomposed as a sum of characteristic elements that form a *spectral basis* for that element [219, 220],

$$A = \sum_i (\lambda_i + \vartheta_i) \varepsilon_i, \quad (\text{A.17})$$

where $\lambda_i \in \mathbb{C}$ are complex *eigenvalues* of A with degeneracy n_i , ε_i are disjoint idempotents that partition unity $\sum_i \varepsilon_i = 1$, and $\vartheta_i^{n_i} = 0$ are *nilpotents* of a degree n_i that matches the associated eigenvalue degeneracy. The spectral eigenvalues, idempotents, and nilpotents can all be obtained from the characteristic polynomial of A . For a *normal* element that satisfies $A^*A = AA^*$, then all characteristic nilpotents are identically zero $\vartheta_i = 0$ and the idempotents are Hermitian $\varepsilon_i = \varepsilon_i^*$ [219]. Therefore, the elements of a Lie algebra \mathfrak{g} , which are anti-Hermitian and therefore normal, will be completely characterized by sets of disjoint idempotents that partition unity and their associated spectra.

Such a spectral decomposition for a normal element holds even for a continuum of idempotents [23], so any Lie algebra element $-i\vec{G} \in \mathfrak{g}$ will have a spectral decomposition in terms of a Hermitian idempotent-valued measure,

$$-i\vec{G} = -i \int d\varepsilon(x) G(x), \quad (\text{A.18})$$

where $G(x)$ is its characteristic *spectrum*, and $d\varepsilon$ is the corresponding *spectral idempotent-valued measure*. Since \vec{G} is Hermitian, its spectrum $G(x)$ will be real and its idempotents will also be Hermitian. It then follows that $\vec{G}^n = \int d\varepsilon(x) G^n(x)$ and higher powers will be entirely characterized by powers of the spectrum. The idempotent structure thus completely determines the bases of any grade in the enveloping algebra $\mathcal{E}(\mathfrak{g})$ and provides an equivalent *spectral basis* for the algebra [91, 218, 221].

Evidently, any elements of the algebra that can be expanded using the same spectral idempotent-valued measure will commute. Moreover, if there are two mutually commuting subsets of the Lie algebra $\{-i\vec{G}_i^{(1)}\}$ and $\{-i\vec{F}_j^{(2)}\}$ of the Lie algebra such that $[\vec{G}_i^{(1)}, \vec{F}_j^{(2)}] = 0$ but do not themselves commute $[\vec{G}_i^{(1)}, \vec{G}_j^{(1)}] \neq 0$ and $[\vec{F}_i^{(2)}, \vec{F}_j^{(2)}] \neq 0$ unless $i = j$, then it follows that their respective partitions of unity are completely independent. That is, the vector space of the Lie algebra *factors* into two independent Lie algebras as a direct sum $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$, so the enveloping algebra has the corresponding structure of a direct product of the enveloping algebra of each factor $\mathcal{E}(\mathfrak{g}) = \mathcal{E}(\mathfrak{g}_1) \otimes \mathcal{E}(\mathfrak{g}_2)$. Each independent idempotent-valued measure is thus a decomposition of the identity that is constrained to a single factor of this product, implying that the identity has a more complete decomposition as a product of identities and thus a product of measures,

$$1 = 1_1 1_2 = \left[\int d\varepsilon_1 \right] \left[\int d\varepsilon_2 \right] = \iint d^2(\varepsilon_1 \varepsilon_2). \quad (\text{A.19})$$

Reducing the Lie algebra to its independent factors in this way is a useful way to understand its structure. When a factor cannot further be reduced into factors, it is called an *irreducible* factor of the Lie algebra. Each irreducible factor can define its own independent Lie group and enveloping algebra, but a product

of factors is a larger space that contains more structure than each of the factors considered separately. The treatment of quantum spin in Section A.8 provides an illustrative example of such a decomposition.

A.5 Trace

The idempotent-valued measures (A.16) permit a natural *trace* measure to be defined on the algebra. The trace is a linear measure $\text{Tr}(A + B) = \text{Tr}(A) + \text{Tr}(B)$ that obeys a cyclic property $\text{Tr}(AB) = \text{Tr}(BA)$. Noting that any primitive idempotent ε has the reduction property (A.14), we can define the trace as the cyclic linear measure that assigns a constant value $c \in \mathbb{R}$ to any primitive idempotent $\text{Tr}(\varepsilon) = c$. Any element A then has a well-defined trace by exploiting a partition of unity,

$$\text{Tr}(A) = \text{Tr}\left(\sum_n \varepsilon_n A\right) = \sum_n \text{Tr}(\varepsilon_n A \varepsilon_n) = \sum_n \lambda_n \text{Tr}(\varepsilon_n) = c \sum_n \lambda_n, \quad (\text{A.20})$$

which will be uniquely defined up to the choice of constant c . Moreover, the trace is invariant under group automorphisms $\text{Tr}(\text{Ad}_g(A)) = \text{Tr}(g^* A g) = \text{Tr}(g g^* A) = \text{Tr}(A)$, so it will not depend on the choice of partition of unity used to compute it. In particular, if we choose the spectral decomposition of A then the constants $\lambda_n = A_n$ become eigenvalues and we can understand the trace as a sum of the intrinsic eigenvalues of A .

For a continuous set of idempotents, this definition can be generalized by taking a limiting procedure of splitting the idempotents until the primitive idempotent relation $(d\varepsilon(x)) A (d\varepsilon(x)) = \lambda(x) d\varepsilon(x)$ will be approximately satisfied with an

error term that vanishes in the limit. The trace will then have the limiting form,

$$\mathrm{Tr}(A) = \int \mathrm{Tr}((d\varepsilon(x)) A (d\varepsilon(x))) = \int \mathrm{Tr}(d\varepsilon(x))\lambda(x) = c \int dx \lambda(x). \quad (\text{A.21})$$

The resulting measure $\mathrm{Tr}(d\varepsilon(x)) = c dx$ must be translation-invariant on the indexing set, so will be a multiple of the *Lebesgue measure*. More generally, it will be a *Lebesgue-Stieltjes measure* that can have isolated singular points corresponding to primitive idempotents in the continuum [218]; the relation (A.20) is then a special case when the indexing measure is purely discrete.

Since a complete idempotent-valued measure will factor into a product of measures if the Lie algebra is reducible, the trace will also factor into a product of traces for each factor, called *partial traces*,

$$\mathrm{Tr}(A) = \mathrm{Tr}_1(\mathrm{Tr}_2(A)) = \mathrm{Tr}_2(\mathrm{Tr}_1(A)). \quad (\text{A.22})$$

For elements that are a product of the factors, then these partial traces decouple into the product of partial traces applied to each factor, $\mathrm{Tr}(A^{(1)}B^{(2)}) = \mathrm{Tr}_1(A^{(1)})\mathrm{Tr}_2(B^{(2)})$.

The trace also imbues the algebra with a natural inner product,

$$\langle A, B \rangle = \mathrm{Tr}(A^* B), \quad (\text{A.23})$$

known as the *Frobenius* (or *Hilbert-Schmidt*) inner product [23]. This inner product is invariant under group automorphisms, $\langle \mathrm{Ad}_g(A), \mathrm{Ad}_g(B) \rangle = \langle A, B \rangle$, and reflects the rigid manifold geometry that is inherent to a Lie group. The trace and inner product also permit a natural classification of the elements of the algebra.

Trace-class:	$\ T\ _1 < \infty$
Hilbert-Schmidt:	$\ S\ _2 < \infty$
Compact:	$C = \sum_n c_n \varepsilon_n, \lim_{n \rightarrow \infty} c_n = 0$
Bounded:	$ \langle T, B \rangle < \infty$

Table A.3: Classification of algebraic elements. Here, $T, S, C, B \in \mathcal{E}(\mathfrak{g})$ such that the letters distinguish the different classes. The classes satisfy the general inclusion relations, $\{T\} \subset \{S\} \subset \{C\} \subset \{B\} \subset \mathcal{E}(\mathfrak{g})$. Elements of $\mathcal{E}(\mathfrak{g})$ that are not bounded are *unbounded*.

bra, as well as a Type classification of the various permitted algebras; these are detailed in Appendix A.6 for completeness.

A.6 Algebraic Norms and Classification

The algebraic trace outlined in (A.20) and (A.21) gives the embedding algebra $\mathcal{E}(\mathfrak{g})$ a natural trace norm,

$$\|A\|_1 = \text{Tr}(|A|) = c \int dx |\lambda(x)|, \quad (\text{A.24})$$

Similarly, the Hilbert-Schmidt inner product (A.23) defines the *Hilbert-Schmidt* norm for the embedding algebra,

$$\|A\|_2 = \sqrt{\langle A, A \rangle}. \quad (\text{A.25})$$

The trace, inner product, and their associated norms allow us to further classify the elements of the algebra, as shown in Table A.3. The trace-class and bounded elements are examples of a special class of *-algebra called W^* -algebras, or *von Neumann algebras* [93]. Their dual relation under the Hilbert-Schmidt inner product permits a non-commutative generalization of a measure space. As such, von

Type <i>I</i> :	$\text{Tr}(\varepsilon) = cn, n \in \mathbb{N}$
—Type I_N :	$\text{Tr}(1) = cN, N \in \mathbb{N}$
—Type I_{\aleph_0} (I_∞):	$\text{Tr}(1) = c\aleph_0$
Type <i>II</i> :	$\text{Tr}(\varepsilon) = cx, x \in \mathbb{R}, x > 0$
—Type II_1 :	$\text{Tr}(1) = cX, X \in \mathbb{R}, X > 0$
—Type II_{\aleph_1} (II_∞):	$\text{Tr}(1) = c\aleph_1$
Type <i>III</i> :	$\text{Tr}(\varepsilon) = c\infty$

Table A.4: Classification of algebraic factors. Here ε is an arbitrary idempotent of the factor, \aleph_0 is the cardinality of the natural numbers, \aleph_1 is the cardinality of the real numbers, and ∞ is an infinity of unspecified cardinality.

Neumann algebras form a natural setting for discussing non-commutative probability theory and quantum operations [17–27, 81–83, 222, 223]. Hence, the development in Chapters 2 and 3 are more rigorously set within the context of these von Neumann algebras embedded within the enveloping algebra for the Lie group.

The irreducible factors of the enveloping algebra can also be classified according to the possible values that the partial traces (A.22) can take on the idempotents of each factor, as shown in Table A.4 [93]. Type *I* factors have countable numbers of primitive idempotents. Type *II* factors have continuous sets of idempotents that can form non-divergent measure spaces using the trace. Type *III* factors are pathological since they always diverge under the standard trace, but are thankfully rare; Connes has developed more advanced methods using non-standard traces for dealing with the divergences of these factors [224].

The factors I_N and II_1 are called *finite* factors since they can always be normalized by choosing the trace constant to be $c = 1/N$ or $c = 1/X$, respectively. The other factors are called *infinite*¹. The probability theory developed in Chap-

¹It is worth noting, however, that the infinite Type I_{\aleph_0} and Type II_{\aleph_1} factors can also be effectively normalized by extending the scalar field to an extension of the real numbers that sensibly admits infinitesimals and inverse infinitesimals, such as Robinson’s non-standard reals \mathbb{R}^* [79, 80], or Connes’ noncommutative infinitesimals [224].

ters 2 and 3 rigorously belongs to the setting of normalizable Type *I* and Type *II* factors.

A.7 Representations and Hilbert Space

I would like to make a confession which may seem immoral: I do not believe absolutely in Hilbert space any more. After all, Hilbert space (as far as quantum mechanical things are concerned) was obtained by generalizing Euclidean space, footing on the principle of ‘conserving the validity of all formal rules’ . . . Now we begin to believe that it is not the vectors which matter, but the lattice of all linear (closed) subspaces. Because: 1) The vectors ought to represent the physical states, but they do it redundantly, up to a complex factor, only 2) and besides, the states are merely a derived notion, the primitive (phenomenologically given) notion being the qualities which correspond to the linear closed subspaces. But if we wish to generalize the lattice of all linear closed subspaces from a Euclidean space to infinitely many dimensions, then one does not obtain Hilbert space, but that configuration which Murray and I called ‘case II_1 .’ (The lattice of all linear closed subspaces of Hilbert space is our ‘case I_∞ .’)

John von Neumann (1935), as quoted in [225]

The dominant formalism for quantum mechanics is not the algebraic formulation that we have developed; it is the state-vector formulation in a separable Hilbert space that was carefully developed in 1932 by von Neumann [11] as a possible rigorous implementation of the physical postulates and formal manipulations outlined by Dirac in 1930 [10]. In the Hilbert space formalism, the observable algebra is treated as an algebra of linear operators that act on state-vectors, which are understood to be fundamental aspects of the theory.

However, by 1935 von Neumann had already renounced Hilbert space as an appropriate foundation of the theory in favor of the abstract algebraic approach [225]. This change in heart brought him closer to the group algebra approach that Weyl had already been advocating by 1930 [213]. Dirac later noted in 1965 that state-vectors in a Hilbert space could become pathological in quantum electrody-

namics when an algebraic approach could still obtain sensible results [226]. Indeed, it has become clear in recent decades that an abstract algebraic approach will generally be required to sensibly describe infinite dimensional systems [23, 227–241]. In these cases, one is forced to abandon the state-vectors in favor of more general state-functionals over the abstract algebra of observables. We use this approach preferentially in the main text. The algebra takes on a fundamental role, while Hilbert space is demoted to a *representation* space for that algebra.

That being said, these observations have not detracted from the utility of Hilbert space as a representation space for calculations and rigorous proofs. Indeed, working within a concrete representation of such an associative $*$ -algebra can have its advantages. However, it is conceptually important that Hilbert space is not an intrinsic part of quantum mechanics; it is an auxiliary vector space that can be introduced to provide a convenient representation for the associative algebra of the symmetry groups inherent to quantum mechanics.

A.7.1 Hilbert Space

Any associative algebra \mathfrak{A} can be manipulated abstractly by specifying the algebraic products. However, one can also make a *representation* of the algebra by mapping its elements to linear operators that act on an auxiliary vector space \mathfrak{H} such that the algebraic product becomes the operator product. This vector space \mathfrak{H} is often assumed to have an inner product $(\vec{v}, \vec{w}) \in \mathbb{C}$, where $\vec{v}, \vec{w} \in \mathfrak{H}$, and contain all its limit points under the inner product norm, $\lim_{n \rightarrow \infty} \|\vec{v}_n - \vec{v}\| = 0 \implies \lim_{n \rightarrow \infty} \vec{v}_n = \vec{v} \in \mathfrak{H}$, where $\|\vec{v}\|^2 = (\vec{v}, \vec{v})$. Such a complete inner product space is called a *Hilbert space*. A *separable* Hilbert can be spanned by a countable set of vectors.

Under such a representation, $\pi : \mathfrak{A} \rightarrow \mathcal{O}(\mathfrak{H})$, the algebraic product will correspond to an operator product, $\pi(AB) = \pi(A)\pi(B) = \hat{A}\hat{B}$, where the hats indicate a Hilbert space operator. The representation is *faithful* if each operator representation \hat{A} can be inverted to retrieve a single corresponding algebraic element A . The representation is *irreducible* if there are no nontrivial subspaces of \mathfrak{H} that are invariant under the action of all operators in the representation $\pi(\mathfrak{A})$ of the algebra \mathfrak{A} .

The primary reason to adopt such a representation is that vector spaces and their linear operators are mathematical tools that are well understood and well developed. By transporting algebraic questions into a Hilbert space setting, certain calculations may become more transparent. However, any algebraically meaningful results will be independent of the chosen representation and *in principle* obtainable without using any particular representation. The physics is contained in the structure of the algebra, not in the representation of that structure.

A Lie group \mathfrak{G} has two natural associative algebras that one can represent on a Hilbert space: its adjoint map composition algebra $\text{Ad}_{\mathfrak{G}}$, and its universal enveloping algebra $\mathcal{E}(\mathfrak{g})$. A representation of the former is called an *adjoint representation* of the group. A representation of the latter is called a *unitary representation* of the group. We will now briefly review both.

Adjoint Representation

The adjoint map Ad_g produces a faithful representation of a Lie group element $g \in \mathfrak{G}$ acting as an inner automorphism operator on its own Lie algebra \mathfrak{g} as the auxiliary vector space. The operator product corresponds to the composition of adjoint maps. Specifically, since the Lie algebra is already a vector space it can

be spanned by a particular basis $\{\vec{D}_j = -i\vec{G}_j\}$. One can thus represent that basis as column vectors that have 1 in the j^{th} place and 0 elsewhere. Then Ad_g can be represented as a matrix acting on that column vector such that the matrix product produces the composite transformation². It follows that Ad_g^* is represented as the inverse matrix of Ad_g .

The induced adjoint map $\text{ad}[-i\vec{G}_j]$ is also represented as an operator acting on the Lie algebra. The operator product corresponds to the successive nesting of the Lie bracket. The matrix of $\text{ad}[-i\vec{G}_j]$ in this representation has particularly simple components $(\text{ad}[-i\vec{G}_j])_k^\ell = c_{jk}^\ell$ that are specified entirely by the structure constants of the algebra (A.1). Computing a matrix exponential of the representation of $\text{ad}[-i\vec{G}_j]$ produces the matrix representation of $\text{Ad}_{\exp_\epsilon(-i\vec{G}_j)}$ according to (A.13). Hence, the structure constants determine the entire group action on its own Lie algebra.

Unitary Representation

A representation $\pi : \mathcal{E}(\mathfrak{g}) \rightarrow \mathcal{O}(\mathfrak{H})$ of the enveloping algebra of a Lie group on an auxiliary Hilbert space \mathfrak{H} must preserve the algebraic product and map the involution to the *operator adjoint* with respect to the Hilbert space inner product. That is, $\pi(AB) = \pi(A)\pi(B) = \hat{A}\hat{B}$ and $\pi(A^*) = \hat{A}^\dagger$ where $A, B \in \mathcal{E}(\mathfrak{g})$, $\hat{A}, \hat{B} \in \mathcal{O}(\mathfrak{H})$, and the operator adjoint is defined via the inner product on \mathfrak{H} according to $(\vec{v}, \hat{A}\vec{w}) = (\hat{A}^\dagger\vec{v}, \vec{w})$ for any vectors $\vec{v}, \vec{w} \in \mathfrak{H}$ in the domain of \hat{A} . The inner product on \mathfrak{H} is chosen such that the operator trace faithfully matches the invariant algebraic trace $\text{Tr}(A) \mapsto \text{Tr}(\hat{A}) = \sum_n (\vec{v}_n, \hat{A}\vec{v}_n)$ up to an arbitrary constant. Such a representation is guaranteed to exist for $\mathcal{E}(\mathfrak{g})$ by the *Gel'fand-*

²The Bloch sphere representation of a qubit is an adjoint representation of this form.

Naimark-Segal (GNS) construction [23].

It follows that in any such representation the group elements will be *unitary* operators $g \mapsto \hat{U}$ that satisfy $\hat{U}^\dagger \hat{U} = \hat{U} \hat{U}^\dagger = \hat{1}$, where $1 \mapsto \hat{1}$ is the identity operator. Hence, any representation of the enveloping algebra will be a *unitary representation* for the group itself. It also follows that the unitary representations of propagators $U_\epsilon \mapsto \hat{U}_\epsilon = \exp(-i\epsilon \hat{G})$ will have self-adjoint operators as their generators $\hat{G}^\dagger = \hat{G}$. Indeed, *Stone's theorem* guarantees from the converse perspective that any one-parameter group of unitary operators acting on a Hilbert space must be generated via the exponentiation of self-adjoint operators in precisely this manner [23].

Furthermore, since a unitary representation preserves the algebraic product, the idempotents of the algebra $\epsilon^2 = \epsilon$ must correspond to *projection operators* on the Hilbert space, $\epsilon \mapsto \hat{\Pi}$. Moreover, each orthogonal basis of Hilbert space vectors $\{\vec{v}_n\} \in \mathfrak{H}$ will be in one-to-one correspondence with a disjoint set of idempotents $\{\epsilon_n\} \mapsto \{\hat{\Pi}_n\}$ that partition unity $\sum_n \epsilon_n = 1 \mapsto \hat{1} = \sum_n \hat{\Pi}_n$. Hence, separable Hilbert spaces naturally and faithfully represent factors that have countable sets of primitive idempotents. Additionally, the Hilbert-Schmidt inner product between two idempotents that correspond to vectors in the Hilbert space will map directly to the *complex square* of the inner product between those vectors, $\langle \epsilon_n, \epsilon_m \rangle = \text{Tr}(\epsilon_n^* \epsilon_m) \mapsto \text{Tr}(\hat{\Pi}_n \hat{\Pi}_m) = |(\vec{v}_n, \vec{v}_m)|^2$. The nonsquared inner product of the Hilbert space only helps to represent the algebraic product.

Irreducible representations correspond to tensor products of the representations of irreducible factors. That is, a Lie algebra $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$ produces an algebra with multiple factors $\mathcal{E}(\mathfrak{g}) = \mathcal{E}(\mathfrak{g}_1) \otimes \mathcal{E}(\mathfrak{g}_2)$ that has an irreducible representation $\pi_1(\mathcal{E}(\mathfrak{g}_1)) \otimes \pi_2(\mathcal{E}(\mathfrak{g}_2))$ on a tensor product of Hilbert spaces $\mathfrak{H}_1 \otimes \mathfrak{H}_2$ corresponding

to the irreducible representations of each factor.

For factors with continuous sets of idempotents there is no direct mapping between those idempotents and the countable basis vectors of a separable Hilbert space. However, an idempotent-valued measure (A.16) that partitions unity can be mapped to a *projection-valued measure* (PVM) on $\int d\varepsilon \mapsto \int d\hat{\Pi}$ that partitions the identity operator on the Hilbert space. It follows that the spectral idempotent decomposition of a Hermitian generator (A.18) maps to the spectral projection decomposition of a self-adjoint operator $\vec{G} = \int d\varepsilon(x) G(x) \mapsto \int d\hat{\Pi}(x) G(x) = \hat{G}$. Indeed, such a decomposition is guaranteed from the converse perspective by the *spectral theorem* for self-adjoint operators on a Hilbert space [23]. An orthonormal basis of Hilbert space vectors in such a case will correspond to a countable set of spectral idempotents of some *compact* algebraic element. The continuum of idempotents will correspond instead to the orthogonal vectors of Φ^* in a *rigged Hilbert space*, or *Gel'fand Triplet*, $\Phi \subset \mathfrak{H} \subset \Phi^*$ [91, 227–229]; the larger space Φ^* contains functionals that act sensibly only on a subset $\Phi \subset \mathfrak{H}$.

A.7.2 Dirac Notation

As we have seen, orthogonal state-vectors of a separable Hilbert space $\vec{v}_n \in \mathfrak{H}$ correspond to disjoint algebraic idempotents $\varepsilon_n \in \mathcal{E}(\mathfrak{g})$ in the enveloping algebra for a Lie group. However, each such idempotent is directly represented by a projection operator $\varepsilon_n \mapsto \hat{\Pi}_n$ corresponding to a one-dimensional subspace $[v_n] \subset \mathfrak{H}$ such that $\hat{\Pi}_n[v_n] = [v_n]$. This subspace contains not just a single vector \vec{v}_n but rather an *equivalence class* of vectors $\vec{v}_n \in [v_n] \subset \mathfrak{H}$ such that the equivalence class $[v_n]$ is invariant under multiplication by a nonzero complex scalar, $\forall z \in \mathbb{C} \setminus \{0\}, z[v_n] = [v_n]$. This subspace excludes the zero vector, so is a ray in

the *projective Hilbert space* $\mathcal{P}(\mathfrak{H})$. For calculations, one typically chooses a particular representative $\vec{v}_n \in [v_n] \subset \mathfrak{H}$, normalizes it $\vec{v}_n \rightarrow \vec{v}_n/(\vec{v}_n, \vec{v}_n)^{1/2}$ for later convenience, and treats any remaining overall phase factor as relative to an arbitrary fixed choice $e^{i\phi}\vec{v}_n \sim \vec{v}_n$. This normalization procedure effectively fixes the trace constant c in (A.20) to be unity.

Each state-vector $\vec{v} \in [v] \subset \mathfrak{H}$ is uniquely associated with a linear functional $v^\dagger : \mathfrak{H} \rightarrow \mathbb{C}$, known as the *dual* of \vec{v} , that is defined by the inner product $v^\dagger(\vec{w}) = (\vec{v}, \vec{w})$. To facilitate computations, Dirac [10] introduced a popular notational simplification that denotes state vectors by a ‘ket’ symbol $|v\rangle = \vec{v}$, and their dual functionals by a ‘bra’ symbol $\langle v| = v^\dagger$. The normalized inner product is written as a conjoined ‘bra-ket’ $\langle v|w\rangle = (\vec{v}, \vec{w}) \in \mathbb{C}$, such that $\langle v|v\rangle = 1$. A one-dimensional projection can then be written as the dyad $\hat{\Pi}_v = |v\rangle\langle v|$, such that its action on a vector takes the form $\hat{\Pi}_v |w\rangle = |v\rangle\langle v|w\rangle$. Spectral expansions can thus be written, $\hat{A} = \sum_n A_n |a_n\rangle\langle a_n|$. A matrix representation of these symbols using a particular reference basis can be constructed such that a ket will correspond to a column vector, a bra to the complex transpose of the ket (a row vector), the inner product will correspond to the matrix product between a row and a column, and the dyad to the matrix product between a column and row.

For factors with a continuum of idempotents represented on a rigged Hilbert space, the Dirac notation is extended to represent the non-normalizable orthogonal elements in the larger space Φ^* . Specifically an idempotent-valued measure maps to a projection-valued measure, $\int d\varepsilon \mapsto \int d\hat{\Pi} = \int dx |x\rangle\langle x|$, such that the individual dyads formally decompose into δ -normalized elements of Φ^* , $\langle x|x'\rangle = \delta(x - x')$, where $\delta(x - x')$ is the Dirac delta distribution. The dyad absorbs the inverse infinitesimal measure of a primitive idempotent in the continuum

such that $(d\varepsilon)^2 = (d\varepsilon) \mapsto (dx |x\rangle \langle x|)(dx' |x'\rangle \langle x'|) = dx |x\rangle (dx' \delta(x - x') \langle x'|) = dx |x\rangle \langle x|$. Hence, each ket $|x\rangle$ has units of the inverse square root of the infinitesimal Lebesgue-Stieltjes measure dx .

In the main text, we express results using Dirac notation in a (rigged) Hilbert space when it is particularly clear or convenient to do so. However, we also strive to emphasize the algebraic foundations of the quantum theory and acknowledge the operator formalism as a convenient representation of the algebraic structure.

A.8 Example: Quantum Spin

As an illuminating example of how the enveloping algebra of a Lie group can be embedded in a larger algebra with additional structure, we will briefly review the natural algebra for describing spacetime. This algebra is a maximal division algebra that is constructed to preserve the Minkowski inner product $\eta(\cdot, \cdot) : \mathbb{R}^4 \times \mathbb{R}^4 \rightarrow \mathbb{R}$ on spacetime, known as an *orthogonal Clifford algebra* [89]. We shall find that it has an even-graded subalgebra that is precisely the enveloping algebra $\mathcal{E}(\mathfrak{g})$ for the $\text{Spin}(1,3)$ Lie group. This group completely encapsulates the notion of quantum spin, including the relativistic Dirac spinors, the nonrelativistic Pauli spinors, and the complex phase of the wavefunction.

A.8.1 Spacetime Algebra

Similarly to the enveloping algebra $\mathcal{E}(\mathfrak{g})$ for a Lie group \mathfrak{G} in (A.6), the orthogonal Clifford algebra of spacetime can be constructed as a quotient algebra from the

graded tensor algebra $\mathcal{T}(\mathbb{R}^4)$ of \mathbb{R}^4 and the two-sided ideal,

$$\mathcal{J}_M = \{A(\vec{v} \otimes \vec{v} - \eta(\vec{v}, \vec{v}))B \mid \vec{v} \in \mathbb{R}^4, A, B \in \mathcal{T}(\mathbb{R}^4)\}, \quad (\text{A.26})$$

that faithfully represents the Minkowski pseudonorm $\eta(\vec{v}, \vec{v}) = \|\vec{v}\|^2$ of a 4-vector \vec{v} as its simple algebraic square $\vec{v}\vec{v} = \|\vec{v}\|^2$. The *spacetime algebra* is then defined as the quotient $\mathcal{T}(\mathbb{R}^4)/\mathcal{J}_M$.

This abstract construction leads to an intuitive decomposition of the algebraic product between two (grade-1) vectors,

$$\vec{v}\vec{w} = \eta(\vec{v}, \vec{w}) + \vec{v} \wedge \vec{w}, \quad (\text{A.27})$$

into a symmetric part $(\vec{v}\vec{w} + \vec{w}\vec{v})/2 = \eta(\vec{v}, \vec{w})$ that gives the Minkowski inner product (producing a grade-0 scalar), and an antisymmetric part $(\vec{v}\vec{w} - \vec{w}\vec{v})/2 = \vec{v} \wedge \vec{w}$ that gives the *wedge product* familiar from differential forms (producing a grade-2 bivector), which is a suitable generalization of the vector cross-product to any dimension.

Using an orthonormal basis $\{\vec{\gamma}_\mu\}_{\mu=0}^3$ for the vectors of spacetime, the algebraic product produces $2^4 = 16$ independent graded basis elements for the spacetime algebra [90, 92, 215–217],

$$\{1, \vec{\gamma}_0, \vec{\gamma}_1, \vec{\gamma}_2, \vec{\gamma}_3, \vec{\gamma}_{01}, \vec{\gamma}_{02}, \vec{\gamma}_{03}, \vec{\gamma}_{12}, \vec{\gamma}_{23}, \vec{\gamma}_{31}, \vec{\gamma}_{012}, \vec{\gamma}_{123}, \vec{\gamma}_{230}, \vec{\gamma}_{301}, \gamma_{0123}\}, \quad (\text{A.28})$$

where $\vec{\gamma}_{\mu\nu} = \vec{\gamma}_\mu \vec{\gamma}_\nu = \vec{\gamma}_\mu \wedge \vec{\gamma}_\nu$ gives a convenient shorthand for the basis elements of higher grade. All the products between these graded basis elements are determined by the Minkowski inner product and the fact that flipping the product order of

two orthogonal vectors simply reverses the sign, $\vec{\gamma}_{\mu\nu} = -\vec{\gamma}_{\nu\mu}$. We use the signature such that $\vec{\gamma}_0^2 = 1$ and $\vec{\gamma}_j^2 = -1$ for $j = 1, 2, 3$, so $\vec{\gamma}_0$ is a timelike basis vector and $\vec{\gamma}_j$ are spacelike basis vectors.

A general element of spacetime algebra, known as a *multivector*, is a linear combination of these 16 basis elements over the reals; one can project a multivector onto any particular grade with a grade projection $\langle A \rangle_\xi$ where $\xi = 0, 1, 2, 3, 4$ are the various grades. The grade-0 projection onto the scalars of the algebra obeys the cyclic property $\langle AB \rangle_0 = \langle BA \rangle_0$ and provides a natural definition for the *trace* on the algebra (A.20), up to a constant. The basis elements of nonzero grade will be traceless.

The algebra possesses a natural involution, known as *reversion*, which simply reverses the order of all basis-element products, $\vec{\gamma}_{\mu\nu\delta}^* = \vec{\gamma}_{\delta\nu\mu}$. As a consequence, it satisfies the property $(AB)^* = B^*A^*$ for any multivector product. Notably, every grade-2 or grade-3 element is anti-Hermitian under this reversion, while elements of grades 0,1,4 are Hermitian. The pseudonorm of any basis element is its reversed square $||\vec{\gamma}_{\mu\nu}||^2 = \vec{\gamma}_{\mu\nu}^* \vec{\gamma}_{\mu\nu}$; as a result, the pseudonorm of any multivector is the scalar part of its reversed square $||A||^2 = \langle A^*A \rangle_0$. The inner product between any two multivectors is similarly the scalar part of their reversed product $\langle A, B \rangle = \langle A^*B \rangle_0$, which has the same form as the Hilbert-Schmidt inner product (A.23) with the grade-0 projection playing the role of the trace. Each element with nonzero pseudonorm has an inverse under the product—e.g., the inverse of a vector is simply $\vec{\gamma}^{-1} = \vec{\gamma}/||\vec{\gamma}||^2$. The existence of this inverse for vectors (and general multivectors) is one of the main advantages of the spacetime algebra.

It is easy to verify that the basis element of grade-4 γ_{0123} squares to -1 , commutes with elements of even grade, anti-commutes with elements of odd grade,

and is Hermitian under reversion $\gamma_{0123}^* = \gamma_{0123}$. We will denote it as I and call it the *pseudoscalar* of the spacetime algebra. Multiplication by the pseudoscalar converts an element of grade- k to an element of grade- $(4 - k)$, which is equivalent to the Hodge star operation in differential forms. In particular, a trivector such as $\vec{\gamma}_{123} = I\vec{\gamma}_0$ can be written using this Hodge duality as the pseudoscalar multiplying a vector, so will have the properties of *pseudovectors* (also known as axial vectors). The six bivectors similarly split into two classes of three bivectors that are Hodge duals of each other.

A.8.2 Spinor Algebra

The even-graded subalgebra of spacetime algebra, known as its *spinor algebra*, has $2^3 = 8$ basis elements,

$$\{1, \vec{\gamma}_{01}, \vec{\gamma}_{02}, \vec{\gamma}_{03}, \vec{\gamma}_{12}, \vec{\gamma}_{23}, \vec{\gamma}_{31}, I\}, \quad (\text{A.29})$$

that can be further partitioned into the scalars $\{1, I\}$ that commute with the whole spinor algebra, the three anti-Hermitian bivectors $\{\vec{\gamma}_{01}, \vec{\gamma}_{02}, \vec{\gamma}_{03}\}$ with -1 pseudonorm, and the three anti-Hermitian bivectors $\{\vec{\gamma}_{12}, \vec{\gamma}_{23}, \vec{\gamma}_{13}\}$ with $+1$ pseudonorm.

These six bivectors obey the commutation relations,

$$[\vec{\gamma}_{23}, \vec{\gamma}_{31}] = 2\vec{\gamma}_{12} \quad [\vec{\gamma}_{31}, \vec{\gamma}_{12}] = 2\vec{\gamma}_{23} \quad [\vec{\gamma}_{12}, \vec{\gamma}_{23}] = 2\vec{\gamma}_{31}, \quad (\text{A.30})$$

$$[\vec{\gamma}_{01}, \vec{\gamma}_{02}] = -2\vec{\gamma}_{12} \quad [\vec{\gamma}_{02}, \vec{\gamma}_{03}] = -2\vec{\gamma}_{23} \quad [\vec{\gamma}_{03}, \vec{\gamma}_{01}] = -2\vec{\gamma}_{31}, \quad (\text{A.31})$$

$$[\vec{\gamma}_{23}, \vec{\gamma}_{02}] = 2\vec{\gamma}_{03} \quad [\vec{\gamma}_{31}, \vec{\gamma}_{03}] = 2\vec{\gamma}_{01} \quad [\vec{\gamma}_{12}, \vec{\gamma}_{01}] = 2\vec{\gamma}_{02}, \quad (\text{A.32})$$

which are the Lie bracket relations for the Lie algebra of the *Lorentz group* of spacetime boosts and rotations. Exponentiating this algebra produces the simply

connected Spin(1,3) Lie group, which is a double-covering of the identity component of the Lorentz group. Hence, the spinor algebra is the enveloping algebra $\mathcal{E}(\mathfrak{g})$ for the Spin(1,3) Lie group, and naturally appears as the even-graded subalgebra of the spacetime algebra.

Given a 4-vector $\vec{v} = \sum_{\mu} v^{\mu} \vec{\gamma}_{\mu}$, and a Spin(1,3) group element $\exp(\theta \vec{B}/2)$, where \vec{B} is any linear combination of the 8 basis elements of the spinor algebra, then the automorphism $\exp(\theta \vec{B}^*/2) \vec{v} \exp(\theta \vec{B}/2)$ produces a Lorentz transformation of \vec{v} with an rotation angle θ in the spacetime plane corresponding to \vec{B} . The bivectors with negative pseudonorm generate (hyperbolic) boost rotations upon exponentiation. The bivectors with positive pseudonorm generate (elliptic) spatial rotations. The scalars generate phase-rotations upon exponentiation, which do not cancel in the automorphism since $I^* = I$.

The Spin(1,3) Lie algebra can be written as a direct sum of the three sets of generators. However, only the phase generators $\{1, I\}$ and the rotation generators $\{\vec{\gamma}_{jk}\}$ are closed under Lie bracket relations and can form proper subgroups upon exponentiation. Hence, the enveloping algebra is a semi-direct product and not a true direct product of independent factors as discussed in Section A.4. Nevertheless, one can neglect the boost generators as an approximation to produce a subgroup Spin(3)=SU(2) of the Spin(1,3) group, which is a simply connected double-covering of the spatial rotation group SO(3) that is equivalent to the quaternion algebra. Neglecting the spatial rotation generators as a further approximation produces the subgroup Spin(1) = SU(1) of the Spin(3) group, which is simply the algebra of complex numbers.

There are 8 graded basis elements of the spinor algebra, so 8 real components are necessary to fully parametrize it. We can, however, construct a complex

spectral basis of four idempotents as in (A.18) by examining the characteristic polynomials of the generators. As an example, consider a bivector $\vec{A} = a\vec{\gamma}_{23} + b\vec{\gamma}_{01}$ with elements in both bivector sets. It has a characteristic polynomial,

$$(\vec{A})^4 - 2(a^2 - b^2)(\vec{A})^2 + 4a^2b^2(a^2 - b^2)^2 = 0, \quad (\text{A.33})$$

that has four roots $(\pm a \pm ib)$. Performing the partial fraction decomposition of the inverse of this polynomial, as outlined in [219], produces a set of four idempotents that decompose unity and form a spectral basis for \vec{A} . Every spectral basis for an element with nonzero pseudonorm³ will be an appropriate Spin(1,3) group automorphism of the following four primitive idempotents that partition unity as discussed in Section A.4,

$$\varepsilon_{\pm, \pm} = \frac{1}{4}(1 \pm \vec{\gamma}_0)(1 \pm i\vec{\gamma}_{23}), \quad (\text{A.34})$$

where $i^* = -i$ is an introduced anti-Hermitian scalar imaginary that commutes with the entire spacetime algebra. The scalar imaginary makes the idempotent factor $(1 \pm i\vec{\gamma}_{23})/2$ properly Hermitian under the algebraic involution.

Thus, the enveloping algebra for the Spin(1,3) Lie algebra can be represented on a complex Hilbert space of dimension 4, known as the *Dirac spinor* basis. In that representation, the spacetime basis vectors $\vec{\gamma}_\mu$ have the form of the complex Dirac γ -matrices (motivating the notation) with their matrix products forming the higher-grade basis elements. The pseudoscalar I has the form of the scalar imaginary $-i$ multiplying the Dirac γ_5 -matrix.

³For contrast, lightlike elements in the algebra have zero pseudonorm and cannot be boosted, so their spectral basis elements will have a nilpotent factor multiplying a purely rotational idempotent such as $(1 \pm i\vec{\gamma}_{23})/2$.

If the boost generators are neglected then performing the spectral decomposition produces only two elements in the complex spectral basis, $(1 \pm i\vec{\gamma}_{23})/2$. Hence, the enveloping algebra for the $\text{Spin}(3)=\text{SU}(2)$ Lie algebra can be represented on a complex Hilbert space of dimension 2, known as the *Pauli spinor* basis. In that representation, the quantity $i\vec{\gamma}_{23}$ becomes the Hermitian Pauli σ_1 -matrix, while the pseudoscalar I reduces to the scalar imaginary $-i$.

If the rotation generators are also neglected, then only the scalar part of the algebra remains. Hence, the enveloping algebra for the $\text{Spin}(1)=\text{SU}(1)$ Lie algebra can be represented on a complex Hilbert space of dimension 1, known as the *Schrödinger spinor* basis. Under this representation I is simply replaced by the scalar imaginary $-i$. Hence, a complex phase will be the only manifestation of the relativistic spin stemming from the Lie group $\text{Spin}(1,3)$ of possible group transformations for spacetime when boosts and rotations can be neglected.

Interpreting the spectral idempotents that partition unity as mutually exclusive propositions, as outlined in Chapters 2 and 3, provides the statistical interpretation for quantum spin. In most non-collider laboratory situations boosts can be neglected, so spin will be effectively two-valued and well-described by the subalgebra of Pauli spinors, as used in the main text.

B Sufficient Conditions for the Weak Value

In Section 3.5 we showed that the post-selected conditioned average (3.40) produces the quantum weak value (3.43) unless the contextual values being averaged have poles larger than $1/\epsilon^n$. For most experiments, these higher-order poles will not appear, so this caveat does not spoil the convergence. However, it is instructive to examine how these poles arise.

For sake of discussion, we consider the following additional assumptions:

1. The probability observables $E_y(\epsilon) = \sum_{y'} M_{y,y'}^\dagger(\epsilon) M_{y,y'}(\epsilon)$ commute with the observable F_X .
2. The equality $F_X = \sum_y f_Y(\epsilon; y) E_y(\epsilon)$ is satisfied, where the CVs $f_Y(\epsilon; y)$ are selected according to the pseudo-inverse prescription.
3. The minimum nonzero order in ϵ for all $|M|_{y,y'}(\epsilon)$ is ϵ^n such that assumption (3) can also be satisfied for some CVs by the truncation to order ϵ^n . That is, for all y, y' , $|M|_{y,y'} = c_{y,y'} 1_X + |M|_{y,y'}^{(n)} \epsilon^n + O(\epsilon^{n+1})$, where $\sum_{y'} c_{y,y'}^2 = P_Y(y)$ is the detector probability in absence of interaction, and some of the $|M|_{y,y'}^{(n)}$ may vanish.

First, we note that F_X commutes with $\{E_y(\epsilon)\}$ by assumption (1). As such, we will replace the CV definition $F_X = \sum_y f_Y(\epsilon; y) E_y(\epsilon)$ with an equivalent matrix equation,

$$\vec{f}_X = \mathcal{S} \vec{f}_Y, \quad (\text{B.1a})$$

$$\mathcal{S} = \begin{pmatrix} \langle E_y(\epsilon) \rangle_x & \cdots \\ \vdots & \ddots \end{pmatrix}. \quad (\text{B.1b})$$

The pseudoinverse is constructed from the singular value decomposition $\mathcal{S} = \mathcal{U} \Sigma \mathcal{V}^T$ as $\mathcal{S}^+ = \mathcal{V} \Sigma^+ \mathcal{U}^T$, where \mathcal{U} and \mathcal{V} are orthogonal matrices such that $\mathcal{U}^T \mathcal{U} = \mathcal{V} \mathcal{V}^T = 1$, Σ is the singular value matrix composed of the square roots of the eigenvalues of $\mathcal{S} \mathcal{S}^T$, and Σ^+ is composed of the inverse nonzero elements in Σ^T .

Next, we note that the truncation of the matrix \mathcal{S} to order ϵ^n has the form,

$$\mathcal{S}' = \mathcal{P} + \epsilon^n \mathcal{S}_n, \quad (\text{B.2})$$

where $\mathcal{P} = (P_Y(y) \vec{1}, \cdots)$ is a matrix whose rows are identical and whose columns contain the interaction-free detector probabilities $P_Y(y)$, and $\mathcal{S}_n = (\vec{E}_1^{(n)}, \cdots)$ is a matrix whose rows all sum to zero. Furthermore, since the solution to the equation $\vec{f}_X = \mathcal{S}' \vec{f}_Y$ is assumed to exist by assumption (3), then the dimension of the detector, N , must be greater than or equal to the dimension of the system, M .

We can then show that the singular values of the truncated matrix \mathcal{S}' have maximum leading order ϵ^n unless an exotic condition is met. To see this, note that the singular values of \mathcal{S}' are $\sigma_k = \sqrt{\lambda_k}$, where λ_k are M eigenvalues of

$\mathcal{H} = \mathcal{S}^T \mathcal{S}$, with its other $N - M$ eigenvalues being zero. This matrix has the form $\mathcal{H} = \mathcal{P}^T \mathcal{P} + \epsilon^n (\mathcal{S}_n^T \mathcal{P} + \mathcal{P}^T \mathcal{S}_n) + \epsilon^{2n} \mathcal{S}_n^T \mathcal{S}_n$, where $(\mathcal{P}^T \mathcal{P})_{ij} = M P_Y(i) P_Y(j)$ is $M \|\vec{p}\|^2$ times the projection operator onto the probability vector $\vec{p} = (P_Y(y), \dots)$. We will use \mathcal{H} to determine the singular values of \mathcal{S}' .

Differentiating the eigenvalue relation $\mathcal{H}(\epsilon^n) \vec{u}_k(\epsilon^n) = \lambda_k(\epsilon^n) \vec{u}_k(\epsilon^n)$ with respect to ϵ^n produces the following deformation equation that describes how the eigenvalues of \mathcal{H} continuously change with increasing ϵ^n ,

$$\dot{\lambda}_k(\epsilon^n) = 2(\mathcal{P} \vec{u}'_k(\epsilon^n)) \cdot (\mathcal{S}_n \vec{u}_k(\epsilon^n)) + 2\epsilon^n \|\mathcal{S}_n \vec{u}_k(\epsilon^n)\|^2. \quad (\text{B.3})$$

Integrating this equation produces the following perturbative expansion of the eigenvalues for small ϵ ,

$$\begin{aligned} \lambda_k(\epsilon^n) = & \lambda_k(0) + 2\epsilon^n (\mathcal{P} \vec{u}'_k(0)) \cdot (\mathcal{S}_n \vec{u}_k(0)) + \epsilon^{2n} \|\mathcal{S}_n \vec{u}_k(0)\|^2 \\ & + \epsilon^{2n} [(\mathcal{P} \vec{u}'_k(0)) \cdot (\mathcal{S}_n \vec{u}_k(0)) + (\mathcal{P} \vec{u}_k(0)) \cdot (\mathcal{S}_n \vec{u}'_k(0))] + O(\epsilon^{3n}), \end{aligned} \quad (\text{B.4})$$

where $\vec{u}'_k(0) = \partial_{\epsilon^n} \vec{u}_k(\epsilon^n)|_{\epsilon=0}$.

Since $\mathcal{H}(0) = \mathcal{P}^T \mathcal{P}$ is a projection operator, $\lambda_1(0) = M \|\vec{p}\|^2$ is its only nonzero eigenvalue with associated eigenvector $\vec{u}_1(0) = \vec{p} / \|\vec{p}\|$. Hence, $\sigma_1(\epsilon^{2n}) \approx \sqrt{M} \|\vec{p}\| > 0$ to leading order. For $k \neq 1$, $\lambda_k(0) = 0$ and $\vec{u}_k(0)$ span the degenerate $(N - 1)$ -dimensional subspace orthogonal to $\vec{u}_1(0)$. Furthermore, by direct computation we can show for $k \neq 1$ that $\vec{u}'_k(0) = -\vec{u}_1(0) \left[\vec{1} \cdot \mathcal{S}_n \vec{u}_k(0) \right] / M \|\vec{p}\|$. Therefore, for $k \neq 1$ the expansion (B.4) simplifies to,

$$\lambda_k(\epsilon^n) = \epsilon^{2n} \left[\|\mathcal{S}_n \vec{u}_k(0)\|^2 - |\vec{1} \cdot \mathcal{S}_n \vec{u}_k(0)|^2 \right] + O(\epsilon^{3n}). \quad (\text{B.5})$$

Hence, the term of order ϵ^{2n} can vanish only if $\mathcal{S}_n \vec{u}_k(0) = \vec{0}$ or if $\mathcal{S}_n \vec{u}_k(0) = \beta \vec{1}$ for some constant β .

Suppose $\mathcal{S}_n \vec{u}_k(0) = 0$. It then follows that $\mathcal{H}(\epsilon^{2n}) \vec{u}_k(0) = 0$ since $\vec{u}_k(0)$ is orthogonal to $\vec{u}_1(0) \propto \vec{p}$. Therefore, $\vec{u}_k(0)$ is an eigenvector of $\mathcal{H}(\epsilon^{2n})$ with eigenvalue 0 for any ϵ . Since \mathcal{H} is symmetric, its eigenvectors form an orthogonal set for any ϵ , so we must have the identification $\vec{u}_k(\epsilon^{2n}) = \vec{u}_k(0)$. In this case, $\lambda_k(\epsilon^n) = 0$. Hence, the only way that $\lambda_k(\epsilon^n)$ can have a leading order greater than ϵ^{2n} is if its corresponding zeroth order eigenvector satisfies the relation $\mathcal{S}_n \vec{u}_k(0) = \beta \vec{1}$.

If we assume this exotic condition does not hold, then we can show that the pseudoinverse solution \vec{f}_Y to (B.1a) cannot have poles larger than $1/\epsilon^n$. To see this, note that in order to satisfy (B.1a), we have the equivalent condition for each component of $\mathcal{U}^T \vec{f}_X = \Sigma \mathcal{V}^T \vec{f}_Y$,

$$(\mathcal{U}^T \vec{f}_X)_k = \Sigma_{kk} (\mathcal{V}^T \vec{f}_Y)_k. \quad (\text{B.6})$$

Therefore, all singular values Σ_{kk} corresponding to nonzero components of $\mathcal{U}^T \vec{f}_X$ must also be nonzero; we shall call these the *relevant* singular values. Singular values which are not relevant will not contribute to the solution $\vec{f}_Y = \mathcal{V} \Sigma^+ \mathcal{U}^T \vec{f}_X$. We can see this since $(\vec{f}_Y)_j = (\mathcal{V} \Sigma^+ \mathcal{U}^T \vec{f}_X)_j = \sum_k \mathcal{V}_{jk} \Sigma_{kk}^+ (\mathcal{U}^T \vec{f}_X)_k$, so any zero element of $\mathcal{U}^T \vec{f}_X$ will eliminate the inverse irrelevant singular value Σ_{kk}^+ from the solution for $(\vec{f}_Y)_j$.

Since the orthogonal matrices \mathcal{U} and \mathcal{V} do not contain any poles, and since \vec{f}_X is ϵ -independent, then the only poles in the solution $\vec{f}_Y = \mathcal{S}^+ \vec{f}_X = \mathcal{V} \Sigma^+ \mathcal{U}^T \vec{f}_Y$ must come from the inverses of the relevant singular values in Σ^+ . If a singular value Σ_{kk} has leading order ϵ^m , then its inverse $\Sigma_{kk}^+ = 1/\Sigma_{kk}$ has leading order $1/\epsilon^m$; therefore, to have a pole of order higher than $1/\epsilon^n$ then there must be at

least one relevant singular value with a leading order greater than ϵ^n . However, if that were the case then the truncation \mathcal{S}' of \mathcal{S} to order ϵ^n could not satisfy (B.6) since to that order it would have a relevant singular value of zero according to the previous argument, contradicting assumption (3) about needing to satisfy the CV definition with the minimum nonzero order in ϵ .

To summarize, given the reasonable additional assumptions (1) and (2) about the form of the weak interaction, and provided that the truncated equation is satisfiable according to assumption (3), then the pseudoinverse solution $\vec{f}_Y = \mathcal{S}^+ \vec{f}_X$ can have no pole with order higher than $1/\epsilon^n$ unless the exotic condition $\mathcal{S}_n \vec{u}_k(0) = \beta \vec{1}$ is met for some k . Hence, the majority of laboratory conditioned averages should approximate weak values as the measurement becomes ambiguous.

C Edge Channel Interaction

Phase

It may not be apparent that an extended Coulomb interaction between scattering excitations in adjacent edge states can result in an additional joint phase accumulation γ without destroying the phase coherence, as claimed in Chapter 5. Indeed, in momentum-space the Coulomb interaction explicitly involves energy exchange between the adjacent excitations, which would seem to imply that phase disruption would occur. To assuage such concerns, we shall solve a simple model of Coulomb edge state interaction for the two-excitation amplitude in the co-propagating region. We shall see that it is sufficient to keep the total energy of both excitations constant in order to obtain a joint phase accumulation over the interaction length; the fact that the excitations may exchange energy between them does not disrupt the joint phase coherence.

Consider the longitudinal part of a two-particle amplitude $\psi(x_1, x_2, t)$ that describes chiral copropagation at a speed v along linear edge channels separated by a fixed distance d . In the absence of any Coulomb interaction between the

channels, ψ must satisfy the effective Schrödinger equation,

$$i\hbar\partial_t\psi = -\frac{i\hbar v}{2}(\partial_{x_1} + \partial_{x_2})\psi. \quad (\text{C.1})$$

For a fixed total energy E , the general solution of this equation has the form,

$$\psi_E(x_1, x_2, t) = \psi_0(t)\psi_1(x_1, x_2)\chi(x_2 - x_1), \quad (\text{C.2a})$$

$$\psi_0(t) = \psi_0 \exp\left(-i\frac{E}{\hbar}t\right), \quad (\text{C.2b})$$

$$\psi_1(x_1, x_2) = \exp\left(\frac{iE}{\hbar v}(x_1 + x_2)\right), \quad (\text{C.2c})$$

where χ is an arbitrary function of the difference of the coordinates.

By choosing an initial boundary condition to be a product state of excitation scattering states at distinct energies,

$$E_p^\pm = E(1 \pm p), \quad (\text{C.3})$$

where $p \in [-1, 1]$ such that the energy matching condition $E_p^+ + E_p^- = 2E$ is satisfied and E is correctly quantized, we fix χ to find the general product form for a fixed joint energy of $2E$,

$$\psi_p(2E, x_1, x_2, t) = \xi(E_p^+, x_1, t)\xi(E_p^-, x_2, t), \quad (\text{C.4a})$$

$$\xi(E, x, t) = \xi_0 \exp\left(-i\frac{E}{\hbar}\left(t - \frac{2x}{v}\right)\right). \quad (\text{C.4b})$$

As expected, the channel amplitudes are completely decoupled in the absence of interaction and independently phase-coherent.

For a low-biased source, each single-particle energy will be approximately the

Fermi energy $E_p^+ \approx E_p^- \approx E_F$, and the propagation speed will be the Fermi velocity $v = v_F$. Therefore, in the absence of interaction each particle will accumulate a dynamical phase,

$$\phi(E_F, L) = \frac{2E_F L}{\hbar v_F}, \quad (\text{C.5})$$

after a propagation length L , leading to a total joint dynamical phase of $4E_F L / \hbar v_F$.

If the co-propagating excitations are instead allowed to interact via a screened Coulomb potential, the effective Schrödinger equation must be modified to,

$$i\hbar\partial_t\tilde{\psi} = -\frac{i\hbar v}{2}(\partial_{x_1} + \partial_{x_2})\tilde{\psi} + \frac{\alpha e^2}{r}e^{-r/\lambda}\tilde{\psi}, \quad (\text{C.6a})$$

$$r = \sqrt{d^2 + |x_2 - x_1|^2}, \quad (\text{C.6b})$$

where λ is the screening length, r is the interaction distance that depends in the interacting region on the difference $|x_2 - x_1|$ between the coordinates as well as the distance d between the edge channels, and α is the Coulomb interaction constant in appropriate units. This linear equation decouples in the coordinates $y_1 = x_1 + x_2$ and $y_2 = x_2 - x_1$, so it may still be exactly solved.

For a fixed joint energy E , the general solution has the form,

$$\tilde{\psi}_E(x_1, x_2, t) = \psi_0(t)\tilde{\psi}_1(x_1, x_2)\tilde{\chi}(x_2 - x_1), \quad (\text{C.7a})$$

$$\tilde{\psi}_1(x_1, x_2) = \exp\left(i\tilde{k}(E, x_1, x_2)(x_1 + x_2)\right), \quad (\text{C.7b})$$

$$\tilde{k}(E, x_1, x_2) = \frac{1}{\hbar v}\left(E - \frac{\alpha e^2}{r}e^{-r/\lambda}\right) \quad (\text{C.7c})$$

where ψ_0 is the same as in (C.2) and $\tilde{\chi}$ is another arbitrary function of the difference of coordinates. The Coulomb potential thus gives an effective position-

dependent shift to the joint wave-number \tilde{k} for the amplitude, which will affect the dynamical phase accumulation for the joint amplitude.

If we demand that for $r \gg \lambda$ the general solution (C.7) should reduce to the noninteracting solution (C.4), then we find the simplest form,

$$\tilde{\psi}_p(x_1, x_2, t) = \psi_p(2E, x_1, x_2, t) e^{-i\gamma(x_1, x_2)}, \quad (\text{C.8a})$$

$$\gamma(x_1, x_2) = \frac{\alpha e^2}{\hbar r} e^{-r/\lambda} \frac{x_1 + x_2}{v}, \quad (\text{C.8b})$$

where ψ_p is the noninteracting solution (C.4). The net effect of the Coulomb interaction between the channels is thus to contribute a position-dependent (but energy- and time-independent) phase $\gamma(x_1, x_2)$ that entangles the coordinates of the channels. Any remaining correction factor $\tilde{\chi}(x_2 - x_1)$ to this simple form must satisfy $\tilde{\chi}(0) = 1$, so we will safely neglect it in what follows.

The excitations are collected at ohmic drains at *fixed* positions $x_1 = L_1$ and $x_2 = L_2$ of the coordinates, so the detected phase $\gamma(L_1, L_2)$ will be fixed by the geometry and stable for any pair detection. Scattering states with fixed energy such as (C.8) are stationary and extended throughout the interaction region, which explains the geometric nature of the interaction-induced phase. The square of the wave-function $|\tilde{\psi}_p(2E, L_1, L_2, t)|^2$ indicates the (typically small) probability that the excitations will be detected *simultaneously* at any particular t at the drains. However, by using a coincidence post-selection or by engineering a correlated initial scattering state, one can in principle restrict the bulk of the measured detections to be coincident.

If we further assume that before and after a co-propagation length L where the channels are a fixed distance d apart both edge channels rapidly split away from

one another, then to good approximation the Coulomb interaction only affects the region of length L . After the interaction region, the equation of motion for the state will effectively revert to (C.1), restoring noninteracting dynamical phase accumulation similar to (C.5). Hence, the amplitude for jointly detecting the excitations at ohmic drain positions $L_1 > L$ and $L_2 > L$ will contain an extra dynamical phase that is accumulated by the joint state only within the interaction region,

$$\tilde{\psi}_p(L_1, L_2, t) = \psi_p(2E, L_1, L_2, t) e^{-i\gamma}, \quad (\text{C.9a})$$

$$\gamma = \gamma(L, L) = \frac{\alpha e^2}{\hbar d} e^{-d/\lambda} \frac{2L}{v}. \quad (\text{C.9b})$$

Moreover, the joint state may be further scattered after accumulating the joint interaction phase but before the joint detection, as indicated in the main text. We see that for such simultaneous detection the interaction phase γ will be linear in the interaction length L and therefore should be tunable in principle.

If the detections are not simultaneous, then one excitation will be detected in a drain at a time t_1 , followed by the second excitation at a later time t_2 . The joint state will therefore be successively collapsed by the drain detections, which will introduce an additional relative phase factor due to the discrepancy in detection time. Specifically, for $x_1, x_2 > L$ the joint state will have the form,

$$\tilde{\psi}_p(x_1, x_2, t) = \psi_p(2E, x_1, x_2, t) e^{-i\gamma}, \quad (\text{C.10})$$

with the accumulated interaction phase γ as in (C.9). Detection of the first

excitation at L_1 at t_1 then collapses the joint state to,

$$\tilde{\psi}'_p(x_2, t) = \frac{\psi_p(2E, L_1, x_2, t)}{\sqrt{P(L_1, t_1)}} e^{-i\gamma}, \quad (\text{C.11a})$$

$$P(L_1, t_1) = \int dx_2 |\tilde{\psi}_p(L_1, x_2, t_1)|^2. \quad (\text{C.11b})$$

Evolving the remaining single particle state to time t_2 and then detecting the second excitation at L_2 produces the amplitude,

$$\tilde{\psi}''_p = \frac{\psi_p(2E, L_1, L_2, t_1)}{\sqrt{P(L_1, t_1)}} e^{-i(\gamma + E(t_2 - t_1)/\hbar)}. \quad (\text{C.12})$$

The only difference between the sequential detection amplitudes and the joint detection amplitude (C.9) is the extra temporal phase that is accumulated between the two detections. Notably, the extra temporal phase factor appears as a global phase that should not affect the final statistics, in contrast to the geometric interaction phase γ , which can be exposed by further scattering before the sequential detection as in the main text. Allowing for fluctuations in γ as in Section 5.2.7 will account for geometric uncertainty in the interaction length, as well as the pair injection frequency of the source.

D Joint Weak Value Moments

Here we derive the general equations (6.33) and (6.41) for the von Neumann detector. All generalized conditioned detector moments can be determined through conditioned characteristic functions following a similar derivation to (6.32),

$${}_f\langle e^{i\lambda x} \rangle = \frac{\text{Tr} \left((\hat{P}_f \otimes e^{i\lambda \hat{x}}) \hat{U}_T \hat{\rho} \hat{U}_T^\dagger \right)}{\text{Tr} \left((\hat{P}_f \otimes \hat{1}_d) \hat{U}_T \hat{\rho} \hat{U}_T^\dagger \right)} = \frac{\text{Tr}_S \left(\hat{P}_f e^{i\lambda g \hat{A}} \mathcal{X}_\lambda(\hat{\rho}) \right)}{\text{Tr}_S \left(\hat{P}_f \hat{\rho}'_s \right)}, \quad (\text{D.1a})$$

$${}_f\langle e^{i\lambda p} \rangle = \frac{\text{Tr} \left((\hat{P}_f \otimes e^{i\lambda \hat{p}}) \hat{U}_T \hat{\rho} \hat{U}_T^\dagger \right)}{\text{Tr} \left((\hat{P}_f \otimes \hat{1}_d) \hat{U}_T \hat{\rho} \hat{U}_T^\dagger \right)} = \frac{\text{Tr}_S \left(\hat{P}_f \mathcal{P}_\lambda(\hat{\rho}) \right)}{\text{Tr}_S \left(\hat{P}_f \hat{\rho}'_s \right)}, \quad (\text{D.1b})$$

where we have used the Weyl relation [23], $e^{ia\hat{x}} e^{-ib\hat{p}/\hbar} = e^{iab} e^{-ib\hat{p}/\hbar} e^{ia\hat{x}}$, and have defined the post-interaction reduced state $\hat{\rho}'_s = \text{Tr}_D \left(\hat{U}_T \hat{\rho} \hat{U}_T^\dagger \right)$, as well as the λ -dependent operations,

$$\mathcal{X}_\lambda(\hat{\rho}) = \frac{1}{2} \text{Tr}_D \left(\hat{U}_T (e^{i\lambda \hat{x}} \hat{\rho} + \hat{\rho} e^{i\lambda \hat{x}}) \hat{U}_T^\dagger \right), \quad (\text{D.2a})$$

$$\mathcal{P}_\lambda(\hat{\rho}) = \frac{1}{2} \text{Tr}_D \left(\hat{U}_T (e^{i\lambda \hat{p}} \hat{\rho} + \hat{\rho} e^{i\lambda \hat{p}}) \hat{U}_T^\dagger \right), \quad (\text{D.2b})$$

Computing derivatives of the characteristic functions produces the conditioned

detector moments,

$${}_f\langle x^n \rangle = \frac{\partial^n}{\partial (i\lambda)^n} {}_f\langle e^{i\lambda x} \rangle \Big|_{\lambda=0}, \quad (\text{D.3a})$$

$${}_f\langle p^n \rangle = \frac{\partial^n}{\partial (i\lambda)^n} {}_f\langle e^{i\lambda p} \rangle \Big|_{\lambda=0}. \quad (\text{D.3b})$$

This procedure is similar in spirit to the full counting statistics approach employed in [184].

The first two moments generalize (6.33) and are given explicitly by,

$${}_f\langle x \rangle = \text{Re} \langle x \rangle^w + g \text{Re} \langle A \rangle^w, \quad (\text{D.4a})$$

$${}_f\langle p \rangle = \text{Re} \langle p \rangle^w, \quad (\text{D.4b})$$

$${}_f\langle x^2 \rangle = \text{Re} \langle x^2 \rangle^w + 2g \text{Re} \langle xA \rangle^w + g^2 \text{Re} \langle A^2 \rangle^w, \quad (\text{D.4c})$$

$${}_f\langle p^2 \rangle = \text{Re} \langle p^2 \rangle^w, \quad (\text{D.4d})$$

in terms of the Heisenberg evolved joint post-selection $\hat{P}_T = \hat{U}_T^\dagger (\hat{P}_f \otimes \hat{1}_d) \hat{U}_T$ and the joint weak values,

$$\langle x \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{1}_s \otimes \hat{x}) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (\text{D.5a})$$

$$\langle A \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{A} \otimes \hat{1}_d) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (\text{D.5b})$$

$$\langle p \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{1}_s \otimes \hat{p}) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (\text{D.5c})$$

$$\langle x^2 \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{1}_s \otimes \hat{x}^2) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (\text{D.5d})$$

$$\langle Ax \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{A} \otimes \hat{x}) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (\text{D.5e})$$

$$\langle A^2 \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{A}^2 \otimes \hat{1}_d) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}, \quad (\text{D.5f})$$

$$\langle p^2 \rangle^w = \frac{\text{Tr} \left(\hat{P}_T (\hat{1}_s \otimes \hat{p}^2) \hat{\rho} \right)}{\text{Tr} \left(\hat{P}_T \hat{\rho} \right)}. \quad (\text{D.5g})$$

D.1 Detector Wigner Function

Assuming an initial product state $\hat{\rho} = \hat{\rho}_i \otimes \hat{\rho}_d$, we can compute the operations (D.2) as follows. After computing the detector trace in the p -basis and inserting two complete x -basis sets, the \mathcal{P}_λ operation takes the form

$$\begin{aligned} \mathcal{P}_\lambda(\hat{\rho}) &= \iiint \frac{dp dx dx'}{2\pi\hbar} \langle x' | \hat{\rho}_d | x \rangle e^{-i\frac{p}{\hbar}(x-x' - g \text{ad}^*[\hat{A}] + \hbar\lambda)}(\hat{\rho}_i), \quad (\text{D.6}) \\ &= \iint dx dx' \langle x' | \hat{\rho}_d | x \rangle \delta(x - x' - g \text{ad}^*[\hat{A}] + \hbar\lambda)(\hat{\rho}_i), \\ &= \int dz \widetilde{W}_d(z, g \text{ad}^*[\hat{A}] - \hbar\lambda)(\hat{\rho}_i). \end{aligned}$$

Here we have changed integration variables to $z = x - x'$ and $y = (x + x')/2$, and have noted that $\widetilde{W}_d(z, y) = \langle z - y/2 | \hat{\rho}_d | z + y/2 \rangle$ is the Fourier-transformed Wigner function of the detector (6.40).

Performing a similar computation for \mathcal{X}_λ yields,

$$\begin{aligned}\mathcal{X}_\lambda(\hat{\rho}) &= \iiint \frac{dpdx dx'}{2\pi\hbar} \frac{1}{2} (e^{i\lambda x'} + e^{i\lambda x}) \langle x' | \hat{\rho}_d | x \rangle e^{-i\frac{p}{\hbar}(x-x'-g\text{ad}^*[\hat{A}])}(\hat{\rho}_i), \quad (\text{D.7}) \\ &= \iint dx dx' \frac{1}{2} (e^{i\lambda x'} + e^{i\lambda x}) \langle x' | \hat{\rho}_d | x \rangle \delta(x - x' - g\text{ad}^*[\hat{A}])(\hat{\rho}_i), \\ &= \int dz e^{i\lambda z} \widetilde{W}_d(z, g\text{ad}^*[\hat{A}]) \cos\left(\frac{\lambda g}{2}\text{ad}^*[\hat{A}]\right)(\hat{\rho}_i).\end{aligned}$$

Taking derivatives with respect to $(i\lambda)$ produces the expressions (6.41) in the main text for the first moments. Setting $\lambda = 0$ in either $\mathcal{P}_\lambda(\hat{\rho}_i)$ or $\mathcal{X}_\lambda(\hat{\rho}_i)$ produces the post-interaction reduced system state $\hat{\rho}'_s = \mathcal{E}(\hat{\rho}_i)$.

The operation $\text{ad}^*[\hat{A}]$ is linear, so any analytic function of $\text{ad}^*[\hat{A}]$ may be defined via its Taylor series in the same manner as an analytic function of a matrix. Indeed, to more rigorously perform the above derivations one can work in the adjoint matrix representation discussed in Appendix A.7. After expanding the expressions into the eigenbasis of the matrix of $\text{ad}^*[\hat{A}]$ and regularizing any singular functions into limits of well-behaved analytic functions, the above computations can be performed for each eigenvalue, summed back into a matrix, and then mapped back into the operator form shown. For unbounded \hat{A} then one must also carefully track the domains to ensure that the resulting expressions properly converge, as they should for any physically sensible result.

D.2 Hermite-Gauss superpositions

The Wigner distribution for an arbitrary superposition of Hermite-Gauss modes $|\psi\rangle = \sum_m c_m |h_m\rangle$ can be computed to find (suppressing arguments for compact-

ness),

$$W = \sum_{m,n=0}^{\infty} \frac{c_m c_n^*}{\sqrt{m! n!}} \frac{(-1)^m e^{i(m-n)\phi}}{\pi \hbar} D_n^m[\sqrt{2G}] e^{-G}, \quad (\text{D.8})$$

where $G(x, p) = x^2/2\sigma^2 + 2p^2\sigma^2/\hbar^2$ and $\phi(x, p) = \tan^{-1}(-2p\sigma^2/\hbar x)$. The polynomial sequence $D_n^m(x)$ has the generating function,

$$\exp(z\bar{z} - x(z - \bar{z})) = \sum_{m,n=0}^{\infty} \frac{z^m \bar{z}^n}{m! n!} D_n^m(x), \quad (\text{D.9})$$

and the explicit form,

$$D_n^m(x) = \sum_{k=0}^{\min(m,n)} \frac{m! n! (-1)^{m-k}}{(m-k)! (n-k)! k!} x^{m+n-2k}. \quad (\text{D.10})$$

Notably, the diagonal elements of this sequence are the Laguerre polynomials, $D_m^m(x) = m! L_m(x^2)$. These results can be obtained by using the generating function for the Hermite polynomials $\exp(2xz - z^2) = \sum_{n=0}^{\infty} H_n(x) z^n/n!$, as well as the identities $H_n(x) = \exp(-\partial_{2x}^2)(2x)^n$ and $\int dx e^{-x^2} H_n(x) H_m(x) = \delta_{m,n} m! \sqrt{\pi} 2^m$.

Computing the reduced system state $\hat{\rho}'_s$ using this Wigner function yields,

$$\hat{\rho}'_s = \sum_{m,n=0}^{\infty} \frac{c_m c_n^*}{\sqrt{m! n!}} D_n^m \left[\sqrt{\epsilon} \text{ad}^*[\hat{A}] \right] e^{\epsilon \mathcal{L}[\hat{A}]}(\hat{\rho}_s), \quad (\text{D.11})$$

where $\epsilon = (g/2\sigma)^2$, and $\mathcal{L}[\hat{A}] = \text{ad}^*[-i\hat{A}]^2/2$ is the Lindblad decoherence operation. Notably, the functional form of the Wigner distribution (D.8) is still largely preserved in Eq. (D.11).

The detector averages can also be computed from this Wigner function. The weak value $\text{Re} \langle x \rangle^w$ will vanish by symmetry; the weak value $\text{Re} \langle p \rangle^w$ involves the

derivative $i\hbar\partial_{\text{gad}^*[\hat{A}]} \hat{\rho}'_s$; and, the weak value $\text{Re} \langle A \rangle^w$ involves the state $\hat{\rho}'_s$ directly. When $c_m = 1$ with the rest of the coefficients zero, then these generalizations reduce to the results (6.86) and (6.87) presented in the main text.